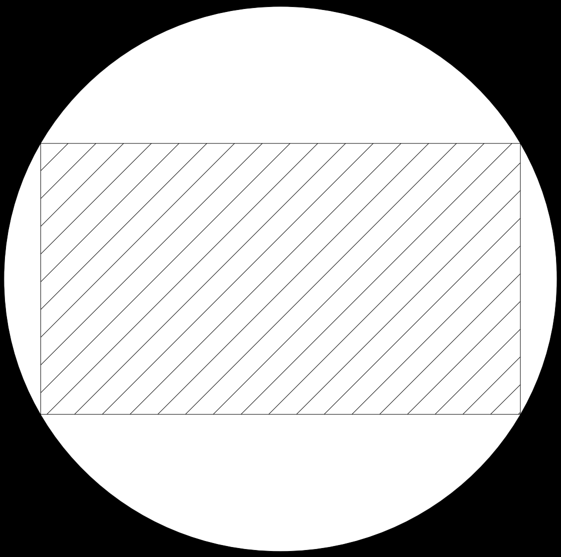
Noise in Quantum and Classical Computation

&

Non-locality



Falk Unger

Noise in Quantum and Classical Computation

&

Non-locality

ILLC Dissertation Series DS-2008-06

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Plantage Muidergracht 24
1018 TV Amsterdam
phone: +31-20-525 6051
fax: +31-20-525 5206
e-mail: illc@science.uva.nl
homepage: http://www.illc.uva.nl/

Noise in Quantum and Classical Computation

&

Non-locality

Academisch Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor
promoties ingestelde commissie, in het openbaar
te verdedigen in de Agnietenkapel
op donderdag 18 september 2008, te 10.00 uur

door

Falk Peter Unger

geboren te Dresden, Duitsland.

Promotiecommissie:

Promotor:	prof. dr. H.M. Buhrman
Overige Leden:	prof. dr. R.E. Cleve
	prof. dr. R.D. Gill
	prof. dr. A. Schrijver
	dr. L. Torenvliet
	dr. R.M. de Wolf

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

The investigations were financially supported by the EU project QAP (IST-2005-15848) and by the Netherlands Organisation for Scientific Research (NWO) in the framework of the Innovational Research Incentives Scheme (in Dutch: Vernieuwingsimpuls), Vici-project 639.023.302.

Printed and bound by PrintPartners Ipskamp.

ISBN: 90-6196-547-0

The results in this thesis are based on the following articles

Chapter 3 The results in this chapter are based on an unpublished manuscript: Falk Unger, **Erasure noise threshold for fault-tolerant computation**, unpublished

Chapter 4 Julia Kempe, Oded Regev, Falk Unger and Ronald de Wolf, **Upper bounds on the noise threshold for fault-tolerant quantum computing**, *ICALP 2008*

Chapter 5 Harry Buhrman, Richard Cleve, Monique Laurent, Noah Linden, Lex Schrijver and Falk Unger, **New Limits on Fault-Tolerant Quantum Computation**, *Proceedings of 47th IEEE FOCS*, 2006

Chapter 6 Falk Unger, **Noise threshold for universality of 2-input gates**, *Proceedings of IEEE International Symposium on Information Theory, 2007*, accepted to *IEEE Transactions on Information Theory*

Chapter 8 Gilles Brassard, Harry Buhrman, Noah Linden, André Allan Methot, Alain Tapp, and Falk Unger, **Limit on nonlocality in any world in which communication complexity is not trivial**, *Physical Review Letters* 96(25), 2006.

Chapter 7 R. Cleve, W. Slofstra, F. Unger, and S. Upadhyay, **Strong parallel repetition theorem for quantum XOR proof systems**, In *Special Issue of 22nd IEEE Conference on Computational Complexity*, 2007.

The author also coauthored the following papers, which are beyond the scope of this thesis

- Falk Unger, **On small hard leaf languages**, In *Proceedings of the 30th International Symposium on Mathematical Foundations of Computer Science*, pages 781–792. Springer, 2005.
- Harry Buhrman, Leen Torenvliet, and Falk Unger, **Sparse self-reducible sets and polynomial size circut lower bounds**, In *Proceedings of the 23rd International Symposium on Theoretical Aspects of Computer Science*, pages 455–468. Springer, 2006.
- H. Buhrman, M. Christandl, F. Unger, S. Wehner, and A. Winter, **Implications of superstrong nonlocality for cryptography**, *Proceedings of the Royal Society A*, 462(2071):1919–1932, 2006.

to my parents

Brita and Hans-Jürgen

Contents

Acknowledgments

This thesis is the culmination of four very enjoyable years at CWI during which I was a PhD student in the INS4 group. I would like to thank all people at CWI for making it the place it is. In particular, I would like to thank my advisor Harry Buhrman for his guidance and the faith he had in me from the start. We had many scientific discussions, but I am also grateful for his advice on non-technical scientific matters. Almost literally the same can be said about Ronald de Wolf, who shared an office with me for four years, and Ben Toner who joined our office in 2006. They were always happy to help me with all kinds of small and big issues which come up during a PhD student's day. I would like to thank them for a lot of fun in the most "social office" at INS4.

During my PhD I had the pleasure to work with many great people, who provided many interesting ideas and patiently explained many difficult things to me. Some of this work resulted in publications. I want to thank my coauthors Gilles Brassard, Harry Buhrman, Matthias Christandl, Richard Cleve, Julia Kempe, Monique Laurent, Noah Linden, Andre Methot, Oded Regev, Lex Schrijver, William Slofstra, Alain Tapp, Leen Torenvliet, Sarvagya Upadhyay, Stephanie Wehner, Andreas Winter and Ronald de Wolf for all the hard work they put into our joint work. I had many interesting scientific (and fortunately also non-scientific) conversations with other people, which have not resulted in publications. Among them were Scott Aaronson, Dorit Aharonov, Andris Ambainis, Manuel Ballester Sánchez, Hartwig Bosse, Jop Briët, Serge Fehr, Daniel Gottesman, Nebojša Gvozdenović, Peter Harremoës, Aram Harrow, Sophie Laplante, Troy Lee, Lasse Leskelä, Ashwin Nayak, Tobias Osborne, Krzysztof Pietrzak, Sandu Popescu, David Poulin, Daniel Preda, Ben Reichardt, Renato Renner, Nitin Saxena, Pranab Sen, Christian Schaffner, Leonard Schulman, Robert Špalek, Mario Szegedy, Ben Toner, John Tromp, Umesh Vazirani, Paul Vitányi, Shengyu Zhang and many other people who kindly shared their insights and expertise with me.

I also managed to have something like a social life. Robert, Troy and Ronald

I want to thank for being more than just friends to hang out with. I also had the fortune to live in five different places in Amsterdam which allowed me to stay with a sizeable number of flatmates: Nebojša, Dennis, Matthieu, Eike, Hartwig, Lasse, Keit and Sumanta. You and all other friends I thank for the nice dinners, parties, Salsa, wind-surfing, etc. and all the things traditionally a family provides you with. Last but not least I want to mention my "Swiss guards" Christian and Krzysztof, who will stand by me during my defense. All of you made these four years a great time.

I would probably not be writing this if I had not gotten continuous support from my family. I thank my parents Brita and Hans-Jürgen for their love and support and my brothers Claus, Ulf and Jörg.

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” 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 $10N$ 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[1]–, then arbitrarily long quantum computation is possible. The overhead of these schemes (the number of additional resources needed in the larger system) is relatively moderate and manageable. This means that if it is possible to build and operate on quantum systems with small enough errors it is possible to efficiently implement any (noise free) quantum algorithm. Thus, faults in quantum system are not an unsurmountable problem and the task of building large quantum computers becomes a "mere" engineering problem.

Lower bounds on quantum fault-tolerance threshold

Unfortunately, as many other engineering problems, this is not an easy one. Initial fault-tolerant schemes were proven to tolerate noise on the order of 10^{-6}, and have been substantially improved in the past decade. The best rigorous lower bounds on the fault-tolerance threshold—which we state without reference to the specific assumptions used—are on the order of 0.1% [7, 6, 5, 82], which is 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

[1] Of 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 ρ outputs ρ 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 ρ 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 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 $\mathbb{I}/d$ with probability p.

The proof in Chapter 3 is relatively simple, but nevertheless the best general[2] 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 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.

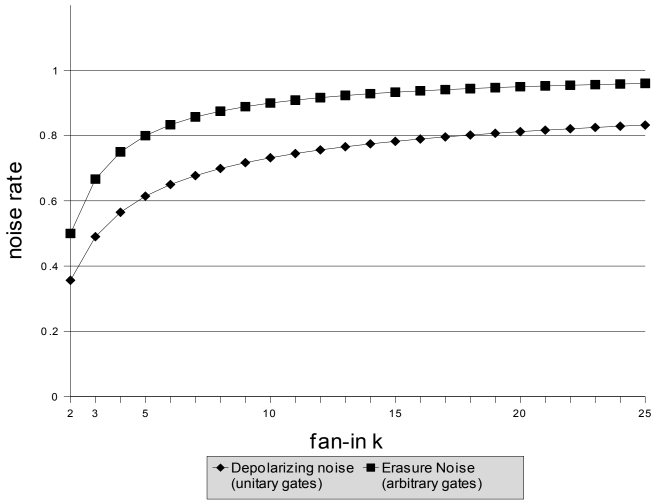
Of special interest from an experimental point of view is the case $k = 2$, for which our bound becomes about 35.7%. Furthermore, for the case in which the only allowed two-qubit gate is the controlled-NOT (CNOT) gate, we can improve our bound further to about 29.3%, as we show in Section 4.5. This case 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.[3] We believe that it is still a reasonable

[2] In the sense that gates may perform any physical operation; only the number of wires going into a gate is restricted.

[3] 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 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 [99] 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 $f : \{0,1\}^k \to \{0,1\}$. 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.

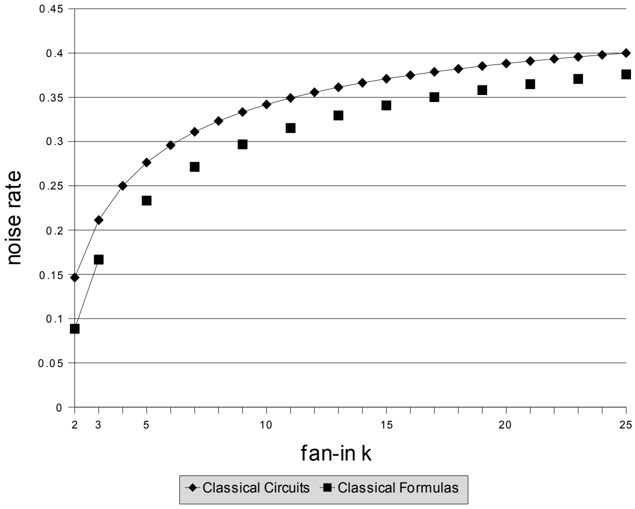
Note that this definition of noise corresponds to replacing the output bit with a uniformly random bit with probability $2p$ and leaving the output bit untouched with probability $1-2p$. Our definition of noise in the classical case is therefore somehow inconsistent with the definition of depolarizing 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].



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 $\beta_k := \frac{1}{2} - \frac{2^{k-2}}{k\binom{k-1}{k/2-1/2}}$ 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 β_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 β_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.

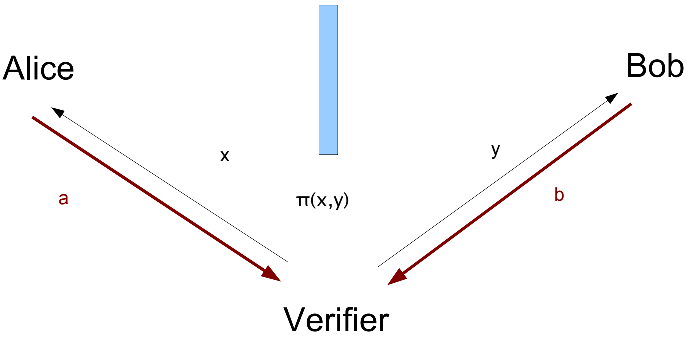
1.2 Entanglement and interactive proof systems

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.



Alice and Bob win if $x \cdot y = a \oplus b$

Figure 1.3: CHSH game

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.

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, altough 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 wether 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 protocols 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 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.[6]

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 $x \wedge y = a \oplus b$. 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

[6] Physicist say that otherwise our theory is not *causal*, see also page 117.

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 [96, 97] 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 $\frac{3+\sqrt{6}}{6} \approx 90.8\%$ 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* $\frac{3+\sqrt{6}}{6} \approx 90.8\%$ *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.

Chapter 2

Preliminaries

In this chapter we will discuss some general techniques and results which will be needed later. In later chapters we will provide pointers to this chapter whenever needed.

Section 2.1 is intended to set up our linear algebra notation, and in particular introduce the famous Dirac notation. Some more facts and notation can be found in Appendix A but it is very concise and mostly intended as a reminder. For an introduction to linear algebra the reader can consult for example [55].

In Section 2.2 we will summarize all prerequisites from quantum mechanics necessary to understand this thesis.

In Sections 2.3 and 2.4 we will give formal definitions of some notions in computational complexity theory and communication complexity theory, although an intuitive understanding of these concepts will be sufficient to read this thesis. Computational complexity theory will be uses to partially motivate the results in Chapter 7 and all the reader needs to know about communication complexity is that there are functions which have high communication complexity.

Section 2.5 from this chapter will summarize facts about the Bloch sphere, which will be a convenient tool in Chapters 4 and 5 for characterizing 1-qubit operations.

Section 2.6 discusses Semidefinite Programming, which will be needed in Chapter 7 to characterize XOR games.

2.1 Linear algebra notation

Hilbert space We denote the d-dimensional complex vector space by $\mathbb{C}^d$. It consists of all column vectors with d complex entries. For vectors elements $\phi, \psi \in$

$\mathbb{C}^d$

$$\phi = \begin{pmatrix} \phi_0 \\ \phi_1 \\ \vdots \\ \phi_{d-1} \end{pmatrix}, \quad \psi = \begin{pmatrix} \psi_0 \\ \psi_1 \\ \vdots \\ \psi_{d-1} \end{pmatrix}$$

we can define a complex inner product by

$$\langle \phi, \psi \rangle = \sum_{i=0}^{d-1} \phi_i^* \psi_i,$$

where ϕ_i^* is the complex conjugate of ϕ_i. This makes $\mathbb{C}^d$ a complex *Hilbert space*, since d is finite. From this inner product we define the *standard norm* (or *Euclidean norm*) as $||\phi|| = \sqrt{\langle \phi, \phi \rangle}$.

Dirac notation It will be convenient to write vectors $\phi \in \mathbb{C}^d$ in *Dirac notation* by sandwiching ϕ in between "$|$" and "$\rangle$" and write $|\phi\rangle$ instead of ϕ. We will denote the conjugate transpose of a vector $|\phi\rangle$ by $\langle\phi|$, which is the row vector with entries

$$\langle \phi | = \left(\phi_0^*, \phi_1^*, \ldots, \phi_{d-1}^* \right).$$

With this notation the inner product $\langle \phi, \psi \rangle$ can be simply written as the matrix product of $\langle\phi|$ and $|\psi\rangle$ as

$$\langle \phi, \psi \rangle = \langle \phi || \psi \rangle.$$

Standard basis It is easy to see that the vectors

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ \ldots \quad |d-1\rangle = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

are linearly independent and orthonormal (since $\langle i||j\rangle$ is 0 if $i \neq j$ and 1 otherwise) with respect to our inner product. We therefore call $\{|0\rangle, |1\rangle, \ldots |d-1\rangle\}$ the *standard basis* or *computational basis*.

Matrices We let $\mathbb{C}^{d\times d}$ be the d^2-dimensional complex vector space of all $d \times d$ matrices with complex entries. For a matrix $A \in \mathbb{C}^{d\times d}$ we let A_{ij} be the entry of A in the i-th row and j-th column. The *identity matrix* is denoted by $\mathbb{I}_d$. For a matrix $A \in \mathbb{C}^{d\times d}$ we let A^T be the transpose of A, which means that $(A^T)_{ij} = A_{ji}$. Similarly, $A^\dagger$ denotes the conjugate transpose and has entries $(A^\dagger)_{ij} = (A_{ji})^*$. We can also equip $\mathbb{C}^{d\times d}$ with an inner product by defining for $A, B \in \mathbb{C}^{d\times d}$

$$\langle A, B \rangle = \mathrm{Tr}(A^\dagger B) = \sum_{i,j=1}^{d} A_{ij}^* B_{ij}.$$

This inner product is called the *Hilbert-Schmidt inner product*.

E_{ij} is the all-zero matrix, except for the entry i, j which is equal to 1. Evidently, the E_{ij} form an orthonormal basis for $\mathbb{C}^{d\times d}$ with respect to our inner product. Later in Chapter 4 we will present a different orthonormal basis (tensor products of Pauli matrices), which is in certain applications more natural.

Positive semidefinite A hermitian matrix $A \in \mathbb{C}^{d\times d}$ is called *positive semidefinite* if for all $\psi \in \mathbb{C}^d$ it holds

$$\langle\psi|A|\psi\rangle \geq 0. \tag{2.1}$$

We also write $A \succeq 0$ for "A is positive semidefinite". We write $A \succeq B$ if $A - B \succeq 0$. Further, we write $A \succ 0$ if the above inequality is strict. We then say that A is positive definite.

2.2 Quantum states, operations and computation

Quantum states We model a closed physical system (of finite dimension d) abstractly by attaching to it the d-dimensional Hilbert space $\mathcal{H}_d = \mathbb{C}^d$. The set of all linear operators ρ mapping $\mathcal{H}_d$ into itself is written $\mathbb{B}(\mathcal{H}_d)$. It is isomorphic to $\mathbb{C}^{d\times d}$, the set of all $d \times d$-matrices. We assume that linear operators are always represented in matrix form. The set of possible *states* the system can be in is the set of all bounded linear operators $\rho \in \mathbb{B}(\mathcal{H})$ which are normalized (i.e. $\mathrm{Tr}(\rho) = 1$) and positive semidefinite $\rho \succeq 0$ (and therefore by definition also hermitian). Any such ρ we call a *density matrix*. For convenience we often do not make a distinction between the physical system and the Hilbert space associated to it and speak of "the physical system $\mathcal{H}_d$".

Measurements In order to gain some information about the state of a system one can perform a *measurement*. A measurement on system $\mathcal{H}$ is given by a set of measurement operators $\mathcal{M} = \{M_1, \ldots, M_m\} \subset \mathbb{B}(\mathcal{H})$, which have the property that $\sum_i M_i^\dagger M_i = \mathbb{I}_d$. The outcome of the measurement $\mathcal{M}$ is one of the labels $1, \ldots, m$. If the system is in state ρ before the measurement then outcome $1 \leq i \leq m$ occurs with probability

$$p_i = \mathrm{Tr}(M_i^\dagger M_i \rho)$$

and if outcome i occurred, the new state after the measurement is

$$\rho_i = \frac{M_i \rho M_i^\dagger}{\mathrm{Tr}(M_i^\dagger M_i \rho).}$$

Note that in general a measurement is a non-reversible operation, i.e. it is in general impossible to "undo" the measurement and obtain the premeasurement state.

General evolution The most general way to change the state of a quantum system is by applying a *completely positive trace-preserving map*, short *CPTP map*. They can change the dimensionality of a system, though we will be mostly concerned with operations that preserve dimensionality. A CPTP map $\mathcal{E}$ mapping a d-dimensional system into a d'-dimensional system is given by Kraus operators $\{E_1, \ldots E_m\} \subset \mathbb{C}^{d' \times d}$ with the property that

$$\sum_i E_i^\dagger E_i = \mathbb{I}.$$

Then $\mathcal{E}$ changes the density matrix ρ into

$$\mathcal{E}(\rho) = \sum_i E_i \rho E_i^\dagger.$$

A special case is when $\mathcal{E}$ is given by one Kraus operator $E_1 = U$ only, which necessarily has to be unitary. We then say that ρ evolves (under unitary evolution) into the state

$$\rho \rightarrow U \rho U^\dagger.$$

Obviously, *unitary operations* are reversible.

Pure states A state ρ is called *pure* if $rank(\rho) = 1$, which is equivalent to saying that there is some $\phi \in \mathbb{C}^d$ with the property that $\rho = |\phi\rangle\langle\phi|$. States which are not pure are called *mixed*. Often the short-hand $|\phi\rangle$ is used for $|\phi\rangle\langle\phi|$.

For every density matrix ρ it is possible to find normalized vectors $|\phi_i\rangle \in \mathbb{C}^d$ and non-negative real numbers p_i with $\sum_i p_i = 1$ with

$$\rho = \sum_{i=1}^{d} p_i |\phi_i\rangle\langle\phi_i|. \tag{2.2}$$

For example, since ρ is hermitian and positive semidefinite it can be unitarily diagonalized, and thus it is possible to choose p_i and $|\phi_i\rangle$ to be the eigenvalues respectively eigenvectors of ρ. We call $\{p_i, |\phi_i\rangle\langle\phi_i|\}$ *an ensemble* for ρ. However, in general the choice of an ensemble for a state is not unique.

It is straightforward to see that a quantum system state which is in a probabilistic mixture of states $|\phi_i\rangle\langle\phi_i| \in \mathbb{C}^{d \times d}$, each occurring with probability p_i, has the same measurement statistics as the state $\rho = \sum_{i=1}^{d} p_i |\phi_i\rangle\langle\phi_i|$ for any measurement. This means that it is impossible to distinguish the ensemble $\{p_i, |\phi_i\rangle\langle\phi_i|\}$ from the state ρ by observing the system. (Note that this remains true even after applying arbitrary quantum operations/measurements on the system.) Hence, the description of the state of a system in terms of ensembles and the description in terms of density matrices are absolutely equivalent from a physics point of view.

Note that the (pure) states $|\phi_i\rangle$ of an ensemble for ρ are elements of the d-dimensional Hilbert space $\mathbb{C}^d$ only. Since these d dimensions are enough to describe the pure states of the physical system and as we have seen we may interpret every state as a probabilistic mixture of pure states, the system is called d-dimensional. The reason for describing states in the larger Hilbert space of hermitian matrices in $\mathbb{C}^{d\times d}$ is that it is more suitable when using general CPTP maps as defined above.

Subsystems and entanglement If two physical systems are represented by Hilbert spaces $\mathcal{A} = \mathbb{C}^a$ respectively $\mathcal{B} = \mathbb{C}^b$, then the joint system is represented by the Hilbert space $\mathcal{A} \otimes \mathcal{B} := \mathbb{C}^{ab}$. Its inner product is defined as above. It holds $\mathbb{B}(\mathcal{A} \otimes \mathcal{B}) = \mathbb{B}(\mathcal{A}) \otimes \mathbb{B}(\mathcal{B})$.

The system $\mathcal{A} \otimes \mathcal{B}$ can be in a state $C \in \mathbb{B}(\mathcal{A}) \otimes \mathbb{B}(\mathcal{B})$ for which there are no $A_i \in \mathbb{B}(\mathcal{A})$, $B_i \in \mathbb{B}(\mathcal{B})$ such that $C = \sum_i A_i \otimes B_i$. We then say that C is *entangled*.

We say that a measurement $\mathcal{M} = \{M_1, \dots, M_m\} \subset \mathbb{B}(\mathcal{A} \otimes \mathcal{B})$ *acts only on subsystem A* if there are $M'_1, \dots, M'_m \subset \mathbb{B}(\mathcal{A})$ such that for all $1 \leq i \leq m$: $M_i = M'_i \otimes \mathbb{I}_B$, where $\mathbb{I}_B$ is the identity operator on $\mathcal{B}$. Similarly, we say that a CPTP map $\mathcal{E}$, given by Kraus operators $\{E_1, \dots E_m\} \subset \mathbb{B}(\mathcal{A} \otimes \mathcal{B})$, *acts on subsystem A only* if there are $E'_1, \dots, E'_m \subset \mathbb{B}(\mathcal{A})$ such that for all $0 \leq i \leq m$: $E_i = E'_i \otimes \mathbb{I}_B$.

Qubits We say that the Hilbert space $\mathcal{H}_2 = \mathbb{C}^2$ represents one *qubit* and generally $\mathcal{H}_d = \mathbb{C}^d$ represents one *qudit.* The space $\mathcal{H}_2^{\otimes n}$ represents n qubits.

2.2.1 Quantum circuits and quantum computation

We will only define quantum circuits on qubits, as this is all we will need later. It is straightforward to generalize this to arbitrary d-dimensional qudits.

Quantum circuits A *quantum circuit* on n qubits and of depth T consists of a set of *(quantum) gates* $\mathcal{G}_i$, each of which contains 3 parameters: $W_i \subseteq \{1, \dots, n\}$ (the qubits $\mathcal{G}_i$ acts on), an integer number $1 \leq t_i \leq T$ (the execution time) and a quantum operation $\mathcal{E}_i$ which acts on $|W_i|$ qubits. Further, we require that for each $1 \leq t \leq T$ and every $1 \leq j \leq n$ there is exactly one i with $j \in W_i$ and $t_i = t$. In other words for each qubit and each time there is exactly one gate which acts upon this qubit (which may be the identity gate). The number of qubits $|W_i|$ is called the *fan-in* of gate $\mathcal{G}_i$.

Evolution We fix a unique total order $\prec$ on the set of gates by defining that

$$\mathcal{G}_i \prec \mathcal{G}_j \qquad \longleftrightarrow \qquad (t_i < t_j) \text{ or } (t_i = t_j \text{ and } \min W_i < \min W_j).$$

Without loss of generality we let the gates be numbered from 1 to S and we assume that this numbering is consistent with the ordering $\prec$, i.e.,

$$\mathcal{G}_i \prec \mathcal{G}_j \qquad \longleftrightarrow \qquad i < j.$$

The computation of a quantum circuit on input $\rho \in \mathcal{H}_2^{\otimes n}$ is inductively defined through the following set of quantum states ρ_i

1. $\rho_0 = \rho$

2. If $i < S$ then $\rho_{i+1} = \mathcal{E}'_{i+1}(\rho_i)$.

If $\mathcal{E}_i$ has Kraus operators E_j, we let $\mathcal{E}'_i$ be the quantum operation with Kraus operators $E_j \otimes \mathbb{I}^{\otimes\{1,\ldots,n\}\backslash W_i}$, where the E_j act on the Hilbert space of the qubits W_i and $\mathbb{I}^{\otimes\{1,\ldots,n\}\backslash W_i}$ is the identity matrix on the Hilbert space of the qubits $\{1,\ldots,n\}\backslash W_i$. We define the *output of the computation* to be ρ_S.

If $x \in \{0,1\}^n$ we denote by $\rho_x \in \mathbb{C}^{2^n \times 2^n}$ the output of the computation with input $|bin(x)\rangle$, where $bin(x)$ is the number with binary representation x and $|bin(x)\rangle$ is short for $|bin(x)\rangle\langle bin(x)|$, as defined above. We say that ρ_x is the *output of the computation with classical input x.*

Computation of a function Often it is necessary to give quantum circuits additional work space, e.g. to "store" intermediate results. To take care of that we pad inputs with additional qubits in state $|0\rangle$, which are input-independent. These qubits are called *ancilla qubits.* The precise definition is as follows.

We say that a quantum circuit on m qubits computes a boolean function $f : \{0,1\}^n \mapsto \{0,1\}$, $n \leq m$, with error $0 \leq \epsilon < \frac{1}{2}$ using $m-n$ ancillas, if there is some measurement $\mathcal{M} = \{M_0, M_1\}$ such that for every $x \in \{0,1\}^n$ the outcome of $\mathcal{M}$ on $\rho_{x0^{n-m}}$ is equal to $f(x)$ with probability at least $1-\epsilon$. Formally,

$$\forall_{x\in\{0,1\}^m} : \mathrm{Tr}(M_{f(x)}^{\dagger} M_{f(x)} \rho_{x0^{n-m}}) \geq 1-\epsilon.$$

We say that $\mathcal{M}$ *measures the first qubit* if $M_i = M'_i \otimes \mathbb{I}_{2^{m-1}}$, for some measurement operators $\{M'_0, M'_1\}$ on the first qubit.

Efficient quantum computation A family of boolean functions $f_1, f_2, f_3, \ldots$ with $f_n : \{0,1\}^n \mapsto \{0,1\}$ is said to be *uniformly computable in polynomial-time with bounded error by a quantum computer*, if there is some polynomial $p(\cdot)$, $0 \leq \epsilon < 1/2$, $k > 0$ and a (classical) Turing machine M with the property that

1. For each n the Turing machine $M(n)$ runs in time at most $p(n)$.

2. The output $M(n)$ is a description of a quantum circuit C_n which has m_n qubits and each gate has fan-in at most k.

3. C_n computes f_n with error at most ϵ using $m_n - n$ ancillas and a measurement on the first qubit.

Note that $n \leq m_n \leq p(n)$ and that the depth of C_n is at most $p(n)$. The output $M(i)$ (which is the description of C_i) should be a list of numbers specifying for each gate: W_i, t_i and numbers specifying the Kraus operators. Note that since the running time of $M(n)$ is at most polynomial in n, the number of gates and the number of ancilla qubits are also polynomially bounded.

At first sight it might look that changing the parameters ϵ, k or the type of gates might change the computational power (i.e. set of functions that can be computed). However, this is not the case. Firstly, it is not hard to show [68] that every family of functions which can be computed with error $0 \leq \epsilon < 1/2$, can also be computed with arbitrarily small error $0 < \epsilon'$, with small overhead in the size of the circuit. Secondly, it turns out that there is a finite set of unitary gates $\mathcal{U}$ (with fan-in 2) such that every family of functions which can be efficiently computed in the above way can also be efficiently computed with circuits using only gates from $\mathcal{U}$. However, the polynomial $p(\cdot)$ might change. For example it is possible to choose $\mathcal{U}$ to be the CNOT gate, the Hadamard gate and the $\pi/8$ gate [18], see (5.2) on page 63 and (5.19) on page 74 for definitions of this gate set.

This makes it possible to define the complexity class BQP without reference to a particular gate set and to fix $0 < \epsilon < 1/2$ arbitrarily, see Section 2.3 later.

2.3 Complexity classes

One—or perhaps *the most*—important part of computer science is to determine how difficult computational problems are, in terms of the number of resources (e.g. time) needed to solve the problem. A thorough introduction can be found in [71].

Computational problems are usually phrased as membership problems (with binary yes/no answers), because every computational problem can be reduced to membership problems. A membership problem is modeled by a set $A \subseteq \{0,1\}^*$, a subset of all 0/1-strings. The task is to determine whether a given $x \in \{0,1\}^*$ is in A or not.

We say that the problem $A \subseteq \{0,1\}^*$ is *polynomial-time decidable* if there is a Turing machine M and a polynomial p such that for each input $x \in \{0,1\}^*$ the machine M stops after at most $p(|x|)$ steps and outputs whether $x \in A$ or not. Here $|x|$ is the number of bits of x. The class P consists of all problems A which are polynomial-time decidable. It contains interesting problems, like (the decision version of) Linear Programming, finding the shortest path between two nodes in a graph, deciding whether a number is prime or not and many others.

Another important complexity class is NP, which we define via projection. A problem $A \subseteq \{0,1\}^*$ is in NP if there is some set $B \in \mathrm{P}$ and a $k \in \mathbb{N}$ such that

$$x \in A \quad \longleftrightarrow \quad \exists_{y \in \{0,1\}^{|x|^k}} : xy \in B,$$

where xy is the concatenation of x and y. We say that the problems in NP can be "accepted by a non-deterministic polynomial-time Turing machine".

Changing "polynomial time" into "exponential time" in the above definitions gives the classes EXP respectively NEXP, which are the classes of problems which can be decided/accepted by a deterministic/non-deterministic Turing machine in exponential time. If in the definition of P we do not require that M stops after $p(|x|)$ steps, but only demand that M uses at most $p(|x|)$ many cells of the tape of M, then we get the definition of PSPACE.

The following inclusions are obvious

$$\mathrm{P} \subseteq \mathrm{NP} \subseteq \mathrm{PSPACE} \subseteq \mathrm{EXP} \subseteq \mathrm{NEXP}.$$

All inclusion are conjectured to be strict, but currently there is no proof for any of them. We do know however, that $\mathrm{P} \subset \mathrm{EXP}$ and $\mathrm{NP} \subset \mathrm{NEXP}$, so at least one of the inclusions must be strict.

The class BPP is the class of all problems which can be decided by a probabilistic polynomial-time Turing machine which always outputs the correct result with probability at least 2/3. More formally, a problem $A \subseteq \{0,1\}^*$ is in BPP if there is some set $B \subseteq \{0,1\}^*$ which can be decided in polynomial time and a polynomial $q(\cdot)$ such that

$$\begin{aligned} x \in A \quad &\rightarrow \quad \exists_{S \subseteq \{0,1\}^{q(|x|)}, |S| \geq \frac{2}{3} 2^{q(|x|)}} \forall_{y \in S} : xy \in B \\ x \notin A \quad &\rightarrow \quad \exists_{S \subseteq \{0,1\}^{q(|x|)}, |S| \geq \frac{2}{3} 2^{q(|x|)}} \forall_{y \in S} : xy \notin B. \end{aligned}$$

The choice of the constant 2/3 is not unique. It can be shown that any constant $1/2 < c < 1$ yields the same class. Clearly,

$$\mathrm{P} \subseteq \mathrm{BPP} \subseteq \mathrm{PSPACE}.$$

The first inclusion is conjectured to be an equality [69].

We now define the class BQP [106], the equivalent of BPP for quantum computers. A set $A \subseteq \{0,1\}^*$ is in BQP if its characteristic function

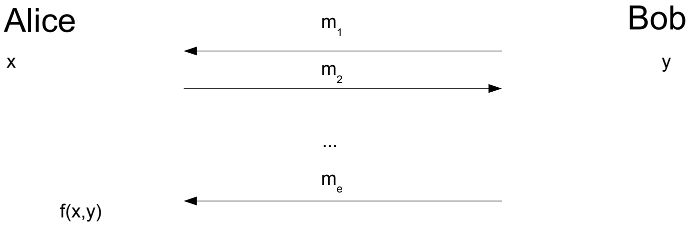
$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

is uniformly computable in polynomial-time with error $\epsilon = 1/3$ by a quantum computer (see end of Section 2.2). The following inclusions are known

$$\mathrm{BPP} \subseteq \mathrm{BQP} \subseteq \mathrm{PSPACE}.$$

2.4 Communication complexity

We briefly review the field of *communication complexity* [63, 28, 103, 19]. Assume Alice and Bob wish to compute some Boolean function $f(x, y)$ of input $x \in \{0,1\}^n$, known to Alice only, and input $y \in \{0,1\}^n$, known to Bob only. The aim is that Alice learns the value $f(x, y)$. To this end they exchange messages, using as little communication as necessary.



Alice and Bob send messages $m_1, \ldots, m_e$ back and forth and Alice outputs $f(x, y)$

Figure 2.1: A communication protocol

Deterministic communication complexity The *deterministic communication complexity* $D(f)$ of $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is the smallest number c such that any protocol which always computes the correct result, needs at least c bits of communication for at least one input pair $x, y \in \{0,1\}^n$. It is clear that this task cannot be accomplished in general without at least *some* communication, unless $f(x, y)$ does not actually depend on y. In fact, it is well-known that there are functions f where $D(f)$ is n. An example is the *inner product* function (see [63]), which is defined as $IP_n(x, y) = \bigoplus_{i=1}^{n}(x_i \wedge y_i)$.

Randomized communication complexity There is a *probabilistic* version of communication complexity, in which there is a source of random bits, to which Alice and Bob both have access. This is called the *public coin* model [105] because

Alice and Bob both see the same random bits.[1] In this model, Alice is not required to learn the value of $f(x,y)$ with certainty. Instead, we shall be satisfied if she can obtain an answer that is correct with probability bounded away from $\frac{1}{2}$. In other words, there must exist some real number $p > \frac{1}{2}$ such that the probability that Alice outputs the correct value $f(x,y)$ is at least p for *all* pairs (x,y) of inputs. The probability is taken over the value of random variables shared between them. The *error probability* of a protocol is defined as $\epsilon = 1 - p$ and we define the *randomized communication complexity* $R_\epsilon(f)$ to be the minimum number of bits needed such that the output is correct with probability at least $1 - \epsilon$. Also in this randomized setting there are "hard" functions. For example, it is known that the inner product function has randomized communication complexity $n - O(\log(1/\epsilon))$, if the outputs have to be correct with probability at least $1 - \epsilon$ (see also [63]).

Entanglement-assisted communication complexity There is another version of communication complexity, introduced in [28], in which Alice and Bob may share entanglement. More precisely, Alice and Bob may hold an arbitrary (entangled) quantum state shared between them. Particularly useful is often an arbitrary amount of *EPR pairs*

$$|\phi^+\rangle = \frac{|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B}{\sqrt{2}},$$

where Alice has access to the qubit(s) with subscript A and Bob has access to the qubit(s) with subscript B. We denote by $R^*_\epsilon(f)$ the number of classical bits needed to communicate in order for one party to compute the correct result.

Shared entanglement between Alice and Bob helps sometimes but not always. Some functions can be computed with exponentially less communication than with a purely classical protocol [23]. However, for other functions, for example the inner product function, the communication cannot be reduced significantly. Alice and Bob cannot succeed with probability $1 - \epsilon > \frac{1}{2}$ if they transmit less than $max(\frac{1}{2}(1-2\epsilon)^2, (1-2\epsilon)^4)n - \frac{1}{2}$ bits, even if they share prior entanglement [29]. In Appendix C we will give a better bound of

$$R^*_\epsilon(IP_n) \geq n - 2\log_2 \tfrac{1}{1-2\epsilon}$$

using a reduction by [104] to a lower bound for the number of qubits needed to communicate with shared entanglement [65].

Worst-case partition communication complexity For functions of the form $f : \{0,1\}^n \rightarrow \{0,1\}$ which depend only on one input string and any $S \subseteq$

[1] There is another model, called *private coin model*, in which Alice and Bob do not have access to the same random bits. But we will not need this model subsequently.

$\{1, \ldots, n\}$ let $D^S(f)$ be the deterministic communication complexity of f if the bits with indices in S are given to Alice and all others to Bob. As in [63] we then define the *worst-case partition communication complexity* as $D^{worst}(f) = \max_{S \subseteq \{1,\ldots,n\}} D^S(f)$. In [91] this is called symmetric communication complexity. The *randomized worst-case partition communication complexity* is defined analogously by replacing "D^{worst}" with "R^{worst}_ϵ" and providing Alice and Bob with shared random bits.

2.5 Bloch sphere

In later chapters it will be convenient to use the *Bloch sphere* representation of 1-qubit states and 1-qubit operations, which we review now (see e.g. Section 4.2 and Chapter 8 in [68]).

2.5.1 Pauli matrices

The following *Pauli matrices* are the basis of the Bloch vector representation

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.3}$$

It is immediately obvious that they span the whole space $\mathbb{C}^{2\times 2}$ and are therefore a basis. They are hermitian and self-inverse. They obey the following commutation relations:

$$\begin{array}{rclcrcl} XY & = & iZ & \qquad & YX & = & -iZ \\ ZX & = & iY & & XZ & = & -iY \\ YZ & = & iX & & ZY & = & -iX \end{array}$$

2.5.2 Bloch-vector representation

For $\mathbf{r} \in \mathbb{R}^3$ define $\mathbf{r} \cdot \sigma = r_x X + r_y Y + r_z Z$, where $\sigma = (X, Y, Z)$ is the vector of Pauli matrices. Then, all 1-qubit density matrices ρ can be uniquely written in the form

$$\rho = \frac{\mathbb{I}_2 + \mathbf{r} \cdot \sigma}{2} = \frac{\mathbb{I}_2 + r_x X + r_y Y + r_z Z}{2},$$

where $\mathbf{r} \in \mathbb{R}^3$ and $||\mathbf{r}|| = \sqrt{r_x^2 + r_y^2 + r_z^2} \leq 1$. We call $\mathbf{r}$ the *Bloch vector* of ρ.

For $\mathbf{n} \in \mathbb{R}^3$ with $||\mathbf{n}|| = 1$ and $\theta \in \mathbb{R}$ we define

$$U_{\mathbf{n}}(\theta) = \exp\left(\frac{-i\theta \mathbf{n} \cdot \sigma}{2}\right) = \mathbb{I}_2 \cos\frac{\theta}{2} - i\mathbf{n} \cdot \sigma \sin\frac{\theta}{2}.$$

We first note that $U_{\mathbf{n}}(\theta) U_{\mathbf{n}}(\theta)^* = \mathbb{I}$, i.e., $U_{\mathbf{n}}(\theta)$ is unitary. Second, let the result of the unitary quantum operation $U_{\mathbf{n}}(\theta)$ applied to state $\rho = \mathbb{I}/2 + \mathbf{r} \cdot \sigma/2$ be

$\rho' = U_{\mathbf{n}}(\theta)^* \rho U_{\mathbf{n}}(\theta) = \mathbb{I}/2 + \mathbf{r}' \cdot \sigma/2$. Then $\mathbf{r}'$ is the image of rotating $\mathbf{r}$ around $\mathbf{n}$ by an angle θ. Third, all 1-qubit unitaries U can be written as

$$U = U_{\mathbf{n}}(\theta)$$

with $\mathbf{n} \in \mathbb{R}^3, \theta \in \mathbb{R}$ and $||\mathbf{n}|| = 1$ (ignoring an unimportant phase factor $\alpha \in \mathbb{C}$ with $|\alpha| = 1$).

Thus, one-qubit states and unitaries are isomorphic to vectors, respectively, rotations in $\mathbb{R}^3$. The set of all rotations in $\mathbb{R}^3$ is the group $SO(3)$.

Arbitrary quantum operations For non-unitary one-qubit quantum operations the picture is a little bit more complicated. We present a characterization of trace-preserving completely-positive maps on one-qubit operations due to Ruskai, Szarek, and Werner [84, Sections 1.2 and 1.3]. They show that any one-qubit CPTP map G can be written as a convex combination of gates of the form $U \circ J \circ V$, where U and V are one-qubit unitary operations (acting on the density matrix by conjugation with some unitary $U, V \in \mathbb{C}^{2\times 2}$), and J is one-qubit map that in the Pauli basis has the form

$$J = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & \lambda_2 & 0 \\ t & 0 & 0 & \lambda_1\lambda_2 \end{pmatrix} \tag{2.4}$$

for some $\lambda_1, \lambda_2 \in [-1, 1]$ and $t = \pm\sqrt{(1-\lambda_1^2)(1-\lambda_2^2)}$. In other words, if a one-qubit state ρ has Bloch vector $\mathbf{r}$, then the Bloch vector $\mathbf{r}'$ of $\rho' = J(\rho)$ is given by

$$\begin{pmatrix} 1 \\ \mathbf{r}' \end{pmatrix} = J \begin{pmatrix} 1 \\ \mathbf{r} \end{pmatrix}.$$

2.6 Semidefinite programming

In Chapter 7 we will use semidefinite programming methods as a tool to characterize quantum XOR games. We briefly review the most important facts here, closely following [17]. A *semidefinite program* (for short SDP) is an optimization problem of the form

$$\begin{array}{rlll} p^* = & \sup & \mathrm{Tr}(CX) & \\ & \text{subject to} & \mathrm{Tr}(A_i X) = b_i, & i = 1, \ldots, m \\ & & X \succeq 0 & \end{array} , \tag{2.5}$$

where $A_1, \ldots, A_m, C \in \mathbb{C}^{d\times d}$ are given hermitian matrices and the constraint $X \succeq 0$ means that $X \in \mathbb{C}^{d\times d}$ is a positive semidefinite hermitian matrix, i.e. $X = X^\dagger$ and for all $|v\rangle \in \mathbb{C}^d$: $\langle v|X|v\rangle \geq 0$. Note that with these constraints

$\forall_i : \mathrm{Tr}(A_iX) \in \mathbb{R}$ and $\mathrm{Tr}(CX) \in \mathbb{R}$. We call (2.5) an *SDP in standard primal form*.

In Chapter 7 we will express the optimal winning probability of a quantum XOR game as the optimal solution p^* of an SDP. Lower bounds on the optimal value of (2.5) can be shown by finding an $\hat{X}$ satisfying the constraints. Then it is clear that the optimal value of (2.5) is at least $\mathrm{Tr}(C\hat{X})$.

However, we also want to be able to give upper bounds on p^*. One way to find upper bounds is by duality. To this end, we first write down the *Lagrangian* of (2.5) in terms of the objective function and the constraints as

$$L(X, y, \Lambda) = \mathrm{Tr}(CX) + \sum_i y_i \left(\mathrm{Tr}(A_iX) - b_i\right) + \mathrm{Tr}(\Lambda X), \tag{2.6}$$

where $y \in \mathbb{R}^m$ and $\Lambda \in \mathbb{C}^{d\times d}$ are called *Lagrange multipliers*. Now, for any $y \in \mathbb{R}^m$ and any hermitian $\Lambda \in \mathbb{C}^{d\times d}$ with $\Lambda \succeq 0$ the value of p^* can be upper bounded by

$$p^* \leq \sup_X L(X, y, \Lambda), \tag{2.7}$$

because every solution $\hat{X}$ for (2.5) satisfies $\mathrm{Tr}(A_i\hat{X}) - b_i = 0$ and $\mathrm{Tr}(\Lambda\hat{X}) \geq 0$. Note that there are no constraints on X, other than $X \in \mathbb{C}^{d\times d}$ and $X = X^\dagger$. Hence, if we can find y, Λ such that the right hand side of (2.7) is small we can also give good upper bounds on p^*. Let us write this out more explicitly

$$\begin{aligned}
\sup_X L(X, y, \Lambda) &= \sup_X \mathrm{Tr}(CX) + \sum_i y_i \left(\mathrm{Tr}(A_iX) - b_i\right) + \mathrm{Tr}(\Lambda X) \\
&= \sup_X \sum_i y_ib_i + \mathrm{Tr}\Big(\big(\sum_i y_iA_i + C + \Lambda\big)X\Big) \\
&= \begin{cases} \sum_i y_ib_i & \text{if} \quad \sum_i y_iA_i + C + \Lambda = 0 \\ \infty & \text{if} \quad \sum_i y_iA_i + C + \Lambda \neq 0. \end{cases}
\end{aligned}$$

Obviously, we are not interested in the trivial upper bound $p^* \leq \infty$. So, to get good upper bounds we have to find y, $\Lambda \succeq 0$ with $\sum_i y_iA_i + C + \Lambda = 0$. Since $\Lambda \succeq 0$, we need to require $\sum_i y_iA_i + C \preceq 0$ and we get the following *dual problem*

$$\begin{array}{rll} d^* = & \inf_{y_1,\ldots,y_m\in\mathbb{R}} & \sum_i y_ib_i \\ & \text{subject to} & \sum_i y_iA_i + C \preceq 0 \end{array} \tag{2.8}$$

From our derivations it immediately follows that

$$p^* \leq d^*$$

a property known as *weak duality*. If this inequality is tight, i.e. $p^* = d^*$, we speak of *strong duality*. There are many results about when strong duality holds [17]. The most important one is *Slater's condition*. It states that if (2.5) is *strictly feasible*, which in our case means that some $X \succ 0$ satisfies the constraints of (2.5), then strong duality always holds. This is the criterion which we will also use later.

Real SDPs versus complex SDPs Most literature deals only with real SDPs, which means that all (entries of all) variables are real numbers. We have given a direct derivation for complex SDPs, because it is more natural for our application in Chapter 7. The complex SDPs we are using can be transformed into real SDPs, using that for any hermitian $X \in \mathbb{C}^{d\times d}$ it holds

$$X \succeq 0 \quad \longleftrightarrow \quad \forall a, b \in \mathbb{R}^d : (a^T, b^T) \underbrace{\begin{pmatrix} Re(X) & -Im(X) \\ Im(X) & Re(X) \end{pmatrix}}_{=:\tilde{X}} \begin{pmatrix} a \\ b \end{pmatrix} \geq 0,$$

where $Re(X)$ is the real part and $Im(X)$ the imaginary part of X. Note that $X \in \mathbb{C}^{d\times d}$ is hermitian if and only if $Re(X) = Re(X^T)$ and $Im(X) = -Im(X^T)$, which is the case if and only if $\tilde{X}$ is symmetric. See also Exercise 4.42 in [17] for a more thorough discussion.

Part I

Limits on noisy quantum and classical computation

Chapter 3

Erasure noise

The results in this chapter are based on an unpublished manuscript

> Falk Unger, **Erasure noise threshold for fault-tolerant computation**, unpublished

It extends and simplifies a result in

> A. Razborov. **An upper bound on the threshold quantum decoherence rate**, *Quantum Information and Computation*, 4(3):222–228, 2004

In this chapter we will present a first upper bound on the noise tolerable for fault-tolerant classical as well as quantum computing. Despite its simplicity, it is currently the best upper bound on storage noise if there are no restrictions on the allowed quantum gates other than their fan-in.

We will consider circuits in which after each time-step, every qubit is "erased" with a certain probability p and prove an upper bound of $1 - 1/k$ on p for fault-tolerant computation, where k is the maximal fan-in[1] of the gates allowed. The result is very general, since all quantum operations with restricted fan-in are allowed (in particular also classical operations).

We prove that for long enough computations it is impossible to distinguish any two input states reliably if the noise is more than $1 - 1/k$. In particular, above this noise rate quantum circuits with single qubit measurements become "useless" after a constant amount of time and polynomial-size circuits (but with arbitrary measurements) become "useless" after a logarithmic amount of time.

Surprisingly, this is tight since for smaller noise rates it is possible to construct circuits of arbitrary depth such that it is possible to distinguish certain input states. However, it is not clear whether it is possible to simulate any quantum circuit efficiently, if the noise is less than $1 - 1/k$. See Section 3.4.

[1] Recall that the fan-in of a gate is the number of its input wires.

3.1 Erasure vs. depolarizing noise

As announced in the introduction of this thesis, our noise bound will apply to *storage noise*, which means that we assume that all gates used are perfect, and after each time-step noise happens on each qubit independently. We will assume that gates can be executed perfectly, i.e. there is no *gate noise*. Further, there are different types of noise which could be applied to stored qubits, for example erasure noise and depolarizing noise. We will first define both and then argue that the upper bound for erasure noise, which we prove later in this chapter, also applies to depolarizing noise.

Erasure noise of rate p is an operation that takes one qubit ρ as input and outputs one qubit *and* one classical bit

$$\rho \mapsto (1-p)|0\rangle\langle 0| \otimes \rho \;+\; p|1\rangle\langle 1| \otimes \psi_0.$$

The classical bit indicates whether an error occurred or not. It is set to 1 with probability p, otherwise 0. If it is equal to 1, then the gate replaces the input qubit by a fixed qubit state ψ_0. If it is 0 it applies the identity to its quantum wire and the output qubit is the same as the input qubit. The exact specification of ψ_0 is unimportant in our case (and most other cases) and can be an arbitrary mixed one-qubit state. The error-indicating bit is sent to the experimenter running the circuit, who can react to erasure errors by letting later quantum gates depend on which erasure errors have happened so far. The precise definitions of our model can be found in the next section.

This error model is often studied and a good model for many kinds of errors happening in quantum mechanical systems. For example a qubit might be stored by an ion, where the lowest energy state represents a $|0\rangle$ and the second lowest a $|1\rangle$. Due to noise processes this ion can shift to some higher excited state with some probability. It is still possible to detect whether the ion is in some higher excited state, in which case we say that the information is lost, or "erased". This is possible without disturbing the state in case no error is detected, i.e. if the qubit is in $|0\rangle$ or $|1\rangle$ (or a superposition thereof). See [49] for more examples.

Recall that for a qubit ρ *depolarizing noise* with probability p is a quantum operation which applies the identity operation with probability $1-p$ and replaces the qubit by the completely mixed state $\mathbb{I}_2/2$ with probability p

$$\rho \mapsto (1-p)\rho + p\mathbb{I}_2/2. \tag{3.1}$$

In the case of erasure noise it is inessential for most applications (including the following noise bound for erasure noise) what ψ_0 exactly is, because in case an error is detected, one can just replace the qubit by the desired state ψ_0. It is immediately clear that erasure noise at rate p is less serious than depolarizing noise at the same rate, because in the former case we are informed when noise happens and could just replace the wire by a completely mixed qubit $\mathbb{I}/2$. Therefore, the upper bound we prove in Theorem 3.3.1 immediately also applies to depolarizing noise.

3.2 Circuit model

We start by explaining the model informally and give formal definitions later. As customary for quantum circuits, we consider *synchronized parallel* circuits, as in Section 2.2. In a synchronized quantum circuit of depth T quantum operations can only happen at discrete *time-steps* $t \in \{1, \ldots, T\}$. The only thing we restrict is the number of qubits $k > 1$ which go into any gate. Otherwise we allow any quantum operation, i.e., completely positive trace-preserving map. In our model time proceeds in discrete time-steps and in each time-step any number of gates can be executed, as long as they act on disjoint sets of wires. We assume that gates can be executed perfectly. After every time-step erasure noise happens on each qubit, the rate of which will determine whether fault-tolerant computation is possible or not.

We further allow *weak classical control*: The operation a gate performs may arbitrarily depend on the classical bits on earlier wires that indicate erasure. But the qubits a gate acts on and the time of its execution are fixed from the start. We call this kind of control *weak* because the most general kind of *classical control*, also allows to choose *on which qubits a gate acts* depending on previous erasure errors and measurements on earlier wires. This stronger kind of classical control is also called *perfect classical control.*

To incorporate erasure noise and weak classical control into our circuit model we have to expand the definitions of circuits from Section 2.2.

Recall from Section 2.2 that a quantum circuit is given by a set of gates and each gate contains a description of which qubits it acts on, the time it is executed and its quantum operation (in terms of Kraus operators).

3.2.1. Definition. [Skeleton graph] A description of a circuit C without the specification of the quantum operations of the gates (the Kraus operators) is called the *topology* of the circuit. From the topology of a circuit we can define its *skeleton graph* S: The circuit graph of a quantum circuit is a graph with $T+1$ parts $1, \ldots, T+1$. The t-th part, $1 \leq t \leq T$, is called G_t and contains the gates at time t. There are only directed edges from G_t to G_{t+1} for any t. For every gate $\mathcal{G}_i \in G_t$ and every gate $\mathcal{G}_j \in G_{t+1}$, operating on qubits W_i respectively W_j, the number of edges between $\mathcal{G}_i$ and $\mathcal{G}_j$ is equal to $|W_i \cap W_j|$, which is the number of output qubits of $\mathcal{G}_i$ which are input qubits of $\mathcal{G}_j$. Such an edge is called a *wire* at time t. The set of all wires at time t is denoted by V_t. The nodes in the first subset G_1 have no incoming edges and correspond to the gates executed in the first time step. Nodes in G_{T+1} have no outgoing edges and represent the output qubits of the circuit after its T computation steps.

When we restrict the gates to fan-in at most k, then every node in the graph has in-degree (and out-degree) at most k.

From the skeleton graph we can define erasure patterns.

3.2.2. Definition. [Error pattern] An error pattern E for a circuit C is a subset of wires of the skeleton graph. If a wire w is in E than we say w that has been *erased*. A particular wire $w \in V_t$ is called *connected* (to the input), if there is a path from the input to w which does not contain any edges in E. In particular, the wire itself must not be erased. We write $conn(w, E)$.

3.2.3. Definition. [Weak classical control] Let M be a map which associates to each gate $\mathcal{G}$ at time $1 \leq t \leq T$ and each error pattern E a quantum operation (which has to act on the same number of qubits as the gate)

$$M(\mathcal{G}, E) \mapsto \hat{\mathcal{E}}_{\mathcal{G},E},$$

where $\hat{\mathcal{E}}_{\mathcal{G},E}$ is given in terms of its Kraus operators. We call $\hat{\mathcal{E}}$ the *intended* or *noise-free* operation. *The* quantum operation of a gate on k qubits is then defined as the operation

$$\mathcal{E}_{\mathcal{G},E} = \hat{\mathcal{E}}_{\mathcal{G},E} \circ \mathcal{N},$$

where $\mathcal{N}$ is the quantum operation that replaces those outputs qubits of $\mathcal{G}$ that are in E (i.e. are erased) by ψ_0.

Further, to each error pattern E our map M associates a two-valued measurement with operators $\{M_{E,0}, M_{E,1}\}$.

Note that via M every error pattern E defines a quantum circuit in the usual sense (as in Section 2.2), which we call C_E. In Definition 3.2.3 we also notice that the Kraus operators which M associates to $\mathcal{G}$ may depend on whether erasures happen on later wires. This is of course physically unreasonable, but since we are proving an upper bound on the tolerable noise, this is not a problem. For the lower bound on the threshold we will give a "physically reasonable" definition in Definition 3.3.4.

3.2.4. Definition. [Erasure noise] With each wire w we associate a probability p_w, which is its probability of being erased. Let $q(E)$ be the probability distribution over error patterns in which each wire w is erased independently of the others with probability p_w.

3.2.5. Definition. [Bias] Let C be a quantum circuit with weak classical control. For each error pattern E of C let ρ_E be the output of C_E on input ρ. The bias of C on the two input states ρ_0 and ρ_1 is

$$\sum_E q(E) \left| \mathrm{Tr}(M_{E,1}^\dagger M_{E,1}(\rho_{0,E} - \rho_{1,E})) \right| .$$

In Theorem 3.3.1 we show that if the rate of erasure noise is higher than $1 - 1/k$ then for any bias δ (a) no measurement on a single qubit can distinguish any two input states with bias δ after some constant amount of time and (b) no measurement on a polynomial number of qubits can distinguish any two input states after some logarithmic amount of time. The first result implies that only for a constant number of functions there is a circuit, that outputs the value of the function on one output qubit with bounded error. The second result implies that functions which need super-logarithmic depth to compute on a quantum computer, cannot be computed if the noise rates are too high. The last result holds even if the final measurement can be arbitrary and arbitrary classical post-processing is allowed.

3.3 Noise threshold

The following theorem was first proved for depolarizing noise by Razborov [79], using a different technique. Our technique extends to erasure noise. Further, we want to point out that a similar argument was already used by Feder [39] for classical noise bounds, but without making the connection to erasure noise.

3.3.1. Theorem. *Consider circuits C with weak classical control, which use arbitrary gates of fan-in at most k. Assume that there is an $\epsilon > 0$ such that on each wire w erasure noise happens with probability at least $p_w \geq 1 - 1/k + \epsilon$, independently of erasures on other wires. Let $\delta > 0$ be the desired output bias and let q be the number of output qubits to be measured. Then there is some $T \in O(\log \frac{q}{\delta})$ such that for any two input states ρ_0 and ρ_1 and for any circuit C that takes at least T steps to compute and uses an arbitrary q-qubit measurement, the bias of C is at most δ.*

In particular this implies that already after a constant amount of time quantum circuits with one final one-qubit measurement are "useless". Further, if we are interested in polynomial-size quantum circuits, then circuits with n input qubits have at most $q \in O(poly(n))$ many output qubits. Hence, already after $T \in O(\log n)$ time-steps the inputs to the circuit are completely indistinguishable.

We first show the following Lemma, which gives a bound on the probability that a wire is connected to the input.

3.3.2. Lemma. *For any circuit C as in Theorem 3.3.1 and $t \geq 1$ let*

$$a_t = \max_{w \in V_t} \mathrm{Pr}_E[conn(w, E)],$$

be the smallest number such that no wire at time t is connected to the input with probability higher than a_t. Then

$$a_t \leq (1 - k\epsilon)^t. \tag{3.2}$$

Proof: The proof is by induction. Clearly, $a_1 \leq 1/k - \epsilon < 1 - k\epsilon$, because the wires at time 1 are erased with probability at least $1 - 1/k + \epsilon$. This proves the base case. For the induction step consider a wire w at time $t + 1$. Let $\mathcal{G}$ be the gate which has output wire w. ($\mathcal{G}$ can also be the 1-qubit identity gate.) Let $v_1, \ldots v_l$ be the input wires of $\mathcal{G}$, with $l \leq k$. Then wire w is connected if and only if one of the wires $v_1, \ldots v_l$ is connected and w itself is not erased. By the union bound we therefore get $a_{t+1} \leq (1/k - \epsilon)ka_t = (1 - k\epsilon)a_t$, which proves the inductive step. ■

The next Lemma shows that if a set of wires is not connected to the input, then the state on its qubits is input-independent.

3.3.3. Lemma. *Fix an error pattern E and some t with $1 \leq t \leq T$ and let $V \subseteq V_t$ be some set of wires which are all not connected to the input. Let ρ_t be the state of the computation of circuit C_E on input ρ after all gates up to time t have been processed. Let $\rho_{V,t} = \mathrm{Tr}_{\{1,\ldots,n\}\setminus V}(\rho_t)$ be the state on the qubits V only. Then $\rho_{V,t}$ does not depend on ρ.*

Proof: The proof is by induction on t. If none of the wires in $V \subseteq V_1$ is connected, then they are all erased. Thus, $\rho_{V,1} = \psi_0^{\otimes|V|}$, which is clearly independent of ρ. This proves the base case.

For $t > 1$ let $V_e \subseteq V \subseteq V_t$ be the set of wires in V which are erased at time t. Let $F \subseteq G_t$ be the set of gates which have an outgoing wire in $V \setminus V_e$. Let $U \subseteq V_{t-1}$ be the set of all wires going into a gate in F. In simple words: U is the set of wires at time $t - 1$ that "lead" into the wires $V \setminus V_e$ at time t. Since the wires $V \setminus V_e$ are neither erased nor connected, none of the wires in U can be connected. This means that by our inductive assumption $\rho_{U,t-1}$ does not depend on ρ and therefore neither $\rho_{V\setminus V_e,t}$. Thus the state $\rho_{V,t} = \rho_{V\setminus V_e,t} \otimes \psi_0^{\otimes|V_e|}$ does not depend on ρ. ■

Proof of Theorem 3.3.1: Let T be the number of steps of C and O be the set of the q measured output qubits. For an error pattern E let $\rho_{E,i}$ be the state of the qubits O at time T of circuit C_E with input ρ_i. The bias of the circuit is

$$\begin{aligned}
& \sum_E q(E) \left|\mathrm{Tr}(M_{E,1}^\dagger M_{E,1}(\rho_{E,0} - \rho_{E,1}))\right| \\
= & \sum_{E, \exists w \in O\,:\, conn(w,E)} q(E) \left|\mathrm{Tr}(M_{E,1}^\dagger M_{E,1}(\rho_{E,0} - \rho_{E,1}))\right| \\
\leq & \sum_{E, \exists w \in O\,:\, conn(w,E)} q(E) \\
= & \sum_{w \in O} \sum_{E:\, conn(w,E)} q(E) \\
\leq & \; q(1 - k\epsilon)^T,
\end{aligned}$$

using Lemmas 3.3.3 for the first equality and Lemma 3.3.2 and the union bound for the last inequality.

Thus, the bias δ can be upper bounded by $q(1-k\epsilon)^T \leq \delta$, which is equivalent to $T \geq \frac{\log\delta - \log q}{\log(1-k\epsilon)}$. This establishes the theorem. ■

We now show that the above Theorem 3.3.1 is essentially tight, by showing that at noise rates less than $1 - 1/k$ it is no longer true. For the upper bound we allowed the quantum operation of gate $\mathcal{G}$ to depend on a "global" error pattern, i.e. it could depend on erasures which happen at later time-steps than the gate itself. For our lower bound we do not want this assumption.

3.3.4. Definition. [Causal] For a quantum circuit let as in Definition 3.2.3

$$M(\mathcal{G}, E) \mapsto \hat{\mathcal{E}}_{\mathcal{G},E},$$

be the map specifying for each gate and erasure pattern the Kraus operators of this gate. We say that this quantum circuit is *causal* (or: can be generated in a causal manner) if for all error patterns E and gates $\mathcal{G}$ it holds that

$$M(\mathcal{G}, E) = M(\mathcal{G}, E \cap (V_1 \cup \cdots \cup V_{t-1})),$$

where t is the time of gate $\mathcal{G}$.

Note that causality means that the quantum operation of gate $\mathcal{G}$ only depends on erasures on wires which happened before $\mathcal{G}$.

3.3.5. Definition. [Efficiency] We say that a quantum circuit with weak classical control can be *efficiently generated* if the description of its skeleton graph, the output of $M(\mathcal{G}, E)$ (Kraus operators of gate G for erasure pattern E) and the measurement operators can be computed efficiently, i.e., in time which is polynomial in the number of input qubits to the circuit.

3.3.6. Theorem. *For every k, T and $\epsilon > 0$ there is a quantum circuit C with weak classical control, a $\delta > 0$ and input states $\rho_0 = |0\rangle^{\otimes k^T}$ and $\rho_1 = |1\rangle^{\otimes k^T}$ with the following properties: C can be efficiently generated, has causal quantum operations and uses gates of fan-in at most k. Further, if each wire in C is subjected to erasure noise with probability at most $1 - 1/k - \epsilon$, then after T steps it is possible to distinguish ρ_0 from ρ_1 with bias δ using a one-qubit measurement.*

Proof: The circuit we construct is a formula: Each gate has k input wires, one output wire and the gates are arranged as a balanced tree of depth T, with the output as the root. Every gate will depend on an erasure pattern in the following way: If one of its input wires $w_1, \ldots, w_k$ is connected then the gate outputs the qubit on the first connected wire. Otherwise output some fixed state, say ψ_0.

The measurement operators for the final 1-qubit measurement are $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$, which corresponds to a measurement in the computational basis. It acts on the output qubit of the last gate. Clearly, this circuit can be generated efficiently and it is causal.[2]

We need the following Lemma, which complements Lemma 3.3.2.

3.3.7. LEMMA. *Let b_t be the minimum probability over all wires w at time t that w is connected, i.e.,*

$$b_t = \min_{w \in V_t} \Pr_E[conn(w, E)].$$

Then there is a constant $D > 0$ such that for all t

$$b_t \geq D.$$

Proof: A wire at time $t+1$ is connected if at least one of the input wires to the preceding gate is connected and no erasure happens after the gate. Therefore we have the recursion

$$\begin{aligned} b_{t+1} &\geq \left(\frac{1}{k} + \epsilon\right)\left(1 - (1 - b_t)^k\right) && (3.3) \\ &\geq \left(\frac{1}{k} + \epsilon\right)\left(kb_t - O(b_t^2)\right) \\ &\geq b_t(1 + k\epsilon - cb_t), && (3.4) \end{aligned}$$

for some $c > 0$. We first note that always $(1 - (1 - b_t)^k) \geq b_t$ and therefore by (3.3) we get (a): $b_{t+1}/b_t \geq 1/k$. Secondly, if $b_t \leq k\epsilon/c$ then (3.4) implies (b): $b_{t+1}/b_t \geq 1$. From (a) and (b) we get

$$b_t \geq \frac{\epsilon}{c}$$

for all t. ■

Continuing with the proof of Theorem 3.3.6, we note that by construction of our circuit every connected wire carries the qubit $|i\rangle$ if ρ_i was input. In particular, this holds for the output wire. As in the proof of Theorem 3.3.1 we can therefore write the bias of the circuit as

$$\begin{aligned} &\sum_E q(E) \left| \mathrm{Tr}\Big(|1\rangle\langle 1|(\rho_{E,0} - \rho_{E,1})\Big) \right| \\ = &\sum_{E, \exists w \in O\ :\ w \text{ is connected}} q(E) \left| \mathrm{Tr}\Big(|1\rangle\langle 1|\big(|0\rangle\langle 0| - |1\rangle\langle 1|\big)\Big) \right| \\ \geq\ & D, \end{aligned}$$

by Lemma 3.3.7. ■

[2] Technically, this circuit is not a quantum circuit as we defined it in Section 2.2, since there we defined that every gate has the same number of input and output wires and that each qubit is acted on by some gate at each time. It is easy to extend our circuit such that it formally matches the definition from Section 2.2: Just output some fixed qubit, say ψ_0, on all the remaining $k-1$ output wires of the gates in the tree and apply the identity operation at all time steps to those qubits which are not operated on by one of our gates.

3.4 Discussion

We have seen an upper bound of $1-1/k$ on the amount of erasure noise that can be tolerated to enable fault-tolerant quantum computation. This bound obviously also applies to classical circuits, although it is less relevant.

Note that in the proof of the upper bound (in particular in the proof of Lemma 3.3.2) we never needed any bound on the out-degree of the nodes in the skeleton graph. Hence, the upper bound actually holds for gates with unbounded fan-out.[3] Further, the proof also applies if the quantum circuit works on arbitrary qudits, where we only formally need to change the dimensions of the Hilbert spaces.

The matching lower bound we have shown is weak in the sense that it only shows that for noise rates less than $1 - 1/k$ Theorem 3.3.1 is no longer true. It is interesting to analyze whether efficient fault-tolerant quantum computing can be possible for noise less than $1-1/k$. We conjecture that fault-tolerant classical computation *is* possible, meaning that every classical circuit can be efficiently simulated in our model if the erasure noise is less than $1 - 1/k$.

A (minor) open problem is to determine what happens at noise rates exactly equal to $1 - 1/k$. We conjecture that Theorem 3.3.1 still holds.

Our upper bound was for erasure noise, which is a very benign error model since errors can be detected, i.e., we are informed when and where errors happen. This suggests that the noise thresholds for other noise models (e.g. depolarizing noise) are much lower. In particular, since for classical circuits[4] the gaps between erasure noise and probabilistic noise are significant: Theorems 3.3.1 and 3.3.6 are still true for erasure noise. However, it follows from [37] that for depolarizing noise Theorem 3.3.1 already holds for noise rates above $1 - 1/\sqrt{k}$. For $k = 2$, the corresponding noise rate of $1 - 1/\sqrt{2} \approx 29.3\%$ is much lower than the noise rate $1 - 1/2 = 1/2$ we get for erasure noise. In fact, results in [36] and Chapter 6 suggest that the true threshold for $k = 2$ should not be $1 - 1/\sqrt{2}$, but rather $(3 - \sqrt{7})/2 \approx 17.7\%$ (where we have adjusted for the different noise model from Chapter 6), which is much smaller than $1 - 1/2$. Thus, for classical circuits it is true that the noise threshold for erasure noise only gives a very crude estimate of the threshold for random non-erasure noise (i.e. depolarizing noise).

In Chapter 4 we will give a better upper bound for quantum circuits with depolarizing storage noise, but we will have to restrict the multi-qubit gates to be unitaries.

[3] We did not define quantum circuits with unbounded fan-out in Section 2.2, but it is straightforward. In our model the fan-in and fan-out of a gate are always the same.

[4] in which gates map computational basis states always into (probabilistic mixtures) of computational basis states

Chapter 4

Perfect 1-qubit operations and noisy k-qubit unitaries

This chapter is based on the paper

> Julia Kempe, Oded Regev, Falk Unger and Ronald de Wolf, **Upper bounds on the noise threshold for fault-tolerant quantum computing**, ICALP 2008

4.1 Introduction

In the previous chapter we have seen an upper bound on the tolerable noise for quantum computation. It applied to erasure noise only, which means that the experimenter is notified whenever an error happens. Of course this is a strong assumption. A more common (and often more realistic) assumption is that errors happen and the experimenter does not know when. Intuitively it is clear that these kinds of errors are harder to correct, since the experimenter does not know where and when errors happen. And it is very plausible that the thresholds for these kinds of noise should be smaller, see also Section 3.4 for more on this. In this chapter we give stronger bounds for the noise model of depolarizing noise on wires and a slightly weaker set of available gates.

We consider circuits consisting of unitary k-qubit gates each of whose input wires is subject to depolarizing noise of strength at least ε_k, as well as arbitrary one-qubit gates that are essentially noise-free. We assume that the output of the circuit is the result of measuring some designated qubit of the final state. The main result is that if the noise ε_k is strictly larger than $1 - \sqrt{2^{1/k} - 1} = 1 - \Theta(1/\sqrt{k})$ the output of any such circuit of large enough (but constant) depth is essentially independent of its input, thereby making the circuit useless. For the important special case of $k = 2$, our bound is $\varepsilon_2 > 1 - \sqrt{\sqrt{2} - 1} \approx 35.7\%$. It is interesting to note that our bound on the threshold behaves like $1 - \Theta(1/\sqrt{k})$. This matches what is known for classical circuits [38, 37], and therefore probably represents the correct asymptotic behavior. In comparison, the bound for erasure noise from Chapter 3 behaves like $1 - 1/k$.

It is known that fault-tolerant quantum computation is impossible (for any positive noise level) without a source of "fresh" qubits. Our model takes care of this by allowing arbitrary one-qubit gates—in particular, this includes gates that take any input, and output a fixed one-qubit state, for instance $|0\rangle$.

By allowing essentially noise-free one-qubit gates, our model addresses the fact that gates on more than one qubit are generally much harder to implement than one-qubit gates. It should also be noted that the exact value of the constant ϵ_1 is inessential and can be chosen to be an arbitrarily small positive constant, see also comments after Theorem 4.2.1.

Note that since our theorem applies to arbitrary starting states, it in particular applies to the case that the initial state is encoded in some good quantum error-correcting code, or that it is some sort of "magic state" [21, 81].[1] Further, we could even allow operations which add/replace arbitrary states on multiple qubits at any time during the computation. To extend our proof to accommodate for this is straightforward. In all these cases, our theorem shows that the computation becomes essentially independent of the input after sufficiently many levels.

Weaknesses of the model Our assumption that all k-qubit gates are mixtures of unitaries does slightly restrict generality. Not every completely-positive trace-preserving map can be written as a mixture of unitaries.[2] However, we believe that it is still a reasonable 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.

A slightly more severe restriction is the assumption that the output consists of just one qubit. Recall that in Chapter 3 we showed that if the noise is above the threshold of $1-1/k$ than after logarithmically many steps no measurement on *all* qubits can distinguish any two input states. However, we believe that in many instances the assumption that there is only one output qubit is still reasonable. For instance, this is the case whenever the circuit is required to solve a decision

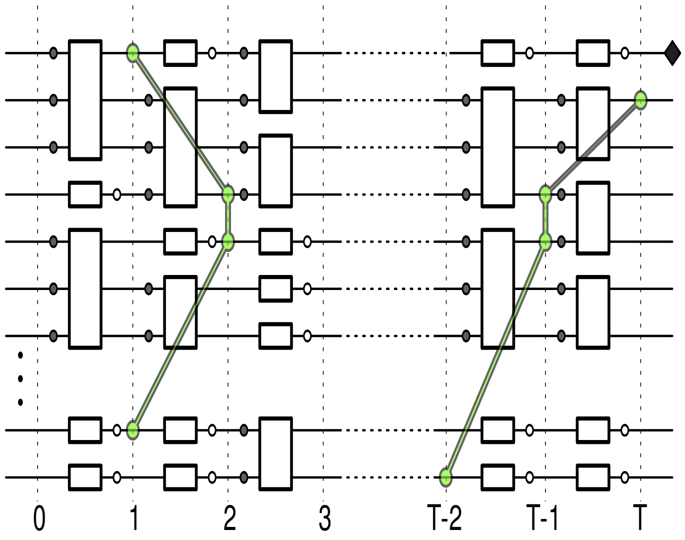
[1] The set of gates STAB (see page 60) is not quantum universal by itself if the only allowed input states are computational basis states. However, if one can additionally create certain "magic states" and allows perfect classical control, then one can perform universal quantum computation [21, 81]. This is the reason to call these states "magic states".

[2] 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 notice that this increases the arity of the gate and moreover the ancilla qubits will be affected by the noise operators that precede the unitary.

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.

4.2 Model and results

Before we state the results, we describe the exact model, recalling definitions from Section 2.2. We consider parallel circuits, composed of n *wires* and T *levels* of gates[3] (see Figure 4.2). We assume that the number of qubits n does not change during the computation. Notice that at each level, all qubits must go through some gate (possibly the identity). For each gate, the number of input qubits is the same as the number of output qubits.



Dark circles denote ε_k-depolarizing noise, and light circles denote ε_1-depolarizing noise. Also marked are two consistent sets (defined in Section 4.4), each containing four qubits. The first has distance 1, the second has distance $T-2$. The output qubit is in the upper right corner.

Figure 4.1: Parallel circuit with $k = 3$ and T levels

We assume the circuit is composed of k-qubit gates that are probabilistic mixtures of unitary operations, as well as arbitrary (i.e., all completely-positive trace-preserving) one-qubit gates. In particular, it is possible to do intermediate

[3] So, we call the parts of the skeleton graph from Definition 3.2.1 “levels”.

one-qubit measurements. We assume the output of the circuit is the outcome of a measurement of a designated output qubit in the computational basis. Finally, we assume that the circuit is subject to noise as follows. Recall that p-depolarizing noise on a certain qubit replaces that qubit by the completely mixed state with probability p, and does not alter the qubit otherwise. Formally, this is described by the superoperator $\mathcal{E}$ acting on a qubit ρ as $\mathcal{E}(\rho) = (1-p)\rho + p\mathbb{I}/2$. We assume that each one-qubit gate is followed by at least ε_1-depolarizing noise on its output qubit, where $\varepsilon_1 > 0$ is an arbitrarily small constant. Thus one-qubit gates can be essentially noise-free. We also assume that each k-qubit gate is preceded by at least ε_k-depolarizing noise on each of its input qubits, where $\varepsilon_k > 1 - \sqrt{2^{1/k} - 1}$.

Main results We prove the following main result

4.2.1. Theorem. *Fix any T-level quantum circuit as above. Then for any two states ρ and τ, the probabilities of obtaining measurement outcome* 1 *at the output qubit starting from ρ and starting from τ, respectively, differ by at most* $2^{-\Omega(T)}$.

In other words, for any $\eta > 0$, the probability of measuring 1 at the output qubit of a circuit running for $T = O(\log(1/\eta))$ levels is independent of the input (up to $\pm\eta$). This makes the output essentially independent of the starting state, and renders long computations "essentially useless".

As pointed out in the introduction, ε_1 can be chosen to be an arbitrarily small but positive constant. The value of ε_1 only affects the constant in the $\Omega(\cdot)$ of Theorem 4.2.1. The reason we require $\epsilon_1 > 0$ is a technicality which simplifies the statement of our result. However, for $\varepsilon_1 = 0$ the statement of Theorem 4.2.1 is just wrong: One could choose input states $\rho := |0\rangle\langle 0| \otimes \rho'$ and $\tau := |1\rangle\langle 1| \otimes \tau'$, do nothing for T levels (i.e., apply noise-free one-qubit identity gates on all wires) and then measure the first qubit in the computational basis. Clearly, from this measurement outcome one can exactly tell which of the two states ρ, τ was input. Nevertheless, it *is* possible to let $\varepsilon_1 = 0$, if we slightly change the model and additionally require that every path from the input to the output qubit goes through enough k-qubit gates. Our proof can easily be adapted to this case.

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 additionally allow controlled-Y and controlled-Z gates.

4.3 Preliminaries

We first recall some definitions.

The Pauli matrices are

$$\mathbb{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \ X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \ Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

and we define $\mathcal{P} = \{\mathbb{I}, X, Y, Z\}$ and $\mathcal{P}_* = \{X, Y, Z\}$. We use $\mathcal{P}^n$ to denote the set of all tensor products of n one-qubit Pauli matrices. For a Pauli matrix $S \in \mathcal{P}^n$ we define its *support*, denoted $\mathrm{supp}(S)$, to be the qubits on which S is not identity. We sometimes use superscripts to indicate the qubits on which certain operators act. Thus $\mathbb{I}^{\mathcal{A}}$ denotes the identity operator applied to the qubits in set $\mathcal{A}$.

The set of all $2^n \times 2^n$ Hermitian matrices forms a 4^n-dimensional real vector space. On this space we consider the Hilbert-Schmidt inner product, given by $\langle A, B\rangle := \mathrm{Tr}(A^\dagger B) = \mathrm{Tr}(AB)$. Note that for any $S, S' \in \mathcal{P}^n$, $\mathrm{Tr}(SS') = 2^n$ if $S = S'$ and $\mathrm{Tr}(SS') = 0$ otherwise. Hence, $\mathcal{P}^n$ is an orthogonal basis of this space. It follows that we can uniquely express any Hermitian matrix δ in this basis as

$$\delta = \frac{1}{2^n} \sum_{S \in \mathcal{P}^n} \widehat{\delta}(S) S$$

where $\widehat{\delta}(S) := \mathrm{Tr}(\delta S)$ are the (real) coefficients.

We now state some easy observations which will be used in the proof of our main result. First, by the orthogonality of $\mathcal{P}^n$, it follows that for any δ,

$$\mathrm{Tr}(\delta^2) = \frac{1}{2^n} \sum_{S \in \mathcal{P}^n} \widehat{\delta}(S)^2.$$

This easily leads to the following observation.

1. Observation (Unitary preserves sum of squares). *For any unitary U and any Hermitian matrix δ, if we denote $\delta' = U\delta U^\dagger$, then*

$$\sum_{S \in \mathcal{P}^n} \widehat{\delta'}(S)^2 = 2^n \mathrm{Tr}(\delta'^2) = 2^n \mathrm{Tr}(U\delta U^\dagger U \delta U^\dagger) = 2^n \mathrm{Tr}(\delta^2) = \sum_{S \in \mathcal{P}^n} \widehat{\delta}(S)^2.$$

This also shows that the operation of conjugating by a unitary matrix, when viewed as a linear operation on the vector of Pauli coefficients, is an orthogonal transformation.

2. Observation (Tracing out qubits). *Let δ be some Hermitian matrix on a set of qubits W. For $V \subseteq W$, let $\delta_V = \mathrm{Tr}_{W\setminus V}(\delta)$. Then,*

$$\widehat{\delta}(S\mathbb{I}^{W\setminus V}) = \mathrm{Tr}(\delta \cdot S\mathbb{I}^{W\setminus V}) = \mathrm{Tr}(\delta_V \cdot S) = \widehat{\delta_V}(S).$$

3. Observation (Noise in the Pauli basis). *Applying p-depolarizing noise $\mathcal{E}$ to the j-th qubit of Hermitian matrix δ changes the coefficients as follows:*

$$\widehat{\mathcal{E}(\delta)}(S) = \begin{cases} \widehat{\delta}(S) & \textit{if } S_j = \mathbb{I} \\ (1-p)\widehat{\delta}(S) & \textit{if } S_j \neq \mathbb{I} \end{cases}$$

In other words, $\mathcal{E}$ "shrinks" by a factor $1-p$ all coefficients that have support on the j-th coordinate.

4. OBSERVATION. *Let ρ and τ be two one-qubit states and let $\delta = \rho - \tau$. Consider the two probability distributions obtained by performing a measurement in the computational basis on ρ and τ, respectively. Then the difference in the probabilities of obtaining the outcome* 1 *given ρ respectively τ is*

$$\frac{1}{2}|\widehat{\delta}(Z)|.$$

Proof: The difference in the probabilities of obtaining the outcome 1 is given by

$$|\mathrm{Tr}((\rho-\tau)\cdot|1\rangle\langle 1|)| = \left|\mathrm{Tr}\left(\delta\cdot\frac{\mathbb{I}-Z}{2}\right)\right| = \frac{1}{2}|\mathrm{Tr}(\delta\cdot Z)| = \frac{1}{2}|\widehat{\delta}(Z)|,$$

where we have used $\mathrm{Tr}(\delta)=0$. ∎

The last observation follows immediately from the convexity of the function x^2.

5. OBSERVATION (CONVEXITY). *Let p_i be any probability distribution, and δ_i a set of Hermitian matrices. Let $\delta = \sum_i p_i\delta_i$. Then*

$$\sum_{S\in\mathcal{P}^n}\widehat{\delta}(S)^2 \le \sum_i p_i \sum_{S\in\mathcal{P}^n}\widehat{\delta_i}(S)^2.$$

Proof: Follows immediately from the convexity of the function x^2. ∎

4.4 Proof of Theorem 4.2.1

In this section we prove Theorem 4.2.1. The rough idea is the following. Fix two arbitrary initial states ρ and τ. Our goal is to show that after applying the noisy circuit, the state of the output qubit is nearly the same with both starting states. Equivalently, we can define $\delta = \rho - \tau$ and show that after applying the noisy circuit to δ, the "state" of the output qubit is essentially 0 (the noisy circuit is a linear operation, and hence there is no problem in applying it to δ, which is the difference of two density matrices). In order to show this, we will examine how the coefficients of δ in the Pauli basis evolve through the circuit. Initially we might have many large coefficients. Our goal is to show that the coefficients of the output qubit are essentially 0. This is established by analyzing the balance between two opposing forces: noise, which shrinks coefficients by a constant factor (as in Observation 3), and gates, which can increase coefficients. As we saw in Observation 1, unitary gates preserve the sum of squares of coefficients. They can, however, "concentrate" several small coefficients into one large coefficient. One-qubit operations need not preserve the sum of squares (a good example is the gate that resets a qubit to the $|0\rangle$ state), but we can still deal with them by using a known characterization of one-qubit gates. This characterization allows us to bound the amount by which one-qubit gates can increase the Pauli coefficients, and very roughly speaking shows that the gate that resets a qubit to $|0\rangle$ is "as bad as it gets".

Before continuing with the proof, we introduce some terminology. From now on we use the term *qubit* to mean a wire at a specific time, so there are $(T+1)n$ qubits (although during the proof we will also consider qubits that are located between a gate and its associated noise). We say that a set of qubits V is *consistent* if we can meaningfully talk about a "state of the qubits of V" (see Figure 4.2). More formally, we define a consistent set as follows. The set of all qubits at time 0 and all its subsets are consistent. If V is some consistent set of qubits, which contains all input qubits IN of some gate (possibly a one-qubit identity gate), then also $(V \setminus IN) \cup OUT$ and all its subsets are consistent, where OUT denotes the gate's output qubits. Note that here we think of the noise as being part of the gate. For a consistent set V and a state (or more generally, a Hermitian matrix) ρ, we denote the state of V when the circuit is applied with the initial state ρ, by ρ_V. In other words, ρ_V is the state one obtains by applying some initial part of the circuit to ρ, and then tracing out from the resulting state all qubits that are not in V.

If v is a qubit, we use $\mathrm{dist}(v)$ to denote its distance from the input, i.e., the level of the gate just preceding it. The qubits of the starting state have $\mathrm{dist}(v) = 0$. For a nonempty set V of qubits we define $\mathrm{dist}(V) = \min\{\mathrm{dist}(v) \mid v \in V\}$, and extend it to the empty set by $\mathrm{dist}(\emptyset) = \infty$. Note that $\mathrm{dist}(V)$ does not increase if we add qubits to V.

In the rest of this section we prove the following lemma, showing that a certain invariant holds for all consistent sets V.

4.4.1. Lemma. *For all $\varepsilon_1 > 0$ and $\varepsilon_k > 1 - \sqrt{2^{1/k} - 1}$ there exists a $\theta < 1$ such that the following holds. Fix any T-level circuit in our model, let ρ and τ be some arbitrary initial states, and let $\delta = \rho - \tau$. Then for every consistent V,*

$$\sum_{S \in \mathcal{P}^V} \widehat{\delta_V}(S)^2 \leq 2 \cdot 2^{|V|} \cdot \theta^{\mathrm{dist}(V)}, \tag{4.1}$$

or equivalently,

$$\mathrm{Tr}(\delta_V^2) \leq 2 \cdot \theta^{\mathrm{dist}(V)}.$$

In particular, if we consider the consistent set V containing the designated output qubit at time T, then we get that $\widehat{\delta_V}(Z)^2 \leq 4\theta^T$. By Observation 4, this implies Theorem 4.2.1.

4.4.1 Proof of Lemma 4.4.1

The proof of the invariant is by induction on the sets V. At the base of the induction are all sets V which only contain qubits at time 0. All other sets are handled in the induction step. In order to justify the inductive proof, we need to provide an ordering on the consistent sets V such that for each V, the proof for V uses the inductive hypothesis only on sets V' that appear before V in the ordering. As will become apparent from the proof, if we denote by latest(V) the maximum time at which V contains a qubit, then each V' for which we use the induction hypothesis has strictly less qubits than V at time latest(V). Therefore, we can order the sets V first in increasing order of latest(V) and then in increasing order of the number of qubits at time latest(V).

Base case

Here we consider the case that V is fully contained within time 0. If $V = \emptyset$ then both sides of the invariant are zero, so from now on assume V is nonempty. In this case dist$(V) = 0$. The matrix δ_V is the difference of two density matrices, say $\delta_V = \rho_V - \tau_V$, and hence $\mathrm{Tr}(\delta_V^2) = \mathrm{Tr}(\rho_V^2) + \mathrm{Tr}(\tau_V^2) - 2\mathrm{Tr}(\rho_V \tau_V) \leq 2$, and the invariant is satisfied.

Induction step

Let V'' be any consistent set containing at least one qubit at time greater than zero. Our goal in this section is to prove the invariant for V''. Consider any of the qubits of V'' located at time latest(V'') and let G be the gate that has this qubit as one of its output qubits. We now consider two cases, depending on whether G is a k-qubit gate or a one-qubit gate.

Case 1: G is a k-qubit gate. Here we consider the case that G is a probabilistic mixture of k-qubit unitaries. First note that by Observation 5 it suffices to prove the invariant for k-qubit unitaries. So assume G is a k-qubit unitary acting on the qubits $\mathcal{A} = \{A_1, \ldots, A_k\}$. Let $\mathcal{A}' = \{A_1', \ldots, A_k'\}$ be the qubits after the ϵ_k-noise but before the gate G and $\mathcal{A}'' = \{A_1'', \ldots, A_k''\}$ the qubits after G (see Figure 4.2). By our choice of G, $\mathcal{A}'' \cap V'' \neq \emptyset$. Define $V' = (V'' \setminus \mathcal{A}'') \cup \mathcal{A}'$ and $V = (V'' \setminus \mathcal{A}'') \cup \mathcal{A}$. Note that V and its subsets are consistent sets with strictly fewer qubits than V'' at time latest(V''), and hence we can apply the induction hypothesis to them.

Recall that our goal is to prove the invariant Eq. (4.1) for V''. To begin, using Observation 2,

$$\sum_{S \in \mathcal{P}^{V''}} \widehat{\delta_{V''}}(S)^2 \leq \sum_{S \in \mathcal{P}^{V'' \cup \mathcal{A}''}} \widehat{\delta_{V'' \cup \mathcal{A}''}}(S)^2. \tag{4.2}$$

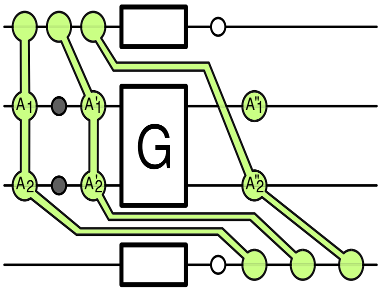


Figure 4.2: An example showing the sets V, V', and V'' for a two-qubit gate G.

Because G (which maps $\delta_{V'}$ to $\delta_{V''\cup\mathcal{A}''}$) is unitary, it preserves the sum of squares of $\widehat{\delta}$-coefficients (see Observation 1), so the right hand side of (4.2) is equal to

$$\sum_{S\in\mathcal{P}^{V'}} \widehat{\delta_{V'}}(S)^2 = \sum_{S\in\mathcal{P}^{V'\setminus\mathcal{A}'}} \sum_{R\in\mathcal{P}^{\mathcal{A}'}} \widehat{\delta_{V'}}(RS)^2.$$

Since the only difference between δ_V and $\delta_{V'}$ is noise on the qubits $A_1,\dots,A_k$, using Observation 3 and denoting $\mu = 1-\epsilon_k$, we get that the above is at most

$$\begin{aligned}
&\sum_{S\in\mathcal{P}^{V\setminus\mathcal{A}}} \sum_{R\in\mathcal{P}^{\mathcal{A}}} \mu^{2|\mathrm{supp}(R)|}\widehat{\delta_V}(RS)^2 \\
=&\sum_{S\in\mathcal{P}^{V\setminus\mathcal{A}}} \sum_{a\subseteq\mathcal{A}} \mu^{2|a|}(1-\mu^2)^{k-|a|} \sum_{R\in\mathcal{P}^{a}\otimes\mathbb{I}^{\mathcal{A}\setminus a}} \widehat{\delta_V}(RS)^2,
\end{aligned}$$

where the equality follows by noting that for any fixed S and any $R\in\mathcal{P}^{\mathcal{A}}$, the term $\widehat{\delta_V}(RS)^2$, which appears with coefficient $\mu^{2|\mathrm{supp}(R)|}$ on the left hand side, appears with the same coefficient $\sum_{a\supseteq\mathrm{supp}(R)}\mu^{2|a|}(1-\mu^2)^{k-|a|} = \mu^{2|\mathrm{supp}(R)|}$ on the right hand side. By rearranging and using Observation 2 we get that the above is equal to

$$\begin{aligned}
&\sum_{a\subseteq\mathcal{A}} \mu^{2|a|}(1-\mu^2)^{k-|a|} \sum_{S\in\mathcal{P}^{(V\setminus\mathcal{A})\cup a}} \widehat{\delta_{(V\setminus\mathcal{A})\cup a}}(S)^2 \\
\le&\sum_{a\subseteq\mathcal{A}} \mu^{2|a|}(1-\mu^2)^{k-|a|} 2\cdot 2^{|(V\setminus\mathcal{A})\cup a|}\cdot\theta^{\mathrm{dist}((V\setminus\mathcal{A})\cup a)}
\end{aligned}$$

where we used the inductive hypothesis. Note that $\mathrm{dist}((V\setminus\mathcal{A})\cup a)\ge\mathrm{dist}(V)$, so the above is

$$\begin{aligned}
&\le 2\cdot 2^{|V\setminus\mathcal{A}|}\cdot\theta^{\mathrm{dist}(V)}\sum_{a\subseteq\mathcal{A}} 2^{|a|}\mu^{2|a|}(1-\mu^2)^{k-|a|} \\
&= 2\cdot 2^{|V\setminus\mathcal{A}|}\cdot\theta^{\mathrm{dist}(V)}((1-\mu^2)+2\mu^2)^k \\
&= 2\cdot 2^{|V\setminus\mathcal{A}|}\cdot\theta^{\mathrm{dist}(V)}(1+\mu^2)^k. \qquad (4.3)
\end{aligned}$$

Note that $|V \setminus \mathcal{A}| \leq |V''| - 1$ and $\mathrm{dist}(V'') - 1 \leq \mathrm{dist}(V)$, so the right hand side is bounded by

$$\leq 2 \cdot 2^{|V''|-1} \cdot \theta^{\mathrm{dist}(V'')-1}(1+\mu^2)^k.$$

Since $\epsilon_k > 1 - \sqrt{2^{1/k} - 1}$, we have that $(1+\mu^2)^k \leq 2\theta$ if θ is close enough to 1, so we can finally bound the last expression by

$$\leq 2 \cdot 2^{|V''|} \cdot \theta^{\mathrm{dist}(V'')}$$

which proves the invariant for V''.

Case 2: G is a one-qubit gate. Before proving the invariant, we need to prove the following property of completely-positive trace-preserving (CPTP) maps on one qubit.

4.4.2. Lemma. *For any CPTP map G on one qubit there exists a $\beta \in [0,1]$ such that the following holds. For any Hermitian matrix δ, if we let δ' denote the result of applying G to δ, then we have*

$$\widehat{\delta'}(X)^2 + \widehat{\delta'}(Y)^2 + \widehat{\delta'}(Z)^2 \leq (1-\beta) \cdot \widehat{\delta}(\mathbb{I})^2 + \beta \cdot (\widehat{\delta}(X)^2 + \widehat{\delta}(Y)^2 + \widehat{\delta}(Z)^2).$$

Proof: The proof is based on the characterization of trace-preserving completely-positive maps on one qubit gates given in Section 2.5.2 on page 28, which we recall now. Any one-qubit gate G can be written as a convex combination of gates of the form $U_1 \circ J \circ U_2$. Here U_1 and U_2 are one-qubit unitaries (acting on the density matrix by conjugation), and J is a one-qubit map that in the Pauli basis has the form

$$J = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & \lambda_2 & 0 \\ t & 0 & 0 & \lambda_1\lambda_2 \end{pmatrix}$$

for some $\lambda_1, \lambda_2 \in [-1,1]$ and $t = \pm\sqrt{(1-\lambda_1^2)(1-\lambda_2^2)}$.

First observe that by the convexity of the square function, it suffices to prove the lemma for G of the form $U_1 \circ J \circ U_2$ (with the resulting β being the appropriate average of the individual β's). Next note that since U_1 and U_2 are unitary, they act on the vector of coefficients $(\widehat{\delta}(X), \widehat{\delta}(Y), \widehat{\delta}(Z))$ as an orthogonal transformation, and hence leave the sum of squares invariant. This shows that it suffices to prove the lemma for a map J as above. For this map,

$$\widehat{\delta'}(X)^2 + \widehat{\delta'}(Y)^2 + \widehat{\delta'}(Z)^2 = \lambda_1^2\widehat{\delta}(X)^2 + \lambda_2^2\widehat{\delta}(Y)^2 + (t\widehat{\delta}(\mathbb{I}) + \lambda_1\lambda_2\widehat{\delta}(Z))^2.$$

Assume without loss of generality that $\lambda_1^2 \geq \lambda_2^2$. Applying Cauchy-Schwarz to the two 2-dimensional vectors $(\pm\sqrt{1-\lambda_1^2}a, \lambda_1 b)$ and $(\sqrt{1-\lambda_2^2}, \lambda_2)$, we get that for any $a, b \in \mathcal{R}$, $(ta + \lambda_1\lambda_2 b)^2 \leq (1-\lambda_1^2)a^2 + \lambda_1^2 b^2$. Hence the above expression is upper bounded by

$$\lambda_1^2\widehat{\delta}(X)^2 + \lambda_1^2\widehat{\delta}(Y)^2 + (1-\lambda_1^2)\widehat{\delta}(\mathbb{I})^2 + \lambda_1^2\widehat{\delta}(Z)^2$$

and we complete the proof by choosing $\beta = \lambda_1^2$. ■

Let A be the qubit G is acting on, and recall that our goal is to prove the invariant for the set V''. Denote by A' the qubit of G after the gate but before the ε_1 noise, and by A'' the qubit after the noise. As before, by our choice of G, we have $A'' \in V''$. Let $\mathcal{A} = \{A\}$, $\mathcal{A}' = \{A'\}$, $\mathcal{A}'' = \{A''\}$. Define $V' = (V'' \setminus \mathcal{A}'') \cup \mathcal{A}'$ and $V = (V'' \setminus \mathcal{A}'') \cup \mathcal{A}$ and notice that $|V| = |V'| = |V''|$. By using Lemma 4.4.2, we obtain a $\beta \in [0,1]$ such that

$$\begin{aligned}
&\sum_{S \in \mathcal{P}^{V''}} \widehat{\delta_{V''}}(S)^2 \\
&\leq \sum_{S \in \mathcal{P}^{V' \setminus \mathcal{A}'}} \left(\widehat{\delta_{V'}}(\mathbb{I}S)^2 + (1-\epsilon_1)^2 \sum_{R \in \mathcal{P}_*^{\mathcal{A}'}} \widehat{\delta_{V'}}(RS)^2 \right) \\
&\leq \sum_{S \in \mathcal{P}^{V \setminus \mathcal{A}}} \left((1 + (1-\epsilon_1)^2(1-2\beta))\widehat{\delta_V}(\mathbb{I}S)^2 + (1-\epsilon_1)^2\beta \sum_{R \in \mathcal{P}^{\mathcal{A}}} \widehat{\delta_V}(RS)^2 \right).
\end{aligned}$$

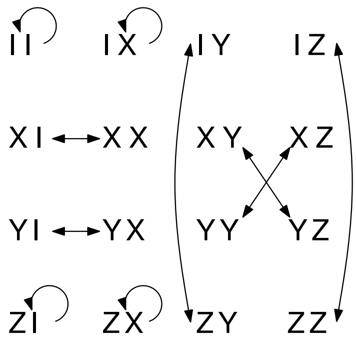
By applying the induction hypothesis to both $V \setminus \mathcal{A}$ and V, we can upper bound the above by

$$\begin{aligned}
&(1 + (1-\epsilon_1)^2(1-2\beta)) \cdot 2 \cdot 2^{|V|-1} \cdot \theta^{\mathrm{dist}(V \setminus \mathcal{A})} + (1-\epsilon_1)^2\beta \cdot 2 \cdot 2^{|V|} \cdot \theta^{\mathrm{dist}(V)} \\
&\leq \frac{1 + (1-\epsilon_1)^2}{2\theta} \cdot 2 \cdot 2^{|V''|} \cdot \theta^{\mathrm{dist}(V'')}
\end{aligned}$$

where we used that $|V| = |V''|$, and $\mathrm{dist}(V'') - 1 \leq \mathrm{dist}(V) \leq \mathrm{dist}(V \setminus \mathcal{A})$. Hence the invariant remains valid if we choose $\theta < 1$ such that $1 + (1-\epsilon_1)^2 \leq 2\theta$.

4.5 Arbitrary one-qubit gates and CNOT gates

In this section we consider the case where CNOT is the only allowed gate acting on more than one qubit. We still allow arbitrary one-qubit gates. The proof follows along the lines of that of Theorem 4.2.1 with one small modification. As before, we will prove that for all $\varepsilon_1 > 0$ and $\varepsilon_2 > 1 - 1/\sqrt{2} \approx 0.293$ the invariant, Eq. (4.1), holds. The proof for the case that G is a one-qubit gate holds without change. We will give the modified proof for the case that G is a CNOT gate. The idea for the improved bound is to make use of the fact that the CNOT gate merely permutes the 16 elements of $\mathcal{P} \otimes \mathcal{P}$, and does not map elements from $\mathbb{I} \otimes \mathcal{P}_*$ to $\mathcal{P}_* \otimes \mathbb{I}$ or vice versa (as illustrated in Figure 4.3). As a result we need



The action of CNOT on $\mathcal{P} \otimes \mathcal{P}$ under conjugation, with the control wire corresponding to the first qubit.

Figure 4.3: Action of CNOT on Pauli group

to apply the induction hypothesis on one less term, which in turn improves the bound.

Assume the CNOT acts on qubits $\mathcal{A} = \{A, B\}$, with $\mathcal{A}' = \{A', B'\}$ and $\mathcal{A}'' = \{A'', B''\}$ as before, where again $\mathcal{A}'' \cap V'' \neq \emptyset$. If both A'' and B'' are contained in V'' then the proof of the general case (cf. Eq. (4.3)) already gives a bound of

$$2 \cdot 2^{|V \setminus \mathcal{A}|} \cdot \theta^{\mathrm{dist}(V)} (1 + \mu^2)^2 \leq 2 \cdot 2^{|V''|-2} \cdot \theta^{\mathrm{dist}(V'')-1} (1 + \mu^2)^2 \leq 2 \cdot 2^{|V''|} \cdot \theta^{\mathrm{dist}(V'')}$$

where the last inequality holds for all $\mu < 1$. Hence it suffices to consider the case that exactly one of A'' and B'' is in V''. Assume without loss of generality that $A'' \in V''$ and $B'' \notin V''$. As before, our goal is to upper bound

$$\sum_{S \in \mathcal{P}^{V''}} \widehat{\delta_{V''}}(S)^2 = \sum_{S \in \mathcal{P}^{V''}} \widehat{\delta_{V'' \cup B''}}(S \mathbb{I}^{B''})^2,$$

where the equality follows from Observation (2). Because of the property of CNOT mentioned above, we can now upper bound this by

$$\sum_{S \in \mathcal{P}^{V' \setminus \mathcal{A}'}} \left(\widehat{\delta_{V'}}(\mathbb{I}^{A'} \mathbb{I}^{B'} S)^2 + \sum_{R \in \mathcal{P}_*^{A'}} \widehat{\delta_{V'}}(R \mathbb{I}^{B'} S)^2 + \sum_{R \in \mathcal{P}_*^{A'} \otimes \mathcal{P}_*^{B'}} \widehat{\delta_{V'}}(RS)^2 \right).$$

This is the crucial change compared to the case of general two-qubit gates (the latter case also includes a term of the form $\sum_{R \in \mathcal{P}_*^{B'}} \widehat{\delta_{V'}}(\mathbb{I}^{A'} RS)^2$). The rest of the proof is similar to the earlier proof. Using the induction hypothesis we can upper bound the above by

$$
\begin{aligned}
&\sum_{S\in\mathcal{P}^{V\setminus\mathcal{A}}}\Big(\widehat{\delta_V}(\mathbb{I}^A\mathbb{I}^B S)^2+\mu^2\sum_{R\in\mathcal{P}^A_*}\widehat{\delta_V}(R\mathbb{I}^B S)^2+\mu^4\sum_{R\in\mathcal{P}^A_*\otimes\mathcal{P}^B_*}\widehat{\delta_V}(RS)^2\Big)\\
\leq\ &(1-\mu^2)\sum_{S\in\mathcal{P}^{V\setminus\mathcal{A}}}\widehat{\delta_{V\setminus\mathcal{A}}}(S)^2+(\mu^2-\mu^4)\sum_{S\in\mathcal{P}^{V\setminus\{B\}}}\widehat{\delta_{V\setminus\{B\}}}(S)^2+\mu^4\sum_{S\in\mathcal{P}^V}\widehat{\delta_V}(S)^2\big)\\
\leq\ &(1-\mu^2)2\cdot 2^{|V\setminus\mathcal{A}|}\theta^{\mathrm{dist}(V\setminus\mathcal{A})}+(\mu^2-\mu^4)2\cdot 2^{|V\setminus\{B\}|}\theta^{\mathrm{dist}(V\setminus\{B\})}+\mu^4\ 2\cdot 2^{|V|}\theta^{\mathrm{dist}(V)}\\
\leq\ &2\cdot 2^{|V''|}\theta^{\mathrm{dist}(V)}\Big(\frac{1+\mu^2}{2}+\mu^4\Big)\\
\leq\ &2\cdot 2^{|V''|}\theta^{\mathrm{dist}(V'')}\Big(\frac{1+\mu^2}{2}+\mu^4\Big)\frac{1}{\theta}.
\end{aligned}
$$

Hence the invariant remains valid as long as $\frac{1+\mu^2}{2}+\mu^4\leq\theta<1$. This can be satisfied as long as $\mu<1/\sqrt{2}$, equivalently $\varepsilon_2>1-1/\sqrt{2}\approx 0.293$.

4.6 Discussion

4.6.1 Comparison with other chapters

In Chapter 3 we have shown an upper bound of $\varepsilon_k=1-1/k$ on erasure noise. On one hand, this result is stronger than the result from this chapter as it allows arbitrary k-qubit gates and not just mixtures of unitaries and it holds for erasure noise instead of depolarizing noise. Further, we saw that the result of an arbitrary n-qubit measurement on the full final state becomes essentially independent of the initial state after $T=O(\log n)$ levels. On the other hand, the bound in this chapter is better for all values of k. Hence the two results are incomparable.

We will see another bound in Chapter 5, which shows that classical circuits can efficiently *simulate* any quantum circuit that consists of perfect, noise-free *stabilizer operations* (meaning Clifford gates (Hadamard, phase gate, CNOT), preparations of states in the computational basis, and measurements in the computational basis), perfect classical control (i.e., the ability to condition future gates on earlier classical measurement outcomes, see page 35) and arbitrary one-qubit unitary gates that are followed by 45.3% depolarizing noise. Hence such circuits are not significantly more powerful than classical circuits. We will also see that this result is tight: If the one-qubits gates have less than 45.3% noise, it is possible to efficiently simulate any (noise-free) quantum circuit. Although this result establishes a tight threshold, it is incomparable to our current result since the result in Chapter 5 applies to a restricted gate set only.

4.6.2 Comments on results and open problems

We believe that a main part of our contribution is introducing a 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.

We only analyze depolarizing noise acting independently on each qubit. Depolarizing noise is the "most symmetric" type of one-qubit noise and therefore a natural choice for our analysis. Also, it is a relatively weak type of noise: it is not adversarial and does not have correlations between the errors occurring on different qubits. However, since we are proving an *upper bound* on the fault-tolerance threshold, this weakness is actually a good thing, making our result stronger. In principle one can extend our results to various other one-qubit noise models, using an analysis similar to the one developed in Lemma 4.4.2. However, not all noise models can actually yield a result like Theorem 4.2.1. For instance, if we have Toffoli gates with only phaseflip errors, then we can do perfect classical computation. Statements like Theorem 4.2.1 are just false in that case.

Open problems In the introduction we mentioned some weaknesses of our model. Of course, we would like to prove a result which does not have these restrictions. Further, it would be interesting to extend the result in a couple of other directions. We now summarize some desirable extensions:

- We should make it work for all possible k-qubit gates (CPTP maps), rather than just mixtures of unitaries.

- We should allow some classical side-processing, where classical outcomes of intermediate measurements can be used by a classical computer and its results can later be fed back into the circuit. Allowing such "classical control" requires a type of theorem different from the one we have now: if initial states ρ and τ were bits 0 and 1, respectively, we could just measure this right at the start, store the bit in the classical part without noise, and feed it back into the circuit only at the last step, yielding distinguishable final states. Furthermore, if we allow classical control (and classical side-processing) then it is clearly possible to compute any function just in the classical part of the circuit. Hence, a statement like ours is just not true.

  To get a noise bound for the model with classical control, one would need to show that if the noise in the quantum hardware is above a certain threshold, then not all problems in BQP could be solved *efficiently*, where BQP is the class of problems that can be efficiently solved with a *noise-free* quantum computer (see also Section 2.3).

- We should relax the assumption that the final output is determined by a measurement on one or a few qubits of the final state. Often in fault-tolerant schemes one encodes each "logical qubit" in a large block of physical qubits, and measures all qubits in that block to obtain the final outcome of the computation. If only some of the qubits in the final measurement are faulty the final result can still be recovered by applying classical post-processing. Our results cannot rule out this approach.

- Last but not least, our upper bounds on the fault-tolerance threshold are still higher than one would expect, and we would like to decrease them further.

Chapter 5

Perfect stabilizer operations and noisy 1-qubit unitaries

This chapter is based on the paper

> Harry Buhrman, Richard Cleve, Monique Laurent, Noah Linden, Lex Schrijver and Falk Unger, **New Limits on Fault-Tolerant Quantum Computation**, *Proceedings of 47th IEEE FOCS*, 2006

In this chapter we prove another noise bound for certain interesting classes of gates. We show that quantum circuits using perfect Clifford operations (CNOT, Hadamard, S, X, Y, Z and measurements in the computational basis) and noisy 1-qubit unitaries cannot be made fault-tolerant if the depolarizing noise on the 1-qubit gates is at least $\hat{\theta} = (6 - 2\sqrt{2})/7 \approx 45\%$. We further show that if we additionally allow noise-free measurements in the computational basis, perfect classical control and perfect classical side processing, then above the noise rate of $\hat{\theta}$ the circuits become efficiently simulatable on a classical computer. This last result is tight, since at lower noise rates this gate set becomes *quantum universal*, that is, it is possible to simulate any other quantum circuit efficiently. A corollary of our approach is that circuits consisting only of gates from the Clifford group cannot be universal for classical computation.

5.1 Introduction

In the previous chapters we have proved very general upper bounds on the noise rates for which fault-tolerant computing is possible. They were very general, because we allowed all possible gates (Chapter 3) respectively all unitary gates (Chapter 4) with bounded fan-in. All multi-qubit gates had the same amount of noise.

Restricted gate sets

However, this might not reflect the actual properties of a physical system. It might be hard in practice to physically perform all possible quantum operations. And among those which can be implemented it might be that certain quantum operations, including storing qubits, might have lower noise rates or even no noise at all. This particular situation was studied by Bravyi and Kitaev in [21], where they argue that some realistic proposals for topological quantum computation have exactly these properties. They consider an example, where perfect stabilizer operations, perfect classical control and perfect classical side-computation is allowed (see definitions in Section 5.2). Then they prove that if it is additionally possible to create certain pure 1-qubit states, then universal fault-tolerant quantum computation becomes possible. Because of this these states were called "magic" states. However, they also show that it is not necessary to be able to prepare these magic states perfectly. It is possible to distill better magic states from several copies of noisy magic states. Thus, perfect stabilizer operations and noisy magic states are enough for universal quantum computation. We will follow a similar path in this chapter.

Model and its computational power

We will assume that we can perfectly implement the set of all stabilizer operations Stab, which includes Clifford gates Clifford (see Section 5.2.2), preparation of computational basis states and measurements in the computational basis (see Section 5.2.3). The Gottesman-Knill Theorem says that this set of gates can be efficiently simulated classically (see also [2]), and therefore it seems plausible that it is not sufficient for universal quantum computation. We prove this intuition later rigorously in Corollary 5.3.2. On the other hand, it is known that Clifford alone together with *any* other 1-qubit unitary gate, not generated by the gates in Clifford, form a universal set of gates for quantum computation [89, 66]. We show, however, that such additional 1-qubit gates should not be too noisy.

Main results

More precisely, let Clifford* be Clifford augmented with arbitrary 1-qubit unitary gates with depolarizing error at least $\hat{\theta} = (6 - 2\sqrt{2})/7 \approx 45\%$. Then this set of gates is not capable of computing arbitrary functions and therefore is not even classically universal, which is proved in Theorem 5.4.3. In particular, fault-tolerant (quantum) computation cannot be performed if there is at least this level of noise. Our second result in Theorem 5.4.4 states that circuits with arbitrary classical control and that use gates from Stab and 1-qubit unitaries with noise at least $\hat{\theta}$ can be simulated efficiently on classical computers. This last result is tight, as we explain in Section 5.5, based on results from [83, 21]. Their results imply that at noise rates less than $\hat{\theta} \approx 45.3\%$ it is possible to do efficient universal quantum computation if perfect stabilizer operations, perfect classical control and perfect classical side computation are available.

On the way, we give a characterization of the convex closure of all 1-qubit Clifford operations (Lemma 5.4.1).

Outline of proof ideas

We first show in Section 5.3 that the set of all Clifford operations is not universal, i.e., that it is is impossible to compute every function with bounded error. In particular, in Corollary 5.3.2, we show that a boolean function which can be computed by Clifford circuits can be written as the parity of a subset of input bits (complementing results in [2]). The argument uses results from communication complexity.

We then show in Lemma 5.4.1 that all 1-qubit unitaries with noise at least $\hat{\theta}$ can be seen as probabilistic mixtures of 1-qubit Clifford operations. In the proof we first compute the smallest polytope P that contains all 1-qubit Clifford gates. Then we show that any 1-qubit unitary with noise at least $\hat{\theta}$ lies inside P. Together with the fact that Clifford operations alone are not universal this establishes the first result Theorem 5.4.3.

The same Lemma together with the Gottesman-Knill theorem implies our second result Theorem 5.4.4.

Best gate

It is interesting to point out that among all 1-qubit unitary gates, the so-called $\pi/8$-gate (see Section 5.6) is the gate that requires the most noise to render it incapable of universal quantum computation by our approach. That is, augmenting the Clifford gates Clifford with other gates (e.g., $\pi/16$-gates), our approach will yield stronger bounds on the tolerable noise level.

5.1.1 Organization

This chapter is organized as follows: In the beginning of Section 5.2 we introduce some notation and review some standard facts about Bloch-sphere representations from Section 2.5.1 and explain how depolarizing noise acts on the Bloch sphere. We then introduce the Clifford group and stabilizer operations in Section 5.2.2. Section 5.3 contains the result that the gate set Clifford cannot be universal and only allows to compute parity functions, see Corollary 5.3.2. The proof uses a reduction to communication complexity (introduced in Section 2.4) and the fact that there are functions with non-trivial communication complexity. Section 5.4 shows that gates from Clifford*, together with all stabilizer operations and perfect classical control are classically simulatable and thus probably not quantum-universal. It can be read independently of the preceding section. Section

5.5 shows how results from [83, 21] imply a lower bound on $\hat{\theta}$. In Section 5.6 we then discuss some possible extensions, including different noise models and show that the $\pi/8$-gate is in some sense the most fault-tolerant gate.

5.2 Preliminaries and notation

Recall from Chapter 2 that E_{ij} is the all-zero matrix, except for the entry i, j which is equal to 1. For matrices $A, B \in \mathbb{R}^{3\times 3}$ we define as before the inner product $\langle A, B\rangle$ as:

$$\langle A, B\rangle = \mathrm{Tr}(A^T B) = \sum_{i,j\in\{1,2,3\}} a_{ij}b_{ij}.$$

The following fact is used repeatedly: $\langle A, BC\rangle = \langle B^T A, C\rangle$ for $A, B, C \in \mathbb{R}^{3\times 3}$.

Recall Section 2.5 where we saw that 1-qubit states $\rho \in \mathbb{C}^{2\times 2}$ are isomorphic to vectors $\mathbf{r} \in \mathbb{R}^3$ via

$$\rho = \frac{\mathbb{I}_2 + \mathbf{r}\cdot\sigma}{2} = \frac{\mathbb{I}_2 + r_x X + r_y Y + r_z Z}{2},$$

and 1-qubit unitary operations $U \in \mathbb{C}^{2\times 2}$ are isomorphic to rotations $R \in SO(3)$ via

$$U_{\mathbf{n}}(\theta) = \exp\left(\frac{-i\theta\mathbf{n}\cdot\sigma}{2}\right) = \mathbb{I}_2\cos\frac{\theta}{2} - i\mathbf{n}\cdot\sigma\sin\frac{\theta}{2},$$

where $\mathbf{n} \in \mathbb{R}^3$ with $||\mathbf{n}|| = 1$ is the axis and $\theta \in \mathbb{R}$ the angle of the rotation R. We introduce some notation reflecting this isomorphism. For unitary $U \in \mathbb{C}^{2\times 2}$ we let $R_U \in SO(3)$ be the corresponding rotation matrix. We get a reverse operation (up to phase factors) by fixing one mapping $f : SO(3) \to \mathbb{C}^{2\times 2}$ with the property that for all unitary $U \in \mathbb{C}^{2\times 2}$ it holds that $f(R_U) = \alpha U$ for some $\alpha \in \mathbb{C}$, $|\alpha| = 1$. We then write $U_R = f(R)$.

This can be extended to probabilistic mixtures of quantum operations. Let $\{p_i\}$ be a probability distribution, i.e., $\sum_i p_i = 1$ and $0 \leq p_i$, and let $U_i \in \mathbb{C}^{2\times 2}$ be a 1-qubit unitary with corresponding Bloch representation $R_i \in \mathbb{R}^{3\times 3}$. Then the quantum operation E in which each U_i is applied with probability p_i has Bloch-representation $R_E = \sum_i p_i R_i$.

5.2.1 Noise

The noise model we consider is again *depolarizing noise*. We repeat its definition from page 34. A 1-qubit state ρ to which depolarizing noise p is applied, becomes

$$\rho \mapsto (1-p)\rho + p\mathbb{I}/2.$$

Thus, with probability $1-p$ the state is not changed, and with probability p the state is replaced with the completely mixed state.

It is not hard to see that applying depolarizing noise p to $\rho = \mathbb{I}/2 + \mathbf{r} \cdot \sigma/2$ yields $\rho' = \mathbb{I}/2 + \mathbf{r}' \cdot \sigma/2$, with $\mathbf{r}' = (1-p)\mathbf{r}$. So, this noise shrinks the Bloch vector of a state to $(1-p)$ of its original length.

We say that a 1-qubit gate implements the unitary operation U with noise p if it transforms states ρ into

$$(1-p)U\rho U^\dagger + p\mathbb{I}/2. \tag{5.1}$$

This quantum operation can be seen as a two-stage process, in which first U and then depolarizing noise is applied. Let $R_U \in \mathbb{R}^{3\times 3}$ be the rotation matrix corresponding to the unitary U. Then this noisy quantum operation has Bloch-representation $(1-p)R_U$, i.e., it rotates a Bloch vector and scales it by a factor $1-p$.

For 1-qubit gates and depolarizing noise, the two representations are (up to unimportant global phase factors) equivalent. See Section 8.3 in [68] for more details.

5.2.2 Clifford group

The (n-qubit) *Clifford group* contains all unitary operations that can be written as a product of tensor products of S, H and CNOT_2^1 (see equation (5.2)).

$$\mathrm{CNOT}_2^1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \qquad H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \qquad S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \tag{5.2}$$

We denote the set of all operations which can be generated in this way by Clifford. In particular, the Clifford group contains also the Pauli group ($S^2 = Z$, $HS^2H = X$ and $ZX = iY$), which is the tensor product of all Pauli operators $\mathbb{I}, X, Y, Z$.

Let Clifford* be the set of gates consisting of Clifford and arbitrary 1-qubit gates followed by depolarizing noise at least $\hat{\theta} = (6 - 2\sqrt{2})/7$.

Bloch-vector representation of Clifford operations For a state with Bloch vector $\mathbf{r}$ we get:

$$S\left(\frac{1}{2}\mathbb{I} + \frac{r_x}{2}X + \frac{r_y}{2}Y + \frac{r_z}{2}Z\right)S^* = \frac{1}{2}\mathbb{I} - \frac{r_y}{2}X + \frac{r_x}{2}Y + \frac{r_z}{2}Z$$

Let R_S be the Bloch representation of S. Then R_S rotates Bloch vectors around the z-axis by $\pi/2$. In particular, the x-axis is mapped to y and y to $-x$. For the Hadamard-gate we similarly have

$$H\left(\frac{1}{2}\mathbb{I} + \frac{r_x}{2}X + \frac{r_y}{2}Y + \frac{r_z}{2}Z\right)H^* = \frac{1}{2}\mathbb{I} + \frac{r_z}{2}X - \frac{r_y}{2}Y + \frac{r_x}{2}Z.$$

So the Bloch representation R_H of H negates the y-coordinate of a Bloch vector and swaps the x and z-coordinates, i.e., it is a rotation by π around the axis $(1, 0, 1)/\sqrt{2}$.

We define $\mathcal{C}$ as the set of matrices which can be generated from R_S and R_H. A $C \in \mathcal{C}$ is called a *Clifford (rotation) matrix*. It is not hard to see that $\mathcal{C}$ contains exactly those rotations which map axes to axes (or their opposite). Those C have in each row and column exactly one non-zero entry, which must be either $+1$ or -1, and $det(C) = 1$. Note that $\mathcal{C}$, being isomorphic to the 1-qubit Clifford group, is a group under matrix multiplication. Examples of Clifford matrices are

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$$

In Appendix B we need (and explain) more details about Clifford rotation matrices, which are only necessary for one technical result in Lemma 5.4.1, which can alternatively be obtained by computer software [46].

5.2.3 Stabilizer operations and the Gottesman-Knill Theorem

We conclude this section by defining a few more terms. The set of stabilizer operations , denoted by Stab, contains all operations generated by the Clifford group Clifford and additionally preparations of computational basis states and measurements in the computational basis. The *Gottesman-Knill Theorem* says that this set of gates can be efficiently simulated classically, see also [2].

Classical control (see page 35) means that later gates may arbitrarily depend on earlier measurement outcomes. In particular, this means that arbitrary *classical side computation* is allowed. In [2] it is shown that quantum circuits using only operations from Stab in which perfect classical control is allowed are also efficiently simulatable on a classical computer, i.e., the simulation can be done with at most a polynomial overhead over the number of quantum operations and the number of (classical) operations needed for the classical control.

5.3 The power of Clifford circuits

The main idea of this section is as follows. Assume we have a Clifford circuit C (i.e. a circuit composed of the gates in (5.2)) with n classical input bits $x = x_1, \ldots, x_n$ and one dedicated output qubit that, when measured in the computational basis, yields the output of the computation of C on x. Suppose now that the input is partitioned over two parties, Alice and Bob, such that Alice has bits $S \subseteq \{1, \ldots, n\}$ of x and Bob has bits $\{1, \ldots, n\} \backslash S$. We first show how Alice, with the help of Bob, can compute the value of C on x with just a single classical bit of communication (Lemma 5.3.1) for any partition S. Recall that in Section 2.4 we defined the worst-case partition communication complexity of $f : \{0,1\}^n \to \{0,1\}$ as $D^{worst}(f) = \max_{S\subseteq\{1,\dots,n\}} D^S(f)$, where $D^S(f)$ is the (deterministic) communication complexity of f when the bits in S are given to Alice and all others to Bob. Hence, Clifford circuits can at the very best compute only those functions that require a single bit of communication for any partition of the inputs; it is well known that most functions require more than one bit of communication, see Section 2.4.

We are now ready to prove the main lemma, which explains the idea of simulating Clifford circuits.

5.3.1. Lemma. *Let $f : \{0,1\}^n \to \{0,1\}$ be a function that is computable with unbounded error*[1] *by a quantum circuit C that uses only gates from* Clifford, *ancillas initialized to $|0\rangle$ and one single-qubit measurement in the computational basis, which determines the output. Then the deterministic worst-case partition communication complexity of f is at most one bit.*

Proof: In the simulation of the circuit C we represent the j-th qubit by two *shares*: a *classical share* consisting of two bits a_j, b_j, and a *quantum share* consisting of 1-qubit. A state $|\psi_C\rangle$ of C will be encoded by

$$|\psi_C\rangle := \bigotimes_j X_j^{a_j} Z_j^{b_j} |\psi\rangle, \tag{5.3}$$

where $|\psi\rangle$ is the state of all quantum shares and the indices j denote the qubits on which the operators X and Z act. We call $|\psi_C\rangle$ the *logical state* (of C).

Assuming that the set of qubits of C is encoded in this manner, the operations H, S, and CNOT can be applied to the logical qubits by *separately* performing operations on the shares that encode them (i.e., the logical qubits do not have to be reconstructed). The reason why this works is because for any Clifford operation $C = H, S, \mathrm{CNOT}_2^1$ and any tensor product of Pauli operators P_1 there is a tensor product of Pauli operators P_2 with $CP_1 = P_2C$. For example, to apply H to the logical qubit i, the two bits that make up its classical share are swapped and H is applied to its quantum share. This works correctly because

$$\begin{aligned} H_j X_j^{a_j} Z_j^{b_j} |\psi\rangle &= H_j X_j^{a_j} H_j H_j Z_j^{b_j} H_j H_j |\psi\rangle \\ &= Z_j^{a_j} X_j^{b_j} H_j |\psi\rangle \\ &= (-1)^{a_j \wedge b_j} X_j^{b_j} Z_j^{a_j} H_j |\psi\rangle, \end{aligned} \tag{5.4}$$

and $(-1)^{a_j\wedge b_j}$ is an irrelevant global phase.

[1] That means, that the output is only correct with probability greater than 1/2, but can go arbitrarily close to 1/2.

To apply S to a logical qubit, the b-part of the classical share is updated to $b := a \oplus b$ and S is applied to its quantum share. This case can be verified by noting that

$$\begin{aligned} S_j X_j^{a_j} Z_j^{b_j} |\psi\rangle &= i^{a_j} X_j^{a_j} S_j Z_j^{a_j} Z_j^{b_j} |\psi\rangle \\ &= i^{a_j} X_j^{a_j} Z_j^{a_j \oplus b_j} S_j |\psi\rangle, \end{aligned} \tag{5.5}$$

where we note that i^{a_j} is a global phase.

To simulate the application of a CNOT_2^1 gate on two logical qubits, with classical shares $a_1 b_1$ and $a_2 b_2$, we update $a_2 := a_1 \oplus a_2$, $b_1 := b_1 \oplus b_2$ and CNOT_2^1 is applied to the two quantum shares. In this case, we omit the details but note that the correctness can be verified using the identities (see also Figure 4.3 on page 54)

$$\begin{aligned} \mathrm{CNOT}_2^1(X \otimes I) &= (X \otimes X)\mathrm{CNOT}_2^1 \\ \mathrm{CNOT}_2^1(I \otimes X) &= (I \otimes X)\mathrm{CNOT}_2^1 \\ \mathrm{CNOT}_2^1(Z \otimes I) &= (Z \otimes I)\mathrm{CNOT}_2^1 \\ \mathrm{CNOT}_2^1(I \otimes Z) &= (Z \otimes Z)\mathrm{CNOT}_2^1. \end{aligned} \tag{5.6}$$

We first describe a *probabilistic* communication protocol for f. Alice operates on the classical shares while Bob operates on the quantum shares.

The initial shares are easy to construct, see also Figure 5.1: for each of Alice's input bits x_j, Alice sets her classical share to $a_j := x_j$,$b_j := 0$ and Bob sets his quantum share to $|0\rangle_j$; for each of Bob's input bits y_j, Alice sets her classical share to $a_j = b_j := 0$ and Bob sets his quantum share to $|y_j\rangle_j$. If the j-th input bit to circuit C is an ancilla qubit (initialized to $|0\rangle$) then Alice sets $a_j := x_j$,$b_j := 0$ and Bob sets the j-th qubit to $|0\rangle_j$. Note that the logical state encoded in this way is $|x\rangle|y\rangle|0\dots0\rangle$, where $|0\dots0\rangle$ denotes the ancilla qubits.

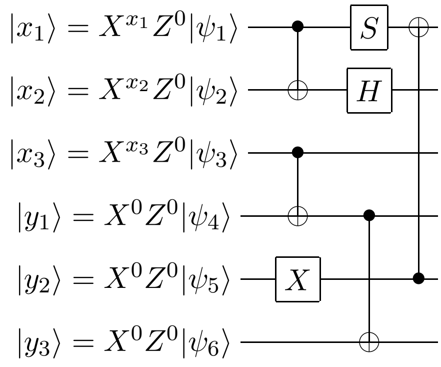

$|x_1\rangle = X^{x_1}Z^0|\psi_1\rangle$
$|x_2\rangle = X^{x_2}Z^0|\psi_2\rangle$
$|x_3\rangle = X^{x_3}Z^0|\psi_3\rangle$
$|y_1\rangle = X^0Z^0|\psi_4\rangle$
$|y_2\rangle = X^0Z^0|\psi_5\rangle$
$|y_3\rangle = X^0Z^0|\psi_6\rangle$
S
H
X

Figure 5.1: Distributed representation of input bits

With this representation, Alice and Bob can simulate the execution of circuit C on input $|x\rangle|y\rangle|0\dots0\rangle$ *without any communication* as explained above. In particular, they can obtain the shares of the output qubit of C, which without loss of generality we assume to be the first qubit of C. For Bob to obtain the result of measuring the (logical) output qubit, Alice sends the first bit of her classical share, a_1, to Bob, who applies X^{a_1} to his quantum share and measures it in the computational basis. Alice need not send b_1, the second bit of the classical share, since Bob is performing a measurement in the computational basis.

Finally, to obtain a *deterministic* communication protocol for f, we note that Bob need not actually manipulate quantum information; rather, he can simulate his quantum registers and his operations with high enough precision on a classical computer. Then, upon receipt of the classical bit from Alice, he can exactly determine the output probabilities of his measurement to determine which outcome is more likely. ■

The next Corollary characterizes exactly all functions computable by Clifford circuits. From Lemma 5.3.1 we get that this set is very limited and far from being universal.

5.3.2. Corollary. *All functions $f : \{0,1\}^n \to \{0,1\}$ which can be computed by a Clifford circuit, can be written in the form*

$$f(x_1 \dots x_n) = c \oplus \bigoplus_{j \in S} x_j,$$

where $S \subseteq [n]$ is a subset of the input bits not depending on the input bits and $c \in \{0,1\}$.

Proof: It is clear that all functions f of this form can be computed by a Clifford circuit. We now also prove the reverse.

Let $f : \{0,1\}^n \to \{0,1\}$ be a function which can be computed by a Clifford circuit C. Then we can simulate this circuit as in Lemma 5.3.1, where we give Alice the whole input, i.e., with the notation before Lemma 5.3.1 this means $S = \{1,\dots,n\}$.

Inspecting the proof of Lemma 5.3.1 we see that in each step Alice always updates her a_i's and b_i's by computing the parity of two bits. So, the final bit a_1 she sends over is just the parity of some of the input bits. Thus we can write $a_i = \bigoplus_{j \in S} x_j$, for some $S \subseteq [n]$. Bob initializes all his quantum bits to $|0\rangle$, so he starts with the state $|\psi^0\rangle = |0 \dots 0\rangle$. Further, Bob just applies the circuit C to his state and measures the i-th qubit of $X^{a_i} C|\psi^0\rangle$ in the computational basis.

It is known that the probability for measuring 1 in a Clifford circuit is either 0, 1/2 or 1 (see [68] page 463). It cannot be 1/2 in our case, because that would mean that the circuit does not compute f. So, measuring the i-th bit of $C|\psi^0\rangle$ yields a bit $c \in \{0,1\}$ with certainty. But this means that $f(x) = c \oplus a_i = c \oplus \bigoplus_{j \in S} x_j$. ■

We mention that Aaronson and Gottesman proved [2] that there is a log-space machine which transforms a Clifford circuit C into a classical circuit C'

consisting only of CNOT and NOT gates, with the property that C accepts the all zero state $|0\rangle^{\otimes n}$ iff C' accepts the (classical) all zero input. Our corollary extends this slightly: For every Clifford circuit C computing a boolean function, there is an equivalent (for classical inputs) classical circuit which uses only NOT- and CNOT-gates. Using the result from [2] we see that we can compute the bit c in the proof of Corollary 5.3.2 in log-space and it is also clear that the circuit Alice uses to compute a_i can be computed in log-space.

5.3.3. Remark. It is trivial to extend Lemma 5.3.1 to functions with m output bits, if the communication complexity of the function is also higher than m, resulting in a scheme that uses m bits of communication.

5.4 Simulating 1-qubit unitaries by Clifford gates

We want to extend Lemma 5.3.1, by replacing Clifford with Clifford*. We show in Lemma 5.4.1 how probabilistic mixtures of Clifford gates can be used to simulate any single qubit unitary gate that has noise $\hat{\theta}(\approx 45\%)$. The proof relies on solving an optimization problem related to the Clifford polytope, defined as the convex hull of the set $\mathcal{C} \subseteq \mathbb{R}^{3\times 3}$ of Clifford rotation matrices in $\mathbb{R}^3$. Here, the matrices $\mathcal{C}$ are the 1-qubit Clifford gates in Bloch sphere representation.

Combining Lemmas 5.3.1 and 5.4.1, we get that for all circuits with Clifford*-gates and any distribution of its input bits among Alice and Bob, the output of the circuit can be obtained with a single bit of communication (Lemma 5.4.2). Using the fact that there are functions which require communication more than one bit, we get our main result (Theorem 5.4.3): The set of gates in Clifford* cannot be universal. We also generalize our result to the case that the inputs are quantum states.

We first show how one can simulate arbitrary 1-qubit gates with depolarizing noise $\hat{\theta} = (6 - 2\sqrt{2})/7$ with a probabilistic mixture of Clifford operations.

5.4.1. Lemma. *Let U be a 1-qubit unitary and E_U be the following noisy version of it*

$$\rho \mapsto E_U(\rho) = (1-\hat{\theta})U\rho U^* + \hat{\theta}\mathbb{I}/2,$$

for any $\rho \in \mathbb{C}^{2\times 2}$. Then there is a probability distribution $\{p_C\}$ over $\mathcal{C}$ such that for all $\rho \in \mathbb{C}^{2\times 2}$ we have

$$E_U(\rho) = \sum_{C\in\mathcal{C}} p_C U_C \rho U_C^*$$

and U_C is a Clifford operation corresponding to the Clifford rotation matrix C.

Proof: Using Section 2.5.1 and Section 5.2 the lemma can be reformulated equivalently in Bloch representation: For any $S \in SO(3)$ there is a probability distribution $\{p_C\}$ over $\mathcal{C}$ such that

$$(1-\hat{\theta})S = \sum_{C \in \mathcal{C}} p_C C. \tag{5.7}$$

We will prove this latter statement. Define the *Clifford polytope*

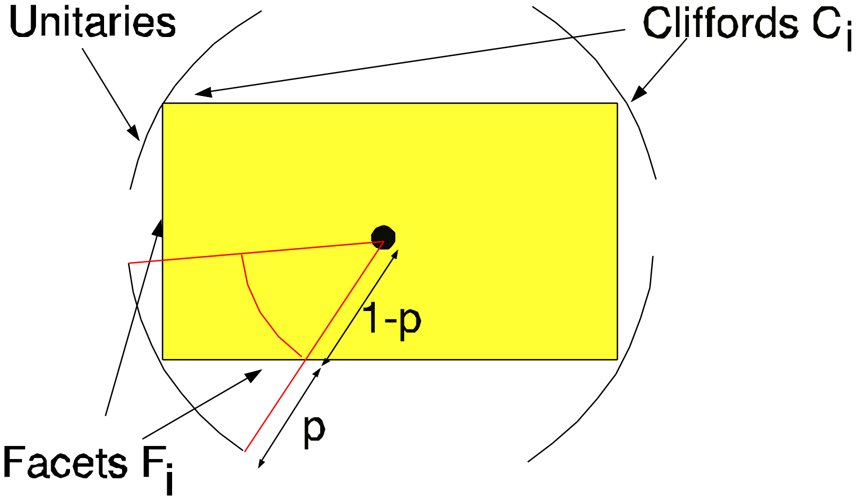


Figure 5.2: The polytope P (schematically in two dimensions)

Rotation matrices (corresponding to 1-qubit unitaries) are depicted by patches of a circle. The polytope P spanned by the Clifford operators C_i is depicted by the rectangle, with facets $F_i \in \mathcal{F}$. For every rotation matrix S on the circle there is a smallest value p such that shrinking S by a factor $(1-p)$ gives a point inside P. Then $\hat{\theta}$ is the maximum of such p over all rotation matrices S.

$$P := \mathrm{conv}(\mathcal{C}) = \left\{ S \mid S = \sum_{C \in \mathcal{C}} p_C C, p_C \geq 0, \sum_{C \in \mathcal{C}} p_C = 1 \right\} \tag{5.8}$$

as the convex hull of the 24 Clifford rotation matrices in $\mathbb{R}^{3\times 3}$. We have to prove (see also Figure 5.2)

$$(1-\hat{\theta})S \in P \quad \text{for any } S \in SO(3). \tag{5.9}$$

For this we use the fact that the Clifford polytope can be alternatively described by its facet description:

$$P = \left\{ S \in \mathbb{R}^{3\times 3} \mid \langle F, S\rangle \leq 1 \text{ for all } F \in \mathcal{F} \right\}, \tag{5.10}$$

where

$$\mathcal{F} := \left\{ C_1 B C_2 | C_1, C_2 \in \mathcal{C}, B \in \{B_1, B_1^T, B_2\} \right\}, \tag{5.11}$$
$$B_1 := \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \; B_2 := \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

One can use the software from [46] for computing the facet description (5.10); we will give a direct proof in Appendix B. In view of (5.10), our claim (5.9) is equivalent to

$$(1-\hat{\theta})\langle F, S\rangle \leq 1 \text{ for all } S \in SO(3), \; F \in \mathcal{F}. \tag{5.12}$$

Let $F \in \mathcal{F}$ be of the form $F = C_1 B C_2$ where $C_1, C_2 \in \mathcal{C}$. As $\langle F, S\rangle = \langle C_1^T S C_2^T, B\rangle$ and $C_1^T S C_2^T \in SO(3)$, (5.12) is equivalent to

$$\langle S, B\rangle \leq \frac{1}{1-\hat{\theta}} = 2\sqrt{2} - 1 \text{ for all } B \in \{B_1, B_2\}, \; S \in SO(3). \tag{5.13}$$

The case $B = B_1$ is easy to handle: For $S \in SO(3)$, $\langle S, B_1\rangle = \sum_{i=1}^{3} S_{i1} \leq \sqrt{3} < 2\sqrt{2} - 1$. We now show (5.13) for $B = B_2$. Write $S \in \mathbb{R}^{3\times 3}$ as

$$S = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}. \tag{5.14}$$

It is well-known that it is necessary and sufficient for $S \in SO(3)$ that the column vectors $\mathbf{a} = (a_1, a_2, a_3)^T, \mathbf{b} = (b_1, b_2, b_3)^T$ and $\mathbf{c} = (c_1, c_2, c_3)^T$ satisfy

$$\mathbf{a}^T\mathbf{b} = 0, \; \mathbf{c} = \mathbf{a} \times \mathbf{b}, \; \mathbf{a}^T\mathbf{a} = 1, \; \mathbf{b}^T\mathbf{b} = 1, \tag{5.15}$$

where × denotes the vector product, defined as

$$\mathbf{a} \times \mathbf{b} := (a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1)^T.$$

Recall that, for $\mathbf{a}, \mathbf{b}, \mathbf{c}$ as in (5.15), $\mathbf{a} = \mathbf{b} \times \mathbf{c}$ and $\mathbf{b} = \mathbf{c} \times \mathbf{a}$. Using $c_3 = a_1 b_2 - a_2 b_1$, we obtain $\langle B, S\rangle = a_1 - a_2 + b_1 + b_2 - a_1 b_2 + a_2 b_1$. Therefore our task is now to prove that the optimum value of the program

$$\begin{array}{ll} \max & f := a_1 - a_2 + b_1 + b_2 - a_1 b_2 + a_2 b_1 \\ \text{s.t.} & g_1 := a_1^2 + a_2^2 + a_3^2 = 1 \\ & g_2 := b_1^2 + b_2^2 + b_3^2 = 1 \\ & g_3 := a_1 b_1 + a_2 b_2 + a_3 b_3 = 0 \end{array} \tag{5.16}$$

is at most $2\sqrt{2}-1$; we in fact show that $\max f = 2\sqrt{2}-1$. For this, consider a global maximizer (a,b) to the program (5.16). Then, the Karush-Kuhn-Tucker conditions have to be satisfied, since the gradient vectors $\{\nabla g_i(a,b) \mid i=1,2,3\}$ are linearly independent; see, e.g., Theorem 12.1 in [70]. (Here the gradient vector $\nabla g_i(a,b) = (\frac{\partial}{\partial_{a_1}}, \frac{\partial}{\partial_{a_2}}, \frac{\partial}{\partial_{a_3}}, \frac{\partial}{\partial_{b_1}}, \frac{\partial}{\partial_{b_2}}, \frac{\partial}{\partial_{b_3}})^T g_i(a,b)$ consists of the partial derivatives with respect to the six variables $a_1,\ldots,b_3$.) That is, there exist scalars $\lambda_1,\lambda_2,\lambda_3$ for which

$$\nabla f(a,b) + \sum_{i=1,2,3} \lambda_i \nabla g_i(a,b) = 0.$$

Equivalently, considering the partial derivatives first with respect to (a_1,a_2,a_3) and then with respect to (b_1,b_2,b_3)

$$\begin{array}{ccccccc} \begin{pmatrix} 1-b_2 \\ -1+b_1 \\ 0 \end{pmatrix} & + & 2\lambda_1\mathbf{a} & & + & \lambda_3\mathbf{b} & = & 0 \\ \begin{pmatrix} 1+a_2 \\ 1-a_1 \\ 0 \end{pmatrix} & & & + & 2\lambda_2\mathbf{b} & + & \lambda_3\mathbf{a} & = & 0. \end{array}$$

Multiplying the first and the second line by $\mathbf{c}^T = (\mathbf{a}\times\mathbf{b})^T$ (recall that $\mathbf{c}\perp\mathbf{a},\mathbf{b}$) we get

$$\begin{aligned} 0 &= c_1(1-b_2) + c_2(-1+b_1) = c_1 - c_2 + a_3 \\ 0 &= c_1(1+a_2) + c_2(1-a_1) = c_1 + c_2 + b_3. \end{aligned}$$

Adding (resp. subtracting) these equations yields $2c_1 = -a_3 - b_3$ and $2c_2 = a_3 - b_3$. Squaring these two equations and then adding them gives $2a_3^2 + 2b_3^2 = 4c_1^2 + 4c_2^2$. Since the rows and columns in S are normalized, we get $2(1-c_3^2) = 4(1-c_3^2)$, from which we conclude $c_3^2 = 1$ and, therefore, $a_3 = b_3 = c_1 = c_2 = 0$. This implies $a_1^2 + b_1^2 = 1 = a_1^2 + a_2^2$ and thus $|b_1| = |a_2|$. Similarly one can establish $|a_1| = |b_2|$. On the basis of this observation we distinguish three cases.

1. $a_1 = b_2 = 0$. Then, $|a_2| = |b_1| = 1$ and $f = -a_2 + b_1 + a_2 b_1 \leq 1$.

2. $a_1 \neq 0$ and $a_1 = -b_2$. From $a^T b = 0$ we have $a_1(b_1 - a_2) = 0$, which gives $a_2 = b_1$. Then, $f = a_1 - a_2 + a_2 - a_1 + a_1^2 + a_2^2 = 1$.

3. $a_1 \neq 0$ and $a_1 = b_2$. From $a^T b = 0$ we have $a_1(b_1 + a_2) = 0$, which gives $a_2 = -b_1$. Then, $f = a_1 - a_2 - a_2 + a_1 - a_1^2 - a_2^2 = 2(a_1 - a_2) - 1$, which (under the condition $a_1^2 + a_2^2 = 1$) is clearly maximized by $a_1 = -a_2 = 1/\sqrt{2}$. Therefore, we find $\max f = 2\sqrt{2} - 1$.

Thus, we have shown that the optimum value of the program (5.16) is equal to $2\sqrt{2}-1$, which concludes the proof. ■

5.4.2. Lemma. *Let $f : \{0,1\}^n \to \{0,1\}$ be a function and K a quantum circuit for f with error probability at most $\epsilon > 0$ which uses only gates from* Clifford* *and one final single qubit measurement in the computational basis. Then the randomized worst-case partition communication complexity (defined in Section 2.4 on page 27) of f is at most one bit, i.e., $R_\epsilon^{worst}(f) \leq 1$.*

Proof: From Lemma 5.3.1 we know how two parties, Alice and Bob, can simulate perfect Clifford gates. From Lemma 5.4.1 we know how they can jointly simulate the other noisy 1-qubit gates in Clifford*, where they use shared randomness to make sure that they always simulate the same Clifford gate. ■

We can now prove an upper bound on the noise in fault-tolerant quantum computation.

5.4.3. Theorem. *The set of gates from* Clifford *together with 1-qubit gates with depolarizing noise more than $\hat{\theta} \approx 45\%$ and one single-qubit measurement is not sufficient for arbitrary classical computation.*

Proof: The result follows by Lemma 5.4.2 and the fact that there are functions with communication complexity greater than 1, for any bounded error. ■

In fact we have that none of the functions f with $R_\epsilon(f) > 1$ can be computed by Clifford* circuits with error at most ϵ. From Corollary 5.3.2 we also get that the functions computable by Clifford* are always probabilistic mixtures of parity functions.

If we additionally allow perfect stabilizer operations Stab and perfect classical control, we can state the following theorem.

5.4.4. Theorem. *Any computation using*

1. *gates* Clifford*
2. *perfect stabilizer operations* Stab
3. *perfect classical control and classical side computation*

in which the input state is a computational basis state, can be simulated by a classical computer with at most polynomial overhead.

Proof: Follows immediately by Lemma 5.4.1 and the Gottesman-Knill Theorem. ■

5.5 Lower bound on $\hat{\theta}$

We now discuss wether it is possible to improve Theorems 5.4.3 and 5.4.4.

Theorem 5.4.3 states that fault-tolerant quantum computing is not possible if we have depolarizing noise at least $\hat{\theta} \approx 45\%$ on 1-qubit gates even if we can use perfect gates from Clifford in our fault-tolerant circuit design. Is this optimal? Could it be that with less than $\hat{\theta}$ noise on the single-qubit gates and perfect gates from Clifford still no fault-tolerant circuit design is possible? This is still an open question since we do not know if Theorem 5.4.3 is tight.

In contrast to this, the second result (Theorem 5.4.4) is tight, which was pointed out by Ben Reichardt [80]. The argument builds upon magic-state distillation, introduced in [21], and goes as follows. Assume we have at our disposal noisy $\pi/8$-gates T', with depolarizing noise strictly less than $\tilde{\theta}$, i.e. $T'(\rho) = (1-p)T\rho T^\dagger + p\mathbb{I}/2$ with $p < \hat{\theta}$, where T is the perfect $\pi/8$ gate, see equation (5.19). Then apply T' to the second half of an EPR-pair and measure the observable $Z \otimes Z$, which can be implemented as a measurement in the computational basis with additional gates from Clifford. If the outcome is -1 throw away the state and do the experiment again. If the outcome is $+1$, apply a CNOT from the first to the second qubit, which gives

$$\frac{1}{2}\left(\mathbb{I} + \frac{1-p}{1-p/2}\frac{1}{\sqrt{2}}X + \frac{1-p}{1-p/2}\frac{1}{\sqrt{2}}Y\right) \otimes |0\rangle\langle 0|. \tag{5.17}$$

Using the result from [83, 82] an arbitrary supply of qubits in the state of the first qubit of (5.17) can be used to distill magic states in the H-direction, which together with stabilizer operations is sufficient for quantum computation. We do not know if this also holds for gates other than the $\pi/8$-gate.

5.6 Discussion and extensions

In this section we will discuss certain extensions and generalizations of our results.

Best gates

From the proof of Lemma 5.4.1 we see that the rotation matrix S which achieves the optimal value, is

$$\begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{5.18}$$

Multiplying from the right by the Clifford-matrix $diag(1,-1,-1)$ we get a rotation around the z-axis by $\pi/4$. The $\pi/8$-gate

$$T = \begin{pmatrix} \exp(-i\pi/8) & 0 \\ 0 & \exp(i\pi/8) \end{pmatrix} \tag{5.19}$$

performs a rotation of $\pi/4$ around the z-axis. So, the $\pi/8$-gate and its symmetric versions are the ones which need the most depolarizing noise to be simulated by gates from Clifford.

Worst case noise

In Lemma 5.4.1 we asked with how much depolarizing noise all 1-qubit unitary gates are equivalent to probabilistic mixtures of Clifford gates. Similarly to [99] one can also ask how much arbitrary noise is needed to make every gate a mixture of Cliffords. More precisely what is the value $\tilde{\theta} = \sup_{U \in SU(2)} p_U$, where p_U is the infimum of all p such that there is a completely positive trace-preserving 1-qubit quantum operation $\mathcal{E}_U$ with the property that the noisy implementation of U

$$U' : \rho \mapsto (1-p)U\rho U^\dagger + p\mathcal{E}_U(\rho)$$

becomes a probabilistic mixture of Clifford operation.

In this section we will provide some bounds on $\tilde{\theta}$. Let $K \in SU(2)$ be any operation that in Bloch representation maps the X-eigenstate $v_X = (1,0,0)$ to $u = \frac{1}{\sqrt{3}}(1,1,1)$. Note that a probabilistic mixture of 1-qubit Clifford operations $C = \sum_i p_i C_i$ can map v_X only into the octahedron $\mathcal{O}$ spanned by $v_X = (1,0,0)$, $v_Y = (0,1,0)$ and $v_Z = (0,0,1)$ and their negatives $-v_X, -v_Y, -v_Z$ (see also [21]). Note that the state of $\mathcal{O}$ which is closest to u is $\frac{1}{3}(1,1,1) = \frac{1}{\sqrt{3}}u$ and their distance is $||u - u/\sqrt{3}||_2 = 1 - \frac{1}{\sqrt{3}}$. The Bloch-state which is furthest away from u is $-u$. All three of these states lie on a line. With this it is clear that the state u_{noise} which needs the smallest noise p, such that $(1-p)u + pu_{noise}$ is inside the octahedron, is $-u$ and the optimal p is $\frac{1}{2}(1 - \frac{1}{\sqrt{3}})$. This implies $21\% \approx \frac{1}{2}(1 - \frac{1}{\sqrt{3}}) \leq \tilde{\theta}$.

To get an upper bound, recall that by Lemma 5.4.1 for any gate $U \in SU(2)$ the operation

$$U' : \rho \mapsto (1-p)U\rho U^\dagger + p\mathbb{I}/2$$

is a Clifford operation, if $p \geq \hat{\theta}$. Setting

$$\mathcal{E}_U(\rho) = \frac{1}{3}\left(XU\rho U^\dagger X + YU\rho U^\dagger Y + ZU\rho U^\dagger Z\right)$$

and noting that for any 1-qubit density matrix ρ it holds

$$\frac{\mathbb{I}}{2} = \frac{1}{4}\left(\rho + X\rho X + Y\rho Y + Z\rho Z\right)$$

we can rewrite the action of U' also as

$$U' : \rho \mapsto (1 - \frac{3}{4}p)U\rho U^\dagger + \frac{3}{4}p\mathcal{E}_U(\rho).$$

Thus, $\tilde{\theta} \leq \frac{3}{4}\hat{\theta} \approx 34\%$. Note that this is certainly not tight, since all gates, apart from the $\pi/8$-gate (and its symmetric versions), need less than $\hat{\theta}$ depolarizing noise to make it a probabilistic mix of Clifford operations, which implies they need less than $\frac{3}{4}\hat{\theta}$ worst case noise. However, as follows from [99], the worst case noise for the $\pi/8$-gate(s) is only $\frac{1}{2} - \frac{1}{2\sqrt{2}} \approx 15\%$.

We leave it as an interesting open question to determine the precise value of $\tilde{\theta}$.

Different noise models

The approach we have taken can in principle also be applied to other noise models: For any 1-qubit noise operation $\mathcal{E}$, with Bloch representation $S_\mathcal{E}$ we can compute the minimum value θ such that for all rotations $R \in \mathbb{R}^{3\times 3}$ the noisy version $(1-\theta)R+\theta S_\mathcal{E}$ is inside the Clifford polytope P, defined in equation (5.8). However, the actual optimization problems might not be as easy as for depolarizing noise, since depolarizing noise with probability p corresponds to multiplying with $(1-p)$ in Bloch-representation.

In principle, a similar approach might be possible to calculate how well one can approximate arbitrary (unitary) gates given a gate set S other than CLIFFORD* under a certain noise model. If S is not universal, this will also give new noise bounds.

Allowing some perfect unitaries

Our threshold theorem says the following. Let $f : \{0,1\}^n \to \{0,1\}$ be a function which requires more than one bit of communication in order to compute it, when the input bits are partitioned over Alice and Bob. There is no quantum circuit consisting of *perfect* Clifford operations and single qubit gates with noise $\hat{\theta}$ ($\approx 45\%$) that can compute f. We can strengthen this result to allow a small number of perfect single-qubit gates as well: Assume that f requires m bits of communication to be computed, i.e., the randomized worst-case partition communication complexity $R_\epsilon^{worst}(f)$ is at least m. Then there is no quantum circuit that uses perfect Clifford operations, s perfect single-qubit gates, and single qubit gates with noise $\hat{\theta}$ that computes f, for $2s + 1 < m$ with error at most ϵ. We get this strengthening by changing the simulation of a Clifford circuit in Lemma 5.3.1 in the following way: Whenever Alice and Bob want to perform a perfect single qubit gate on some qubit, Alice sends her classical share a, b of that specific qubit to Bob. Note that Bob now has complete control over this qubit and can perform the perfect gate on that qubit. They then proceed as in Lemma 5.3.1.

By the end of the simulation Alice has sent $2s+1$ bits to Bob and he will be able to compute f, contradicting that the communication complexity of f is at least $m > 2s+1$.

Quantum inputs

Lemma 5.3.1 can actually be extended to the case where Alice and Bob get quantum states as inputs and they are provided with entanglement. The statement is as follows: Suppose they have a quantum circuit C as in Lemma 5.3.1, but they get a quantum state $\rho \in \mathbb{C}^{2^n \times 2^n}$. Let $p_{\rho,i}$, $i = 0, 1$, be the probability that C (which uses one 1-qubit measurement in the computational basis to determine the output) outputs i on input ρ. Now, let ρ be arbitrarily partitioned between Alice and Bob, that is, Alice gets the qubits with indices $S \subseteq \{1, \ldots, n\}$ of ρ and Bob the rest and they both know S, but they do not know what ρ is. Then it is possible with one classical bit of communication from Alice to Bob that Bob outputs i with probability exactly $p_{\rho,i}$.

To see how this works let without loss of generality Alice's input qubits be $S = \{1, \ldots, m\}$, $m \leq n$. Alice and Bob then need to share m EPR pairs. We use an "aborted" teleportation scheme to set up a representation as in Lemma 5.3.1, equation (5.3): Call ρ_i the i-th qubit of ρ. Recall that during the standard protocol (see e.g. [68], page 26) of teleporting ρ_i to Bob, Alice measures at some point two classical bits a_i, b_i. In the standard protocol for teleportation she then sends these two bits to Bob, who applies $X^{a_i} Z^{b_i}$ on his share of the i-th EPR-pair and this then contains ρ_i.

Now, in our aborted version of teleportation Alice does not send the bits a_i, b_i $(1 \leq i \leq m)$. She keeps them and for $m < i \leq n$ she additionally initializes bits a_i, b_i to $a_i = b_i = 0$, so that Alice ends up with $2n$ classical bits in total. With this protocol, Alice and Bob obtain the correct representation of the state ρ from Lemma 5.3.1. More precisely, if $\rho' \in \mathbb{C}^{2^n \times 2^n}$ is the state in the $n-m$ qubits given initially to Bob and the m qubits from his shares of the EPR-pairs (after the "aborted" teleportation), then

$$\rho = \left(\bigotimes_{i=1}^{m} X_i^{a_i} Z_i^{b_i} \right)^{\dagger} \rho' \left(\bigotimes_{i=1}^{m} X_i^{a_i} Z_i^{b_i} \right),$$

where the subscript i means that the operator acts on the Hilbert space of Bob's share of the i-th EPR-pair. Note that ρ' is completely in Bob's hands. This is the same representation as in equation (5.3), just that now the quatum share ρ' is some *mixed* state, which is not known to Alice and Bob. This is necessary since we also assumed that the logical input state ρ, which is encoded in this way, can also be some arbitrary mixed state.

From here, they can then run the same protocol as in the proof Lemma 5.3.1, where in the end Alice sends one classical bit to Bob. Of course, this time Bob *has* to do the final measurement and can not just classically simulate the quantum computation since we assumed the state ρ to be arbitrary and not known to Alice and Bob. It is clear that the outcome of his final measurement will have the correct distribution $p_{\rho,i}$.

Chapter 6

Classical 2-input gates

This chapter is based on the paper

> Falk Unger, **Noise threshold for universality of 2-input gates**, presented at *IEEE International Symposium on Information Theory, 2007*, published in *IEEE Transactions on Information Theory*, 54(8), Aug 2008 (see [95])

6.1 Introduction

It is a common belief that the hardware of quantum computers will be faulty and that some kind of fault-tolerant architectures will be needed. This is reflected in the large amount of work on quantum fault-tolerance schemes. On the other hand, the situation for classical computers seems very different at first sight. The error probability of gates in modern computer chips is so small that the problem of error correction is still mostly ignored in modern computer processors. However, the situation might change in the future.

Moore's law versus faulty components

In order to make classical computers faster and faster, hardware engineers have been increasing the number of gates on computer chips steadily. The rate of this increase is roughly given by Moore's law [64]. In order to increase the number of gates, they have to be made smaller and smaller, a process known as hardware miniaturization. However, there are physical limits to the possible extent of this miniaturization, and the closer one gets to these limits, the less robust and more error-prone the components become [16, 15]. It is estimated that the time when processor architects face these limitations is within the next decade [34]. Gates can fail in (at least) two ways. The first is that they do not work at all, which could be due to manufacturing faults. For example, it could be that a gate always outputs 0. The second is that they work most of the time correctly, and fail sometimes. This type of errors is called "soft errors" by hardware engineers. In this chapter we deal with faults of the second type. We assume a probabilistic error model in which the probability of a fault of any gate is independent of the input and whether other faults have occurred.

We establish a noise threshold for computation by formulas which use gates of fan-in at most 2 (see definitions in Section 6.2).

Some known noise thresholds

We have already given a detailed overview of known classical noise bounds in Chapter 1. Recall from Chapter 1 (see also Figure 1.2 on page 9) that it is impossible to compute all boolean functions with bounded error by circuits using gates of fan-in at most k, if each gate fails independently with probability at least $\epsilon > \frac{1}{2} - \frac{1}{2\sqrt{k}}$ [37]. For formulas with gates of fan-in k and k odd, we know the tight bound $\beta_k = \frac{1}{2} - \frac{2^{k-2}}{k\binom{k-1}{k/2-1/2}}$. Tight here means that if all gates fail 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 β_k (which does not need to be the same for all gates), *universal computation* is not possible.

6.1.1 Noise threshold for fan-in 2 gates

For formulas of gates with even fan-in much less is known. In this chapter we deal with noisy gates of fan-in at most 2. Evans and Pippenger [36] showed that all functions can be computed by formulas with noisy NAND-gates with fan-in 2, if each NAND-gate fails with probability exactly ϵ, for any $0 \leq \epsilon < \beta_2 = \frac{3-\sqrt{7}}{4}$. The main result of this chapter is that this result is tight.

6.1.1. THEOREM. *Assume $\Delta > 0$. Functions that are computable with bias Δ by a formula in which all gates have fan-in at most 2 and fail independently with probability at least $\beta_2 = (3 - \sqrt{7})/4$, depend on at most a constant number of input bits.*

Together with the first mentioned result from [36] this gives the exact threshold for formulas with gates of fan-in 2. In [36] there is already a weaker version of our Theorem 6.1.1, which we extend in the following ways: (1) We allow all gates of fan-in 2, whereas in [36] the upper bound is only established for the case that all gates are NAND-gates. (2) We prove that if the noise is exactly β_2, then no universal bounded-error computation is possible. (3) In contrast to our result, the upper bound in [36] only applies to "soft" inputs. They show that gates with noise more than β_2 cannot increase the bias. More precisely, if the inputs to the formula are noisy themselves and have bias at most $\Delta > 0$, then the output of the formula cannot have larger bias than Δ. This left open the case where the input bits are not noisy and either 0 or 1, which is the case we care about most. Our argument shows that even with perfect inputs fault-tolerant computation is not possible for noise at least β_2.

To prove Theorem 6.1.1 we introduce a new technique, which is also applicable in the even fan-in case. We expect that it can be extended to other (even) fan-in cases.

6.1.2 Outline of the proof

For any function $f : \{0,1\}^n \to \{0,1\}$ we will choose an input bit x_i on which f depends, and fix all other bits such that flipping x_i flips the value of f. Assume that there is a formula F with noisy gates that fail independently with probability at least β_2. Then, for each gate in the formula F with input wires A and B and output wire C we can define $a = \frac{1}{2}\mathrm{Pr}[A = 0 \mid x_i = 0] + \frac{1}{2}\mathrm{Pr}[A = 0 \mid x_i = 1]$ and $\delta_a = \mathrm{Pr}[A = 0 \mid x_i = 0] - \mathrm{Pr}[A = 0 \mid x_i = 1]$ and analogously for B and C. The variable a can be seen as the average probability of A being 0. We call δ_a the *bias* of A with respect to the two input settings $x_i = 0$ and $x_i = 1$.

To prove our result one could attempt the following, which will turn out to not quite work (but we then show how to fix that): For an ϵ-noisy gate with fan-in 2, input wires A, B and output wire C, we would like to show that if the noise ϵ is at least the threshold β_2 then for any $\delta > 0$ there is some $0 \leq \theta < 1$ such that if $|\delta_a| \geq \delta$ or $|\delta_b| \geq \delta$ then

$$|\delta_c| \leq \theta \max\{|\delta_a|, |\delta_b|\} \tag{6.1}$$

This would mean that the bias goes down exponentially with the number of computation steps, until it reaches δ. Further, it is easy to show that for any $d > 0$ there is a function f such that any formula computing f has one input bit x_i on which f depends and the number of computation steps on any path from x_i to the output bit is at least d. Hence, the bias cannot be bounded away from zero for all f and x_i.

Unfortunately, (6.1) is not always true. Sometimes the bias can actually go up.[1] We use a more sophisticated approach, showing that the bias goes down "on average": We define a *potential function* q, which is positive and bounded on $[0,1]$. Instead of showing (6.1) we show that for any $\delta > 0$ there is some $0 \leq \theta < 1$ such that if $|\delta_a| \geq \delta$ or $|\delta_b| \geq \delta$ then

$$|\delta_c| q(c) \leq \theta \max\{|\delta_a| q(a), |\delta_b| q(b)\}. \tag{6.2}$$

[1] An easy example is an OR-gate with noise $\epsilon = 1/10$, $\delta_a = \delta_b = 1/10$ and $a = b = 8/10$, for which $\delta_c = (a\delta_b + b\delta_a)(1 - 2\epsilon) = 0.128 > 1/10$.

and (6.2) holds for $\theta = 1$ even if $|\delta_a|, |\delta_b| < \delta$. Since q is bounded, this implies that for any arbitrarily small constant $\delta > 0$ the bias of any formula becomes $O(\delta)$ after a constant number of computation steps. We can then proceed as above.

6.1.3 Organization

Section 6.2 contains all further definitions. The main proof is in Section 6.4. In Section 6.3 we prove (6.2), in the main Lemma 6.3.6. This is the most technical part of this chapter. Section 6.5 contains some remarks on our particular choice of q.

6.2 Definitions

A *circuit* is represented by a directed acyclic graph with one unique sink. We call its nodes *gates* and its edges *wires*. The sink is called the *output gate*. Each *gate* has a certain number of input wires (incoming edges), which is called the *fan-in* of the gate. The wires can take boolean values 0 or 1. A gate computes an output bit as a boolean function of its input bits. Gates with no incoming edges correspond to input bits. The output of the output gate determines the output of the circuit.

A *formula* is a particular type of circuit in which the gates are connected in a tree, with the output gate at the root and the input bits at the leaves. In particular, this mean that each gate has exactly one output wire. Each input bit to the function we want to compute can appear more than once as an input bit to the formula.

A (perfect) PARITY-gate with input bits x_1 and x_2 outputs 0 if $x_1 = x_2$ and 1 otherwise. A (perfect) OR-gate outputs 0 if $x_1 = x_2 = 0$ and 1 otherwise.

We call a gate *ϵ-noisy* if it outputs the correct result with probability $1-\epsilon$ and with probability ϵ it outputs the opposite. We say that a formula F with noisy gates computes the function f with bias $\Delta > 0$ if for all $x \in f^{-1}(0)$, $y \in f^{-1}(1)$: $\Pr[F(x) = 0] \geq \Delta + \Pr[F(y) = 0]$. If f can be computed with some bias $\Delta > 0$ we also say that f is *computable with bounded-error*.

A function $f : \{0,1\}^n \to \{0,1\}$ *depends* on the i-th input bit x_i if there is some setting of the other bits, such that flipping x_i flips the function value. The number of bits that f depends on is denoted by $d(f)$.

In a formula, we define the *depth* of a wire A, denoted by $depth(A)$, as the number of 2-input gates on a path from A to the output wire. Gates with fan-in 1 (NOT gates and identity gates) or fan-in 0 (input gates) are not counted and are assumed to be noise-free.

For the definition of the quantities a and δ_a for a wire A we refer to Section 6.1.2.

6.3 Bias reduction for noisy gates

We define the constant $x_0 = 1/(2-4\beta_2) = (1+\sqrt{7})/6 \approx 0.61$. It will turn out later that an OR-gate with input wires A, B performs best when $a \approx x_0$ and $b \approx x_0$. Further, we define the *potential function*

$$\begin{aligned} q(x) &= \left(\tfrac{29}{2}+2\sqrt{7}\right)\left(x-\tfrac{1}{2}\right)^4 \\ &\qquad + \left(\tfrac{5\sqrt{7}}{2}-\tfrac{13}{4}\right)\left(x-\tfrac{1}{2}\right)^2 - \tfrac{\sqrt{7}}{2}+\tfrac{73}{32} \qquad (6.3) \\ &\approx 19.79(x-0.5)^4 + 3.36(x-0.5)^2 + 0.96, \end{aligned}$$

which is given in Figure 6.1. This is a bi-quadratic function in $(x-1/2)$. Further,

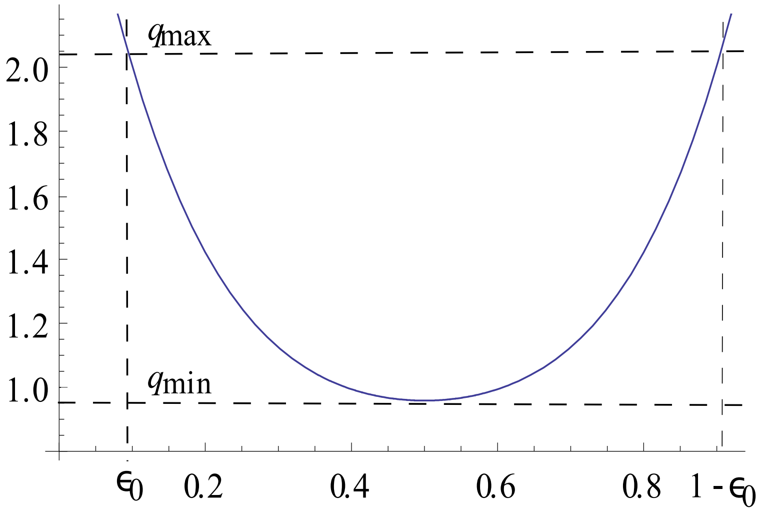


Figure 6.1: Graph of $q(x)$

q is symmetric around $1/2$ and convex. In $[\beta_2, 1-\beta_2]$ the potential function q is bounded between $q_{min} = q(1/2) = -\sqrt{7}/2 + 73/32 > 0.9$ and $q_{max} = q(\beta_2) = (247+8\sqrt{7})/128 < 2.1$.

For any $\epsilon \leq 1/2$ we define the function

$$\eta_\epsilon(x) = (1-2\epsilon)x + \epsilon.$$

If x is the probability that some variable is 0, then $\eta_\epsilon(x)$ is the probability that this variable is 0 after it has been flipped with probability ϵ.

In the rest of this section we establish inequality (6.8) in Lemma 6.3.6, from which the proof of the main theorem will follow relatively straightforwardly. The proof of this inequality is quite technical and so at first reading the reader might just want to read the statement of Lemma 6.3.6 and then move immediately to Section 6.4, where we establish the main result. Inequality (6.8) can also be checked with the help of a computer (e.g. using Mathematica [1]), but in the remainder of this section we will prove it rigorously.

6.3.1 Technical lemmas

The following technical results are used in the proof of Lemma 6.3.6.

6.3.1. Proposition. *For all a, b with $\beta_2 \leq a, b \leq 1 - \beta_2$ it holds that*

$$q(a)q(b) - (1 - 2\beta_2)(aq(a) + bq(b))q(\eta_{\beta_2}(ab)) \geq 0. \tag{6.4}$$

Proof: We write $a = x_0 + s_a$ and $b = x_0 + s_b$. Without loss of generality let $|s_b| \geq |s_a|$ and choose $-1 \leq k \leq 1$ s.t. $s_a = ks_b$. Then the lhs of (6.4) can be written as

$$\sum_{i=0}^{11} r_i(k) s_b^{i+2}. \tag{6.5}$$

The reason why (6.5) only starts with a quadratic term in s_b is our special choice of q, see Section 6.5.2 for more on this. The first coefficient is easily computed

$$r_0(k) = \left(3 - \tfrac{3\sqrt{7}}{4}\right)\left(k^2 + 1\right).$$

This function attains its minimum value of $3 - 3\sqrt{7}/4 \approx 1.02$ at $k = 0$. Therefore, there is a $\kappa > 0$ s.t. for $a, b \in [x_0 - \kappa, x_0 + \kappa]$ the lhs of (6.4) is non-negative. We show that $\kappa = 0.02$ is a solution.

The absolute value of the other coefficients for $-1 \leq k \leq 1$ can be bounded by $|r_1(k)| \leq 5$, $|r_2(k)| \leq 31$, $|r_3(k)| \leq 18$, $|r_4(k)| \leq 68$, $|r_5(k)| \leq 326$ and for all other $|r_i(k)| \leq 5000$. Therefore, if $|s_b| \leq 1/50$, (6.5) is at least

$$s_b^2 \left(1.02 - 5(0.02) - 31(0.02)^2 - 18(0.02)^3 - \dots\right) \geq 0.90 s_b^2 \geq 0.$$

This proves the case $x_0 - 1/50 \leq a, b \leq x_0 + 1/50$. For all other $|a - x_0| \geq 1/50$ or $|b - x_0| \geq 1/50$ the proposition follows from Fact 6.3.2 with $\mu = 0$. ■

We now state some bounds on polynomials. They are similar in spirit to (6.4), with the crucial difference that these bounds are not tight. This is convenient, because there are numerical techniques for finding global optima of multivariate polynomials up to arbitrary precision. See [72] for an overview. We have used the computer algebra program Mathematica [1]. We used an accuracy of 10^{-10} and rounded the results in such a way that the bounds given *are rigorous.*[2]

6.3.2. Fact. For all a, b with $\beta_2 \leq a, b \leq 1 - \beta_2$ with $|a - x_0| \geq 1/50$ and $0 \leq \mu \leq \xi := (1 - \beta_2 - a)(1 - \beta_2 - b)$ it holds that

$$q(a)q(b) - (1 - 2\beta_2)(aq(a) + bq(b))q(\eta_{\beta_2}(ab + \mu)) > 0.0003. \tag{6.6}$$

[2] Even more simple, one could bound the first derivatives and check all values of the polynomials on a small enough grid.

Proof: Notice that μ only appears in the term $q(\eta_{\beta_2}(ab+\mu))$. For $0 \leq \mu \leq \xi$ we notice that by convexity of q and linearity of η_{β_2} it follows that $q(\eta_{\beta_2}(ab+\mu)) \leq \max\{q(\eta_{\beta_2}(ab)), q(\eta_{\beta_2}(ab+\xi))\}$. Thus, (6.6) is minimized for $\mu = 0$ or $\mu = \xi$. For $\mu = 0$ the lhs of (6.6) is lower bounded by 0.0003 and for $\mu = \xi$ by 0.01. ■

6.3.3. Fact. For all a, b, μ with $\beta_2 \leq b \leq 1-\beta_2$, $1/2 \leq a \leq 1-\beta_2$ and $|\mu| \leq \xi := 2(1-\beta_2-a)(1-\beta_2-b)$ it holds that

$$\begin{aligned}&q(a)q(b)\\ &\quad -\big((2a-1)q(a)+(2b-1)q(b)\big)(1-2\beta_2)q\big(\eta_{\beta_2}\big(ab+(1-a)(1-b)+\mu\big)\big)\\ &\geq 0.45.\end{aligned}$$

Proof: For $\mu = \xi$ the term is lower bounded by 0.48 and for $\mu = -\xi$ by 0.55. Using convexity of q as above the fact follows. ■

6.3.4. Fact. For all a, b, μ with $\beta_2 \leq b \leq 1-\beta_2$, $\beta_2 \leq a \leq 1/2$ and $|\mu| \leq \xi := 2(a-\beta_2)(1-\beta_2-b)$ it holds that

$$\begin{aligned}&q(a)q(b)\\ &\quad -\big(-(2a-1)q(a)+(2b-1)q(b)\big)(1-2\beta_2)q\big(\eta_{\beta_2}(ab+(1-a)(1-b)+\mu)\big)\\ &\geq 0.48.\end{aligned}$$

Proof: For $\mu = \xi$ the term is lower bounded by 0.51 and for $\mu = -\xi$ by 0.48. The fact then follows by convexity of q as above. ■

6.3.5. Fact. Let a, b, μ with $\beta_2 \leq a, b \leq 1-\beta_2$ Then

$$q(a) - (1-2\beta_2)bq(\eta_{\beta_2}(ab+\mu)) \quad > \quad 0.22$$

holds if (a) $a \leq 1/2$ and $-(a-\beta_2)(1-\beta_2-b) \leq \mu \leq 0$ or (b) $1/2 \leq a$ and $-(1-\beta_2-a)(1-\beta_2-b) \leq \mu \leq 0$.

Proof: For $\mu = 0$ (and all $\beta_2 \leq a \leq 1-\beta_2$) the term is lower bounded by 0.23. For both cases $a \leq 1/2$, $\mu = -(a-\beta_2)(1-\beta_2-b)$ and $1/2 \leq a$, $\mu = -(1-\beta_2-a)(1-\beta_2-b)$ the term is lower bounded by 0.22. Using convexity of q as above the fact follows. ■

6.3.2 Main Lemma

We can state our main Lemma. Recall from Section 6.1.2 that we chose one input bit x_i of the function f we want to compute and set all the other input bits such that flipping x_i flips the output value of the function. See also the notation in the proof of our main theorem in Section 6.4.

6.3.6. Lemma. *Let $\beta_2 \leq \epsilon \leq 1/2$. Assume an ϵ-noisy OR-gate or PARITY-gate, with input wires A and B and output wire C. Let*

$$\begin{array}{rcccl} \beta_2 & \leq & \Pr[A=0 \mid x_i=0] & \leq & 1-\beta_2 \\ \beta_2 & \leq & \Pr[A=0 \mid x_i=1] & \leq & 1-\beta_2, \end{array} \tag{6.7}$$

and let the same be true for B. Define a, b, c and $\delta_a, \delta_b, \delta_c$ for A, B, C as in Section 6.1.

1. *The following inequality holds for $\theta = 1$:*

$$|\delta_c| q(c) \leq \theta \max\{|\delta_a| q(a), |\delta_b| q(b)\}. \tag{6.8}$$

2. *For any $\delta > 0$ there is a $0 \leq \theta < 1$ such that if $|\delta_a| \geq \delta$ or $|\delta_b| \geq \delta$, then (6.8) is still true for this θ.*

Proof: We consider the OR-gate first. We have

$$\begin{array}{rcl} \Pr[C=0 \mid x_i=0] & = & \eta_\epsilon\left((a+\delta_a/2)\left(b+\delta_b/2\right)\right) \\ \Pr[C=0 \mid x_i=1] & = & \eta_\epsilon\left((a-\delta_a/2)\left(b-\delta_b/2\right)\right), \end{array}$$

which implies

$$\begin{array}{rcl} \delta_c & = & (a\delta_b + b\delta_a)(1-2\epsilon) \\ c & = & \eta_\epsilon\left(ab + \delta_a\delta_b/4\right). \end{array}$$

Increasing ϵ decreases $|\delta_c|$ as well as $q(c)$, since the c gets closer to $1/2$ and q decreases towards $1/2$. Thus we may assume $\epsilon = \beta_2$. Further, we may assume $|\delta_a| q(a) \geq |\delta_b| q(b)$. Note that, for $\delta_a = 0$ we then also have $\delta_b = 0$ and the Lemma holds trivially. In the remainder we therefore assume $\delta_a \neq 0$. In fact, we will even assume $\delta_a > 0$: In case $\delta_a < 0$ we can just formally replace every occurrence of δ_a and δ_b with $-\delta_a$ resp. $-\delta_b$. Because of the absolute value signs, this will not change the validity of (6.9). So we have to prove

$$(1-2\epsilon)\left|a\delta_b + b\delta_a\right| q\left(\eta_\epsilon\left(ab+\delta_a\delta_b/4\right)\right) \leq \theta |\delta_a| q(a). \tag{6.9}$$

In the remainder, we will repeatedly use that a and b are bounded between β_2 and $1-\beta_2$ and that in this range, $0.9 < q_{min} \leq q(a) \leq q_{max} < 2.1$, without mentioning it each time. We distinguish the following cases:

$\underline{\boldsymbol{\delta_b > 0}}$: Since we assumed $|\delta_a| q(a) \geq |\delta_b| q(b)$, it is enough to prove (6.9) where we replace the first occurrence of δ_b by $\delta_a q(a)/q(b)$. Cancelling δ_a and multiplying by $q(b)$ we get

$$\theta q(a) q(b) (1-2\beta_2)\left(aq(a) + bq(b)\right) q\left(\eta_{\beta_2}\left(ab+\delta_a\delta_b/4\right)\right) \geq 0. \tag{6.10}$$

In case $|a - x_0| \geq 1/50$ or $|b - x_0| \geq 1/50$, note that $\delta_a\delta_b/4 \leq (1-\beta_2-a)(1-\beta_2-b)$. If we set $\mu = \delta_a\delta_b/4$ and $\theta = 1$, then by Fact 6.3.2 the lhs of (6.10) is greater than 0.0003. This implies the existence of a $\theta < 1$ for (6.10) and settles both parts of the Lemma.

We are left with the case $|a - x_0| < 1/50$ and $|b - x_0| < 1/50$. By (6.7) we can then bound $\delta_a/2 \leq 1-\beta_2-a \leq 1-\beta_2-x_0+1/50 < 0.33$ and similarly $\delta_b/2 < 0.33$, i.e. $(1-2\beta_2)\delta_a\delta_b/4 < 0.1$. We also note that in our case $0.37 < \eta_{\beta_2}(ab) < 0.42$. By convexity, $\min_{0.37 \leq x \leq 0.42} q(x) - q(x+0.1) = q(0.42) - q(0.52) > 0.02$, and thus $q(\eta_{\beta_2}(ab)+0.1) < q(\eta_{\beta_2}(ab)) - 0.02$. This, convexity of q and $(1-2\beta_2)\delta_a\delta_b/4 < 0.1$ imply $q(\eta_{\beta_2}(ab) + (1-2\beta_2)\delta_a\delta_b/4) < q(\eta_{\beta_2}(ab)) - \frac{0.02}{0.1}(1-2\beta_2)\delta_a\delta_b/4$. Noting that $\eta_{\beta_2}(ab) + (1-2\beta_2)\delta_a\delta_b/4 = \eta_{\beta_2}(ab + \delta_a\delta_b/4)$ this becomes

$$q(\eta_{\beta_2}(ab + \delta_a\delta_b/4)) < q(\eta_{\beta_2}(ab)) - (1-2\beta_2)\delta_a\delta_b/20. \tag{6.11}$$

In particular $q(\eta_{\beta_2}(ab + \delta_a\delta_b/4)) < q(\eta_{\beta_2}(ab))$. Plugging the lhs of this into (6.10) and using Proposition 6.3.1 implies (6.10) for $\theta = 1$. This establishes part 1 of the Lemma for $\delta_b > 0$.

Now part 2 of the Lemma. Let $\delta_a \geq \delta$ or $\delta_b \geq \delta$. Consider first the case that δ_b is not too small compared to δ_a, say $\delta_b \geq \delta_a/100$. Together with our assumption $|\delta_a|q(a) \geq |\delta_b|q(b)$ this implies $(1-2\beta_2)\delta_a\delta_b/20 \geq (1-2\beta_2)\delta^2/2000$. With (6.11) we then get $q(\eta_{\beta_2}(ab + \delta_a\delta_b/4)) + c < q(\eta_{\beta_2}(ab))$ for $c = (1-2\beta_2)\delta^2/2000 > 0$ and putting this into (6.4) gives $q(a)q(b) - (1-2\beta_2)(aq(a) + bq(b))(q(\eta_{\beta_2}(ab + \delta_a\delta_b/4)) + c) > 0$. This implies the existence of a $\theta < 1$ for (6.10) and establishes part 2 of the Lemma when $\delta_b \geq \delta_a/100$.

If δ_b is small, i.e. $\delta_b < \delta_a/100$, then we can use that upper bounding the first occurrence of δ_b by $\delta_a q(a)/q(b)$ to get from (6.9) to (6.10) was far from tight. A better bound is $\delta_b < \delta_a q(a)/(10q(b))$, which derives from $q(a)/(10q(b)) \geq q_{min}/(10q_{max}) > 1/100$. Analogously to the derivation of (6.10) we get

$$\theta q(a)q(b) - (1-2\beta_2)\left(aq(a)/10 + bq(b)\right) q\left(\eta_{\beta_2}\left(ab + \delta_a\delta_b/4\right)\right) \geq 0. \tag{6.12}$$

By (6.7), $a > \beta_2$. Thus, $aq(a) > \beta_2 q_{min}$ and also $q\left(\eta_{\beta_2}\left(ab + \delta_a\delta_b/4\right)\right) > q_{min}$. Hence, the lhs of (6.12) is at least $(1-2\beta_2)\beta_2 q_{min}^2 9/10$ smaller than the lhs of (6.10). Since we already proved earlier that (6.10) holds for $\theta = 1$ without the restriction $\delta_b < \delta_a/100$, we conclude that (6.12) holds for some $\theta < 1$. This establishes part 2 of the Lemma for $\delta_b < \delta_a/100$.

$\underline{\boldsymbol{\delta_b \leq 0}}$: It is enough to prove (6.9) where we replace $|a\delta_b + b\delta_a|$ by (a) $|b\delta_a|$ or (b) $|a\delta_b|$. If in case (a) we cancel δ_a and $q(a)$ after the replacement, we see that a $\theta < 1$ must exist if

$$q(a) - (1-2\beta_2)bq(\eta_{\beta_2}(ab + \delta_a\delta_b/4)) \geq \chi, \tag{6.13}$$

for some $\chi > 0$. Note that in case $a \leq 1/2$ we have $-(a-\beta_2)(1-\beta_2-b) \leq \delta_a\delta_b/4 \leq 0$ and in case $1/2 \leq a$ we have $-(1-\beta_2-a)(1-\beta_2-b) \leq \delta_a\delta_b/4 \leq 0$. The Lemma then follows from Fact 6.3.5.

For case (b) we note that $|a\delta_b| \le a\delta_a q(a)/q(b)$. Replacing $|a\delta_b + b\delta_a|$ in (6.9) by $a\delta_a q(a)/q(b)$ and rearranging terms we get exactly the same as (6.13), with a and b swapped. We proceed as in case (a).

We now consider the PARITY-gate. First note, that if the two input wires of a noiseless PARITY gate are independently 0 with probability α resp. β, then the output wire will be 0 with probability $\alpha\beta + (1-\alpha)(1-\beta)$. Thus, in our case

$$\begin{aligned}\Pr[C=0 \mid x_i=0] &= \eta_\epsilon((a+\delta_a/2)\,(b+\delta_b/2)\\ &\qquad +(1-a-\delta_a/2)\,(1-b-\delta_b/2))\\ \Pr[C=0 \mid x_i=1] &= \eta_\epsilon((a-\delta_a/2)\,(b-\delta_b/2)\\ &\qquad +(1-a+\delta_a/2)\,(1-b+\delta_b/2))\end{aligned}$$

which implies

$$\begin{aligned}c &= \eta_\epsilon\,(ab+(1-a)(1-b)+\delta_a\delta_b/2)\\ \delta_c &= ((2a-1)\delta_b+(2b-1)\delta_a)\,(1-2\epsilon)\end{aligned}$$

We need to prove

$$\begin{aligned}&|(2a-1)\delta_b+(2b-1)\delta_a|\,(1-2\epsilon)\times q(\eta_\epsilon\,(ab+(1-a)(1-b)+\delta_a\delta_b/2))\\ &\le \theta|\delta_a|q(a).\end{aligned} \tag{6.14}$$

As for the OR gate we only need to consider $\epsilon = \beta_2$ and may assume $\delta_a \ge 0$ w.l.o.g, because otherwise we can just change the signs of both δ_a and δ_b. Also, w.l.o.g. we assume $|\delta_a|q(a) \ge |\delta_b|q(b)$. If $\delta_a = 0$, then also $\delta_b = 0$ and the Lemma becomes trivial. So we assume $\delta_a > 0$. Further, we may assume $b \ge 1/2$ (and therefore $(2b-1)\delta_a \ge 0$), because formally replacing a and b by $1-a$ and $1-b$ does not change (6.14). We condition on the sign of $2a-1$.

First $2a-1 \ge 0$. It is enough to prove (6.14), where we replace the first occurrence of δ_b by $\delta_a q(a)/q(b)$, since we assumed $|\delta_a|q(a) \ge |\delta_b|q(b)$. Cancelling δ_a and rearranging terms, the existence of a $0 \le \theta < 1$ for (6.14) then follows from

$$\begin{aligned}&q(a)q(b)\\ &\quad -((2a-1)q(a)+(2b-1)q(b))(1-2\beta_2)q(\eta_{\beta_2}(ab+(1-a)(1-b)+\delta_a\delta_b/2))\\ &\ge \chi > 0.\end{aligned}$$

This inequality follows from Fact 6.3.3 by noting that $|\delta_b| \le 2(1-\beta_2-b)$ and $|\delta_a| \le 2(1-\beta_2-a)$.

In case $2a-1<0$ we can proceed similarly, where this time we replace the first occurrence of δ_b by $-\delta_a q(a)/q(b)$ and bound $|\delta_a| \le 2(a-\beta_2)$. The resulting inequality follows from Fact 6.3.4. ■

6.4 Proof of Theorem 6.1.1

Proof: Let f be any function and let F be any formula with noisy gates that fail independently with probability at least β_2. Let F compute f with bias Δ. We show that f depends on at most a constant number of bits, i.e. $d(f) \leq c(\Delta)$, for some function $c(\Delta)$.

Before starting we note the following: Every ϵ-noisy fan-in-2 gate can be constructed from an ϵ-noisy PARITY- or an ϵ-noisy OR-gate, perfect NOT-gates and constant inputs. Hence, we may assume that F is constructed only from perfect NOT-gates, noisy PARITY-gates with fan-in 2 and noisy OR-gates with fan-in 2.

Let x_i be an input bit on which f depends with the additional property that any input wire of F carrying x_i has depth at least $\lceil \log_2 d(f) \rceil$. Because all gates in F have fan-in at most 2, the existence of such x_i is guaranteed. Fix all other input bits such that the output of F changes when flipping x_i.

Set $\bar{d} = \lceil \log_2 d(f) \rceil - 1$ and $\delta = \frac{\Delta}{2q_{max}}$. Let $\theta < 1$ be given by Lemma 8.4.6 for this δ. In case this results in $\theta < 1 - 2\beta_2$, set $\theta = 1 - 2\beta_2$. (The adjustment $\theta \geq 1 - 2\beta_2$ is not really needed, but will later simplify the proof.) We will prove inductively that for any wire C at depth $d \leq \bar{d}$

$$q(c)|\delta_c| \leq \max\{\tfrac{\Delta}{2}, \theta^{\bar{d}-d} q_{max}\}. \tag{6.15}$$

For $d = \bar{d}$ (6.15) holds trivially. Now take any wire C in F with depth $d < \bar{d}$. We distinguish what signal C carries.

Firstly, C can be an input wire carrying an input bit x_j. Then necessarily $i \neq j$, because input wires carrying x_i have depth at least $\bar{d} + 1$. Thus, $\delta_c = 0$ and (6.15) holds.

Secondly, C can be the output of a noiseless NOT-gate, which has input wire B. Note that since we do not count NOT-gates in the depth of a wire, $depth(C) = depth(B)$, $c = 1 - b$ and $\delta_c = -\delta_b$. Then, by symmetry of q around 1/2 we get (6.15) for C from the same statement for B.

Thirdly, C can be the output of gate G, with G either an OR-gate or a PARITY-gate. Let the input wires to G be A and B. If one wire is a constant (i.e. its value is independent of the value of x_i), then gate G is essentially a (noisy) gate with fan-in 1. Hence, G always outputs either a (noisy) 0 or 1, or G is the noisy identity- or the noisy NOT-gate. In the first two cases $\delta_c = 0$. In the last two cases we can easily calculate that $|b - 1/2|(1 - 2\epsilon) = |c - 1/2|$ and $|\delta_c| \leq (1 - 2\beta_2)|\delta_b|$. Because q decreases monotonically towards 1/2 and we chose $\theta \geq 1 - 2\beta_2$, (6.15) holds.

So we are left with the case where both inputs to G are non-constant. Since $d < \bar{d}$, both wires A and B are the output of some noisy gate, so the conditions (6.7) in Lemma 6.3.6 are satisfied. We may assume $|\delta_b|q(b) \leq |\delta_a|q(a)$ w.l.o.g. If $|\delta_a|q(a) \leq \Delta/2$, then by part 1 from Lemma 6.3.6 also $|\delta_c|q(c) \leq \Delta/2$ and (6.15) holds. If $|\delta_a|q(a) > \Delta/2$, then $|\delta_a| > \frac{\Delta}{2q_{max}} = \delta$. Then (6.15) follows from part 2 of Lemma 6.3.6 and the inductive assumption.

Let O be the output wire of F, which by assumption has bias Δ, and as before let o be its average probability of being zero. Because $q(o)\Delta \leq \Delta/2$ is impossible (since $q(o) \geq q_{min} > 1/2$) we get from (6.15): $q(o)\Delta \leq \theta^{\bar{d}} q_{max}$, and further $\Delta \leq \theta^{\lceil \log_2 d(f) \rceil - 1}(q_{max}/q_{min})$, which implies $\frac{\log_2(\Delta q_{min}/q_{max})}{\log_2 \theta} + 1 \geq \log_2 d(f)$. Since θ depends only on Δ, $d(f)$ is upper bounded by the function

$$c(\Delta) := 2\,(\Delta q_{min}/q_{max})^{1/\log_2 \theta}\,.$$

■

6.5 Discussion

We have shown a tight threshold for the noise which is tolerable for computation by formulas with gates of fan-in at most 2. This is the first tight threshold for gates with an even number of input wires. It should be possible to generalize this to other gates with even fan-in, although the proof is probably more tedious.

6.5.1 Circuits vs. formulas

The same bound probably also applies to *circuits* (with gates of fan-in at most 2), where our noise model is that if a gate fails, then *all* its output wires carry the incorrect output. The intuition why circuits should not be more fault-tolerant than formulas is the following. If a gate G has $m > 2$ output wires, then they all carry the same value. But if G fails, then all output wires carry the same "wrong" value. It therefore becomes impossible to detect or correct the error on this gate by just looking at these output wires. A better idea should be to make m copies of G, all having the same inputs. In this case it is less likely that all gates fail at the same time. So, most of the output wires carry the correct output and it is possibly easier to correct the errors.

Proving that circuits are not more fault-tolerant than formulas is an interesting and important open question.

6.5.2 Choice of potential function

So far we have not given any idea of why we chose this particular potential function. In fact, this choice is not unique. The choice of q was determined as follows: (1) It is convenient to choose q symmetric around 1/2, so applying a NOT-gate to wire A does not change the value of $|\delta_a|q(a)$. (2) It is natural to scale q such that $q(x_0) = 1$. (3) After these choices, we have to choose $\frac{d}{dx}q(x)|_{x=x_0} = \frac{1}{2}(-1+\sqrt{7})$. This ensures that (6.5) does not have a linear term in s_b and only starts with the quadratic term, i.e. "$r_{-1}(k) \equiv 0$". (4) We also need $\frac{d^2}{dx^2}q(x)|_{x=x_0} > 16 - 4\sqrt{7} \approx 5.42$, because that makes $r_0(k) > 0$ for $-1 \leq k \leq 1$. The rest of the choices are not so binding.

However, a quadratic function alone is not enough. For (6.5) to be at least 0 one also has constraints on higher derivatives of q. The expression in (6.3) for q is one of the "nicer" possible potential functions. One can also find a possible q by dividing the interval $[\beta_2, 1 - \beta_2]$ into smaller intervals and define q as different quadratic functions in each of these intervals.

Our choice of q was determined by the properties we want q to have. Unfortunately, q does not have an immediate interpretation in "standard" information-theoretic terms. It would be nice to come up with a proof which can be more easily interpreted in standard terms.

Part II

Entanglement and interactive proof systems

Chapter 7

Parallel repetition of quantum XOR games

This chapter is based on the paper

> R. Cleve, W. Slofstra, F. Unger, and S. Upadhyay, **Strong parallel repetition theorem for quantum XOR proof systems**, In *Special Issue of 22nd IEEE Conference on Computational Complexity*, 2007.

7.1 Introduction

7.1.1 Motivation

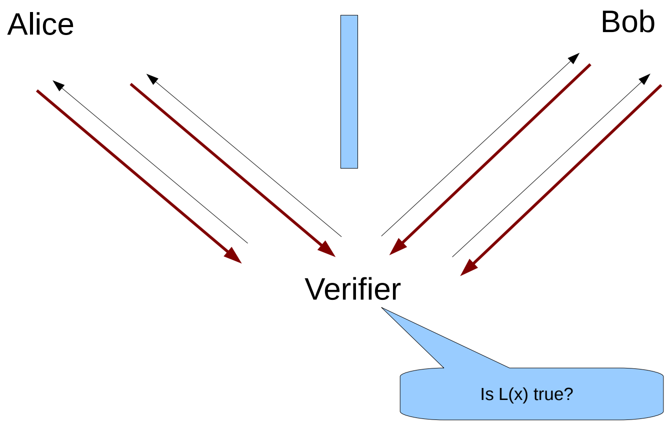
Complexity classes

A large part of theoretical computer science is concerned with classifying the difficulty of computational problems. Usually the difficulty of a problem is defined as the amount of resources (for example time or storage space) a Turing machine needs to solve the problem. The most important class of problems is the well-known class P, that contains all problems which can be solved by a polynomial time Turing machine. Polynomial-time solvable problems are considered to be efficiently solvable. The class P contains interesting problems, like (the decision version) of Linear Programming, finding the shortest path between two nodes in a graph, testing whether a given natural number is prime and many others. Readers not familiar with computational complexity theory can find some concise definitions in Section 2.3 on page 23.

Another approach to define complexity classes does not ask how difficult it is to solve a problem, but rather how much resources are needed to become "convinced" that a solution exists. The most important class defined this way is NP, which contains all problems for which a polynomial time Turing machine can check that a solution to a problem is correct. A particular example is the satisfiability problem SAT, which is the problem of deciding whether for a given Boolean formula there is an assignment of values to the variables which makes the formula evaluate to true. Obviously, if one sees an assignment it is easy to verify (i.e., it is possible to decide in polynomial time in the length of the formula) whether it makes the formula true. In the example of NP "being convinced" means that it is possible to be sure without doubt that a solution exists. It is widely believed that for this strong version of being convinced it is essentially necessary to see a solution.

Interactive proof systems

There is a more liberal version of "being convinced" in which one only wants to be sure that a solution exists beyond "reasonable doubt". Here it might not be necessary to see the solution in order to believe that there is one. This notion can be explained more formally in the framework of *multi-prover interactive proof systems* (MIP proof systems), in which games –the focus of this chapter– show up naturally. An MIP system is a protocol between two provers Alice and Bob and a *verifier*. The computational power of the verifier is usually that of a probabilistic polynomial-time machine. He wants to know if some instance x of a problem has



Verifier interacts with provers and then accepts or rejects x

Figure 7.1: Multi-prover interactive proof systems

a solution (i.e. whether x satisfies some predicate L), and in order to do that he can interact with the two provers by sending them questions generated by some probabilistic procedure and receive their answers. The only restriction on the provers is that they are not allowed to interact once the protocol has started. They see the verifier's input x and they may agree on a strategy which depends on x. At the end of the protocol the verifier outputs a single bit, indicating whether he believes that x has a solution or not. We say that problem C can be decided by an MIP system, if there is a verifier V (i.e. a probabilistic polynomial-time machine) and constants $c > s$, such that for any x it holds: If $x \in C$ then there is a strategy of the provers which makes the verifier V accept with probability at least c and if $x \notin C$ then every strategy of the provers makes the verifier accept with probability at most s. The class of problems which can be characterized this way is called MIP. The parameter c is called the *correctness parameter* and s the *soundness parameter*. This model is very powerful. Every problem in NEXP has an MIP proof system [9]. Recall that NEXP is the class of all problems which can be verified by a Turing-machine whose running time is at most exponential in the input size. It is not hard to see that MIP⊆NEXP, since a NEXP-verifier can just guess the prover's best strategy. Hence, MIP=NEXP.

The number of rounds needed in such MIP systems is only 1, that is, it is enough for the verifier to send one question to Alice and one to Bob and then accept or not depending on the replies. In this way, we can say that for each input x Alice and Bob are playing a game G_x in which they always try to make the verifier accept. If x has a solution then the provers can win G_x with probability at least c, if there is no solution then they can win with probability at most s.

Uncertainty versus acceptance power

Recall (see also Section 2.3) that the class NEXP is incredibly more powerful than NP. This means that the verifier can trade some uncertainty about the correctness of his output (if $c < 1$ or $s > 0$) against the ability to verify "exponentially" harder problems. How much uncertainty do we have to accept such that we get the full expressive power of the class MIP? Surprisingly, it turns out that it is possible to choose $c = 1$ and s an arbitrarily small constant. Note though, that it is *impossible* to choose $s = 0$ in addition to $c = 1$, because then the proof system becomes deterministic and therefore characterizes only NP, which is known to be different from NEXP. In other words, if we are willing to tolerate some small uncertainty in the result, then we can characterize problems in NEXP instead of NP only.

Error-reduction by parallel repetition

Constructing MIP proof systems which achieve $c = 1$ and $s > 0$ arbitrarily small is clearly desirable. In particular because it makes the definition of the class MIP independent of the exact values of c and s. One way (but not the only) to see that there are MIP proof systems which match the above minimal requirements is as follows: Show that there is a proof system, with parameters $c = 1$ and $s < 1$. This is always possible for NEXP and can even be achieved with one-round protocols, see [42]. Then just repeat the same protocol again and again. Accept if the provers have won all instances. Clearly, if the provers had a perfectly convincing strategy for one protocol, then they can just use the same strategy on all instances and win them all. If their strategy allowed them to win only with probability s then the probability for winning T protocols has probability at most s^T and increasing T makes the soundness error go to 0 very quickly. This strategy has one obvious drawback. The verifier might need to involve in a very lengthy discussion, T times longer than the original protocol. This might be inefficient but also theoretically bad, because sequential repetition might not preserve certain properties of a proof system (like Zero-knowledge, see [78]). The other solution is to run all the T protocols in *parallel*, sending the questions of all the protocols at the same time to the respective provers. Clearly, the *expected* number of protocols the provers can win in this case cannot go up. However, for our purposes this is not enough yet, because the provers might use their knowledge about which question are asked in the other protocols in a sophisticated way: For instance, if one protocol can be won with probability s then there could be a collective strategy that wins *all* parallel protocols with probability s and loses all of them with probability $1-s$. Unfortunately, this would not help in reducing the error probability. In fact, we will later see an example in which *two* parallelized protocols have the same winning probability as *one*, see Section 7.6.2. So, does repeating protocols in parallel help to reduce winning probabilities?

The answer is yes and is established by the celebrated Parallel Repetition Theorem by Raz [78], see also [54, 77]. For any protocol that can be won with probability $s < 1$ there is some $s' < 1$, such that k parallel repetitions can be won with probability at most s'^k. This means that the above mentioned procedure for reducing the soundness error to almost 0 works, without increasing the number of rounds needed. And so the exact value of s is inessential, as long as $s > 0$.

XOR proof systems

The main result of this chapter considers a particular variant of this model, called *quantum XOR proof systems.* In *XOR proof systems* the protocol is restricted to one round only and the provers Alice and Bob reply with one bit a resp. b only. The verifier's verdict depends only on the parity $a \oplus b$.[1] If the provers are not allowed to share an entangled state, we speak of a (classical) *XOR proof system* and the complexity class they characterize is called $\oplus$MIP. In Section 7.2.1 we will see that even this restricted class is powerful enough to characterize NEXP. It holds that $\oplus$MIP =NEXP [32] (although only with parameters $c = 12/16 - \epsilon$

[1] $a \oplus b$ is equal to 0 if $a = b$ and otherwise 1.

and $s = 11/16 + \epsilon$ for arbitrarily small ϵ). If the provers are allowed to share an entangled state, we call it a *quantum XOR proof system* and the complexity class they characterize is called $\oplus$MIP*. When entanglement is allowed the complexity can be bounded by $\oplus\text{MIP}^* \subseteq \text{EXP}$ [33, 102], so assuming that NEXP$\neq$EXP, entanglement strictly weakens the expressive power of XOR proof systems.

Main result

Our main result in this chapter is a *perfect parallel repetition theorem* for quantum XOR proof systems, which states that if one protocol can be won with probability p, then k parallel repetitions of the protocol can be won exactly with probability p^k but not more. This means that the optimal collective strategy is to play all protocols individually optimal, i.e., there is no way in which knowledge of the other instances can help.

7.1.2 Organization

More about interactive proof systems with and without entanglement will be explained in Section 7.2.1. This section is meant to acquaint the reader with some further background but will not be needed for later sections. As mentioned in the introduction we focus on proof systems which are based on quantum XOR games. Quantum XOR games will be defined in Section 7.2.2 and in the following section we define parallel repetition. It is enough to read these two sections to understand the main results in Sections 7.4 and 7.5. We explain the connection between Bell inequalities and XOR games in Section 7.2.3.

Section 7.3 contains a characterization of quantum XOR games in terms of semidefinite programs. In Section 7.4 we show that quantum XOR games are additive, in a sense which is defined in the same section. The proof of our parallel repetition theorem in Section 7.5 uses this result and a Fourier transform technique.

7.2 Background and definitions

7.2.1 More on interactive proof systems

In the introduction we saw that proof systems can be very powerful and we explained their use in complexity theory. Interactive proof systems also had a significant impact in cryptography, but we will not explain this connection further (the interested reader is referred to [47]). In this section we present some more details about proof systems which further motivate the results of this chapter. The amount of work done on this is significant, [47] and the forthcoming book [8] might be a good starting point to learn more about classical interactive proof systems. We will not attempt to be in any way complete or reflect the historical line of events and only review certain aspects.

Further, we want to mention that it is also possible to define interactive proof systems in which the verifier and the prover(s) communicate quantum messages. We will not go into this subject in this thesis. The interested reader might find the following articles [58, 100, 60, 56] interesting and can find further pointers to the literature in there.

Number of provers

Let us first see how the number of provers influences the expressive power. It is known that allowing more than two classical provers does not increase the power of these proof systems [14]. However, if there is only one prover, then the resulting class IP exactly captures the problems in PSPACE, or more succinctly: IP=PSPACE [85]. Hence, it seems that at least 2 provers are needed to unleash the full power of the model. Further, the number of rounds needed is at most one, which follows by Raz's parallel repetition theorem [78].[2] A natural question is whether a parallel repetition theorem also holds for any other number of provers. Under the reasonable assumption that PSPACE$\neq \Pi_2^p$ it is impossible that a classical parallel repetition theorem for one prover only holds, since it is known that IP(m)=IP(2)$\subseteq \Pi_2^p$, but PSPACE=IP.

XOR proof systems

As mentioned in the introduction, the simplest case of MIP systems are XOR multi-prover interactive proof systems in which the verifiers reply only with one bit each, and the verifier accepts depending on $a \oplus b$. We defined the corresponding complexity classes $\oplus$MIP (no entanglement) and $\oplus$MIP* (with entanglement). In [32] it is pointed out that results in [13, 53] imply that, in the case of classical provers, these $\oplus$MIP systems are sufficient to recognize every language in NEXP (with soundness probability $s = 11/16 + \epsilon$ and completeness probability $c = 12/16 - \epsilon$, for arbitrarily small $\epsilon > 0$). Thus, although these proof systems appear restrictive, they can recognize the same languages as unrestricted multi-prover interactive proof system. Moreover, in [33, 102] it is shown that any language recognized by a quantum XOR proof system is in EXP, which uses a semidefinite programming characterization due to Tsirelson [94, 32], which is given in Section 7.3. Thus, assuming EXP $\neq$ NEXP, quantum entanglement *strictly weakens* the expressive power of XOR proof systems.

[2] Though we should remark that it is not needed.

Parallel repetition for quantum games

The only other parallel repetition theorem along the lines of [78] for quantum games (where the players share entanglement) we know of is for *unique games* [57]. Unique games are two-prover games where for each pair of questions to the verifier and each answer of Alice, there is always exactly one answer of Bob that makes the verifier accept. Note that XOR games are a particular kind of unique games. The result in [57] is not "perfect" in our sense though, since it does not imply that the trivial strategy (of playing all parallel games independently) achieves the best success probability for winning all games. We do not know about a parallel repetition theorem for general quantum games. A perfect parallel repetition theorem cannot be true in general as was pointed out by Watrous [101], who has shown that there is a binary game (that is not an XOR game) for which the success probability $\omega_q(G)$ of winning one game and the success probability $\omega_q(G \wedge G)$ of winning two games played in parallel is in both cases $2/3$. as in the classical case. This is explained in Section 7.6.2.

This chapter is about parallel repetition of 2-prover games. We do not know about any results for more than two provers.

7.2.2 XOR games

The definition of *XOR interactive proof systems* can be based on *XOR games*, which we define first. For a predicate $f : S \times T \to \{0,1\}$ and a probability distribution π on $S \times T$, define the XOR game $G = (f, \pi)$ operationally as follows.

- The Verifier selects a pair of questions $(s,t) \in S \times T$ according to distribution π.
- The Verifier sends one question to each prover: s to prover Alice and t to prover Bob (who are not allowed to communicate with each other once the game starts).
- Each prover sends a bit back to the Verifier: a from Alice and b from Bob.
- The Verifier accepts if and only if $a \oplus b = f(s,t)$.

A definition that is essentially equivalent to this[3] appears in [32]. In the classical version, the provers have unlimited computing power, but are restricted to possessing classical information; in the quantum version, the provers may possess qubits whose joint state is entangled. In both versions, the communication between the provers and the verifier is classical.

[3] Except that *degeneracies* are allowed, where for some (s,t) pairs, the Verifier is allowed to accept or reject independently of the value of $a \oplus b$. All results quoted here apply to nondegenerate games.

Following [32], for an XOR game G, define its *classical value* $\omega_c(G)$ as the maximum success probability achievable by a classical strategy, i.e., if the provers do not share entanglement. Similarly, define its *quantum value* $\omega_q(G)$ as the maximum success probability achievable by a quantum strategy. It is convenient to define the *bias* of a quantum XOR game as $\varepsilon_q(G) = 2\omega_q(G) - 1$ and similarly $\varepsilon_c(G) = 2\omega_c(G) - 1$ in the classical case.

Using the definition of XOR games it is straightforward to define *XOR interactive proof systems.* A language L has an XOR interactive proof systems (with soundness probability s and completeness probability $c > s$) if it is possible to associate to each x an *efficient* XOR game G_x such that if $x \in L$ then the maximum acceptance probability over the prover's strategies is at least c and if $x \notin L$ then the maximum acceptance probability over prover's strategies is at most s. The game $G_x = (f, \pi)$ is *efficient* if the verifier can be efficiently implemented: More precisely we demand that S and T consist of strings of length polynomial in $|x|$, π can be sampled in time polynomial in $|x|$, and f can be computed in time polynomial in $|x|$. It is clear that the restriction to efficient XOR games is crucial for the definition of $\oplus$-MIP systems. However, our parallel repetition theorem will hold for all XOR games and therefore we will not talk about efficiency anymore.

7.2.3 XOR games and non-locality

It is interesting to note that quantum physicists have, in a sense, been studying quantum XOR games since the 1960s, when John Bell introduced his celebrated results that are now known as Bell inequality violations [12]. An example is the *CHSH* game, named after the authors of [25]. This game will play a prominent role in Chapter 8, but we will quickly explain it here. In this game, $S = T = \{0, 1\}$, π is the uniform distribution on $S \times T$, and $f(s,t) = s \wedge t$. In other words, the provers win if and only if they output bits a and b which satisfy $a \oplus b = s \wedge t$.

The best possible classical strategy succeeds with probability $3/4$, whereas the best possible quantum strategy succeeds with higher probability of $(1+1/\sqrt{2})/2 \approx 0.85$ [25, 93], which can be straight forwardly computed from the characterization in Section 7.3. This difference in success probabilities can be used to show that classical physics cannot explain all physical phenomena.[4]

7.3 Characterization of quantum XOR games

A quantum strategy for an XOR game consists of a bipartite quantum state $|\psi\rangle$ shared by Alice and Bob, a set of observables X_s $(s \in S)$ corresponding to Alice's

[4] Although it should be noted that because of the inaccuracies in today's quantum hardware it is not possible to completely rule out classical theories.

part of the quantum state, and a set of observables Y_t ($t \in T$) corresponding to Bob's part of the state. The bias achieved by this strategy is given by

$$\varepsilon_q(G) = \sum_{s,t} \pi(s,t)(-1)^{f(s,t)} \langle\psi|X_s \otimes Y_t|\psi\rangle .$$

Tsirelson's vector characterization

We make use of a vector characterization of XOR games due to [94] (also pointed out in [32]), which is a consequence of the following.

7.3.1. Theorem (**[94, 32]**). *Let S and T be finite sets, and let $|\psi\rangle$ be a pure quantum state with support on a bipartite Hilbert space $\mathcal{H} = \mathcal{A} \otimes \mathcal{B}$ such that $dim(\mathcal{A}) = dim(\mathcal{B}) = n$. For each $s \in S$ and $t \in T$, let X_s and Y_t be observables on $\mathcal{A}$ and $\mathcal{B}$ with eigenvalues ± 1 respectively. Then there exist real unit vectors x_s and y_t in $\mathbb{R}^{2n^2}$ such that*[5]

$$\langle\psi|X_s \otimes Y_t|\psi\rangle = x_s \cdot y_t,$$

for all $s \in S$ and $t \in T$.
Conversely, suppose that S and T are finite sets, and x_s and y_t are unit vectors in $\mathbb{R}^N$ for each $s \in S$ and $t \in T$. Let $\mathcal{A}$ and $\mathcal{B}$ be Hilbert spaces of dimension $2^{\lceil N/2 \rceil}$, $\mathcal{H} = \mathcal{A} \otimes \mathcal{B}$ and $|\psi\rangle$ be a maximally entangled state on $\mathcal{H}$. Then there exist observables X_s and Y_t with eigenvalues ± 1, on $\mathcal{A}$ and $\mathcal{B}$ respectively, such that

$$\langle\psi|X_s \otimes Y_t|\psi\rangle = x_s \cdot y_t,$$

for all $s \in S$ and $t \in T$.

A proof of this theorem can be found in Appendix D.

Using Theorem 7.3.1, we can characterize Alice and Bob's quantum strategies by a choice of unit vectors $\{x_s\}_{s\in S}$ and $\{y_t\}_{t\in T}$. Using this characterization, the bias becomes

$$\varepsilon_q(G) = \max_{\{x_s\},\{y_t\}} \sum_{s,t} \pi(s,t)(-1)^{f(s,t)}\, x_s \cdot y_t. \tag{7.1}$$

The *cost matrix* for the game is defined as the matrix A with entries $A_{s,t} = \pi(s,t)(-1)^{f(s,t)}$. Note that any matrix A, with the provision that the absolute values of the entries sum to 1, is the cost matrix of an XOR game.

Symmetry considerations and convex combinations of XOR games

We start by some symmetry considerations. If G_1 and G_2 are XOR games with cost matrices A_1 and A_2, then define the convex combination $\lambda G_1 + (1-\lambda)G_2$ to be the XOR game with cost matrix

$$\begin{pmatrix} 0 & \lambda A_1 \\ (1-\lambda)A_2 & 0 \end{pmatrix}.$$

[5] For real-valued vectors x, y we write the inner product $x^T y$ as $x \cdot y$.

This convex combination can be interpreted as the game where, with probability λ, game G_1 is played and, with probability $1-\lambda$, game G_2 is played (and Alice and Bob are informed about which game is occurring). Also, for a game G with cost matrix A, define G^T to be the game with cost matrix A^T. In other words, Alice and Bob switch places to play G^T. The following facts are easy to verify.

7.3.2. Proposition. *1.* $\varepsilon_q(G) = \varepsilon_q(G^T)$.

2. For all $0 \leq \lambda \leq 1$,

$$\varepsilon_q\left(\lambda G_1 + (1-\lambda)\, G_2\right) = \lambda \varepsilon_q(G_1) + (1-\lambda)\, \varepsilon_q(G_2)$$

Value of XOR games as SDP

The bias of a quantum XOR game may be stated as a semidefinite programming problem (SDP). We refer to Boyd and Vandenberghe [98] for a detailed introduction to semidefinite programming. For cost matrix A, the bias is equivalent to the objective value of problem

$$\begin{array}{rl} \max & \operatorname{Tr}\left(A^T U_1^T U_2\right) \\ \text{s.t.} & \operatorname{diag}\left(U_1^T U_1\right) = \operatorname{diag}\left(U_2^T U_2\right) = \bar{e} \end{array}, \tag{7.2}$$

where $\{x_s\}$ and $\{y_t\}$ appear as the columns of U_1 and U_2 respectively. Here $\operatorname{diag}(M)$ denotes the column vector of diagonal entries of the matrix M, and $\bar{e}$ is the column vector $(1,\ldots,1)^T$.

We note that instead of directly analyzing the XOR game G with cost matrix A we can also analyze the XOR game $\frac{1}{2}G + \frac{1}{2}G^T$ with cost matrix

$$B = \begin{pmatrix} 0 & \frac{1}{2}A \\ \frac{1}{2}A^T & 0 \end{pmatrix}.$$

It has useful structural properties, one of them being that it is symmetric, i.e. we are *guaranteed* that B is hermitian. Proposition 7.3.2 implies that $\varepsilon_q(\frac{1}{2}G+\frac{1}{2}G^T) = \varepsilon_q(G)$. Hence, we can express the value of game G in terms of the SDP (P_B) defined by

$$\begin{array}{rl} \max & \operatorname{Tr} BX \\ \text{s.t.} & \operatorname{diag}(X) = \bar{e}, \quad X \succeq 0 \end{array}.$$

The notation $X \succeq Y$ means that the matrix $X - Y$ lies in the cone of positive semidefinite matrices. That (P_B) is equivalent to problem (7.2) follows from the fact that a semidefinite matrix X can be written as $(U_1, U_2)^T(U_1, U_2)$ for some matrices U_1 and U_2.

Dual of SDP

To show that an optimal solution for (P_B) exists, we can examine the Lagrange-Slater dual of (P_B). The dual, denoted by (D_B), is (see Section 2.6)

$$\begin{array}{ll} \min & (x,y)\bar{e} \\ \text{s.t.} & \Delta(x,y) \succeq B \end{array},$$

where $\Delta(x,y)$ denotes the diagonal matrix with entries given by the (row) vectors x, y. Both (P_B) and (D_B) have Slater points—that is, feasible points in the interior of the semidefinite cone and are therefore strictly feasible. Explicitly, the identity matrix is a Slater point for (P_B), and $\bar{e}$ is a Slater point for (D_B). Therefore, by the strong duality theorem, the optimal values of (P_B) and (D_B) are the same and both problems have optimal solutions attaining this value.

7.3.3. Remark. For any XOR game G, the semidefinite programming relaxations of G due to Feige and Lovász [42] have value equal to the quantum value of G, given by equations (7.3) and (7.3). We say more about the relation of our result to the ones in [42] in Section 7.6.3.

7.4 Additivity theorem

For any two XOR games $G_1 = (f_1, \pi_1)$ and $G_2 = (f_2, \pi_2)$, define their *sum (modulo 2)* as the XOR game

$$G_1 \oplus G_2 = (f_1 \oplus f_2, \pi_1 \times \pi_2). \tag{7.3}$$

In this game, the verifier begins by choosing questions $((s_1,t_1),(s_2,t_2)) \in (S_1 \times T_1) \times (S_2 \times T_2)$ according to the product distribution $\pi_1 \times \pi_2$, sending (s_1, s_2) to Alice and (t_1, t_2) to Bob. Alice and Bob then win if and only if their respective outputs, a and b, satisfy $a \oplus b = f_1(s_1,t_1) \oplus f_2(s_2,t_2)$. If G_1 and G_2 have cost matrices A_1 and A_2 respectively, then the cost matrix of $G_1 \oplus G_2$ is $A_1 \otimes A_2$. The next proposition summarizes some simple facts.

7.4.1. Proposition. *1.* $\varepsilon_q(G_1 \oplus G_2) = \varepsilon_q(G_2 \oplus G_1)$

2. For all $0 \le \lambda \le 1$,

$$G_1 \oplus (\lambda G_2 + (1-\lambda)G_3) = \lambda(G_1 \oplus G_2) + (1-\lambda)(G_1 \oplus G_3).$$

A simple way for Alice and Bob (who may or may not share entanglement) to play $G_1 \oplus G_2$ is to optimally play G_1 and G_2 separately, producing outputs a_1, b_1 for G_1 and a_2, b_2 for G_2, and then to output $a = a_1 \oplus a_2$ and $b = b_1 \oplus b_2$ respectively. It is straightforward to calculate that the above method for playing $G_1 \oplus G_2$ succeeds with probability

$$\omega(G_1)\omega(G_2) + (1-\omega(G_1))(1-\omega(G_2)), \tag{7.4}$$

where ω (and later ε) can be indexed by q or c depending on whether Alice and Bob share entanglement. Then it is easy to see that the bias $\varepsilon^{trivial}(G_1 \oplus G_2)$ for this particular strategy of playing $G_1 \oplus G_2$ is

$$\varepsilon^{trivial}(G_1 \oplus G_2) = \varepsilon(G_1)\varepsilon(G_2). \tag{7.5}$$

Is this the optimal way to play $G_1 \oplus G_2$?

The answer is *no* for *classical* strategies. To see why this is so, note that, using this approach for the XOR game *CHSH* $\oplus$ *CHSH*, produces a success probability of 5/8. A better strategy is for Alice to output $a = s_1 \wedge s_2$ and Bob to output $b = t_1 \wedge t_2$. It is straightforward to verify that this latter strategy succeeds with probability 3/4.

Our first result is that the answer is *yes* for *quantum* strategies.

7.4.2. Theorem (**Additivity**). *For any two XOR games G_1 and G_2 an optimal quantum strategy for playing $G_1 \oplus G_2$ is for Alice and Bob to optimally play G_1 and G_2 separately, producing outputs a_1, b_1 for G_1 and a_2, b_2 for G_2, and then to output $a = a_1 \oplus a_2$ and $b = b_1 \oplus b_2$.*

The proof uses the characterization of quantum strategies for individual XOR games as semidefinite programs from Section 7.3.

Recall that we defined the quantum *bias* of an XOR game as as $\varepsilon_q(G) = 2\omega_q(G) - 1$. Then, due to equation (7.5), we already have one part of Theorem 7.4.2.

7.4.3. Proposition. *For two XOR games G_1 and G_2,*

$$\varepsilon_q(G_1 \oplus G_2) \geq \varepsilon_q(G_1)\varepsilon_q(G_2).$$

The nontrivial part of the proof of Theorem 7.4.2 is the reverse inequality.

The next lemma establishes the upper bound for the game $(\frac{1}{2}G_1 + \frac{1}{2}G_1^T) \oplus (\frac{1}{2}G_2 + \frac{1}{2}G_2^T)$ (which we will show afterwards has the same bias as $G_1 \oplus G_2$).

7.4.4. Lemma. *If G_1 and G_2 are XOR games, then*

$$\varepsilon_q((\tfrac{1}{2}G_1 + \tfrac{1}{2}G_1^T) \oplus (\tfrac{1}{2}G_2 + \tfrac{1}{2}G_2^T)) \leq \varepsilon_q(G_1)\varepsilon_q(G_2).$$

Proof: Let G_1 and G_2 be two games with cost matrices A_1 and A_2, respectively, and let

$$B_1 = \begin{pmatrix} 0 & \frac{1}{2}A_1 \\ \frac{1}{2}A_1^T & 0 \end{pmatrix} \text{ and } B_2 = \begin{pmatrix} 0 & \frac{1}{2}A_2 \\ \frac{1}{2}A_2^T & 0 \end{pmatrix}. \tag{7.6}$$

Let (x_1, y_1) and (x_2, y_2) be optimal solutions to (D_{B_1}) and (D_{B_2}), respectively, which implies $\Delta(x_i, y_i) - B_i \succeq 0$ and $\varepsilon_q(G_i) = (x_i, y_i)\bar{e}$, for $i = 1, 2$. It suffices to show that $(x_1, y_1) \otimes (x_2, y_2)$ is a solution to $(\mathrm{D}_{B_1 \otimes B_2})$, since $B_1 \otimes B_2$ is the cost matrix of $(\frac{1}{2}G_1 + \frac{1}{2}G_1^T) \oplus (\frac{1}{2}G_2 + \frac{1}{2}G_2^T)$. Note that, for *arbitrary* B_1 and B_2, $\Delta(x_1, y_1) \succeq B_1$ and $\Delta(x_2, y_2) \succeq B_2$ does *not* imply that $\Delta(x_1, y_1) \otimes \Delta(x_2, y_2) \succeq B_1 \otimes B_2$ (a simple counterexample is when $\Delta(x_1, y_1) = \Delta(x_2, y_2) = 0$ and $B_1 = B_2 = -I$). We make use of the structure of B_1 and B_2 arising from equation (7.6). For each i, $\Delta(x_i, y_i) - B_i \succeq 0$ implies that, for all (row) vectors u, v,

$$\begin{aligned} 0 \quad &\leq \quad \begin{pmatrix} u & v \end{pmatrix} \begin{pmatrix} \Delta(x_i) & -\frac{1}{2}A_i \\ -\frac{1}{2}A_i^T & \Delta(y_i) \end{pmatrix} \begin{pmatrix} u^T \\ v^T \end{pmatrix} \\ &= \quad \begin{pmatrix} u & -v \end{pmatrix} \begin{pmatrix} \Delta(x_i) & +\frac{1}{2}A_i \\ +\frac{1}{2}A_i^T & \Delta(y_i) \end{pmatrix} \begin{pmatrix} u^T \\ -v^T \end{pmatrix}, \end{aligned}$$

which in turn implies that $\Delta\left(x_i, y_i\right) + B_i \succeq 0$ also holds. Therefore,

$$\begin{aligned} (\Delta(x_1, y_1) - B_1) \otimes (\Delta(x_2, y_2) + B_2) \quad &\succeq \quad 0 \quad \text{and} \\ (\Delta(x_1, y_1) + B_1) \otimes (\Delta(x_2, y_2) - B_2) \quad &\succeq \quad 0, \end{aligned}$$

which, by averaging, yields

$$\Delta(x_1, y_1) \otimes \Delta(x_2, y_2) - B_1 \otimes B_2 \succeq 0.$$

Therefore, $(x_1, y_1) \otimes (x_2, y_2)$ is a feasible point in the dual $(\mathrm{D}_{B_1 \otimes B_2})$, which obtains the objective value $\varepsilon_q(G_1)\varepsilon_q(G_2)$. Noting that $(\frac{1}{2}G_1 + \frac{1}{2}G_1^T) \oplus (\frac{1}{2}G_2 + \frac{1}{2}G_2^T)$ has cost matrix $B_1 \otimes B_2$ implies the Lemma. ■

Now we may complete the proof of Theorem 7.4.2. Using Proposition 7.4.3 for line (7.7), Lemma 7.4.4 for line (7.8) and Propositions 7.3.2 and 7.4.1 and some easy algebra for the rest we can derive the following

$$\begin{aligned} &\varepsilon_q\left(G_1 \oplus G_2\right) \\ &\quad \geq \quad \varepsilon_q\left(G_1\right) \varepsilon_q\left(G_2\right) &(7.7)\\ &\quad \geq \quad \varepsilon_q((\tfrac{1}{2}G_1 + \tfrac{1}{2}G_1^T) \oplus (\tfrac{1}{2}G_2 + \tfrac{1}{2}G_2^T)) &(7.8)\\ &\quad = \quad \varepsilon_q\left(\tfrac{1}{4}(G_1 \oplus G_2) + \tfrac{1}{4}(G_1 \oplus G_2^T) + \tfrac{1}{4}(G_1^T \oplus G_2) + \tfrac{1}{4}(G_1^T \oplus G_2^T)\right) \\ &\quad = \quad \varepsilon_q\left(\tfrac{1}{2}\left[\tfrac{1}{2}(G_1 \oplus G_2) + \tfrac{1}{2}(G_1 \oplus G_2^T)\right] + \tfrac{1}{2}\left[\tfrac{1}{2}(G_1 \oplus G_2) + \tfrac{1}{2}(G_1 \oplus G_2^T)\right]^T\right) \\ &\quad = \quad \tfrac{1}{2}\varepsilon_q\left(G_1 \oplus G_2\right) + \tfrac{1}{2}\varepsilon_q\left(G_1 \oplus G_2^T\right). \end{aligned}$$

Therefore $\varepsilon_q(G_1 \oplus G_2) \geq \varepsilon_q(G_1 \oplus G_2^T)$. By symmetry, $\varepsilon_q(G_1 \oplus G_2^T) \geq \varepsilon_q(G_1 \oplus G_2)$, as well, which means that all of the above inequalities must be equalities. This completes the proof of Theorem 7.4.2.

7.5 Parallel repetition theorem

For any sequence of XOR games $G_1 = (f_1, \pi_1)$, ..., $G_n = (f_n, \pi_n)$, define their *conjunction*, denoted by $\wedge_{j=1}^n G_j$, as follows. The verifier chooses questions

$$((s_1, t_1), \ldots, (s_n, t_n)) \in (S_1 \times T_1) \times \cdots \times (S_n \times T_n)$$

according to the product distribution $\pi_1 \times \cdots \times \pi_n$, and sends $(s_1, \ldots, s_n)$ to Alice and $(t_1, \ldots, t_n)$ to Bob. Alice and Bob output bits $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$, respectively, and win if and only if their outputs simultaneously satisfy these n conditions: $a_1 \oplus b_1 = f_1(s_1, t_1)$, $\ldots$, $a_n \oplus b_n = f_n(s_n, t_n)$. (Note that $\wedge_{j=1}^n G_j$ is not itself an XOR game for $n > 1$.)

One way for Alice and Bob to play $\wedge_{j=1}^n G_j$ is to independently play each game optimally. This succeeds with probability $\prod_{j=1}^n \omega(G_j)$. Is this the optimal way to play $\wedge_{j=1}^n G_j$?

The answer is again *no* for classical strategies. It is shown in [11] that[6] $\omega_c(\mathit{CHSH} \wedge \mathit{CHSH}) = 10/16 > 9/16 = \omega_c(\mathit{CHSH})\omega_c(\mathit{CHSH})$.

Our second result is that the answer is *yes* for quantum strategies.

7.5.1. Theorem (**Parallel Repetition**). *For any XOR games $G_1, \ldots, G_n$, we have that $\omega_q(\wedge_{j=1}^n G_j) = \prod_{j=1}^n \omega_q(G_j)$.*

This is a quantum version of Raz's parallel repetition theorem [78] for the restricted class of XOR games. We call it a *perfect* parallel repetition theorem because the probabilities are multiplicative in the exact sense (as opposed to an asymptotic sense, as in [78]). The proof of Theorem 7.5.1 is based on Theorem 7.4.2 combined with Fourier analysis techniques for boolean functions. Section 7.5 contains the proof.

In this section we prove Theorem 7.5.1, which is stated in Section 7.5.

We begin with the following simple probabilistic lemma.

7.5.2. Lemma. *For any sequence of binary random variables $X_1, X_2, \ldots, X_n$,*

$$\frac{1}{2^n} \sum_{M \subseteq [n]} \mathrm{E}\left[(-1)^{\oplus_{j \in M} X_j}\right] = \Pr[X_1 \ldots X_n = 0 \ldots 0].$$

Proof: By the linearity of expectation,

$$\begin{aligned}
&\frac{1}{2^n} \sum_{M \subseteq [n]} \mathrm{E}\left[(-1)^{\oplus_{j \in M} X_j}\right] \\
&= \mathrm{E}\left[\frac{1}{2^n} \sum_{M \subseteq [n]} (-1)^{\oplus_{j \in M} X_j}\right] \\
&= \mathrm{E}\left[\prod_{j=1}^{n} \left(\frac{1 + (-1)^{X_j}}{2}\right)\right] \\
&= \Pr\left[X_1 \ldots X_n = 0 \ldots 0\right],
\end{aligned}$$

[6] After posing this question about $\omega_c(\mathit{CHSH} \wedge \mathit{CHSH})$, the answer was first shown to us by S. Aaronson, who later found that this result was already stated in [11].

where the last equality follows from the fact that

$$\prod_{j=1}^{n}(1+(-1)^{X_j}) \neq 0$$

only if $X_1 \dots X_n = 0 \dots 0$. ■

We introduce the following terminology. For any strategy $\mathcal{S}$—*classical* or *quantum*—and for any game G, define $\omega(\mathcal{S}, G)$ as the success probability of strategy $\mathcal{S}$ on game G. Similarly, define the corresponding bias as $\varepsilon(\mathcal{S}, G) = 2\omega(\mathcal{S}, G) - 1$.

Now let $\mathcal{S}$ be any protocol for the game $\wedge_{j=1}^{n} G_j$. For each $M \subseteq [n]$, define the protocol $\mathcal{S}_M$ (for the game $\oplus_{j \in M} G_j$) as follows.

1. Run protocol $\mathcal{S}$, yielding $a_1, \dots, a_n$ for Alice and $b_1, \dots, b_n$ for Bob.
2. Alice outputs $\oplus_{j \in M} a_j$ and Bob outputs $\oplus_{j \in M} b_j$.

7.5.3. Lemma.

$$\frac{1}{2^n} \sum_{M \subseteq [n]} \varepsilon(\mathcal{S}_M, \oplus_{j \in M} G_j) = \omega(\mathcal{S}, \wedge_{j=1}^{n} G_j).$$

Proof: For all $j \in [n]$, define $X_j = a_j \oplus b_j \oplus f_j(s_j, t_j)$. Then, for all $M \subseteq [n]$, we have $\mathrm{E}[(-1)^{\oplus_{j \in M} X_j}] = \varepsilon(\mathcal{S}_M, \oplus_{j \in M} G_j)$, and $\Pr[X_1 \dots X_n = 0 \dots 0] = \omega(\mathcal{S}, \wedge_{j=1}^{n} G_j)$. The result now follows from Lemma 7.5.2. ■

7.5.4. Corollary.

$$\omega_c(\wedge_{j=1}^{n} G_j) \leq \frac{1}{2^n} \sum_{M \subseteq [n]} \varepsilon_c(\oplus_{j \in M} G_j) \tag{7.9}$$

and

$$\omega_q(\wedge_{j=1}^{n} G_j) \leq \frac{1}{2^n} \sum_{M \subseteq [n]} \varepsilon_q(\oplus_{j \in M} G_j). \tag{7.10}$$

Now, to complete the proof of Theorem 7.5.1, using Theorem 7.4.2, we have

$$\begin{aligned} \frac{1}{2^n} \sum_{M \subseteq [n]} \varepsilon_q(\oplus_{j \in M} G_j) &= \frac{1}{2^n} \sum_{M \subseteq [n]} \prod_{j \in M} \varepsilon_q(G_j) \\ &= \prod_{j=1}^{n} \left(\frac{1 + \varepsilon_q(G_j)}{2} \right) \\ &= \prod_{j=1}^{n} \omega_q(G_j). \end{aligned} \tag{7.11}$$

Combining this with equation (7.10), we deduce $\omega_q(\wedge_{j=1}^{n} G_j) = \prod_{j=1}^{n} \omega_q(G_j)$, which completes the proof of Theorem 7.5.1. ■

7.6 Discussion

A natural question to ask is whether it is possible to extend the proof for other kinds of games. We have already mentioned that a *perfect* parallel repetition cannot hold for classical XOR games, since it does not even hold for classical CHSH games. The next remark gives some more details about how the classical CHSH game behaves under repetition. The second remark in this section will show that a perfect parallel repetition theorem can also not hold for *general* quantum games. We will conclude by explaining the connection of our results to Feige-Lovász games games.

7.6.1 Perfect parallel repetition of classical XOR games

Although equation (7.10) is used to prove a tight upper bound on $\omega_q(\wedge_{j=1}^n G_j)$, equation (7.9) cannot be used to obtain a tight upper bound on $\omega_c(\wedge_{j=1}^n G_j)$ for general XOR games. This is because $\varepsilon_c(\mathit{CHSH}) = \varepsilon_c(\mathit{CHSH} \oplus \mathit{CHSH}) = 1/2$ and it can be shown that $\varepsilon_c(\mathit{CHSH} \oplus \mathit{CHSH} \oplus \mathit{CHSH}) = 5/16$. Therefore, for $G_1 = G_2 = G_3 = \mathit{CHSH}$, the right side of equation (7.9) is $\frac{1}{8}\sum_{M\subseteq[3]} \varepsilon_c(\oplus_{j\in M} G_j) = 34.5/64$, whereas $\omega_c(\wedge_{j=1}^3 G_j)$ must be expressible as an integer divided by 64 (in fact[7], $\omega_c(\wedge_{j=1}^3 G_j) = 31/64$).

7.6.2 Parallel repetition of general games

We give the unpublished proof due to Watrous [101] that there is a binary game G (that is not an XOR game) for which $\omega_q(G) = \omega_q(G \wedge G) = 2/3$. The game used was originally proposed by Fortnow, Feige and Lovász [45, 42], who showed that $\omega_c(G) = \omega_c(G \wedge G) = 2/3$.

The game has binary questions ($S = T = \{0, 1\}$) and binary answers ($A = B = \{0, 1\}$). The operation of the game is as follows. The Verifier selects a pair of questions (s, t) uniformly from $\{(0, 0), (0, 1), (1, 0)\}$ and sends s and t to Alice and Bob, respectively. Then the Verifier accepts the answers, a from Alice and b from Bob, if and only if $s \vee a \neq t \vee b$.

Consider a quantum strategy for this game, where $|\phi\rangle$ is the shared entangled state. We may assume that Alice's behavior is determined by the observables A_0 and A_1, and Bob's behavior is determined by the observables B_0 and B_1 and that all observables have only eigenvalues ±1. On input (s, t), Alice computes a by measuring with respect to A_s, and Bob computes b by measuring with respect to B_t. It is straightforward to deduce that the bias of this strategy is

$$\langle\phi| \left(-\tfrac{1}{3}A_0 \otimes B_0 + \tfrac{1}{3}A_0 \otimes \mathbb{I}_B + \tfrac{1}{3}\mathbb{I}_A \otimes B_0\right) |\phi\rangle \tag{7.12}$$

[7] This was independently calculated by S. Aaronson and B. Toner, by searching over a finite number of deterministic classical strategies.

(curiously, the bias does not depend on A_1 or B_1). We can rewrite $-\frac{1}{3}A_0 \otimes B_0 + \frac{1}{3}A_0 \otimes \mathbb{I}_B + \frac{1}{3}\mathbb{I}_A \otimes B_0 = \frac{1}{3}(\mathbb{I}_A - A_0) \otimes (B_0 - \mathbb{I}_B) + \frac{1}{3}(\mathbb{I}_A \otimes \mathbb{I}_B)$, and the bias becomes

$$\frac{1}{3}\langle\phi|(\mathbb{I}_A - A_0) \otimes (B_0 - \mathbb{I}_B)|\phi\rangle + \frac{1}{3}. \tag{7.13}$$

We note that $\mathbb{I}_A - A_0$ has eigenvalues 0 and 2 and $B_0 - \mathbb{I}_B$ has eigenvalues 0 and -2, from which we can conclude that the hermitian matrix $(\mathbb{I}_A - A_0) \otimes (B_0 - \mathbb{I}_B)$ has no positive eigenvalues. This implies that $\varepsilon_q(G) = 1/3$ and further $\omega_q(G) = 2/3$. Combining this with the fact that $2/3 = \omega_c(G \wedge G) \leq \omega_q(G \wedge G) \leq \omega_q(G)$, we obtain $\omega_q(G \wedge G) = \omega_q(G) = 2/3$.

7.6.3 Feige-Lovász games

For a broad class of games, Feige and Lovász [42] define quantities that are relaxations—and hence upper bounds—of their classical values, and show that one of these quantities satisfies a parallel repetition property analogous to Theorem 7.5.1. For any XOR game G, the Feige-Lovász relaxations of its classical value are equal to the quantum value of G. This was noted first in [40, 41] and an explicit proof appears in the appendix of [30]. It is important to note that, *for general games*, the relationship between their quantum values and the Feige-Lovász relaxations of their classical values are not understood. As far as we know, neither quantity bounds the other for general games.

Using this relation between the value of a quantum XOR game and the value of its Feige-Lovász relaxation combined with Theorem 7.5.1, it follows that for XOR games $G_1, \ldots, G_n$, the quantum value of $\wedge_{j=1}^n G_j$ is also determined by its associated Feige-Lovász relaxation. However, it should be stressed that the parallel repetition property for Feige-Lovász relaxations does not imply our Theorem 7.5.1, since we do not know a priori that for the non-XOR game $\wedge_{j=1}^n G_j$ the same relation between its quantum value and the value of its Feige-Lovász relaxation holds. Our Theorem 7.5.1 shows this.

Chapter 8

Limits on non-locality from communication complexity

This chapter is based on the paper

> Gilles Brassard, Harry Buhrman, Noah Linden, André Allan Methot, Alain Tapp, and Falk Unger, **Limit on nonlocality in any world in which communication complexity is not trivial**, *Physical Review Letters* 96(25), 2006.

8.1 Introduction

Quantum mechanics is a physical theory which is hugely successful in describing very small physical systems, on the scale of atoms. Its foundations were laid out in the 1920s and 1930s. At first quantum mechanics was not accepted immediately by all physicists, because of its counter-intuitive properties and predictions. Most notably, Albert Einstein rejected quantum mechanics with the famous words "Gott würfelt nicht", which translates to "God does not play dice". Nowadays, quantum mechanics is widely accepted, mainly due to the fact that quantum mechanics predicts outcomes of many experiments where classical physics fails. For example, later in this chapter we will describe (an abstract setting of) an experiment, so called Bell inequality violations, which can be explained by quantum mechanics but not by classical mechanics. Because of experiments of this kind, quantum mechanics is nowadays widely believed to accurately describe the world in the "small".

Nevertheless, despite its successes in experiments, quantum mechanics is often considered mysterious, due to its counter-intuitive predictions and intriguing properties. One of these properties is entanglement. In our example, based on a violation of a Bell inequality, we will see that two separated, non-communicating parties possessing entanglement can create correlations which are not attainable without entanglement. However, it follows from the axioms of quantum mechanics that also entanglement does not allow for all (causal[1]) correlations which are in principle conceivable in the physical world. The question we want to address in this chapter is whether this limitation of achievable correlations is more than merely a consequence of quantum mechanics. We will explain why limitations of possible correlations can be seen as a "natural axiom", which every reasonable physical theory (including quantum mechanics) should obey. In our argument we show that under a reasonable assumption about the physical world—namely that communication complexity (explained in Section 2.4 or later in this chapter) is non-trivial—restrictions on achievable correlations *are indeed necessary.*

This chapter is about the axioms of physics. Since they cannot be stated as rigorously as mathematical theories, the first part (in particular the introduction) of this chapter will be less rigorous than the other chapters. After we have set the stage and defined a suitable mathematical model, the proofs presented will be rigorous.

CHSH inequality

We will start by explaining the CHSH inequality, a particular type of *Bell inequality*. Assume two parties, Alice and Bob, share a quantum state $|\psi\rangle$ and they can both perform two different measurements on their respective parts of $|\psi\rangle$, with binary outcomes 0 or 1. Let Alice's choice of the measurement be denoted by $x \in \{0,1\}$ and Bob's by $y \in \{0,1\}$. Let Alice's and Bob's outcomes be $a, b \in \{0,1\}$, respectively. Tsirelson [93] proved a bound on the correlation between Alice's and Bob's outcomes, see also Figure 1.3:

$$\frac{1}{4}\sum_{x,y} \Pr[x \cdot y = a \oplus b] \leq \wp = \frac{1}{2} + \frac{\sqrt{2}}{4} \approx 85\%. \tag{8.1}$$

Here $x \cdot y$ is the logical AND of x and y and $a \oplus b$ is the logical XOR of a and b. This means that $x \cdot y = a \oplus b$ is satisfied if and only if (1) $x = y = 1$ and $a \neq b$ or if (2) x and y are not both 1 and $a = b$.

In quantum mechanics, this bound is actually tight and can be attained, see [97, 96] for a derivation. Furthermore, it is even possible to achieve that

$$\forall_{x,y} \Pr[x \cdot y = a \oplus b] = \wp = \frac{1}{2} + \frac{\sqrt{2}}{4}, \tag{8.2}$$

which means that it is possible to guarantee *optimal worst-case behaviour.*

[1] explained later

However, classical physics, using the so-called local hidden-variable model (LHV), only allows a correlation of up to 3/4. Bell [12] and Clauser, Horne, Shimony and Holt [25] proved for classical LHV-theories

$$\frac{1}{4}\sum_{x,y}\Pr[x\cdot y = a\oplus b] \leq \frac{3}{4}, \tag{8.3}$$

which is known as the *CHSH inequality*. It is a special kind of Bell inequality. It can be used to test local hidden-variable theories because it follows also from CHSH that in a local realistic theory (i.e. under a local hidden-variable model) Alice and Bob cannot succeed with probability greater than $\frac{3}{4}$ if they are space-like separated. If correlations greater than $\frac{3}{4}$ can be detected in an experiment, then we conclude that the world cannot be *local realistic* and in particular not classical.

The CHSH inequality shows up in many different contexts in quantum mechanics and appears under different names. In this chapter we will discuss it in the framework of non-local boxes (whose exact definition is not important at the moment and will be given in the next section) as we focus on the non-local properties of quantum mechanics. The CHSH inequality can also be cast in terms of XOR games, which are the focus of Chapter 7.

Violations of CHSH inequality

Within the framework of non-local boxes it is possible to define correlations for which the left-hand side of equation (8.1) becomes 1, but which do not violate causality (see Section 8.2). This was first observed in 1995 by Popescu and Rohrlich and elaborated on in a series of papers [74, 75, 76]. They asked: Why does Quantum Mechanics not allow a higher correlation in equation (8.5) than $\wp = 1/2 + \sqrt{2}/4$? In fact, they constructed a toy theory with correlations of up to 1 in equation (8.1), which did not exhibit any apparent inconsistency.

This question was later answered by Cleve [26] and van Dam [97, 96] who showed that a maximal correlation of 1 in (8.1) would imply some improbable consequences in the real world, namely that all functions have *trivial communication complexity* of just one bit. Recall from Section 2.4 that for certain functions $f : \{0,1\}^n \times \{0,1\}^n \mapsto \{0,1\}$ Alice and Bob have to communicate essentially n bits to compute the value of f. An example was the inner product function, but in fact this holds for almost all functions. Indeed, trivial communication complexity seems too good to be true.

When we extend the notion of "trivial" communication complexity to bounded error protocols (either with public random coins or shared entanglement) the inner product function remains nontrivial according to quantum mechanics: In Appendix C we use an argument from [104] to show that Alice and Bob cannot succeed at computing the inner product function on n bits with probability $1-\epsilon$

if they transmit fewer than $n - 2\log_2 \frac{1}{1-2\epsilon}$ bits, even if they share prior entanglement. Hence, if we assume that Nature does not offer "free lunch" by allowing to compute functions on distributed inputs with trivial communication, then the above results by Cleve and van Dam imply that Nature also cannot allow all possible CHSH-correlations.

However, their argument only excludes CHSH-correlations which achieve the maximum value of 1. In this chapter we will improve this and answer a stronger question: *Considering that values larger than $\wp$ on the right hand side of (8.1) would not violate causality, why do the laws of quantum mechanics only allow correlations of up to $\wp$, but not something better?* In fact, could it even be that this "magic" value of $\wp$ is not only a consequence of quantum mechanics but rather a *defining* property and should be seen as an axiom [75]? In this chapter we attempt to give a partial answer to this question by generalizing van Dam's and Cleve's result: If correlations of more than $\frac{3+\sqrt{6}}{6} \approx 90.8\%$ were possible, still all functions with shared inputs could be computed with bounded error with trivial communication complexity (i.e., with just just one bit of communication).

Ideally, we would want to show that for all correlations higher than $\wp \approx 85\%$, communication complexity becomes trivial. This indeed would imply that the value of $\wp$ may be taken as an axiom of any (reasonable) physical theory. It is an interesting open problem to determine whether our result can be extended up to the value $\wp$.

For the precise statement of our result and our proof we use the more modern framework of non-local boxes, introduced by Popescu and Rohrlich [74, Eq. (7)]. This will be explained in the next section. We then prove our main result, Theorem 8.3.1.

8.2 Non-local boxes

A *non-local box* (NLB) is an imaginary device shared between Alice and Bob, who can be arbitrarily far apart. It has an input-output port at Alice's and another one at Bob's. Whenever Alice feeds a bit x into her input port, she instantaneously gets a uniformly distributed random output bit a, locally uncorrelated with anything else, including her own input bit. The same applies to Bob, whose input and output bits we call y and b, respectively. The "magic" appears in the form of a correlation between the pair of outputs and the pair of inputs: the exclusive-or of the outputs is always equal to the logical AND of the inputs: $a \oplus b = x \cdot y$. Much like the correlations that can be established by use of quantum entanglement, this device is atemporal: Alice gets her output as soon as she feeds in her input, regardless of whether (and when) Bob feeds in *his* input, and vice versa. These devices are also known under the name "Popescu-Rohrlich box" or PR-box. The name *non-local box* derives from the fact that an operation (i.e., choosing an input bit and receiving an output bit) on Alice's side has an

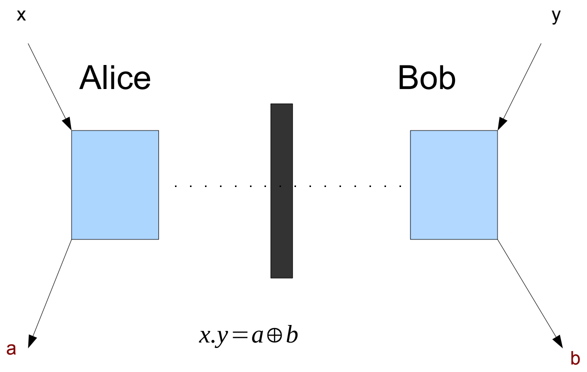


Figure 8.1: Non-local box

instantaneous effect on Bob's side as well. In particular, depending on the values x and a which Alice observes, there is a deterministic function f computing Bob's output $b := f(y)$. The function f is either the identity function, logical negation, constant zero or constant one.

Also inspired by entanglement, this is a *one-shot* device: the correlation appears only as a result of the first pair of inputs fed in by Alice and Bob.

NLBs cannot be used by Alice and Bob to signal instantaneously to one another, i.e., they are *non-signalling*. This is because the outputs that can be observed are purely random from a local perspective.[2] Hence, NLBs are *causal*: they cannot make an effect precede its cause in the context of special relativity. We are interested in the question of how well the correlation of NLBs can be *approximated*. An *approximation of an NLB with success probability* p is an NLB with the property that

$$\forall_{x,y} \sum_{x,y} \Pr[x \cdot y = a \oplus b] \geq p. \tag{8.4}$$

Alternatively, we can define approximate NLBs operationally, by saying that an approximate NLB is "obtained" from a (perfect) NLB by flipping Bob's output bit with probability $1 - p$.

[2] Remember that Bob's output is "created" even if Alice has not yet input a bit into her input port and vice versa. It is even possible to demand that locally the output bit a of an NLB is always uniformly random, if b is not yet determined; and vice versa.

Although originally presented differently, the CHSH inequality can be recast in terms of imperfect NLBs. The availability of shared entanglement allows Alice and Bob to approximate NLBs with success probability

$$\wp = \cos^2 \tfrac{\pi}{8} = \tfrac{2+\sqrt{2}}{4} \approx 85.4\% \,.$$

Tsirelson proved the optimality of the CHSH inequality, which translates into saying that quantum mechanics does not allow for a success probability greater than $\wp$ at the game of simulating NLBs [93]. See also [24] for an information-theoretic proof of the same result.

In the next section we attempt to show *why* the axioms of quantum mechanics are such that they do not allow to approximate NLBs with success probability greater than $\wp$.

8.3 Main result

Our main theorem is stated below and proved in the rest of this chapter. It shows that even the availability of imperfect NLBs would dramatically change the picture of communication complexity discussed in Section 2.4: It would make the randomized communication complexity of all functions trivial. Indeed, most computer scientists would consider a world in which randomized communication complexity is trivial to be as surprising as a modern physicist would find the violation of causality.

8.3.1. Theorem. *In any world in which it is possible, without communication, to implement an approximate NLB that works correctly with probability greater than* $\frac{3+\sqrt{6}}{6} \approx 90.8\%$*, i.e.*

$$\forall_{x,y} \Pr[x \cdot y = a \oplus b] > \frac{3+\sqrt{6}}{6} \approx 90.8\%,$$

every *Boolean function has trivial probabilistic communication complexity of just one bit.*

8.4 Proof

To prove this theorem, we introduce the notion of *distributed computation* and the notion of *bias* for such computations. Then, we explain how compute any Boolean function with small bias and show how to amplify this “natural” bias by having Alice and Bob calculate it many times and taking the majority. We determine how imperfect a majority gate can be and still increase the bias. Finally, we construct a majority gate with the use of NLBs, and we determine to what extent we can allow *them* to be faulty.

8.4.1 Distributed computation

8.4.1. Definition. A bit c is *distributed* if Alice has bit a and Bob bit b such that $c = a \oplus b$.

8.4.2. Definition. A Boolean function f is *distributively computed* by Alice and Bob if, given inputs x and y, they can produce a distributed bit equal to $f(x, y)$.

8.4.3. Definition. A Boolean function is *biased* if it can be distributively computed without any communication and with probability strictly greater than $\frac{1}{2}$.

8.4.4. Lemma. *Provided Alice and Bob are allowed to share random variables,* all *Boolean functions are biased.*

Proof: Let $f : \{0,1\}^n \times \{0,1\}^n \mapsto \{0,1\}$ be an arbitrary Boolean function and let Alice and Bob share a uniformly distributed random variable $z \in \{0,1\}^n$. Upon receiving her input x, Alice produces $a = f(x, z)$. Bob's strategy is to test whether $y = z$. If so, he produces $b = 0$; if not, he produces a uniformly distributed random bit b. In the lucky event that $y = z$, the bit distributed between Alice and Bob is correct since $a \oplus b = f(x, z) \oplus 0 = f(x, y)$. In all other cases, the distributed bit $a \oplus b$ is uniformly random. Summing up, the distributed bit is correct with probability $\frac{1}{2^n} + (1 - \frac{1}{2^n})\frac{1}{2} = \frac{1}{2} + \frac{1}{2^{n+1}} > \frac{1}{2}$. ■

8.4.5. Definition. A Boolean function has *bounded bias* if it can be distributively computed without any communication and with probability bounded away from $\frac{1}{2}$, with probability at least $\frac{1}{2} + \epsilon$ for some $\epsilon > 0$.

The difference between bias and *bounded* bias is that the probability of being correct in the former case can come arbitrarily close to $\frac{1}{2}$ as the input size increases. In the latter case, there must be some fixed $p > \frac{1}{2}$ such that the probability of being correct is at least p no matter how large the inputs are.

8.4.6. Lemma. *Any Boolean function that has bounded bias has trivial probabilistic communication complexity of one bit.*

Proof: Assume Boolean function f has bounded bias. For all inputs x and y, Alice and Bob can produce bits a and b, respectively, without communication, such that $a \oplus b = f(x, y)$ with probability at least $p > \frac{1}{2}$. If Bob transmits his single bit b to Alice, she can compute $a \oplus b$, which is correct with bounded error probability. ■

8.4.2 Bias Amplification

8.4.7. Definition. The *non-local majority* problem consists in computing the distributed majority of three distributed bits. More precisely, let Alice have bits x_1, x_2, x_3 and Bob have y_1, y_2, y_3. The purpose is for Alice and Bob to compute a and b, respectively, such that

$$a \oplus b = \mathrm{Maj}(x_1 \oplus y_1, x_2 \oplus y_2, x_3 \oplus y_3)\,,$$

where $\mathrm{Maj}(u, v, w)$ denotes the bit occurring the most among u, v and w. The computation must be achieved without any communication between Alice and Bob.

Von Neumann proved a statement rather similar to Lemma 8.4.8 below in 1956, albeit not in the context of distributed computation [67]. A more general result appears also in [38]. We sketch the proof nevertheless for the sake of completeness.

8.4.8. Lemma. *For any q such that $\frac{5}{6} < q \le 1$, if Alice and Bob can compute non-local majority with probability at least q, then every Boolean function has bounded bias.*

Proof: Let f be an arbitrary Boolean function, fix Bob's input size, and consider any $p > \frac{1}{2}$ so that Alice and Bob can distributively compute f with probability at least p. We know from Lemma 8.4.4 that such a p exists (although it may depend on the input size). Let Alice and Bob apply their distributed computational process three times, with independent random choices and shared random variables each time. This produces three distributed bits such that each of them is correct with probability at least p. Now, let Alice and Bob compute the non-local majority of these three outcomes with correctness probability at least q, which we assumed they can do. Because the overall result will be correct either if most of the distributed outcomes were correct and the distributed majority calculation was performed correctly, or if most of the distributed outcomes were wrong and the distributed majority calculation was performed incorrectly, the probability that the distributed majority as computed yields the correct value of f is at least

$$h(p) = q(p^3 + 3p^2(1-p)) + (1-q)(3p(1-p)^2 + (1-p)^3).$$

Define

$$s = \frac{1}{2} + \sqrt{\frac{6q-5}{8q-4}} > \frac{1}{2}\,.$$

With this definition $h(s) = s$ and $h(\frac{1}{2}) = \frac{1}{2}$, see also Figure 8.4.2. Further,

$$\tfrac{\partial}{\partial p} h(p) = 6(1-p)p(2q-1)$$

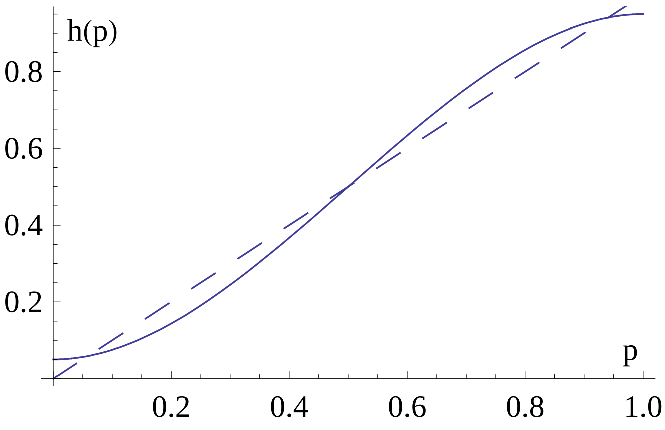


Figure 8.2: h(p) for q=0.95

is positive for $0 < p < 1$. Thus, $p < h(p) < s$ provided $\frac{1}{2} < p < s$. Because of this and the fact that $h(p)$ is continuous over the entire range $\frac{1}{2} < p < s$, iteration of the above process can boost the probability of distributively computing the correct answer arbitrarily close to s. This proves that f has bounded bias because, given any fixed value of $q > \frac{5}{6}$, we can choose an arbitrary constant $t < s$ such that $t > \frac{1}{2}$ and distributively compute f with probability at least t, independently of the input size. ■

8.4.9. Definition. The *non-local equality* problem consists in distributively deciding if three distributed bits are equal. More precisely, let Alice have bits x_1, x_2, x_3 and Bob have y_1, y_2, y_3. The purpose is for Alice and Bob to compute a and b, respectively, such that

$$a \oplus b = \begin{cases} 1 & \text{if } x_1 \oplus y_1 = x_2 \oplus y_2 = x_3 \oplus y_3 \\ 0 & \text{otherwise} . \end{cases}$$

The computation of a and b must be achieved without any communication between Alice and Bob.

8.4.10. Lemma. *Non-local equality can be computed using only two (perfect) non-local boxes.*

Proof: The goal is to obtain a and b such that:

$$a \oplus b \;=\; (x_1 \oplus y_1 = x_2 \oplus y_2) \wedge (x_2 \oplus y_2 = x_3 \oplus y_3). \tag{8.5}$$

First, Alice and Bob compute locally $x' = \overline{x_1} \oplus x_2$, $y' = y_1 \oplus y_2$, $x'' = \overline{x_2} \oplus x_3$ and $y'' = y_2 \oplus y_3$. Then (8.5) becomes equivalent to $(x' \oplus y') \wedge (x'' \oplus y'') = a \oplus b$. Hence, it is sufficient to show how Alice and Bob can compute the AND of the distributed bits $x' \oplus y'$ and $x'' \oplus y''$.

By distributivity of the AND over the exclusive-or,

$$(x' \oplus y') \wedge (x'' \oplus y'') = (x' \wedge x'') \oplus (x' \wedge y'') \oplus (x'' \wedge y') \oplus (y' \wedge y'').$$

Using two non-local boxes, Alice and Bob can compute distributed bits $a' \oplus b'$ and $a'' \oplus b''$ with $a' \oplus b' = x' \wedge y''$ and $a'' \oplus b'' = x'' \wedge y'$. Setting $a = (x' \wedge x'') \oplus a' \oplus a''$ and $b = (y' \wedge y'') \oplus b' \oplus b''$ yields (8.5), as desired. ■

8.4.11. LEMMA. *Non-local majority can be computed using only two (perfect) non-local boxes.*

Proof: Let x_1, x_2, x_3 be Alice's input and y_1, y_2, y_3 be Bob's. For $i \in \{1,2,3\}$, let $z_i = x_i \oplus y_i$ be the i^{th} distributed input bit. By virtue of Lemma 8.4.10, Alice and Bob use their two NLBs to compute the non-local equality of their inputs, yielding a and b so that $a \oplus b = 1$ if and only if z_1, z_2 and z_3 are equal. Finally, Alice produces $a' = \overline{a} \oplus x_1 \oplus x_2 \oplus x_3$ and Bob produces $b' = b \oplus y_1 \oplus y_2 \oplus y_3$. Let

$$z = a' \oplus b' = (\overline{a} \oplus b) \oplus (z_1 \oplus z_2 \oplus z_3)$$

be the distributed bit computed by this protocol. Four cases need to be considered, depending on the number ℓ of 1s among the z_i's:

1. if $\ell = 0$, then $a \oplus b = 1$ and $z_1 \oplus z_2 \oplus z_3 = 0$;
2. if $\ell = 1$, then $a \oplus b = 0$ and $z_1 \oplus z_2 \oplus z_3 = 1$;
3. if $\ell = 2$, then $a \oplus b = 0$ and $z_1 \oplus z_2 \oplus z_3 = 0$;
4. if $\ell = 3$, then $a \oplus b = 1$ and $z_1 \oplus z_2 \oplus z_3 = 1$.

We see that $z = 0$ in the first two cases and $z = 1$ in the last two, so that $z = \text{Maj}(z_1, z_2, z_3)$ in all cases. ■

We are now ready to prove our main theorem.

Proof of Theorem 8.3.1: Assume NLBs can be approximated with some probability p of yielding the correct result. Using these approximate NLBs, we can compute non-local majority with probability $q = p^2 + (1-p)^2$ since the protocol given in the proof of Lemma 8.4.11 succeeds precisely if none or both of the NLBs behave incorrectly. The result follows from Lemmas 8.4.6 and 8.4.8 because $q > \frac{5}{6}$ whenever $p > \frac{3+\sqrt{6}}{6}$. ■

8.4.12. COROLLARY. *In any world in which probabilistic communication complexity is nontrivial, non-local boxes cannot be implemented without communication, even if we are satisfied in obtaining the correct behaviour with probability* $\frac{3+\sqrt{6}}{6} \approx 90.8\%$*.*

8.5 Discussion

In conclusion, we have shown that in any world in which communication complexity is nontrivial, there is a bound on how much nature can be non-local. This bound, which is an improvement over previous knowledge that non-local boxes could not be implemented exactly [96, 97, 26], approaches the actual bound $\wp \approx 85.4\%$ imposed by quantum mechanics. The obvious open question is to close the gap between these probabilities. A proof that nontrivial communication complexity forbids non-local boxes to be approximated with probability greater than $\wp$ would be very interesting, as it would render Tsirelson's bound [93] inevitable, making it a candidate for a new information-theoretic axiom for quantum mechanics [20]. We will finish with some remarks.

8.5.1 One NLB for majority

Neither non-local majority nor non-local equality can be solved exactly with a single non-local box. Otherwise, entanglement could approximate that NLB well enough to solve the non-local majority problem with probability $\wp \approx 0.854 > \frac{5}{6}$ of being correct [25]. It would follow from Lemmas 8.4.6 and 8.4.8 that all Boolean functions have trivial entanglement-assisted communication complexity. But we know this not to be the case for the inner product, as we stated earlier in Section 2.4 and prove in Appendix C.

8.5.2 Fault-tolerance threshold

Quite surprisingly, our results also give bounds on the maximum admissible error for purely classical fault-tolerant computation. This subject was discussed already in more detail in earlier chapters, see in particular Chapter 6. Here we only want to explain this interesting connection.

Suppose that we could transform any classical circuit into a fault-tolerant version that would work with probability bounded away from $\frac{1}{2}$ even if each gate failed independently with probability $\frac{1}{4}$. Assume furthermore that the fault-tolerant circuit C is composed only of unary and binary gates, i.e., gates with at most two input wires. In the proof of Lemma 8.4.10, we showed how to simulate *distributed* AND gates by use of two NLBs. Similarly, it is easy to see that all other binary gates can be computed distributively with at most two NLBs. (Several gates require no NLBs at all, such as the unary NOT gate and the binary XOR gate, also known as the CNOT gate or PARITY gate.) Now, quantum mechanics provides us with NLBs with correctness probability $\wp$, which yields distributed gates that are correct with probability $(1-\wp)^2 + \wp^2 = \frac{3}{4}$ if two NLBs are needed (e.g. the AND gate), or better if no NLBs are needed.

This allows us to use the assumed fault-tolerant circuit C in a distributed way and conclude that all Boolean functions have bounded bias, and therefore trivial quantum probabilistic communication complexity. But this is impossible since most Boolean functions, for example the inner product, require $\Omega(n)$ bits of communication even if Alice and Bob share entanglement and are satisfied with a probability of correct answer bounded away from $\frac{1}{2}$. It follows that circuits cannot in general be fault-tolerant if all gates have at most two input wires and the gates fail with probability $\frac{1}{4}$ or more, even if NOT and XOR gates are perfect.

As an interesting coincidence, the best known upper bound on the error threshold, due to Evans and Schulman [37], states that fault-tolerance is impossible in general for circuits with gates of fan-in at most 2 which fail with probability $1 - \wp = \frac{2-\sqrt{2}}{4}$ or worse.

Appendix A

Some more facts about Linear algebra

In the following we will present some more facts about linear algebra, which might help as a reminder. We use the same notation as in Section 2.1 and continue from there.

Linear independence A set $|\phi_1\rangle, \ldots, |\phi_m\rangle \in \mathbb{C}^d$ is called *linearly independent* if the only way to choose $\alpha_i \in \mathbb{C}$ such that $\sum_{i=1}^m \alpha_i |\phi_i\rangle = 0$ is to choose $\alpha_i = 0$ for all i.

Rank The *rank* of a matrix $A \in \mathbb{C}^{d\times d}$ is the largest number of linearly independent rows of A.

Inverse matrix If for some matrix $A \in \mathbb{C}^{d\times d}$ there exists some $B \in \mathbb{C}^{d\times d}$ with the property that $AB = \mathbb{I}$ then we call B the *inverse* of A and denote it by A^{-1}. Note that if $AB = \mathbb{I}$ then also $BA = \mathbb{I}$ [55]. $A \in \mathbb{C}^{d\times d}$ is *invertible* if and only if A has full rank d.

Unitary matrix A matrix A is called *unitary* if $AA^\dagger = \mathbb{I}$. The following conditions for $A \in \mathbb{C}^{d\times d}$ are equivalent:

1. A is unitary

2. $\forall \phi, \psi \in \mathbb{C}^d : \langle A\phi, A\psi\rangle = \langle \phi, \psi\rangle$ (inner-product preserving)

3. $\forall \phi \in \mathbb{C}^d : ||A\phi|| = ||\phi||$ (norm-preserving),

with $\langle \cdot, \cdot \rangle$ and $|| \cdot ||$ as defined in Section 2.1.

Unitary diagonalization A matrix $A \in \mathbb{C}^{d\times d}$ can be *unitarily diagonalized*, if there is some matrix $U \in \mathbb{C}^{d\times d}$ and a matrix Λ whose off-diagonal entries are all zero with the property that

$$A = U^\dagger \Lambda U.$$

The values on the diagonal of Λ are called the *eigenvalues* of A and the columns of U the corresponding *eigenvectors*.

A matrix A is called *normal* if $AA^\dagger = A^\dagger A$. It turns out that precisely all normal matrices can be diagonalized in this way.

Hermiticity A matrix A is called *hermitian* if $A = A^\dagger$. Note that every hermitian matrix is normal and therefore can be unitarily diagonalized as $A = U^\dagger \Lambda U$ as above. From $A = A^\dagger$ it follows that $U^\dagger \Lambda U = U^\dagger \Lambda^\dagger U$, and then further $\Lambda = \Lambda^\dagger$. This means that hermitian matrices only have real eigenvalues.

Tensor products If $\mathcal{A} = \mathbb{C}^{a\times a}$ and $\mathcal{B} = \mathbb{C}^{b\times b}$, then

$$\mathcal{A} \otimes \mathcal{B} := \mathbb{C}^{ab\times ab}$$

is called the *tensor product* of $\mathcal{A}$ and $\mathcal{B}$.

For elements $A \in \mathcal{A}$ and $B \in \mathcal{B}$ we define

$$A \otimes B = \begin{pmatrix} A_{11}B & \dots & A_{1a}B \\ & \dots & \\ A_{a1}B & \dots & A_{aa}B \end{pmatrix}$$

as the *tensor product* of A and B. The tensor product enjoys many nice properties, for example

$$\begin{aligned} (A \otimes B)^* &= A^* \otimes B^* \\ A \otimes (B + C) &= A \otimes B + A \otimes C \\ (A \otimes B) \otimes C &= A \otimes (B \otimes C) \\ (A \otimes B)(C \otimes D) &= (AC) \otimes (BD). \end{aligned}$$

We will write $A^{\mathcal{A}} \otimes B^{\mathcal{B}}$ if it is otherwise not clear from the context on which space A and B act.

(partial) Trace The *trace* of a matrix $A \in \mathbb{C}^{d\times d}$ is

$$\mathrm{Tr}(A) = \sum_{i=1}^{d} A_{ii},$$

i.e., it is the sum of all entries on the diagonal of A. Note that for unitarily diagonalizable matrices $\mathrm{Tr}(A)$ is the sum of all its eigenvalues. If $A \in \mathcal{A}$ and $B \in \mathcal{B}$ we define the operator

$$\mathrm{Tr}_{\mathcal{A}}(A \otimes B) = B \cdot \mathrm{Tr}(A).$$

Requiring that $\mathrm{Tr}_{\mathcal{A}}(\cdot)$ is linear uniquely defines this operator. Note that $\mathrm{Tr}_{\mathcal{A}}(\cdot)$ maps from $\mathcal{A} \otimes \mathcal{B}$ to $\mathcal{B}$. The operation $\mathrm{Tr}_{\mathcal{A}}(\cdot)$ is called the *partial trace over* A or just "tracing out system A".

Similarly, one can define $\mathrm{Tr}_{\mathcal{B}}(\cdot)$ to be the unique linear operator with the property that

$$\mathrm{Tr}_{\mathcal{B}}(A \otimes B) = A \cdot \mathrm{Tr}(B).$$

Appendix B

Convex hull of all 1-qubit Clifford operations

We now show that the Clifford polytope P defined in (5.8) is equivalent to the polytope defined by (5.10) and (5.11). In principle, this proof can be carried out by a computer, using for example the software cplex [46]. We give an explicit proof.

Define the polyhedron

$$Q := \{S \in \mathbb{R}^{3\times 3} \mid \langle F, S\rangle \leq 1 \text{ for all } F \in \mathcal{F}\},$$

where $\mathcal{F}$ is as in (5.11). Our objective is to show the equality $P = Q$.

First, let us prove the easy inclusion $P \subseteq Q$. For this, let $C \in \mathcal{C}$ and $F \in \mathcal{F}$ be of the form $F = C_1 B C_2$ with $C_1, C_2 \in \mathcal{C}$ and $B \in \{B_1, B_1^T, B_2\}$. Then, $\langle F, C\rangle = \langle B, C_1^T C C_2^T\rangle$. As $C_1^T C C_2^T \in \mathcal{C}$, it suffices to verify that $\langle B, C\rangle \leq 1$ for any $C \in \mathcal{C}$ and $B = B_1, B_2$. (We have used here the fact that $\mathcal{C}$ is a group which is closed under transposing matrices.) For $C \in \mathcal{C}$, the inequality $\langle B_1, C\rangle \leq 1$ is obvious and the inequality $\langle B_2, C\rangle \leq 1$ can be checked by direct inspection.

The reverse inclusion $Q \subseteq P$ follows from the following result.

B.0.1. Theorem. *Any facet of the polytope P is defined by an inequality of the form $\langle F, S\rangle \leq 1$ where $F \in \mathcal{F}$.*

The rest of the Appendix is devoted to the proof of this result. We first need to go in more detail into the structure of the Clifford matrices.

B.1 Preliminaries about the Clifford matrices

Each matrix $C \in \mathcal{C}$ corresponds to a "signed permutation" (σ, s), where $\sigma \in \mathrm{Sym}(3)$ and $s \in \{\pm 1\}^3$. Namely, C has nonzero entries precisely at the $(\sigma(i), i)$-positions with $C_{\sigma(i),i} = s_i$ for $i = 1, 2, 3$; we then also denote C as $C_{\sigma,s}$. The condition $\det(C) = 1$ translates into $s_1 s_2 s_3 = \mathrm{sign}(\sigma)$; that is, $s_1 s_2 s_3 = 1$ if σ is an even permutation (i.e., one of $\sigma_1 := (1,2,3)$, $\sigma_2 := (2,3,1)$, $\sigma_3 := (3,1,2)$) and $s_1 s_2 s_3 = -1$ if σ is an odd permutation (i.e., one of $\sigma_4 := (1,3,2)$, $\sigma_5 := (2,1,3)$, $\sigma_6 := (3,2,1)$). Thus the set $\mathcal{C}$ of Clifford matrices is naturally partitioned into six subclasses

$$\mathcal{C} = \bigcup_{\sigma \in \mathrm{Sym}(3)} \mathcal{C}_\sigma, \text{ where } \mathcal{C}_\sigma := \{C_{\sigma,s} \mid s \in \{\pm 1\}^3,\ s_1 s_2 s_3 = \mathrm{sign}(\sigma)\}$$

with $|\mathcal{C}_\sigma| = 4$. For convenience we display in the table below the six subclasses $\mathcal{C}_\sigma$; the nonzero entries are indicated by $*$.

Even permutations	Odd permutations
$\sigma_1 = (1,2,3) \quad \mathcal{C}_{\sigma_1} : \begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & * \end{pmatrix}$	$\sigma_4 = (1,3,2) \quad \mathcal{C}_{\sigma_4} : \begin{pmatrix} * & 0 & 0 \\ 0 & 0 & * \\ 0 & * & 0 \end{pmatrix}$
$\sigma_2 = (2,3,1) \quad \mathcal{C}_{\sigma_2} : \begin{pmatrix} 0 & 0 & * \\ * & 0 & 0 \\ 0 & * & 0 \end{pmatrix}$	$\sigma_5 = (2,1,3) \quad \mathcal{C}_{\sigma_5} : \begin{pmatrix} 0 & * & 0 \\ * & 0 & 0 \\ 0 & 0 & * \end{pmatrix}$
$\sigma_3 = (3,1,2) \quad \mathcal{C}_{\sigma_3} : \begin{pmatrix} 0 & * & 0 \\ 0 & 0 & * \\ * & 0 & 0 \end{pmatrix}$	$\sigma_6 = (3,2,1) \quad \mathcal{C}_{\sigma_6} : \begin{pmatrix} 0 & 0 & * \\ 0 & * & 0 \\ * & 0 & 0 \end{pmatrix}$

Table 1

The following observation can be directly verified and will be useful for the proof.

6. Observation. *Let $\sigma \in \mathrm{Sym}(3)$. Then, $\sum_{C \in \mathcal{C}_\sigma} C = 0$. Moreover, for any position $(\sigma(i), i)$ corresponding to a nonzero entry for matrices in $\mathcal{C}_\sigma$ and for $d \in \{\pm 1\}$, there exist $C, C' \in \mathcal{C}_\sigma$ with $C + C' = 2dE_{\sigma(i),i}$, which implies $dE_{\sigma(i),i} \in P$. Thus, $\pm E_{i,j} \in P$ for any $i, j = 1, 2, 3$.*

We now proceed with the proof of Theorem B.0.1. Let $\langle F, S \rangle \leq b$ be an inequality defining a facet of P, where $F \in \mathbb{R}^{3\times 3}$ and $b \in \mathbb{R}$. That is, the inequality $\langle F, S \rangle \leq b$ is valid for P, which means that $\langle F, S \rangle \leq b$ holds for any $S \in P$, and the set

$$\mathcal{R}_F := \{C \in \mathcal{C} \mid \langle F, C \rangle = b\}$$

contains nine affinely independent matrices. Without loss of generality, we may assume that $b = 1$. Indeed, $b \geq 0$ since $0 \in P$. Moreover, $b > 0$ for, otherwise, we would have $F_{ij} = 0$ for all $i, j = 1, 2, 3$, implying $F = 0$, in view of Observation 6. Thus, by rescaling, we can now assume that the facet is of the form $\langle F, S \rangle \leq 1$. We sometimes speak of the "facet F" for short. Our objective is to show that $F = C_1 B C_2$ for some $C_1, C_2 \in \mathcal{C}$, $B \in \{B_1, B_1^T, B_2\}$.

Call $F, F' \in \mathbb{R}^{3\times 3}$ *equivalent* if $F' = C_1 F C_2$ for some $C_1, C_2 \in \mathcal{C}$. Then, as $\mathcal{C}$ is a group, $\langle F', S\rangle \leq 1$ defines a facet of P if and only if $\langle F, C\rangle \leq 1$ does. Moreover, $\mathcal{R}_{F'} = C_1 \mathcal{R}_F C_2 = \{C_1 C C_2 \mid C \in \mathcal{R}_F\}$. This property will be used repeatedly throughout the proof as it permits to exploit symmetry and to reduce the number of cases we need to check.

The proof is based on a detailed inspection of the structure of the set $\mathcal{R}_F$. We begin with collecting several properties of the matrix F and the set $\mathcal{R}_F$.

7. Observation. *$|\mathcal{R}_F \cap \mathcal{C}_\sigma| \leq 3$ for any $\sigma \in \mathrm{Sym}(3)$.*

Proof: If $\mathcal{C}_\sigma \subseteq \mathcal{R}_F$, then $\langle F, C\rangle = 1$ for any $C \in \mathcal{C}_\sigma$, which implies $4 = \sum_{C\in\mathcal{C}_\sigma} \langle F, C\rangle$, contradicting the fact that $\sum_{C\in\mathcal{C}_\sigma} C = 0$ by Observation 6. ■

8. Observation. *If $F_{ij} = d \in \{-1, 1\}$, then all $C \in \mathcal{C}$ with $C_{ij} = d$ belong to R_F.*

Proof: Let $C \in \mathcal{C}$ with $C_{ij} = d$. There exists $C' \in \mathcal{C}$ with $C + C' = 2dE_{ij}$. Summing $\langle F, C\rangle \leq 1$ and $\langle F, C'\rangle \leq 1$ yields $\langle F, C + C'\rangle \leq 2$. As $\langle F, C + C'\rangle = 2dF_{ij} = 2$, we have the equalities $\langle F, C\rangle = \langle F, C'\rangle = 1$, which implies $C \in \mathcal{R}_F$. ■

9. Observation. *Let $C \neq C' \in \mathcal{R}_F \cap \mathcal{C}_\sigma$ (for some $\sigma \in \mathrm{Sym}(3)$) and assume that $C_{\sigma(i),i} = C'_{\sigma(i),i} = d \in \{\pm 1\}$ for some $i \in \{1, 2, 3\}$. Then, $F_{\sigma(i),i} = d$ and $F_{\sigma(j),j} + \mathrm{sign}(\sigma) d F_{\sigma(k),k} = 0$ with $\{j, k\} = \{1, 2, 3\} \setminus \{i\}$.*

Proof: Equality $F_{\sigma(i),i} = d$ follows from the fact that $C + C' = 2dE_{\sigma(i),i}$. Then, $1 = \langle F, C\rangle$ implies $0 = F_{\sigma(j),j} C_{\sigma(j),j} + F_{\sigma(k),k} C_{\sigma(k),k}$. Using $C_{\sigma(i),i} C_{\sigma(j),j} C_{\sigma(k),k} = \mathrm{sign}(\sigma)$, we find $F_{\sigma(j),j} + \mathrm{sign}(\sigma) d F_{\sigma(k),k} = 0$. ■

Our last observation is an easy corollary of the former two observations.

10. Observation. *If $F_{\sigma(i),i} = d \in \{\pm 1\}$ (for some $\sigma \in \mathrm{Sym}(3)$), then $F_{\sigma(j),j} + \mathrm{sign}(\sigma) d F_{\sigma(k),k} = 0$, where $\{i, j, k\} = \{1, 2, 3\}$.*

One can verify that, for $F = B_1, B_1^T$, $|\mathcal{R}_F \cap \mathcal{C}_\sigma| = 2$ for all $\sigma \in \mathrm{Sym}(3)$ while, for $F = B_2$, $|\mathcal{R}_F \cap \mathcal{C}_\sigma| = 3$ for $\sigma = \sigma_1, \sigma_5$. Based on this observation we now distinguish two cases: Either, $|\mathcal{R}_F \cap \mathcal{C}_\sigma| \leq 2$ for all $\sigma \in \mathrm{Sym}(3)$ (in which case we show that F is equivalent to B_1 or B_1^T), or $|\mathcal{R}_F \cap \mathcal{C}_\sigma| = 3$ for some $\sigma \in \mathrm{Sym}(3)$ (in which case we show that F is equivalent to B_2).

B.2 The case $|\mathcal{R}_F \cap \mathcal{C}_\sigma| = 3$ for some $\sigma \in \mathrm{Sym}(3)$

Using symmetry, we may assume that $|\mathcal{R}_F \cap \mathcal{C}_\sigma| = 3$ for the (odd) permutation $\sigma = \sigma_4$. We prove this in detail to show how this kind of symmetry argument works. To shorten notation we will use $+, -$ for 1 respectively -1 whenever we explicitly write out matrices.

Define the matrices

$$C_1 = \begin{pmatrix} - & 0 & 0 \\ 0 & 0 & - \\ 0 & - & 0 \end{pmatrix}, \; C_2 = \begin{pmatrix} 0 & + & 0 \\ 0 & 0 & + \\ + & 0 & 0 \end{pmatrix}, \; C_3 = \begin{pmatrix} 0 & 0 & + \\ + & 0 & 0 \\ 0 & + & 0 \end{pmatrix} \tag{B.1}$$

lying, resp., in $\mathcal{C}_{\sigma_4}, \mathcal{C}_{\sigma_3}, \mathcal{C}_{\sigma_2}$.

Our assumption is that $|\mathcal{R}_F \cap \mathcal{C}_{\sigma_i}| = 3$ for some $i = 1, \dots, 6$; we show that one can replace F by another equivalent facet F' in such a way that $i = 4$ holds. For this, suppose first $i = 2, 3$. As the mapping $X \mapsto XC_i$ maps $\mathcal{C}_{\sigma_i}$ to $\mathcal{C}_{\sigma_1}$, we can replace the facet F by $F' := FC_i$ and then we find $|\mathcal{R}_{F'} \cap \mathcal{C}_{\sigma_1}| = 3$ since $\mathcal{R}_{F'} = \mathcal{R}_F C_i$. Thus we may assume $|\mathcal{R}_F \cap \mathcal{C}_{\sigma_1}| = 3$. As the mapping $X \mapsto XC_1$ maps $\mathcal{C}_{\sigma_1}$ to $\mathcal{C}_{\sigma_4}$, replacing the facet F by $F' := FC_1$, we find $|\mathcal{R}_{F'} \cap \mathcal{C}_{\sigma_4}| = 3$. Thus we can now assume $|\mathcal{R}_F \cap \mathcal{C}_{\sigma_i}| = 3$ for some $i = 4, 5, 6$. If $i = 5$, as the mapping $X \mapsto XC_3$ maps $\mathcal{C}_{\sigma_5}$ to $\mathcal{C}_{\sigma_4}$, replace F by $F' := FC_3$; if $i = 6$, the mapping $X \mapsto XC_2$ maps $\mathcal{C}_{\sigma_6}$ to $\mathcal{C}_{\sigma_4}$ and one can replace F by $F' := FC_2$; in both cases we get back to the case when $|\mathcal{R}_{F'} \cap \mathcal{C}_{\sigma_4}| = 3$.

Thus we now assume $|\mathcal{R}_F \cap \mathcal{C}_{\sigma_4}| = 3$. Moreover, we may assume that the following matrices from $\mathcal{C}_{\sigma_4}$

$$\begin{pmatrix} + & 0 & 0 \\ 0 & 0 & - \\ 0 & + & 0 \end{pmatrix}, \begin{pmatrix} - & 0 & 0 \\ 0 & 0 & + \\ 0 & + & 0 \end{pmatrix}, \begin{pmatrix} - & 0 & 0 \\ 0 & 0 & - \\ 0 & - & 0 \end{pmatrix} \tag{B.2}$$

belong to $\mathcal{R}_F$. (To see this, replace if necessary F by FC, where $C \in \mathcal{C}_{\sigma_1}$.) Using Observation 9, we obtain $F_{11} = -1, F_{23} = -1, F_{32} = 1$. From this we get by Observation 8 that also the matrices

$$\begin{pmatrix} - & 0 & 0 \\ 0 & - & 0 \\ 0 & 0 & + \end{pmatrix}, \begin{pmatrix} - & 0 & 0 \\ 0 & + & 0 \\ 0 & 0 & - \end{pmatrix} \in \mathcal{C}_{\sigma_1}, \tag{B.3}$$

$$\begin{pmatrix} 0 & 0 & + \\ + & 0 & 0 \\ 0 & + & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & - \\ - & 0 & 0 \\ 0 & + & 0 \end{pmatrix} \in \mathcal{C}_{\sigma_2}, \tag{B.4}$$

$$\begin{pmatrix} 0 & + & 0 \\ 0 & 0 & - \\ - & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & - & 0 \\ 0 & 0 & - \\ + & 0 & 0 \end{pmatrix} \in \mathcal{C}_{\sigma_3} \tag{B.5}$$

$$\tag{B.6}$$

belong to $\mathcal{R}_F$. By Observation 9, we also obtain $F_{22} = F_{33}$, $F_{12} = F_{31}$ and $F_{13} = -F_{21}$.

1. Claim. *There exists also an even permutation $\sigma \in \mathrm{Sym}(3)$ for which $|\mathcal{R}_F \cap \mathcal{C}_\sigma| = 3$.*

Proof: Assume for contradiction that, for $i = 1, 2, 3$, the set $\mathcal{R}_F \cap \mathcal{C}_{\sigma_i}$ contains only the respective two matrices from (B.3)-(B.5). Choose a subset $\mathcal{B} \subseteq \mathcal{R}_F$ consisting of nine affinely independent matrices and such that $\mathcal{R}_F \cap \mathcal{C}_{\sigma_4} \subseteq \mathcal{B}$. We have $|\mathcal{B} \cap \mathcal{C}_{\sigma_1}| \leq 1$, since the two matrices in (B.3) are affinely dependent with the last two matrices in (B.2). Similarly, $|\mathcal{B} \cap \mathcal{C}_{\sigma_2}| \leq 1$, $|\mathcal{B} \cap \mathcal{C}_{\sigma_3}| \leq 1$. As $|\mathcal{B}| = 9$, we deduce that $|\mathcal{B} \cap \mathcal{C}_{\sigma_5}| \geq 2$ or $|\mathcal{B} \cap \mathcal{C}_{\sigma_6}| \geq 2$. Assume first that $|\mathcal{B} \cap \mathcal{C}_{\sigma_5}| \geq 2$. Say, $C \neq C' \in \mathcal{R}_F \cap \mathcal{C}_{\sigma_5}$. Then C and C' have the same nonzero entry $d \in \{-1, 1\}$ in some position (k, l). By Observation 9 this yields $F_{kl} = d$. Now, there is also an even permutation σ for which $k = \sigma(l)$. By Observation 8 we then deduce that at least two matrices from $\mathcal{C}_\sigma$ must be in $\mathcal{R}_F$, which contradicts our assumption. The other case $|\mathcal{B} \cap \mathcal{C}_{\sigma_6}| \geq 2$ goes analogously. ■

It is sufficient to consider the case $|\mathcal{R}_F \cap \mathcal{C}_{\sigma_1}| = 3$. Indeed, if $|\mathcal{R}_F \cap \mathcal{C}_{\sigma_2}| = 3$, then one may replace F by $C_3 F C_4$ with C_3 as in (B.1) and

$$C_4 := \begin{pmatrix} 0 & 0 & + \\ - & 0 & 0 \\ 0 & - & 0 \end{pmatrix},$$

since the mapping $X \mapsto C_3 X C_4$ maps $\mathcal{C}_{\sigma_2}$ to $\mathcal{C}_{\sigma_1}$ and preserves the set of three matrices from (B.2), as well as the set of 6 matrices from (B.3)-(B.5) (namely, (B.3) → (B.5) → (B.4) → (B.3)). One can handle the case when $|\mathcal{R}_F \cap \mathcal{C}_{\sigma_3}| = 3$ in the same way.

The set $\mathcal{R}_F$ already contains the matrices

$$D_1 := \begin{pmatrix} - & 0 & 0 \\ 0 & - & 0 \\ 0 & 0 & + \end{pmatrix}, \; D_2 := \begin{pmatrix} - & 0 & 0 \\ 0 & + & 0 \\ 0 & 0 & - \end{pmatrix}$$

from $\mathcal{C}_{\sigma_1}$ (displayed in (B.3)). The remaining two matrices of $\mathcal{C}_{\sigma_1}$ are

$$D_3 := \begin{pmatrix} + & 0 & 0 \\ 0 & + & 0 \\ 0 & 0 & + \end{pmatrix}, \; D_4 := \begin{pmatrix} + & 0 & 0 \\ 0 & - & 0 \\ 0 & 0 & - \end{pmatrix}.$$

If $D_4 \in \mathcal{R}_F$, one may replace the facet F by $F' := D_2 F D_1$ to obtain that $D_1, D_2, D_3 \in \mathcal{R}_{F'}$, since the mapping $X \mapsto D_2 X D_1$ maps the set $\{D_1, D_2, D_4\}$ to $\{D_1, D_2, D_3\}$ and leaves the set of 3 matrices from (B.2) invariant as well as the set of 6 matrices from (B.3)-(B.5). Thus we may assume that $D_3 \in \mathcal{R}_F$.

By Observation 9, we find that $F_{33} = F_{22} = 1$. As $F_{22} = 1$, Observation 10 implies that $F_{31} = F_{13}$. Similarly, $F_{33} = 1$ implies that $F_{12} = F_{21}$. Putting all equations together we obtain $F_{12} = F_{21} = -F_{13} = -F_{31} = -F_{12}$, implying they are all zero. Thus

$$F = \begin{pmatrix} - & 0 & 0 \\ 0 & + & - \\ 0 & + & + \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} B_2 \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}. \tag{B.7}$$

B.3 The case $|\mathcal{R}_F \cap \mathcal{C}_\sigma| \leq 2$ for all $\sigma \in \mathrm{Sym}(3)$

Let again $\mathcal{B} \subseteq \mathcal{R}_F$ consist of nine affinely independent matrices. As $|\mathcal{R}_F| \geq 9$, $|\mathcal{R}_F \cap \mathcal{C}_\sigma| = 2$ for at least three permutations σ. W.l.o.g. we can assume that two of those permutations are odd permutations and that they are equal, say, to σ_4 and σ_6 (replacing if necessary F by an equivalent facet). Further we may assume $\mathcal{R}_F$ contains the following two matrices of $\mathcal{C}_{\sigma_4}$:

$$\begin{pmatrix} - & 0 & 0 \\ 0 & 0 & - \\ 0 & - & 0 \end{pmatrix}, \begin{pmatrix} - & 0 & 0 \\ 0 & 0 & + \\ 0 & + & 0 \end{pmatrix} \in \mathcal{R}_F. \tag{B.8}$$

This can be seen using the following two mappings $X \mapsto C_2XC_2$ (with C_2 defined as in (B.1)) and $X \mapsto CX$ (with $C \in \mathcal{C}_{\sigma_1}$) which permit to map any subset of size 2 of $\mathcal{C}_{\sigma_4}$ to any other such subset and which preserve $\mathcal{C}_{\sigma_6}$ as well. We choose the basis $\mathcal{B}$ containing the two matrices of (B.8). ¿From Observation 9 we find $F_{11} = -1$ and $F_{23} = -F_{32} \neq \pm 1$; the latter inequality follows from the fact that $|\mathcal{R}_F \cap \mathcal{C}_{\sigma_4}| = 2$ combined with Observation 8. As $F_{11} = -1$, by Observation 8,

$$\begin{pmatrix} - & 0 & 0 \\ 0 & - & 0 \\ 0 & 0 & + \end{pmatrix}, \begin{pmatrix} - & 0 & 0 \\ 0 & + & 0 \\ 0 & 0 & - \end{pmatrix} \in \mathcal{R}_F \cap \mathcal{C}_{\sigma_1} \tag{B.9}$$

and Observation 9 implies $F_{22} = F_{33} \neq \pm 1$. At most one of the two matrices in (B.9) belongs to $\mathcal{B}$ since they are affinely dependent with the matrices in (B.8). Say, $|\mathcal{B} \cap \mathcal{C}_{\sigma_1}| = 1$.

Let us now examine which two matrices of $\mathcal{C}_{\sigma_6}$ belong to $\mathcal{R}_F$. Set

$$C_5 := \begin{pmatrix} + & 0 & 0 \\ 0 & - & 0 \\ 0 & 0 & - \end{pmatrix} \in \mathcal{C}_{\sigma_1},\ X_1 := \begin{pmatrix} 0 & 0 & - \\ 0 & - & 0 \\ - & 0 & 0 \end{pmatrix} \in \mathcal{C}_{\sigma_6}.$$

The two mappings $X \mapsto XC_5$ and $X \mapsto C_5X$ preserve the set of matrices in (B.8) and permit to map any other matrix of $\mathcal{C}_{\sigma_6}$ to the matrix X_1. Therefore we can assume w.l.o.g. that $X_1 \in \mathcal{R}_F \cap \mathcal{C}_{\sigma_6}$. The second matrix of $\mathcal{R}_F \cap \mathcal{C}_{\sigma_6}$ does not have entry -1 at the position $(2,2)$ since, otherwise, $F_{22} = -1$ contradicting an earlier claim. Hence the second matrix in $\mathcal{R}_F \cap \mathcal{C}_{\sigma_6}$ is

$$X_2 := \begin{pmatrix} 0 & 0 & + \\ 0 & + & 0 \\ - & 0 & 0 \end{pmatrix}, \text{ or } X_3 := \begin{pmatrix} 0 & 0 & - \\ 0 & + & 0 \\ + & 0 & 0 \end{pmatrix}.$$

1. Consider first the case when $X_2 \in \mathcal{R}_F \cap \mathcal{C}_{\sigma_6}$. Then, $F_{31} = -1$ and $F_{22} = -F_{13} \neq \pm 1$. As $F_{31} = -1$, we have

$$\begin{pmatrix} 0 & + & 0 \\ 0 & 0 & - \\ - & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & - & 0 \\ 0 & 0 & + \\ - & 0 & 0 \end{pmatrix} \in \mathcal{R}_F \cap \mathcal{C}_{\sigma_3} \tag{B.10}$$

and $F_{12} = F_{23} \neq \pm 1$. As $\mathcal{B}$ contains at most three of the matrices X_1, X_2 and in (B.10), we must have $|\mathcal{B} \cap \mathcal{C}_{\sigma_2}| = 2$ or $|\mathcal{B} \cap \mathcal{C}_{\sigma_5}| = 2$. We obtained earlier that $F_{33} = F_{22} = -F_{13} \neq \pm 1$ and $F_{12} = F_{23} = -F_{32} \neq \pm 1$. In other words, the second and third columns of F contain no entry ± 1. On the other hand, the two matrices from $B \cap \mathcal{C}_{\sigma_i}$ $(i = 2, 5)$ have one common nonzero entry which therefore is located in the first column, at the position $(2, 1)$. This implies $F_{21} = \pm 1$.

(a) If $F_{21} = 1$, then Observation 10 implies $F_{12} = F_{33}$ and $F_{13} = -F_{32}$. Combining with the former relations on entries of F, we find

$$F = \begin{pmatrix} - & 0 & 0 \\ + & 0 & 0 \\ - & 0 & 0 \end{pmatrix}. \tag{B.11}$$

(b) If $F_{21} = -1$, then in the same way we find

$$F = \begin{pmatrix} - & 0 & 0 \\ - & 0 & 0 \\ - & 0 & 0 \end{pmatrix}. \tag{B.12}$$

In both cases we find that F is equivalent to B_1.

2. Consider now the case when $X_3 \in \mathcal{R}_F \cap \mathcal{C}_{\sigma_6}$. Then, $F_{13} = -1$, $F_{22} = -F_{31} \neq \pm 1$,

$$\begin{pmatrix} 0 & 0 & - \\ - & 0 & 0 \\ 0 & + & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & - \\ + & 0 & 0 \\ 0 & - & 0 \end{pmatrix} \in \mathcal{R}_F \cap \mathcal{C}_{\sigma_2} \tag{B.13}$$

and $F_{21} = F_{32} \neq \pm 1$. In the same way as in the first case one finds that F is equivalent to B_1^T.

This proves that the facet description of polytope P from Lemma 5.4.1 is indeed given by (5.10) and (5.11).

Appendix C

Classical entanglement-assisted communication complexity of inner product

We show here that the classical communication complexity of the inner product function under shared entanglement is

$$R^*_\epsilon(IP_n) \geq n - 2\log_2 \tfrac{1}{1-2\epsilon}, \tag{C.1}$$

where $\epsilon > 0$ is (a lower bound on) the desired error probability. The proof is via a reduction from the quantum communication complexity of inner product with entanglement [65]

$$Q^*_\epsilon(IP_n) \geq \tfrac{1}{2}n - \log_2 \tfrac{1}{1-2\epsilon}. \tag{C.2}$$

Here $Q^*_\epsilon(IP_n)$ is the minimum number c such that there is a protocol which uses arbitrary entanglement, communicates at most c qubits and for every $x, y \in \{0,1\}^n$ it correctly outputs $IP_n(x, y)$ with probability at least $1 - \epsilon$. The bound (C.1) already appears in [65], but only for the model of one-way communication. We show that this bound even holds without a restriction on the number of rounds.

The problem of proving (C.1) came up in a discussion with Ronald de Wolf who pointed out the following easy reduction to me [104].

Proof: Assume there is an entanglement-assisted protocol P for IP_n which needs at most $\bar{c}$ classical bits to compute IP_n correctly with probability at least $1 - \epsilon$. We will show that $\bar{c}$ must be at least as large as the right-hand side of (C.1).

Consider the following protocol for computing IP_{kn} for inputs $x, y \in \{0,1\}^{kn}$: Chop x into k blocks x^i, $1 \leq i \leq k$, of length n each. Do the same for y. Run P to compute all instances $IP_n(x^i, y^i)$ and then output $\bigoplus_{i=1}^k IP_n(x^i, y^i)$. It is straightforward to prove (by induction on k) that this protocol $P^{\otimes k}$ will correctly compute $IP_{kn}(x, y)$ with probability

$$\Pr[P^{\otimes k}(x, y) = IP(x, y)] \geq \frac{1}{2} + \frac{1}{2}(1 - 2\epsilon)^k. \tag{C.3}$$

Further, by assumption protocol $P^{\otimes k}$ uses at most $k\bar{c}$ bits of classical communication, and it is easy to design $P^{\otimes k}$ such that it never uses more than $\bar{c}$ *rounds* of communication.

If $P^{\otimes k}$ sends c_i classical bits in the i-th round, then this can be simulated with $\lceil c_i/2 \rceil$ many qubits using superdense coding. Simulating the whole protocol $P^{\otimes k}$ in this way results in a protocol which never communicates more than

$$\sum_{i=1}^{\bar{c}} \lceil \tfrac{c_i}{2} \rceil \leq \sum_{i=1}^{\bar{c}} \tfrac{c_i+1}{2} \leq (k + \tfrac{1}{2})\bar{c} \tag{C.4}$$

qubits, where we used that $P^{\otimes k}$ needs $\sum_{i=1}^{\bar{c}} c_i \leq k\bar{c}$ classical bits. From this and equations (C.3) and (C.2) it then follows that

$$\begin{aligned} (k + \tfrac{1}{2})\bar{c} + 1 &\geq \tfrac{1}{2}kn - \log_2 \tfrac{1}{1-2(\frac{1}{2}+\frac{1}{2}(1-2\epsilon)^k)} \\ &= \tfrac{1}{2}kn - k \log_2 \tfrac{1}{1-2\epsilon}. \end{aligned}$$

Since this inequality has to hold for any k, we can conclude that $\bar{c} \geq n - 2\log_2 \frac{1}{1-2\epsilon}$, which proves our claim. ■

Appendix D

Tsirelson's vector characterization of XOR games

We now give the proof of Theorem 7.3.1. The proof is originally due to Tsirelson [94]. In the form we use it here it first appeared in [32]. For convenience we repeat its statement here.

D.0.1. Theorem (7.3.1, **[94, 32]**). *Let S and T be finite sets, and let $|\psi\rangle$ be a pure quantum state with support on a bipartite Hilbert space $\mathcal{H} = \mathcal{A} \otimes \mathcal{B}$ such that $dim(\mathcal{A}) = dim(\mathcal{B}) = n$. For each $s \in S$ and $t \in T$, let X_s and Y_t be observables on $\mathcal{A}$ and respectively $\mathcal{B}$ with eigenvalues ± 1. Then there exist real unit vectors x_s and y_t in $\mathbb{R}^{2n^2}$ such that*

$$\langle\psi|X_s \otimes Y_t|\psi\rangle = x_s \cdot y_t,$$

for all $s \in S$ and $t \in T$.
Conversely, suppose that S and T are finite sets, and x_s and y_t are unit vectors in $\mathbb{R}^N$ for each $s \in S$ and $t \in T$. Let $\mathcal{A}$ and $\mathcal{B}$ be Hilbert spaces of dimension $2^{\lceil N/2 \rceil}$, $\mathcal{H} = \mathcal{A} \otimes \mathcal{B}$ and $|\psi\rangle$ be a maximally entangled state on $\mathcal{H}$. Then there exist observables X_s and Y_t with eigenvalues ± 1, on $\mathcal{A}$ and $\mathcal{B}$ respectively, such that

$$\langle\psi|X_s \otimes Y_t|\psi\rangle = x_s \cdot y_t,$$

for all $s \in S$ and $t \in T$.

The following version of the proof seems not to have appeared in the literature before and I am grateful to Ben Toner [92] and Richard Cleve [27] for sharing it with me.

Proof: "→" We start with the first part. Define

$$\begin{aligned} |x_s\rangle &= X_s \otimes I|\psi\rangle \in \mathbb{C}^{n^2} && \text{(D.1)} \\ |y_t\rangle &= I \otimes Y_t|\psi\rangle \in \mathbb{C}^{n^2} && \text{(D.2)} \end{aligned}$$

and define[1] $x_s, y_t \in \mathbb{R}^{2n^2}$ as

$$\begin{aligned} x_s &= (Re(x_{s,1}), \ldots, Re(x_{s,n}), Im(x_{s,1}), \ldots, Im(x_{s,n}))^T && \text{(D.3)} \\ y_t &= (Re(y_{t,1}), \ldots, Re(y_{t,n}), Im(y_{t,1}), \ldots, Im(y_{t,n}))^T && \text{(D.4)} \end{aligned}$$

By construction we then have

$$x_s \cdot y_t = \langle\psi|X_s \otimes Y_t|\psi\rangle.$$

Further,

$$x_s \cdot x_s = \langle\psi|(X_s \otimes I)^\dagger (X_s \otimes I)|\psi\rangle = \langle\psi|\psi\rangle = 1,$$

since X_s has eigenvalues ±1. Similarly, $y_t \cdot y_t = 1$, which establishes the first direction of the proof.

"←" Recall the definition of the Pauli matrices

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \mathbb{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{D.5}$$

which obey the commutation relations $XY = -YX$, $ZX = -XZ$ and $YZ = -ZY$. Further, $X^2 = Y^2 = Z^2 = \mathbb{I}_2$.

For $i = 1, \ldots, N$ we define $G_i \in \mathbb{C}^{2^{\lceil N/2\rceil} \times 2^{\lceil N/2\rceil}}$ by setting

$$\begin{aligned} G_{2k+1} &= Z^{\otimes k} \otimes Y \otimes \mathbb{I}_2^{\otimes(\lceil N/2\rceil - k - 1)}, \quad \text{if } i = 2k+1 \text{ with } k \in \mathbb{N} \\ G_{2k} &= Z^{\otimes k} \otimes X \otimes \mathbb{I}_2^{\otimes(\lceil N/2\rceil - k - 1)}, \quad \text{if } i = 2k \text{ with } k \in \mathbb{N} \end{aligned}$$

The G_i are hermitian and satisfy the relations

$$\{G_i, G_j\} = G_i G_j + G_j G_i = 2\delta_{i,j}. \tag{D.6}$$

(Remark: The G_i are generators of a Clifford algebra). We set

$$\begin{aligned} X_s &= x_{s,1} G_1 + \cdots + x_{s,N} G_N \\ Y_t &= y_{t,1} G_1^T + \cdots + y_{t,N} G_N^T \end{aligned}$$

and note that they are also hermitian. Then by (D.6) we get that

$$X_s^2 = \sum_{i,j=1}^{N} x_{s,i} x_{s,j} G_i G_j = \sum_{i=1}^{N} x_{s,i} x_{s,i} \mathbb{I}_2^{\otimes\lceil N/2\rceil} = \mathbb{I}_{2^{\lceil N/2\rceil}}$$

and similarly

$$Y_t^2 = \mathbb{I}_{2^{\lceil N/2\rceil}}.$$

[1] For convenience we will use superscripts for the entries of a vector in this proof, whereas in the rest of this thesis we use subscripts for the entries of vectors.

Hence, all X_s and all Y_t have eigenvalues ±1. For the maximally entangled state

$$|\psi\rangle = 2^{-\lceil N/2\rceil/2} \sum_{i=1}^{2^{\lceil N/2\rceil}} |i\rangle_{\mathcal{A}} |i\rangle_{\mathcal{B}}$$

we calculate

$$\begin{aligned}
\langle\psi|X_s \otimes Y_t|\psi\rangle &= \sum_{i,j=1}^{2^{\lceil N/2\rceil}} x_{s,i} y_{t,j} \langle\psi|G_i \otimes G_j|\psi\rangle \\
&= 2^{-\lceil N/2\rceil} \sum_{i,j=1}^{N} x_{s,i} y_{t,j} \underbrace{\mathrm{Tr}(G_i G_j)}_{=2^{\lceil N/2\rceil}\delta_{i,j}} \\
&= \sum_{i=1}^{N} x_{s,i} y_{t,i} \\
&= x_s \cdot y_t
\end{aligned}$$

This proves the second part of the theorem. ■

Bibliography

[1] Available at http://www.wolfram.com/.

[2] S. Aaronson and D. Gottesman. Improved simulation of stabilizer circuits. Technical report, arXiv:quant-ph/0406196 v4, 2004.

[3] D. Aharonov and M. Ben-Or. Polynomial simulations of decohered quantum computers. *Proceedings of the 37th IEEE FOCS*, pages 46–55, 1996.

[4] D. Aharonov and M. Ben-Or. Fault tolerant quantum computation with constant error. In *Proceedings of 29th ACM STOC*, pages 176–188, 1997. quant-ph/9611025.

[5] P. Aliferis. *Level Reduction and the Quantum Threshold Theorem.* PhD thesis, Caltech, 2007. quant-ph/0703264.

[6] P. Aliferis. Threshold lower bounds for Knill's Fibonacci scheme. quant-ph/0709.3603, 22 Sep 2007.

[7] P. Aliferis, D. Gottesman, and J. Preskill. Accuracy threshold for postselected quantum computation. quant-ph/0703264, 28 Mar 2007.

[8] S. Arora and B. Barak. *Complexity Theory: A Modern Approach.* Princeton University Press. preliminary version available at: http://www.cs.princeton.edu/theory/complexity/.

[9] L. Babai, L. Fortnow, and C. Lund. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity*, 1(1):3–40, 1991.

[10] A. Barenco, C.H. Bennett, R. Cleve, D.P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. Smolin, and H. Weinfurter. Elementary gates for quantum computation. *Physical Review A*, 52:3457–3467, 1995. quant-ph/9503016.

[11] J. Barrett, D. Collins, L. Hardy, A. Kent, and S. Popescu. Quantum nonlocality, bell inequalities and the memory loophole. *Physical Review A*, 66:042111, 2002.

[12] J. S. Bell. On the Einstein-Podolsky-Rosen Paradox. *Physics*, 1(3):195–200, 1964.

[13] M. Bellare, O. Goldreich, and M. Sudan. Free bits, PCPs, and non-approximability — towards tight results. *SIAM Journal on Computing*, 27(3):804–915, 1998.

[14] M. Ben-Or, S. Goldwasser, J. Kilian, and A. Wigderson. Multi-prover interactive proofs: How to remove intractability assumptions. In *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing*, pages 113–131, 1988.

[15] S. Borkar. Designing reliable systems from unreliable components: The challenges of transistor variability and degradation. *IEEE Micro*, 25(6):10–16, 2005.

[16] P. Bose. Designing reliable systems with unreliable components. *IEEE Micro*, 26(5):5–6, 2006.

[17] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[18] P. O. Boykin, M. Pulver T. Mor, V. Roychowdhury, and F. Vatan. A new universal and fault-tolerant quantum basis. *Information Processing Letters*, 75:101–107, 2000. quant-ph/9906054.

[19] G. Brassard. Quantum communication complexity. *Foundations of Physics*, 33(11):1593–1616, 2003.

[20] G. Brassard. Is information the key? *Nature Physics*, 1:2–6, 2005.

[21] S. Bravyi and A. Kitaev. Universal quantum computation with ideal clifford gates and noisy ancillas. *Phys. Rev. A 71, 022316*, 2005.

[22] D. Bruss, D. DiVincenzo, A. Ekert, C. Fuchs, C. Macchiavello, and J. Smolin. Optimal universal and state-dependent quantum cloning. *Physical Review A*, 43:2368–2378, 1998.

[23] H. Buhrman, R. Cleve, and A. Wigderson. Quantum vs. classical communication and computation. In *Proceedings of 30th ACM STOC*, pages 63–68, 1998. quant-ph/9802040.

[24] H. Buhrman and S. Massar. Causality and Cirel'son bounds. 2004.

[25] J. Clauser, M. Horne, A. Shimony, and R. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23:880–884, 1969.

[26] R. Cleve, 1996. unpublished.

[27] R. Cleve, 2007. personal communication.

[28] R. Cleve and H. Buhrman. Substituting quantum entanglement for communication. *Physical Review A*, 56(2):1201–1204, Aug 1997.

[29] R. Cleve, W. van Dam, M. Nielsen, and A. Tapp. Quantum entanglement and the communication complexity of the inner product function. In *Proceedings of 1st NASA QCQC conference*, volume 1509 of *Lecture Notes in Computer Science*, pages 61–74. Springer, 1998. quant-ph/9708019.

[30] R. Cleve, W. Slofstra, F. Unger, and S. Upadhyay. Strong parallel repetition theorem for quantum XOR proof systems. In *Special Issue of 22nd IEEE Conference on Computational Complexity*, 2007.

[31] R. Cleve, P. Høyer, B. Toner, and J. Watrous. Consequences and limits of nonlocal strategies. In *Ninteenth IEEE Conference on Computational Complexity*, pages 236–249. 2004.

[32] R. Cleve, P. Høyer, B. Toner, and J. Watrous. Consequences and limits of nonlocal strategies. In *IEEE Conference on Computational Complexity*, pages 236–249, 2004.

[33] R. Cleve, P. Høyer, B. Toner, and J. Watrous. Consequences and limits of nonlocal strategies. June 2004. Presentation given at IEEE Conference on Computational Complexity.

[34] B. Colwell. Computer architecture beyond moore's law. St. Petersburg, 8-12 June 2006. International Computer Science Symposium in Russia.

[35] D. Deutsch. Quantum theory, the Church-Turing principle, and the universal quantum Turing machine. In *Proceedings of the Royal Society of London*, volume A400, pages 97–117, 1985.

[36] W. Evans and N. Pippenger. On the maximum tolerable noise for reliable computation by formulas. *IEEE Trans. Inform. Theory*, 44(3):1299–1305, 1998.

[37] W. Evans and L. Schulman. Signal propagation and noisy circuits. *IEEE Trans. Inform. Theory*, 45(7):2367–2373, 1999.

[38] W. Evans and L. Schulman. On the maximum tolerable noise of k-input gates for reliable computation by formulas. *IEEE Trans. Inform. Theory*, 49(11):3094–3098, 2003.

[39] T. Feder. Reliable computation by networks in the presence of noise. *IEEE Trans. Inform. Theory*, 35(3):569–571, 1989.

[40] U. Feige and M. Goemans. Approximating the value of two power proof systems, with applications to max 2sat and max dicut. In *Proceedings of the 3rd Israel Symposium on the Theory of Computing Systems (ISTCS'95)*, pages 182–189, 1995.

[41] U. Feige, G. Kindler, and R. O'Donnell. Understanding parallel repetition requires understanding foams. In *Proceedings of the Twenty-Second Annual IEEE Conference on Computational Complexity*, pages 179–192, 2007.

[42] U. Feige and L. Lovász. Two-prover one-round proof systems: Their power and their problems (extended abstract). In *Twenty-Fourth Annual ACM Symposium on Theory of Computing*, pages 733–744, 1992.

[43] J. Fern. An upper bound on quantum fault tolerant thresholds, 2008.

[44] R. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6/7):467–488, 1982.

[45] L. Fortnow. *Complexity theoretic aspects of interactive proof systems.* PhD thesis, Massachusetts Institute of Technology, 1989.

[46] K. Fukuda. *Software* `cdd+`. available from http://www.ifor.math.ethz.ch/ fukuda.

[47] O. Goldreich. A taxonomy of proof systems. pages 109–134, 1997.

[48] D. Gottesman. *Stabilizer Codes and Quantum Error Correction.* PhD thesis, Caltech, 1997. quant-ph/9702052.

[49] M. Grassl, T. Beth, and T. Pellizzari. Codes for the quantum erasure channel. *Physical Review A*, 56:33–38, July 1997.

[50] L. K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of 28th ACM STOC*, pages 212–219, 1996.

[51] B. Hajek and T. Weller. On the maximum tolerable noise for reliable computation by formulas. *IEEE Trans. Inform. Theory*, 37(2):388–391, 1991.

[52] A. Harrow and M. Nielsen. How robust is a quantum gate in the presence of noise? *Phys. Rev. A 68, 012308*, 2003.

[53] J. Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001.

[54] T. Holenstein. Parallel repetition: simplifications and the no-signaling case. *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*, 2007. Manuscript available at http://www.arxiv.org/abs/cs/0607139.

[55] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, 1985.

[56] J. Kempe, H. Kobayashi, K. Matsumoto, and T. Vidick. Using entanglement in quantum multi-prover interactive proofs. In *IEEE Conference on Computational Complexity*, 2008.

[57] J. Kempe, O. Regev, and B. Toner. The unique games conjecture with entangled provers is false, 2007.

[58] A. Kitaev and J. Watrous. Parallelization, amplification, and exponential time simulation of quantum interactive proof systems. In *Proceedings of 32nd ACM STOC*, pages 608–617, 2000.

[59] A. Yu. Kitaev. Quantum computations: Algorithms and error correction. *Russian Mathematical Surveys*, 52(6):1191–1249, 1997.

[60] A. Yu. Kitaev, A. H. Shen, and M. N. Vyalyi. *Classical and Quantum Computation.* American Mathematical Society, 2002.

[61] M. Knill. Quantum computing with realistically noisy devices. *Nature*, 434:39–44, 2005.

[62] M. Knill, R. Laflamme, and W. Zurek. Threshold accuracy for quantum computation. quant-ph/9610011, 15 Oct 1996.

[63] E. Kushilevitz and N. Nisan. *Communication Complexity.* Cambridge University Press, 1997.

[64] G. Moore. Cramming more components onto integrated circuits. *Electronics*, 38(8), 1965.

[65] A. Nayak and J. Salzman. Limits on the ability of quantum states to convey classical messages. *Journal of the ACM*, 53(1):184–206, 2006.

[66] G. Nebe, E. M. Rains, and N. J. A. Sloane. The invariants of the clifford groups. *Designs, Codes and Cryptography*, 24(99), 2001. math.CO/0001038.

[67] J. von Neumann. Probabilistic logics and the synthesis of reliable organisms from unreliable components. In C. E. Shannon and J. McCarthy, editors, *Automata Studies*, volume 3, pages 43–99. Princeton University Press, Princeton, 1956.

[68] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, 2000.

[69] N. Nisan and A. Wigderson. Hardness vs. randomness. *Jounal of Computer Systems and Sciences*, 49:149–167, 2 1994.

[70] J. Nocedal and S.J. Wright. *Numerical Optimization.* Springer Press, New York, 1999.

[71] C. M. Papadimitriou. *Computational complexity.* Addison-Wesley, 1994.

[72] P. Parrilo and B. Sturmfels. Minimizing polynomial functions. In Saugata Basu and Laureano Gonzalez-Vega, editors, *Algorithmic and quantitative real algebraic geometry*, volume 60, pages 83–100. American Mathematical Society, 2003.

[73] N. Pippenger. Reliable computation by formulas in the presence of noise. *IEEE Trans. Inform. Theory*, 34(2):194–197, 1988.

[74] S. Popescu and D. Rohrlich. Quantum nonlocality as an axiom. In *Foundations of Physics*, volume 24(3), pages 379–385. 1994.

[75] S. Popescu and D. Rohrlich. Nonlocality as an axiom for quantum theory: The dilemma of einstein, podolsky and rosen, 60 years later. In *International symposium in honour of Nathan Rosen.* 1996.

[76] S. Popescu and D. Rohrlich. Causality and nonlocality as axioms for quantum mechanics. In *Proceedings of the Symposium of Causality and Locality in Modern Physics and Astronomy: Open Questions and Possible Solutions.* 1997. quant-ph/9709026.

[77] A. Rao. Parallel repetition in projection games and a concentration bound. In *Proceedings of 40th ACM STOC*, 2008.

[78] R. Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, 1998.

[79] A. Razborov. An upper bound on the threshold quantum decoherence rate. *Quantum Information and Computation*, 4(3):222–228, 2004. quant-ph/0310136.

[80] B. Reichardt. personal communicaton. 2005.

[81] B. Reichardt. Quantum universality from Magic States Distillation applied to CSS codes. 4:251–264, 2005.

[82] B. Reichardt. *Error-Detection-Based Quantum Fault Tolerance Against Discrete Pauli Noise.* PhD thesis, UC Berkeley, 2006. quant-ph/0612004.

[83] B. Reichardt. Quantum universality by distilling certain one- and two-qubit states with stabilizer operations. quant-ph/0608085, 2006.

[84] M. B. Ruskai, S. Szarek, and E. Werner. An analysis of completely-positive trace-preserving maps on $\mathcal{M}_2$. *Linear Algebra and its Applications*, 347:159–187, 2002. quant-ph/0101003.

[85] A. Shamir. IP=PSPACE. *J. ACM*, 39(4):869–877, 1992.

[86] P. Shor. Scheme for reducing decoherence in quantum memory. *Phys. Rev. A 52, 2493*, 1995.

[87] P. Shor. Fault-tolerant quantum computation. In *Proc. 37th Annual Symposium on Foundations of Computer Science*, pages 56–65. IEEE Computer Society Press, 1996.

[88] P. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, 1997. Earlier version in FOCS'94. quant-ph/9508027.

[89] R. Solovay. talk at mathematical sciences research institute.

[90] A. Steane. Multiple particle interference and quantum error correction. In *Proceedings of the Royal Society of London*, volume A452, pages 2551–2577, 1996. quant-ph/9601029.

[91] M. Szegedy. Functions with bounded symmetric communication complexity, programs over commutative monoids, and acc. *Journal of Computer and System Sciences*, 47:405–423, 1993.

[92] B. Toner, 2007. personal communication.

[93] B. Tsirelson. Quantum generalizations of bell's inequality. *Letter in Mathematical Physics*, 4:93–100, 1980.

[94] B. S. Tsirel'son. Quantum analogues of the Bell inequalities: The case of two spatially separated domains. Journal of Soviet Mathematics, pages 36:557–570, 1987.

[95] Falk Unger. Noise threshold for universality of two-input gates. *IEEE Transactions on Information Theory*, 54(8):3693–3698, 2008.

[96] W. van Dam. *Nonlocality & Communication Complexity.* PhD thesis, University of Oxford, Department of Physics, 2000.

[97] W. van Dam. Impossible consequences of superstrong nonlocality. 2005. quant-ph/0501159.

[98] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38:49–95, 1996.

[99] S. Virmani, S. Huelga, and M. Plenio. Classical simulability, entanglement breaking, and quantum computation thresholds. *Physical Review A*, 71(042328), 2005. quant-ph/0408076.

[100] J. Watrous. PSPACE has constant-round quantum interactive proof systems. In *Proceedings of 40th IEEE FOCS*, pages 112–119, 1999. cs.CC/9901015.

[101] J. Watrous, 2004.

[102] S. Wehner. Entanglement in interactive proof systems with binary answers. In *STACS*, pages 162–171, 2006.

[103] R. de Wolf. Quantum communication and complexity. *Theoretical Computer Science*, 287(1):337–353, 2002.

[104] R. de Wolf, 2008. personal communication.

[105] A. C. Yao. Some complexity questions related to distributive computing. *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*, pages 209–213, 1979.

[106] A. C. Yao. Quantum circuit complexity. pages 352–361, 1993.

Index

Samenvatting

Quantum computers lijken mogelijkheden te hebben die verder gaan dan die van klassieke computers. Een voorbeeld dat belangrijk is voor de cryptografie, is dat quantum computers veel sneller grote getallen kunnen factoriseren dan mogelijk lijkt op klassieke machines.

Om daadwerkelijk een quantum computer te kunnen bouwen, is het noodzakelijk om voldoende preciese hardware te bouwen, wat een grote uitdaging is. In Deel I van dit proefschrift bewijzen we ondergrenzen op de noodzakelijke precisie van de hardware die nodig is voor quantum computers. We presenteren ook analoge resultaten voor klassieke computers.

Eén middel dat quantum computers wel, maar klassieke computers niet hebben, is entanglement. In Deel II bestuderen we bepaalde algemene aspecten van entanglement in termen van XOR-spelen en niet-localiteit.

Part I: Limits on fault-tolerant quantum and classical computation

Op dit moment zijn quantum algoritmes alleen snel in theory, aangezien we ver verwijderd zijn van het daadwerkelijk bouwen van quantum computers die groot genoeg zijn om grote instanties van die problemen (zoals factorisatie) op te lossen. Dit is ondanks een krachtige inspanning van de natuurkundigen de afgelopen tien jaar. De reden hiervoor is dat quantum computers uit heel kleine componenten moeten worden opgebouwd om quantum effecten te kunnen laten zien. Het bouwen en manipuleren van zulke kleine componenten is moeilijk, en het is waarschijnlijk onmogelijk om dit geheel foutloos te doen. Niet goed functionerende componenten bevatten “ruis”. Het is tamelijk verassend dat het nog steeds mogelijk is om willekeurig lange quantum berekeningen te doen zelfs wanneer de componenten wat ruis bevatten. Helaas is het op dit moment onmogelijk om hardware te bouwen die voldoende precies is voor grootschalige quantum berekeningen.

In Deel I van dit proefschrift laten we bepaalde minimum vereisten zien voor de precisie van quantum hardware, door bovengrenzen te bewijzen op de hoeveelheid ruis die quantum computers kunnen tolereren. We laten de volgende bovengrenzen zien op de tolereerbare hoeveelheid ruis:

1. In Hoofdstuk 3 behandelen we circuits waarin de poorten ("gates") fan-in hoogstens k hebben, waarbij elke ingaande draad ("input wire") met kans $1 - 1/k$ verwijderd wordt (zogenaamde "erasure noise"). We laten zien dat het al na een constante hoeveelheid tijd onmogelijk wordt om twee verschillende begintoestanden van elkaar te onderscheiden met een meting op één qubit. Voor circuits van polynomiale grootte kan na logarithmische tijd zelfs een meting op *alle* qubits de twee begintoestanden niet meer van elkaar onderscheiden.

2. In Hoofdstuk 4 analyseren we circuits die opgebouwd zijn uit vrijwel perfecte 1-qubit poorten, en willekeurige unitaire k-qubit poorten waarvan elk van de k ingaande draden $1 - \sqrt{2^{1/k} - 1}$ depolariserende ruis ondergaat. We laten zien dat na constante tijd geen enkele één-qubit meting twee verschillende begintoestanden nog kan onderscheiden. Voor het interessante geval $k = 2$ is onze bovengrens op de ruis 35.7%.

3. In Hoofdstuk 5 laten we zien dat circuits die zijn opgebouwd uit zogenaamde "stabilizer gates" (Hadamard, Phase, CNOT, metingen in de computationele basis, en preparaties van computationele basis-toestanden) en willekeurige 1-qubit poorten met depolariserende ruis van minstens $\hat{\theta} = (6 - 2\sqrt{2})/7 \approx 45\%$, efficiënt door een klassieke computer gesimuleerd kunnen worden.

In Hoofdstuk 6 analyseren we ruis in klassieke computers. In moderne computers treden zo weinig fouten op dat de problemen van fout-correctie en fout-tolerantie grotendeels genegeerd worden. Echter, wanneer de grootte van nieuwe hardware-componenten blijft krimpen, dan zullen fouten steeds waarschijnlijker worden, en wordt het belangrijk om hier goed mee om te gaan.

4. We laten een drempelwaarde zien op de tolereerbare hoeveelheid ruis voor berekeningen met formules bestaande uit poorten met 2 inputs, en ϵ ruis per poort: fout-tolerante klassieke berekening is mogelijk dan, en slechts dan, als $\epsilon < (3 - \sqrt{7})/4 \approx 8.856\%$.

Part II: Entanglement and interactive proof systems

In Hoofdstuk 7 bestuderen we spelen ("games") tussen een "verificator" en twee "bewijzers". Tussen de twee bewijzers kan entanglement bestaan. In deze spelen stuurt de verificator vragen naar de bewijzers, die het spel winnen als ze deze vragen correct beantwoorden. Spelen zoals deze vormen de basis van alle interactieve bewijssystemen met meerdere bewijzers, maar ze hebben ook andere toepassingen. Voor spelen van een bepaald type, namelijk zogenaamde "XOR spelen", laten we een perfecte parallelle repetitie stelling zien: de maximale kans waarmee de bewijzers een aantal simultaan gespeelde XOR spelen kunnen winnen, is precies gelijk aan het product van de maximale kansen waarmee ze de individuele spelen kunnen winnen. Dit is een opmerkelijke eigenschap van entanglement, aangezien zo'n stelling niet waar is wanneer de bewijzers *klassiek* zijn. Bovendien zijn quantum XOR spelen de enige soort spelen waarvoor op dit moment een perfecte parallelle repetitie stelling bekend is.

In Hoofdstuk 8 analyseren we een ander aspect van entanglement. Entanglement staat twee gescheiden partijen toe om niet-locale correlaties te hebben, d.w.z. correlaties die niet verklaard kunnen worden door een lokale klassieke theorie met verborgen variabelen. Tsirelson heeft een bovengrens laten zien op de sterkte van dit soort correlaties, met behulp van de axioma's van de quantummechanica. Zijn grens staat bekend als "Tsirelson's bound". In Hoofdstuk 8 laten we zien dat een zwakkere versie van Tsirelson's bound kan worden afgeleid uit bepaalde algemene plausibele aannames over de wereld, zonder de axioma's van de quantummechanica zelf te gebruiken. Het doel hierbij is om bepaalde verrassende gevolgen van de quantummechanica te verklaren, met gebruik van zwakkere aannames over de werkelijke wereld. De aanname die we gebruiken is de volgende: wanneer twee gescheiden partijen die elk een deel van de invoer bezitten samen een functie willen berekenen, dan moeten ze over het algemeen meer dan één bit communiceren.

Abstract

Quantum computers seem to have capabilities which go beyond those of classical computers. A particular example which is important for cryptography is that quantum computers are able to factor numbers much faster than what seems possible on classical machines.

In order to actually build quantum computers it is necessary to build sufficiently accurate hardware, which is a big challenge. In part I of this thesis we prove lower bounds on the accuracy of the hardware needed to do quantum computation. We also present a similar result for classical computers.

One resource that quantum computers have but classical computers do not have is entanglement. In Part II of this thesis we study certain general aspects of entanglement in terms of quantum XOR games and non-locality.

Part I: Limits on fault-tolerant quantum and classical computation

At the moment, quantum algorithms are only fast in theory, since we are a long way from building quantum computers large enough to solve large instances of these problems (for example factoring). This is despite a decade-long, concentrated effort by experimental physicists. The reason is that quantum computers must be built from very small components in order to exhibit quantum properties. Building and operating on these small components is hard and is probably not possible without faults. Faulty devices are also called "noisy". Rather surprisingly, it is still possible to do arbitrarily long quantum computation even if the physical devices used are not perfect but slightly noisy. Unfortunately, currently it is not possible to build hardware which is accurate enough to allow large-scale quantum computation.

In Part I of this thesis we show minimum requirements on the accuracy of quantum hardware, by proving upper bounds on the amount of noise tolerable for fault-tolerant quantum computation. We show the following upper bounds on the tolerable noise rates:

1. In Chapter 3 we consider circuits with arbitrary gates of fan-in at most k, in which each wire is subject to more than $1 - 1/k$ erasure noise. We show that already after a constant amount of time it is impossible to distinguish any two input states by a single-qubit measurement. For polynomial-size circuits it is impossible to distinguish any two input states by measurements on *all* qubits after time which is logarithmic in the size of the circuit.

2. In Chapter 4 we analyze circuits built with almost perfect 1-qubit gates and arbitrary k-qubit unitaries in which all incoming wires are subject to at least $1 - \sqrt{2^{1/k} - 1}$ depolarizing noise. We show that after a constant amount of time no single-qubit measurement can distinguish any two input states. For the interesting case $k = 2$ our bound becomes 35.7%.

3. In Chapter 5 we show that circuits built from stabilizer gates (Hadamard gate, Phase gate, CNOT, measurements in the computational basis, preparation of computational basis states) and arbitrary 1-qubit gates with depolarizing noise at least $\hat{\theta} = (6-2\sqrt{2})/7 \approx 45\%$ can be efficiently simulated on a classical computer.

In Chapter 6 we analyze noise in classical computation. Faults happen in modern computers so rarely that the problem of error-correction and fault-tolerance is nowadays essentially ignored. However, if hardware engineers continue to shrink the size of components, faults will become more likely and it will be important to know how to cope with them.

4. We show a threshold on the tolerable noise for computation by formulas with ϵ-noisy 2-input gates: Fault-tolerant classical computation is possible if and only if $\epsilon < (3 - \sqrt{7})/4 \approx 8.856\%$.

Part II: Entanglement and interactive proof systems

In Chapter 7 we consider games played between a verifier and two (possibly entangled) provers. In these games the verifier sends questions to the provers, who win if they can answer them correctly. Games like this are the basis of all multi-prover interactive proof systems, but they also have many other applications. For certain types of games, namely quantum XOR games, we show a perfect parallel repetition theorem: The provers' optimal success probability for winning a collection of quantum XOR games played simultaneously is equal to the product of the success probabilities of the individual games. This is a remarkable feature of entanglement, since for *classical* XOR games a perfect parallel repetition theorem does not hold. Further, quantum XOR games are the only kind of games which are currently known to obey a perfect parallel repetition theorem.

In Chapter 8 we analyze a different feature of entanglement. Entanglement allows two separated parties to exhibit non-local correlations, i.e., correlations

that cannot be explained by any classical local hidden-variable model. Tsirelson proved an upper bound on the strength of these correlations, using the quantum mechanical axioms. His bound is known as Tsirelson's bound. In Chapter 8 we show that a weaker version of Tsirelson's bound can be derived from some general plausible assumptions about the world, without invoking the axioms of quantum mechanics themselves. The aim is to explain certain surprising consequences of quantum mechanics, using plausible assumptions about the real world. The assumption we use is the following: Two separated parties Alice and Bob need to communicate in general more than one bit, in order to compute the value of a Boolean function for which some of the input bits are in Alice's possession and some in Bob's.

Titles in the ILLC Dissertation Series:

ILLC DS-2001-01: **Maria Aloni**
Quantification under Conceptual Covers

ILLC DS-2001-02: **Alexander van den Bosch**
Rationality in Discovery - a study of Logic, Cognition, Computation and Neuropharmacology

ILLC DS-2001-03: **Erik de Haas**
Logics For OO Information Systems: a Semantic Study of Object Orientation from a Categorial Substructural Perspective

ILLC DS-2001-04: **Rosalie Iemhoff**
Provability Logic and Admissible Rules

ILLC DS-2001-05: **Eva Hoogland**
Definability and Interpolation: Model-theoretic investigations

ILLC DS-2001-06: **Ronald de Wolf**
Quantum Computing and Communication Complexity

ILLC DS-2001-07: **Katsumi Sasaki**
Logics and Provability

ILLC DS-2001-08: **Allard Tamminga**
Belief Dynamics. (Epistemo)logical Investigations

ILLC DS-2001-09: **Gwen Kerdiles**
Saying It with Pictures: a Logical Landscape of Conceptual Graphs

ILLC DS-2001-10: **Marc Pauly**
Logic for Social Software

ILLC DS-2002-01: **Nikos Massios**
Decision-Theoretic Robotic Surveillance

ILLC DS-2002-02: **Marco Aiello**
Spatial Reasoning: Theory and Practice

ILLC DS-2002-03: **Yuri Engelhardt**
The Language of Graphics

ILLC DS-2002-04: **Willem Klaas van Dam**
On Quantum Computation Theory

ILLC DS-2002-05: **Rosella Gennari**
Mapping Inferences: Constraint Propagation and Diamond Satisfaction

ILLC DS-2002-06: **Ivar Vermeulen**
A Logical Approach to Competition in Industries

ILLC DS-2003-01: **Barteld Kooi**
Knowledge, chance, and change

ILLC DS-2003-02: **Elisabeth Catherine Brouwer**
Imagining Metaphors: Cognitive Representation in Interpretation and Understanding

ILLC DS-2003-03: **Juan Heguiabehere**
Building Logic Toolboxes

ILLC DS-2003-04: **Christof Monz**
From Document Retrieval to Question Answering

ILLC DS-2004-01: **Hein Philipp Röhrig**
Quantum Query Complexity and Distributed Computing

ILLC DS-2004-02: **Sebastian Brand**
Rule-based Constraint Propagation: Theory and Applications

ILLC DS-2004-03: **Boudewijn de Bruin**
Explaining Games. On the Logic of Game Theoretic Explanations

ILLC DS-2005-01: **Balder David ten Cate**
Model theory for extended modal languages

ILLC DS-2005-02: **Willem-Jan van Hoeve**
Operations Research Techniques in Constraint Programming

ILLC DS-2005-03: **Rosja Mastop**
What can you do? Imperative mood in Semantic Theory

ILLC DS-2005-04: **Anna Pilatova**
A User's Guide to Proper names: Their Pragmatics and Semanics

ILLC DS-2005-05: **Sieuwert van Otterloo**
A Strategic Analysis of Multi-agent Protocols

ILLC DS-2006-01: **Troy Lee**
Kolmogorov complexity and formula size lower bounds

ILLC DS-2006-02: **Nick Bezhanishvili**
Lattices of intermediate and cylindric modal logics

ILLC DS-2006-03: **Clemens Kupke**
Finitary coalgebraic logics

ILLC DS-2006-04: **Robert Špalek**
Quantum Algorithms, Lower Bounds, and Time-Space Tradeoffs

ILLC DS-2006-05: **Aline Honingh**
The Origin and Well-Formedness of Tonal Pitch Structures

ILLC DS-2006-06: **Merlijn Sevenster**
Branches of imperfect information: logic, games, and computation

ILLC DS-2006-07: **Marie Nilsenova**
Rises and Falls. Studies in the Semantics and Pragmatics of Intonation

ILLC DS-2006-08: **Darko Sarenac**
Products of Topological Modal Logics

ILLC DS-2007-01: **Rudi Cilibrasi**
Statistical Inference Through Data Compression

ILLC DS-2007-02: **Neta Spiro**
What contributes to the perception of musical phrases in western classical music?

ILLC DS-2007-03: **Darrin Hindsill**
It's a Process and an Event: Perspectives in Event Semantics

ILLC DS-2007-04: **Katrin Schulz**
Minimal Models in Semantics and Pragmatics: Free Choice, Exhaustivity, and Conditionals

ILLC DS-2007-05: **Yoav Seginer**
Learning Syntactic Structure

ILLC DS-2008-01: **Stephanie Wehner**
Cryptography in a Quantum World

ILLC DS-2008-02: **Fenrong Liu**
Changing for the Better: Preference Dynamics and Agent Diversity

ILLC DS-2008-03: **Olivier Roy**
Thinking before Acting: Intentions, Logic, Rational Choice

ILLC DS-2008-04: **Patrick Girard**
Modal Logic for Belief and Preference Change

ILLC DS-2008-05: **Erik Rietveld**
Unreflective Action: A Philosophical Contribution to Integrative Neuroscience

ILLC DS-2008-06: **Falk Unger**
Noise in Quantum and Classical Computation and Non-locality