Noise in quantum and classical computation & non-locality

Unger, F.P.

Citation for published version (APA):
Chapter 1

Introduction

Quantum mechanics is a physical theory attempting to describe the world on the smallest scales. Its theoretical foundations were mostly laid out in the 1920’s and 1930’s. It accurately predicts effects which are not explainable by classical theories. These effects can aid in information processing tasks, for example in computation but also communication and cryptography.

The idea to use quantum mechanics to do computation goes back at least to the early 80’s [35, 44]. To compute the value of some function \( f \) on a particular input \( x \), one takes a quantum mechanical system (consisting of a collection of photons, ions or some other suitable objects), and initializes its state depending on \( x \). One then manipulates the system according to some predetermined procedure (a quantum algorithm), which depends on the function \( f \) one wants to compute. In the end one observes (measures) the final state of the quantum state and determines the value \( f(x) \) from it.

It is not at all obvious why this way of computing should have advantages over the way classical computers work. Even today we are far away from a full understanding. However, in 1994 Shor [88] showed that it is possible to factor large numbers on quantum computers quickly, i.e. in polynomial time. The currently known best classical algorithms for factoring numbers are comparatively slow; they run in exponential time. This suggests that quantum computers might have capabilities which go beyond those of classical computers. Apart from its theoretical significance, this result is important since factoring numbers quickly will allow to break many cryptographic protocols, which are for example used on the internet. Factoring is not the only problem for which quantum computers give advantages over classical computers. For example, Grover [50] invented an algorithm to search for an item in a database significantly faster than on a classical computer. See [68] for more examples of quantum algorithms which outperform classical ones.

The full potential offered by quantum mechanics for information processing is still not clear. One reason is a practical one: We simply do not yet know how to
construct the physical devices necessary to practically implement the proposed applications. For example, it is currently impossible to build error-free quantum mechanical devices that are strong enough for large-scale quantum computation. It is likely that also in the future the hardware for building quantum computers will have faults. In Part I we will look at the limitations of quantum computation, when all we have are faulty devices (Chapters 3, 4 and 5). We will also analyze noise in classical computation (Chapter 6).

Apart from the practical problem of building quantum devices, we are only starting to discover what kind of applications are theoretically possible using quantum mechanics. One particular example is that of multi-prover interactive proof systems (MIP systems), which we discuss in Part II in Chapter 7. MIP systems are verification procedures in which a certain number of provers try to convince a verifier of the truth of some statement. It is important that the provers are not allowed to communicate with each other during the protocol. Classical proof systems are relatively well-understood. Much less is known about quantum interactive proof systems, in which the provers may share an entangled state. The reason we know much less is that we do not have a full understanding of entanglement, yet. In Chapter 7 we analyze a special class of quantum MIP systems and their behaviour under simultaneous (or parallel) repetition.

In the last chapter we try to explain another mystery of entanglement: Why does quantum mechanics allow entangled, non-communicating parties to generate certain shared distributions, but certain others not? Or more generally: Why are the quantum mechanical axioms like they are? We give a partial answer to this question. The existence of some of the distributions that are ruled out by quantum mechanics would have some really strange consequences and would make our world very different from how it is. This result partially explains why quantum mechanics is like it is and puts constraints on all physical theories extending it. Incidentally, the techniques of Chapter 8 use fault-tolerant computation.

The author hopes that this brief overview of the results has incited the reader to read on. In the rest of this chapter we will explain the results and some more background in more detail.

1.1 Limits on fault-tolerant computation

Quantum computing with imperfect devices

At the moment we are a long way from building quantum computers large enough to solve large instances of the problems for which we believe quantum algorithms are faster than classical algorithms (e.g. factoring). This is despite a decade-long effort by experimental physicists. The general problem is that the objects carrying the quantum information must be small in order to exhibit quantum mechanical behaviour. Common proposals use photons, ions or other “small”
1.1. Limits on fault-tolerant computation

objects. One reason why large enough quantum computers do not exist yet is that it is hard to manipulate and operate on these small objects faultlessly. Even worse, also if not operated on, the state of the quantum system can deviate over time from its original state if no precautions are taken. If too many faults in the system accumulate over time the final measurement will not give any useful information about the value of the function we want to compute. This problem of manipulating and preserving quantum states makes it hard to build large quantum computers.

Rather surprisingly, there is a way around this. In the mid 90’s Shor, Steane, and others [86, 87, 90, 48] invented a “software solution” for the problem of noise. They showed that quantum error-correcting codes exist, which means it is possible to map the state of a quantum mechanical system $A$ (consisting e.g. of $N$ photons) into some slightly larger system $\tilde{A}$ (consisting e.g. of $10^N$ photons), in such a way that it is possible to store the state of $A$ essentially perfectly even if $\tilde{A}$ is slightly noisy (i.e. some of the photons are not in state they are supposed to be). Later results improved on this and showed that it is not merely possible to store quantum states fault-tolerantly in this way. It was shown [4, 59, 62] that it is even possible to do this encoding in a way which allows to simulate the computation of the noise-free system $A$ with the faulty system $\tilde{A}$. In particular, if the probability that errors happen in the $\tilde{A}$ system is small enough and below some threshold—usually referred to as the fault-tolerance threshold

\[ \text{0.1%} \] , which orders of magnitude smaller than the noise rates currently achievable in the lab. On the other hand, the situation is not as bad as it looks, since these rigorous lower bounds are rather conservative and probably underestimate the true thresholds. The gaps between rigorous lower and upper bounds on the threshold are significant, often by a factor of $10^2$ to $10^3$. For most models the exact values are still

---

1Of course, its value depends crucially on the exact parameters of each particular system. The exact value will not matter for the following qualitative account, though.
unknown. The true thresholds will be somewhere in between. A particular very interesting scheme proposed by Knill [61] was estimated to allow universal quantum computation with gates that have more than 3% of depolarizing errors, and a recent result [43] estimates that the actual threshold is as high as 6.88% for this particular scheme.

Upper bounds on quantum fault-tolerance threshold

In this thesis we try to prove rigorous upper bounds on the tolerable noise level, thereby shrinking the gaps between lower and upper bounds. For one particular model we will show a tight threshold. In the following we list the contributions of this thesis and compare them. The definitions of relevant terms are given in the respective chapters and Chapter 2. In particular, Chapter 2 contains the definitions of efficient quantum computing and quantum circuits.

In Chapters 3 and 4 we will consider quantum circuits with storage noise only, which means that we assume that all gates used are perfect, and after each time-step noise happens on each qubit independently. In contrast, when we talk about gate noise we mean that after the execution of each gate some noise happens, which may be an arbitrary quantum operation applied to all the outgoing wires of the gate coherently. This will be the model in Chapter 5, in which we establish a threshold for a particular set of gates.

The reason for considering storage noise in Chapters 3 and 4 are manyfold: At the current stage of the development it is not clear which proposal for building physical quantum computers will be used eventually. Since each proposed implementation has different noise properties, it currently seems more appropriate to develop general techniques and tools for proving noise bounds, rather than exact results for concrete proposals. The techniques presented in the following are very general and can be easily adapted to gate noise as well. Furthermore, in several proposals for physical implementations of quantum computers, e.g. in ion-traps, storage noise actually seems to be the most severe noise. Finally, in most models storage noise can be seen as a particular kind of gate noise, since the noise on the outgoing wires may be considered to belong to the previous gate.

In Chapter 3 we study erasure noise. Erasure noise (see Chapter 3 for a more precise definition) of rate $p$ is an operation that on input of some quantum state $\rho$ outputs $\rho$ and a classical bit $|0\rangle$ with probability $1 - p$ and with probability $p$ it outputs some fixed quantum state of the same dimension as $\rho$ and a classical bit $|1\rangle$. The classical bit indicates whether an error occurred or not.

The main result from Chapter 3 is that circuits that use gates with at most $k$ input wires, and in which each wire is erased with probability $1 - 1/k$ in each time-step can neither be universal for classical or quantum computation. The proof shows that above this noise rate it is impossible to transmit a single bit from the input to the output, if the output is sufficiently far away from the input. In particular, after a logarithmic amount of time any two input states become
1.1. Limits on fault-tolerant computation

indistinguishable. Further, already after a constant amount of time, any two input states become indistinguishable for measurements which act on one qubit only. The proof works by showing that above this noise rate the output becomes “disconnected” from the input.

A slightly weaker result which applies only to depolarizing noise was obtained by Razborov [79]. Depolarizing noise with probability $p$ is a quantum operation which applies the identity operation with probability $1 - p$ and replaces the state by the completely mixed state $I/d$ with probability $p$.

The proof in Chapter 3 is relatively simple, but nevertheless the best general upper bound on the tolerable noise currently known. Further, we will show that this bound is tight in some sense, for if erasure noise has rate less than $1 - 1/k$, it is possible to transmit a bit from the input to the output. It is likely (though not proven) that below the threshold arbitrary fault-tolerant quantum computation is possible but it is not clear whether efficient fault-tolerant quantum computation is possible.

In Chapter 4 we show that circuits with arbitrary, essentially noise-free 1-qubit gates and unitary $k$-qubit gates are useless for fault-tolerant quantum computation if there is depolarizing noise of more than $1 - \sqrt{2^{1/k} - 1}$ on all the incoming wires of the $k$-qubit gates. “Useless” in this case means that after a constant amount of time it is impossible to distinguish any two input states with bounded error by a single-qubit measurement.

Of special interest from an experimental point of view is the case $k = 2$, for which our bound becomes about 35.7%. Furthermore, for the case in which the only allowed two-qubit gate is the controlled-NOT (CNOT) gate, we can improve our bound further to about 29.3%, as we show in Section 4.5. This case is interesting both theoretically and experimentally. Note also that the CNOT gate together with all one-qubit gates forms a universal set [10]. The same noise-bound applies if we also allow controlled-Y and controlled-Z gates.

The results of Chapter 3 and 4 are summarized in Figure 1.1. The bound $1 - \sqrt{2^{1/k} - 1}$ obtained in Chapter 4 is better than $1 - 1/k$ from Chapter 3 for all $k$. In particular the bound behaves like $1 - \Theta(1/\sqrt{k})$. This matches what is known for classical circuits (see later in this chapter), and therefore probably represents the correct asymptotic behavior.

However, the result in Chapter 4 is weaker than the result in Chapter 3 in certain aspects. Most importantly, we analyze depolarizing noise instead of erasure noise. Further, we assume that all $k$-qubit gates are mixtures of unitaries, which slightly restricts generality. Not every completely-positive trace-preserving map can be written as a mixture of unitaries.\footnote{In the sense that gates may perform any physical operation; only the number of wires going into a gate is restricted.} We believe that it is still a reasonable

\footnote{One can implement an arbitrary gate by a unitary gate acting on the original qubits and additional ancilla qubits in a fixed pure state, but this increases the arity of the gate and moreover the ancilla qubits will be affected by the noise operators that precede the unitary.}
Bounds given apply to (1) quantum circuits with noisy fan-in-\(k\) unitaries with depolarizing noise on input wires and essentially noise free 1-qubit gates and (2) quantum circuits with arbitrary gates and erasure noise on wires.

Figure 1.1: Upper bounds on noise for fault-tolerant quantum computation

assumption. For instance, to the best of our knowledge, all known fault-tolerant constructions can be implemented using such gates (in addition to arbitrary one-qubit gates). Moreover, all known quantum algorithms obtain their speed-up over classical algorithms by using only unitary gates.

Another restriction is the assumption that the output consists of just one qubit. In many instances this is an acceptable assumption. For instance, this is the case whenever the circuit is required to solve a decision problem. Moreover, our results can easily be extended to the case where the output is obtained by a measurement on a small number of qubits, instead of only one.

To prove the results in Chapter 4 we introduce a new technique for obtaining upper bounds on the fault-tolerance threshold. Namely, we use a Pauli basis decomposition in order to track the state of the computation. We believe this framework will be useful also for further analysis of quantum fault-tolerance. A finer analysis of the Pauli coefficients might improve the bounds we achieve here, and possibly obtain bounds that are tailored to other computational models.

Note that the results in Chapters 3 and 4 all apply to arbitrary starting states. In particular they also apply when the initial state is encoded in some good quantum error-correcting code.

The third result is in Chapter 5 and there we do not consider storage noise, but gate noise for a very specific but interesting set of gates. We establish a
1.1. Limits on fault-tolerant computation

A threshold of \( \hat{\theta} = (6 - 2\sqrt{2})/7 \approx 45\% \) for depolarizing noise on 1-qubit unitaries, when additionally noisefree stabilizer operations (CNOT gates, Hadamard gates, \( \pi/4 \)-gates, preparation of computational basis states and measurements in the computational basis) are available. We first prove that in this model with noise rates at least \( \hat{\theta} \) fault-tolerant quantum computation is impossible. We then show a second result, that if one allows additionally classical co-processing and perfect classical control (i.e. later quantum gates may arbitrarily depend on earlier measurement outcomes) at noise rates above \( \hat{\theta} \) the whole computation can be efficiently simulated, using the Gottesman-Knill Theorem. We then explain how it follows from [83, 21] that this last result is tight, i.e., at noise rates less than \( \hat{\theta} \) it is possible to do efficient universal quantum computation.

Other related results

Early results on upper bounds of the threshold decoherence rate were obtained by showing that quantum computers with faulty gates can be simulated efficiently on a classical computer. The first to prove one of these results were Aharonov and Ben-Or [3], who proved an upper bound of 97\% for depolarizing noise. In other words, if the noise has rate is higher than 97\%, then quantum computers cannot be (significantly) faster than classical computers. Later Harrow and Nielsen [52] showed that if 74\% depolarizing noise is applied to each output qubit of each gate, then (faulty) two-qubit gates cannot produce entanglement. They concluded that circuits containing only one- and two-qubit gates with depolarizing noise at least 74\% can be simulated efficiently on a classical computer.

Another result by Virmani, Huelga and Plenio [98] shows that the set consisting of CNOT with depolarizing noise at least 67\% and arbitrary 1-qubit gates is efficiently simulatable classically. In this paper they also introduce the interesting idea that sufficiently noisy 1-qubit gates can be simulated by Clifford gates, which we will also use in Chapter 5. Their strongest results are for a restricted class of gates (ones which are diagonal in the computational basis) and dephasing or worst-case noise. They prove that \( (\sqrt{2} - 1)/\sqrt{2} \approx 29\% \) dephasing noise is enough to make these diagonal gates a mixture of Clifford operations. (They define 1-qubit dephasing noise as \( \rho \mapsto \frac{1}{2}(\rho + Z\rho Z) \).) Note that classical states (states in the computational basis) are not affected by dephasing noise. Therefore even at 100\% dephasing noise it is possible to do universal classical computation, using classical gates. The result in Chapter 5 uses depolarizing noise, which is symmetric in the sense that it is invariant under local basis changes. Our result in Chapter 5 implies that if the depolarizing noise is at least \( \hat{\theta} \approx 45.3\% \) then even universal classical computation is not possible anymore.

Finally, it is known that it is impossible to transmit quantum information through a \( p \)-depolarizing channel for \( p > 1/3 \) [22]. As Razborov [79] noticed, this seems to suggest that quantum computation is impossible with depolarizing noise of strength greater than 1/3, but there is no proof that this is indeed the case.
1.1.1 Limits on fault-tolerant classical computation

Proving noise bounds for quantum fault-tolerance is difficult. However, even classically we do not have an exact understanding of fault-tolerant computation. We believe that a good understanding of fault-tolerant classical computing will also help for a better understanding in the quantum case, as many concepts from classical fault-tolerance also naturally appear in quantum fault-tolerance. One particular example are CSS-codes, the most commonly used quantum error-correcting codes, which are derived from classical error-correcting codes.

Moore’s law and noise in classical computation

But there is another and perhaps even more compelling reason to study classical fault-tolerance. Over the last half century we saw great increases in computational power, by shrinking the size of the components on computer chips. This is known as hardware miniaturization and is roughly governed by Moore’s law [64]. Moore’s law states that the number of gates on computer chips has been increasing roughly exponentially over the last decades. Accordingly, the size of the gates has been shrinking exponentially. However, the size of the components used in modern computer chips are close to the physical limits above which non-faulty behaviour of the components can be guaranteed. It was pointed out [15] that at the current speed of miniaturization we will reach the point within the next decade at which it will be impossible to make the components/gates on chips smaller without making them faulty at the same time. If we want to continue those increases in computational power from the past, it is likely that at some point one has to deal with faulty components.

In Chapter 6 we consider noise in classical computation. Given a set of gates, which sometimes fail, we ask how much noise on the gates is tolerable, such that any function can still be computed with bounded-error. Gates on \( k \) input wires compute boolean functions \( f : \{0,1\}^k \rightarrow \{0,1\} \). A gate fails with probability \( p \), if the output is flipped with probability \( p \). We will assume throughout that gates fail independently of each other.

Note that this definition of noise corresponds to replacing the output bit with a uniformly random bit with probability 2\( p \) and leaving the output bit untouched with probability 1 − 2\( p \). Our definition of noise in the classical case is therefore somehow inconsistent with the definition of depolarizing noise in the quantum case, because depolarizing noise \( p \) means that with probability \( p \) a bit is replaced by the completely mixed state \( \mathbb{I}/2 \) (i.e. a random bit) and with probability 1 − \( p \) nothing happens. In order to compare our noise bounds for classical computation to those for quantum computation\(^4\) it is therefore necessary to multiply the noise bounds for classical computation by a factor of 2. This inconsistency in definitions

\(^4\)which is strictly speaking of course not possible, since in one model we allow quantum gates and in the other only classical gates
is unfortunate, but we also stick to them here since these definitions are standard in the classical respectively quantum literature.

**Fault-tolerance thresholds for classical computation**

The question of noise in computation has been studied already during the infancy of computers. Already in 1956 von Neumann discovered that reliable computation is possible with noisy 3-majority gates if each gate fails independently with probability less than $0.0073$ [67]. The first to prove an upper bound on the tolerable noise was Pippenger [73]. He proved that formulas\(^5\) with gates of fan-in at most $k$, where each gate fails independently with probability at least $\epsilon \geq \frac{1}{2} - \frac{1}{2k}$, are not sufficient for universal computation (i.e. not all functions can be computed with bounded error). Feder proved that this bound also applies to circuits [39]. Later, Feder’s bound was improved to $\frac{1}{2} - \frac{1}{2\sqrt{k}}$ by Evans and Schulman [37].

![Graph showing upper bounds on noise for fault-tolerant classical computation](image)

Bounds given apply to (a) arbitrary classical circuits with gates of fan-in at most $k$, (b) classical formulas with gates of fan-in at most $k$, with $k=2$ or $k$ odd. The bounds in (b) are thresholds.

**Figure 1.2: Upper bounds on noise for fault-tolerant classical computation**

For formulas with gates of fan-in $k$ and $k$ odd, Evans and Schulman [38] proved the tight threshold $\beta_k := \frac{1}{2} - \frac{2^{k-2}}{k(k/2-1/2)}$ on the amount of noise for which fault-tolerant computation is possible. Tight here means that if all gates fail

---

\(^5\)Formulas are circuits in which every gate has exactly one output wire. See Section 6.2 for exact definitions.
independently with the same fixed probability \( \epsilon < \beta_k \), then any function can be bounded-error computed, and if each gate fails with some probability at least \( \beta_k \) (which does not need to be the same for all gates), universal computation is not possible. For \( k = 3 \) the threshold was first established by Hajek and Weller [51].

However, so far it has not been possible to establish thresholds for gates with even fan-in (or even prove their existence), as pointed out in [38]. In particular, the most basic case of fan-in 2, which is most commonly used in modern computer hardware, had been elusive. An intuitive argument why even fan-in is different is that for even fan-in, threshold gates (and in particular majority gates) can never be “balanced”, in the sense that the number of inputs on which they evaluate to 1 cannot be the same as the number of inputs on which they evaluate to 0.

In Chapter 6 we show that fault-tolerant computation with formulas at noise rates more than \( \beta_2 = (3 - \sqrt{7})/4 \approx 8.856\% \) is impossible. Together with a result by Evans and Pippenger [36], which shows that at noise rates less than \( \beta_2 \) fault-tolerant computing is possible (if all gates fail with the same probability), this establishes a threshold. We introduce a new technique, which takes care of the peculiarities in the even fan-in case. We expect that it can be extended to other (even) fan-in cases. We conjecture that our bound also holds for circuits. The results for classical fault-tolerant computation are shown in Figure 1.2.

1.2 Entanglement and interactive proof systems

1.2.1 Repetition of XOR games

Entanglement is probably the most intriguing notion in quantum mechanics and describes the phenomenon that two particles (which can in principle be arbitrarily far away) can in some sense be connected, or “entangled”: Doing something to one particle, seems to have an instantaneous effect on the other particle. This effect cannot be used to transmit information faster than light and therefore does not contradict causality (meaning that an effect cannot precede its cause). Nevertheless, many notable physicists rejected the possibility of those “spooky actions at a distance”, because they contradict the principle of locality (meaning that an object can only be influenced by its immediate surrounding). However, after many experiments during the last half century have verified the predictions of quantum mechanics, it is currently mostly accepted that quantum theory is the best theory we have. It explains phenomena which happen in the “small” world very accurately.

CHSH game

One particular aspect of entanglement is that spatially separated parties who share an entangled quantum state, can produce correlations which parties who only share a classical state cannot achieve. This statement is best explained with
a game, played between a referee—also called verifier—and two more parties, Alice and Bob, see Figure 1.3. Alice and Bob are not allowed to communicate but may share a quantum state (or a classical state). The referee selects two random bits \( x, y \) and sends them to Alice and Bob respectively. They each reply with one bit \( a, b \) respectively. The referee accepts if \( x \cdot y = a \oplus b \). In words: He accepts if \( x = y = 1\) and \( a \neq b \), and he also accepts if \( x \) and \( y \) are not both 1 and \( a = b \). In all other cases he rejects. This particular game is called the CHSH game and is an example of the more general class of XOR games, in which the referee’s decision only depends on the parity of the two output bits from Alice and Bob. These games will be the focus of Chapter 7.

If Alice and Bob share some particular entangled state (an EPR-pair), then they can win the CHSH game with probability \((2 + \sqrt{2})/4 \approx 85\%\). Tsirelson’s bound [93] states that this is the best they can do. On the other hand, if they do not share entangled states, they can win with probability at most \(3/4\). This bound is known as the CHSH-inequality [25], which is a particular kind of Bell-inequality.

Bell-inequalities

The term *Bell-inequality* is generally used for bounds on the winning probability of games in which Alice and Bob do not share entanglement.
Chapter 1. Introduction

So, if Alice and Bob claim to possess entanglement, it is easy for the referee to verify this by just playing sufficiently many CHSH games after each other and checking whether Alice and Bob win close to 85% of the games or not (significantly) more than 3/4. With more and more repetitions of this protocol, the probability that unentangled Alice and Bob can trick the referee into believing that they do share entanglement becomes smaller and smaller and can be made arbitrarily small.

These “games” are not just play. The whole topic of Bell inequalities in physics can be cast in the framework of games. In fact, constructing games in which Alice and Bob share a quantum state and have a higher winning probability than without entanglement is an important way to show that classical physics cannot explain all real-world phenomena. On the other hand, quantum mechanics can explain these phenomena, and hence gives a more accurate description of the world.

Games in computer science

Furthermore, these games are not only relevant for physics but also for computer science. One of the great challenges in computer science is to capture the computational complexity of algorithmic problems, for example by the number of elementary operations needed to solve a problem. In general this is very hard. Interestingly, it turns out that these games (without entanglement) are also a powerful tool to characterize the complexity of many computational problems. For example all problems in NEXP (the class of problems which can be solved by a non-deterministic Turing machine in exponential time) can be characterized using classical XOR games (i.e. XOR games in which Alice and Bob do not share entanglement), see [31, 13, 53]. It is also possible to characterize other complexity classes in this setting and the general area is called the theory of interactive proof systems, from which many beautiful and deep results emerged during the last two decades. The expressive power of entangled games is less understood, although there has been significant progress recently. The introduction of Chapter 7 contains some known results concerning classical as well as quantum interactive proof systems.

Sequential versus parallel repetition

It turns out that also in interactive proof systems it is often necessary to repeat games to boost one’s confidence about the results (see Chapter 7 for some more explanations). However, regardless of whether games are used inside interactive proof systems or for generalized Bell inequality tests in physics, repeating games sequentially (i.e. one after another) might be undesirable for certain reasons. For example, in the case of Bell inequality tests the time needed to execute the whole sequential protocol goes up. In the case of interactive proof systems proto-
Entanglement and interactive proof systems

1.2. Entanglement and interactive proof systems

ocols with several rounds of interaction lose certain desirable structural properties which one-round protocols have (like Zero-knowledge). A different solution would be to play all games in parallel, i.e., sending Alice and Bob the questions for all games at once and then getting all answers back at the same time. This induces a new problem though, since Alice and Bob might not play in each individual game the strategy they would have played in a sequential protocol. Since they get all inputs at once, they can also choose a collective strategy which depends on all inputs. If we care about the probability that Alice and Bob can win all games in one parallel protocol, a collective strategy can indeed help, see Section 7.6.2 on page 110 for an example. The celebrated Parallel Repetition Theorem [78] by Ran Raz addresses one aspect of this problem in the classical case and shows that if the maximum probability to win one game is $c < 1$, then there is some constant $c' < 1$ such that winning $n$ games in parallel has winning probability at most $c'^n$. This means that with sufficiently many parallel repetitions the success probability of winning all games goes down exponentially. For the particular kind of games encountered in classical interactive proof systems, this result is sufficient to make the error probability of the referee negligible. We show an analogous result for XOR games with entanglement in Chapter 7: Let $G_1, \ldots, G_n$ be a number of (possibly different) quantum XOR games and assume that the maximum probabilities to win each game individually are respectively $c_1, \ldots, c_n$. Then the probability to win all games $G_1, \ldots, G_n$ in a parallel protocol is exactly $\prod_{i=1}^{n} c_i$.

What this means is that if Alice and Bob share entanglement, they do not gain anything by correlating their strategies, but rather playing them independently is optimal. We therefore call our theorem a perfect parallel repetition theorem.

This chapter also contains an additivity result for quantum XOR games, which is central in the proof of our parallel repetition theorem proof. The setup for the second result is exactly like for the parallel repetition theorem, but we define that the parallel protocol is won if Alice and Bob lose an even number of individual games, otherwise it is lost. In particular, Alice and Bob win the protocol if they win all individual games, but also if they lose exactly $2, 4, 6, \ldots$ games. Our additivity theorem states that the best strategy for Alice and Bob again is the trivial strategy of playing all games independently.

Our results imply that in the quantum world XOR games behave perfectly natural under composition, which is not always true for XOR games without entanglement.

1.2.2 Limits on non-locality

At the end of the last paragraph we noted that sometimes quantum mechanics behaves more natural and intuitive than classical physics. In Chapter 8, we try to reverse this argument. We start by identifying some natural property any physical theory—in particular quantum mechanics—should have and study its implications. It follows that these implications themselves should hold in.
any reasonable physical theory, and hence we can interpret these implications as “natural” axioms themselves.

None of the predictions of quantum mechanics have been disproved so far, and therefore we assume that its axioms predict (small-scale) physical phenomena accurately. What our kind of approach can add, is that we will not need to look at the axioms of quantum mechanics as just some purely mathematical theory, which happens to describe quantum effects very well, but we can explain how these axioms come about. Even further, these natural axioms should not only hold for quantum mechanics itself, but also for all other theories extending it.

Note, that the whole theory of general relativity is derived from some natural assumptions about the world and it would be interesting to do the same for quantum mechanics. The goal of recovering the axioms of quantum mechanics from some natural assumptions alone is of course very ambitious, and we shall be content with some more modest results.

The non-signalling condition

To illustrate this we go back to the example of CHSH games from Figure 1.3 and consider the kind of correlations Alice and Bob can create. For some arbitrary (quantum) strategy of Alice and Bob, let $P(a, b|x, y)$ be the probability of outputs $a$ and $b$ given the inputs $x$ and $y$. The correlation $P(a, b|x, y)$ obtainable from a shared quantum state is non-signalling, i.e., it is impossible for Alice to gain any information about Bob’s input $y$ by observing her output $a$, without actually communicating with him. More formally, the non-signalling condition means that the marginal distributions satisfy $\forall a,x : P(a|x,0) = P(a|x,1)$ and analogously $\forall b,y : P(b|0,y) = P(b|1,y)$. If Alice (respectively Bob) could manipulate her share of the quantum state in such a way that Bob can gain some knowledge about her input, then she could instantaneously (in particular at a speed faster than light) send information to Bob. This is another natural condition/axiom we require.\footnote{Physicist say that otherwise our theory is not causal, see also page 117.}

However, the non-signalling condition alone will not lead to strong constraints on the axioms of quantum mechanics. For example for the case of CHSH games it is possible to define correlations $P(a, b|x, y)$ which are non-signalling, yet they can be used to achieve the maximum success probability of 1 in the CHSH game, whereas we pointed out before that the maximum success probability in the quantum case is only $\approx 85\%$. Such maximal non-signalling correlations can be operationally defined by saying that the first player who fixes their input bit gets a uniformly random bit as output and once the second player also fixes their input bit, the output bit will be chosen such that $x \land y = a \oplus b$. Note that $a$ (and similarly $b$) is always independent of $x$ and $y$ and, hence, the resulting distribution satisfies the non-signalling requirement and can be used to achieve success.
1.2. Entanglement and interactive proof systems

probability 1 in the CHSH game.

Stronger conditions

So, if we want to have a better characterization of the quantum mechanical axioms, we need to add some more natural assumptions. Popescu and Rohrlich were the first to observe that the non-signalling condition alone allows for higher correlations in the CHSH game than allowed by quantum mechanics [74, 75, 76]. They asked themselves whether there is any fundamental reason—or natural requirement—why nature does not allow arbitrary non-signalling correlations. They constructed a toy theory, which (for certain games) allows arbitrary non-signalling correlations and which is apparently consistent with causality. If there is no obvious formal contradiction, why is quantum mechanics like it is and why (or because of which natural requirements) does nature not allow for stronger correlations?

Cleve [26] and van Dam [95, 96] realized that maximal correlations which are only constrained by the non-signalling condition, would indeed imply a very strange world: In this world two parties Alice and Bob could compute the outcome of any function $f(x, y)$ with just one bit of communication, when the input $x$ is given only to Alice and $y$ only to Bob. This seems too good to be true, and somewhat unreasonable. This result implies that maximal non-signalling correlations of the CHSH type should not exist, under the following natural requirement: There are functions on shared inputs for which the number of bits one needs to communicate in order to compute the value of that function is not trivial (i.e. it should be larger than 1). We may take the non-existence of maximal non-signalling correlations of the CHSH type as a natural requirement for any physical theory.

In Chapter 8 we expand this idea further and prove that even non-signalling correlations which allow to win the CHSH game with probability more than $\frac{3+\sqrt{8}}{6} \approx 90.8\%$ allow for trivial communication complexity, i.e., there is an $\epsilon > 0$ such that every Boolean function on distributed inputs can be evaluated with probability at least $1/2 + \epsilon$ using just a single bit of communication. By the same argument as before these correlations should not exist in nature. This suggests to make the following an axiom of any reasonable physical theory: Instantaneous correlations which can win the CHSH game with probability more than $\frac{3+\sqrt{8}}{6} \approx 90.8\%$ cannot exist.

We do not know if our bound is optimal. Ideally, we would like to lower it to the quantum mechanical bound of $(2 + \sqrt{2})/4 \approx 85\%$. This would mean that the assumption that communication complexity should not be trivial implies Tsirelson’s bound and we would get a tight characterization of the quantum mechanically achievable CHSH type correlations. Incidentally, the techniques we use in this chapter use fault-tolerant techniques and are therefore strongly related to the results in Part I of this thesis.