Quantum Query Complexity and Distributed Computing

ILLC Dissertation Series DS-2004-01

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@wins.uva.nl`
homepage: `http://www.illc.uva.nl/`

Quantum Query Complexity and Distributed Computing

Academisch Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof.mr. P.F. van der Heijden
ten overstaan van een door het college voor
promoties ingestelde commissie, in het openbaar
te verdedigen in de Aula der Universiteit
op dinsdag 27 januari 2004, te 12.00 uur

door

Hein Philipp Röhrig

geboren te Frankfurt am Main, Duitsland.

Promotores: Prof.dr. H.M. Buhrman
Prof.dr.ir. P.M.B. Vitányi

Overige leden: Prof.dr. R.H. Dijkgraaf
Prof.dr. L. Fortnow
Prof.dr. R.D. Gill
Dr. S. Massar
Dr. L. Torenvliet
Dr. R.M. de Wolf

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

The investigations were supported by the Netherlands Organization for Scientific Research (NWO) project “Quantum Computing” (project number 612.15.001), by the EU fifth framework projects QAIP, IST-1999-11234, and RESQ, IST-2001-37559, the NoE QUIPROCONE, IST-1999-29064, and the ESF QiT Programme.

Revision 411
ISBN: 3–933966–04–3

Contents

Acknowledgments

I am indebted to Harry Buhrman for his guidance, advice, and comradeship. Most problems addressed in this thesis were raised by him; his ideas also contributed to many of the solutions. I am also very grateful to Paul Vitányi, who offered me the PhD position and whose pragmatic yet highly competent management style I admire.

Special thanks go to Ronald de Wolf. He was the first real quantum-computing researcher I ever talked to, the night before AQIP 98. His close reading of the thesis improved it a lot; all remaining errors are of course mine. He, John Tromp, and I shared an office—the drie 'heren' frequently were the only people at the institute at night and during the weekend.

I also thank my other coauthors of the papers that are the foundation of this thesis: Andris Ambainis, Yevgeniy Dodis, Lance Fortnow, Peter Høyer, Serge Massar, and Ilan Newman. During my studies, I spent a substantial amount of time as guest of Lov Grover at Bell Labs. What I know about quantum search, I learned from him.

Two long-time mentors deserve special mention here: Prof. Dr. Karl Hensen, my "Vertrauensdozent" in the Studienstiftung, was a constant point of reference outside my specialty. From Prof. Dr. Dr. h.c. mult. Günter Hotz I learned to critically review objectives of research in a wider context.

For many pleasant, instructive, and fruitful scientific discussions, I thank Scott Aaronson, Luis Antunes, Eldar Fischer, Péter Gács, Mart de Graaf, Peter Grünwald, Jaap-Henk Hoepman, Troy Lee, Ashwin Nayak, Daniel Preda, Yaoyun Shi, Robert Špalek, and Arjen de Vries. For introducing me to new worlds, for advice and friendship, and generally a good time I thank Torben Hagerup, Ute Röhrig, Richard Hahnloser, Sebastian Seung, Tarmo Johannes, Dasha Beltsiukova, Liddy Shriver, Markus Jakobsson, and Susanne Wetzel. Thanks to Rudi Cilibrasi and Rudolf Janz for proofreading drafts of this thesis and instructing me to exorcize parentheses; Stefan Manegold and Kolja Sulimma provided valuable advice about printing this thesis.

I am grateful to my parents for their support throughout my studies. And Tzveta, thanks for always stimulating me to finish this thesis and for providing distraction from work!

Hein Röhrig
Amsterdam, December 2003

Publications

The following publications are the base for Chapters 2–6.

- H. Buhrman, L. Fortnow, I. Newman, and H. Röhrig. Quantum property testing. In *Proceedings of 14th SODA*, pages 480–488, 2003, quant-ph/0201117.
- H. Buhrman, I. Newman, H. Röhrig, and R. de Wolf. Robust quantum algorithms and polynomials. Submitted, quant-ph/0309220.
- H. Buhrman, P. Høyer, S. Massar, and H. Röhrig. Combinatorics and quantum nonlocality. *Physical Review Letters*, 91(4):047903, 2003, quant-ph/0209052.
- H. Buhrman and H. Röhrig. Distributed quantum computing. In B. Rovan and P. Vojtas, editors, *Mathematical Foundations of Computer Science 2003*, volume 2747 of *Lecture Notes in Computer Science*, pages 1–20. Springer, 2003.
- A. Ambainis, H. Buhrman, Y. Dodis, and H. Röhrig. Multiparty quantum coin flipping. Submitted, 2003, quant-ph/0304112.

Chapter 1

Introduction

1.1 Computation Is Physical

The last two decades have seen a renewed interest in the relation between computation and physics. In particular, in the early 1980s Feynman [54, 55] pointed out that simulating quantum mechanics appears to be difficult in models of computation that were then thought to represent the strongest feasible form of computation. Moreover, he raised the question whether a quantum-mechanical computer could be more powerful than a "classical" computer, e.g., in analyzing other quantum-mechanical systems. Most physicists believe that the world is quantum-mechanical, or at least is more accurately described by quantum mechanics than by classical theories. In this case, it should in principle be possible to actually build such quantum computers. Conversely, quantum computers can also be regarded as experiments for verifying the predictions of quantum mechanics, and principal obstacles may very well indicate limitations of quantum mechanics. This would be of great interest since to date there is no experimental data contradicting quantum mechanics.

Initially, interest among computer scientists was limited. In part this was due to the similarity of Feynman's proposal to conventional "analog" computers. Deutsch [47] laid the theoretical groundwork for a "digital" variant of quantum computing, and in the 1980s and in the early 1990s a sequence of quantum algorithms appeared [48, 22, 109] that showed that quantum computers are in certain aspects significantly more powerful than classical computers. However, the area really became popular only after Shor presented an efficient quantum algorithm for factoring integers [108], a problem considered so difficult classically that the most important cryptographic systems both in theory and practice rely on its hardness. Further theoretical discoveries included the feasibility of correcting errors in quantum computation. This answered affirmatively the question whether an imperfect real-world quantum mechanical device could benefit from these theoretical advantages; thus, one could say that quantum error correction is an example of a contribution to physics by computer scientists.

1.2 Quantum Mechanics

An atom is not a soccer ball: the measures and laws of classical physics, which describe the motion of a soccer ball in space, fail to explain physical phenomena at the atomic scale. For example, the laws of classical physics do not adequately describe the structure of the atomic nucleus, the wave-particle character of light, and discrete absorption spectra. These phenomena can be explained by the physical theory of quantum mechanics. This theory has been developed from 1925 on chiefly by Heisenberg and Schrödinger. Despite some seemingly "unnatural" model assumptions, quantum mechanics is today accepted by most physicists as the best tool to describe nature on the very small-scale level or where tiny differences of energy are involved. In the limit of many particles and great energy differences, quantum mechanical laws converge to their classical counterparts.

As in classical physics, we would like to model the state of our quantum-mechanical system (say, five hydrogen atoms, or a photon, or two electrons) at time t. That is, we want to describe the system at time t to the extent that, knowing the dynamics, we can predict the behavior of the system at any time in the future. Hence, our model also must say how the system develops in time.

In classical physics, we have a one-to-one correspondence between the state of a system at time t and the result of a complete measurement of the system at time t. This is not the case in quantum mechanics; here the concept of state and measurement differ: in general, the result of a measurement cannot be fully predicted when knowing the state; moreover, the action of measuring will affect the state of system. Figure 1.1 provides a vague sketch of this property. In the following, we are going to present the *postulates* or *axioms* of quantum mechanics as far as they are relevant to quantum computing.

1.2.1 States

1.2.1. Postulate. The state of a quantum system is a nonzero vector in a Hilbert space $\mathcal{H}$, which is called the *state space*.

A Hilbert space is a vector space over the complex numbers with an inner product. In this work we need to consider only Hilbert spaces of finite dimension, which suffice for the *quantization* of classical discrete systems; the term

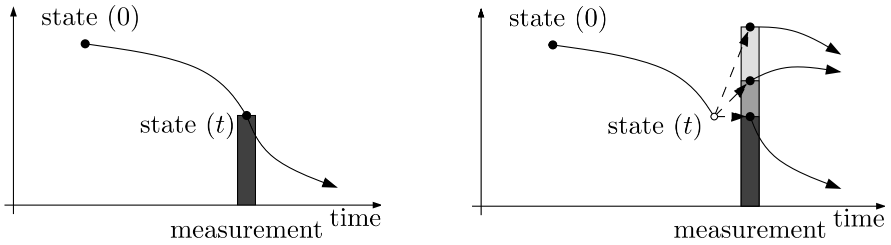


Figure 1.1: Classical (left) vs. quantum physics (right): sketched are *trajectories* in the *state* space. In quantum mechanics a measurement projects the state at random into one of several outcomes. The outcome is observed macroscopically and time evolution resumes from it.

quantization stands for the necessarily informal process of generalizing or embedding a classical system into quantum mechanics. For infinite-dimensional vector spaces, there are some additional requirements to be a Hilbert space, but in the finite-dimensional case, all Hilbert spaces are isomorphic to some $\mathbb{C}^n$ with $n \in \mathbb{N}$ and the additional requirements to guarantee a "nice" topology are automatically given.

We regard vectors from the state space $\mathcal{H} = \mathbb{C}^n$ as n-dimensional column vectors and elements of the *dual space* $\mathcal{H}^*$ as n-dimensional row vectors. The *Dirac notation* defines an elegant shorthand for expressions involving vectors and their duals:

- Column vectors are enclosed by the *ket* sign $|\ \rangle$. Hence, we write

$$|\psi\rangle = \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_n \end{pmatrix} \in \mathcal{H} \ .$$

- Row vectors are enclosed by the *bra* sign $\langle\ |$. Hence, the dual of $|\psi\rangle$ is

$$\langle\psi| = \begin{pmatrix} \psi_1^* & \cdots & \psi_n^* \end{pmatrix} \in \mathcal{H}^*$$

  where $\psi_j^* := \alpha - \mathrm{i}\,\beta$ denotes the complex conjugate of the complex number $\psi_j = \alpha + \mathrm{i}\,\beta$, $\alpha, \beta \in \mathbb{R}$.

The matrix product of a bra and a ket is $\langle\psi| \cdot |\varphi\rangle = \langle\psi|\varphi\rangle$, which is the inner product of $|\psi\rangle$ and $|\varphi\rangle$. The fusion of symbols $\langle\ | \cdot |\ \rangle = \langle\ |\ \rangle$ is the origin of the names bra-ket = bra(c)ket.

The most simple nontrivial quantum system is called a *qubit* and has a two-dimensional state space. Using the convention

$$|0\rangle := \begin{pmatrix}1\\0\end{pmatrix} \quad \text{and} \quad |1\rangle := \begin{pmatrix}0\\1\end{pmatrix}$$

the general state of a qubit can be written as

$$\alpha_0|0\rangle + \alpha_1|1\rangle = \begin{pmatrix}\alpha_0\\\alpha_1\end{pmatrix} \quad \text{with} \quad \alpha_0, \alpha_1 \in \mathbb{C} \text{ and } |\alpha_0|^2 + |\alpha_1|^2 > 0 \ .$$

This is a *superposition* of the states $|0\rangle$ and $|1\rangle$, whereas a classical bit can assume only the two values 0 and 1. We will see further down that collinear vectors represent the same "physical" states and that the *amplitudes* α_0 and α_1 are a measure of how close the qubit is to the classical states 0 and 1. In particular, $|\alpha_0|^2/(|\alpha_0|^2 + |\alpha_1|^2)$ is the probability for observing 0 in a measurement of the qubit. Therefore it is often useful to *normalize* the states to have ℓ_2 norm 1; then $|\alpha_0|^2$ is the probability for observing 0 and $|\alpha_1|^2$ is the probability for observing 1.

What is the state space of two qubits? Again we use the "classical" values to denote the canonical basis vectors, leading to

$$|00\rangle := \begin{pmatrix}1\\0\\0\\0\end{pmatrix} \ , \ |01\rangle := \begin{pmatrix}0\\1\\0\\0\end{pmatrix} \ , \ |10\rangle := \begin{pmatrix}0\\0\\1\\0\end{pmatrix} \ , \ |11\rangle := \begin{pmatrix}0\\0\\0\\1\end{pmatrix}$$

and again each nontrivial linear combination will be a possible state of the two-qubit system. More generally, n qubits have state space $\mathbb{C}^{2^n}$; the state space grows exponentially with the number of qubits. This is an important feature for quantum computation; however, note that even a joint probability distribution on n bits is a vector in $\mathbb{R}^{2^n}$ of ℓ_1 norm 1 and nonnegative components, so that the power of the computational model derives really from the possible operations on vectors in the state space.

The mathematical construct underlying the combination of the two state spaces of two quantum systems into one state space is the *tensor product* or Kronecker product:

$$\begin{aligned}|\psi\rangle \otimes |\varphi\rangle &= (\alpha_0|0\rangle + \alpha_1|1\rangle) \otimes (\beta_0|0\rangle + \beta_1|1\rangle) \\ &= \alpha_0\beta_0|00\rangle + \alpha_0\beta_1|01\rangle + \alpha_1\beta_0|10\rangle + \alpha_1\beta_1|11\rangle \ .\end{aligned}$$

By convention, $|0\rangle \otimes |0\rangle$, $|0\rangle|0\rangle$, and $|00\rangle$ denote the same thing. In general not all the two-qubit states can be obtained as the tensor product of two qubits. We will see an important example, the EPR pair, in Subsection 1.2.4. Such states are called *entangled*.

1.2.2 Evolution

As a physical model quantum mechanics must explain how the system develops as time progresses; this we call the *dynamics* or the *evolution* of the system. A complete description of the system at an arbitrary but fixed time τ_0 is given by the Hilbert-space vector $|\psi(\tau_0)\rangle$ representing the system state at time τ_0. This description should permit us to predict the state of the undisturbed system at all future points in time. Mathematically, this means that we require a function of the time t, the time evolution function, $t \mapsto |\psi(t)\rangle$, where $|\psi(t)\rangle$ is the state of the system at time t.

Knowledge of a state will also at a later time τ be sufficient for the prediction of the states $|\psi(t)\rangle$, $t > \tau$; accordingly, the time derivative at point τ, $\frac{\mathrm{d}}{\mathrm{d}\,t}|\psi(t)\rangle\big|_{t=\tau}$, must be a function $\widehat{H}$ that depends solely on $|\psi(\tau)\rangle$:

$$\frac{\mathrm{d}}{\mathrm{d}\,t}|\psi(t)\rangle\bigg|_{t=\tau} = \widehat{H}(|\psi(\tau)\rangle) \ . \tag{1.1}$$

At first, there is no reason for this differential equation for $|\psi(t)\rangle$ to be anything but arbitrarily complex. However, an enormous simplification is obtained by the following assumption:

1.2.2. Postulate (Superposition Principle). Let state $|\psi\rangle$ at time τ_0 evolve according to the time evolution function into $|\psi'\rangle$ at time $\tau_1 > \tau_0$, and let state $|\varphi\rangle$ at time τ_0 evolve into $|\varphi'\rangle$ at time τ_1. Then linear combinations of $|\psi\rangle$ and $|\varphi\rangle$ evolve into the corresponding linear combinations of $|\psi'\rangle$ and $|\varphi'\rangle$, i.e., for all $w, z \in \mathbb{C}$ we have

$$|\psi\rangle \rightsquigarrow |\psi'\rangle \quad \text{and} \quad |\varphi\rangle \rightsquigarrow |\varphi'\rangle \quad \Longrightarrow \quad (u\,|\psi\rangle + w\,|\varphi\rangle) \rightsquigarrow (u\,|\psi'\rangle + w\,|\varphi'\rangle) \ ,$$

where $\cdot \rightsquigarrow \cdot$ symbolizes the time evolution function from time τ_0 to time τ_1.

The superposition principle is a linearity assumption. It implies that the function $\widehat{H}$ in Eq. (1.1) must be a *linear* function from the Hilbert space $\mathcal{H}$ to itself. Such functions we call *operators*. Hence, $\widehat{H}$ must be an operator on the $|\psi(\tau)\rangle$ argument. This is a strong constraint, since

> “superposed systems evolve in total obliviousness of each of the others, quite *independently* of whether there are any interactions involved. This fact alone might lead us to question the absolute truth of the linearity property. Yet it is very well confirmed for phenomena that remain entirely at the quantum level.” [98, p. 289]

$\widehat{H}$ can, however, depend arbitrarily on the time τ. We rewrite the differential equation (1.1) for the time evolution, letting $\tau_0 = 0$, using t instead of τ and substituting $\frac{-\mathrm{i}}{\hbar}H(t)$ for the linear operator $\widehat{H}(t)$. This way we arrive at the following assumption, which is essentially a reformulation of the superposition principle:

1.2.3. Postulate (Schrödinger Equation). Every quantum-mechanical system has at every time t a uniquely defined self-adjoint operator $H(t)$, the *Hamiltonian*, which describes the "total energy" of the system at time t. For the state $|\psi(t)\rangle$ of the quantum-mechanical system at time t we have the differential equation

$$\mathrm{i}\,\hbar\frac{\mathrm{d}}{\mathrm{d}\,t}|\psi(t)\rangle = H(t)\,|\psi(t)\rangle \quad \text{for } t > 0$$

where $\hbar := h/(2\pi)$ and h is a physical constant.

Let us investigate in more detail this postulate for finite-dimensional state spaces. A *self-adjoint* operator H satisfies $H = H^*$ where the adjoint operator T^* of T is the unique function for which $\langle T^*\psi|\varphi\rangle = \langle\psi|T\varphi\rangle$ for all $|\psi\rangle, |\varphi\rangle \in \mathcal{H}$. If we represent T as a matrix with respect to a fixed basis then T^* is the transposed and componentwise complex conjugated matrix. Why does it make sense to require that the Hamiltonian operator in the Schrödinger equation be self-adjoint? The reason is that we would like the solution of the differential equation $U_t : |\psi(0)\rangle \mapsto |\psi(t)\rangle$ to preserve the ℓ_2 norm of $|\psi(0)\rangle$—we mentioned in Subsection 1.2.1 that collinear vectors are the same state and therefore preserving the norm means that we have fewer redundant degrees of freedom. A self-adjoint operator H is diagonalizable and therefore the power series

$$U := e^{-(\mathrm{i}\,/\hbar)H} = \sum_{k\geq 0}\frac{(-(\mathrm{i}\,/\hbar)H)^k}{k!} \tag{1.2}$$

is convergent. Moreover, U is diagonalizable and all its eigenvalues have absolute value 1 as required for preserving the norm of $|\psi(0)\rangle$. U is a unitary operator: $UU^* = \mathbb{1}$. The same construction implies that for every unitary U we can find a self-adjoint H such that $U = e^{\mathrm{i}\,H}$.

In case the Hamiltonian $H(t)$ is independent of time t, U as defined in Eq. (1.2) yields a solution of the Schrödinger equation via the time evolution function

$$U^{\Delta t} = e^{-(\mathrm{i}\,/\hbar)H\Delta t} \quad ,$$

which is U to the power Δt with Δt the length of the time that the Hamiltonian H acts:

$$U^{\tau_2-\tau_1}|\psi(\tau_1)\rangle = |\psi(\tau_2)\rangle \tag{1.3}$$

In the following we will mostly consider situations where the Hamiltonian is constant for discrete time intervals Δt and changes arbitrarily in between.

Then the state $|\psi(k\Delta t)\rangle$ after k time steps is

$$|\psi(k\Delta t)\rangle = U_k^{\Delta t} \cdots U_2^{\Delta t} U_1^{\Delta t} |\psi(0)\rangle$$

where $U_j = e^{-(\mathrm{i}/\hbar)H_j}$ is the unitary operator induced by the Hamiltonian H_j acting at times t with $(j-1)\Delta t \leq t < j\Delta t$. We choose units so that $\Delta t = 1$ and $\hbar = 1$ and usually only talk in terms of unitary operators and "forget" about the underlying Hamiltonians.

If the Hamilton operator $H(t)$ is not time independent, the evolution of the system from time τ_1 to τ_2 is still unitary[1], i.e., for each τ_1 and τ_2 there exists a unitary operator U_{τ_1,τ_2} such that

$$U_{\tau_1,\tau_2}|\psi(\tau_1)\rangle = |\psi(\tau_2)\rangle \ . \tag{1.4}$$

However, in general we cannot express the different U_{τ_1,τ_2} as powers of a single unitary operator.

Eqs. (1.3) and (1.4) hold even if $\tau_2 < \tau_1$; in other words, the evolution of a quantum mechanical systems is *reversible*. Knowing the state of the system at a given time allows us to determine using the Schrödinger equation the state of the system at any given point in time, future or past. Moreover, if we can control the Hamiltonian and replace H by $-H$, the system will evolve backwards in time! Classically, breaking the cup is much easier than mending it—at this level, quantum mechanics is very far from an ensemble theory or thermodynamics. Computing with individual quantum systems at first appears to pose difficulties rather than opportunities since we cannot even reliably set a qubit to 0.

Postulate 1.2.3 also states that the Hamiltonian $H(t)$ in the Schrödinger equation describes the total energy of the system. In order to demonstrate this property, we first need to introduce the quantum mechanical concept of what information can be obtained when measuring a quantum system.

1.2.3 Observables

At the "atomic" scale, experiments suffer from the fact that every measurement requires interaction of the system to be measured with the measuring apparatus and this interaction disturbs the very sensitive state. Another phenomenon at this scale is that repeating an experiment consisting of a given preparation and measurement does not necessarily lead to the same measurement outcome in every run of the experiment, but to different outcomes that appear to obey some probability distribution specific to the experiment.

[1] for finite-dimensional state spaces or bounded $H(t)$ this is a consequence of the Baker-Campbell-Hausdorff theorem; in the general case the unitarity is a requirement that restricts what choices of $H(t)$ are permitted.

Quantum mechanics models these two phenomena, disturbance and uncertainty about the outcome, by defining the measurement as a process that operates on states and that yields a probabilistic outcome. It is important to note that information about the quantum system can only be obtained via a quantum measurement and that in general, it is not possible to reconstruct the entire state vector with a single measurement. We begin by introducing the most basic kind of measurement and will then give a general postulate that captures all possible measurements, even those involving interaction with other quantum systems.

Probabilities from amplitudes Consider the general state of n qubits

$$|\psi\rangle = \sum_{x\in\{0,1\}^n} \alpha_x |x\rangle \quad \text{with } \alpha_x \in \mathbb{C} \ .$$

By Postulate 1.2.1 we have that $|\psi\rangle \neq 0$ and therefore $\| |\psi\rangle \|^2 = \langle\psi|\psi\rangle = \sum_{x\in\{0,1\}^n} |\alpha_x|^2 > 0$. Note that here we do not assume that the states have norm 1; this permits us to avoid rescaling state vectors after measurements. The most simple form of measurement specifies that when measuring $|\psi\rangle$, we observe x with probability

$$\Pr[\text{observe } x] = \frac{|\alpha_x|^2}{\langle\psi|\psi\rangle} \tag{1.5}$$

and when obtaining measurement outcome x, the system is afterwards in state

$$|\psi'\rangle = |x\rangle \ . \tag{1.6}$$

Eq. (1.5) says that the normalized square of the amplitudes induces a probability distribution over the n-bit binary strings $x \in \{0,1\}^n$ and measuring $|\psi\rangle$ means sampling from this distribution. The outcome is discrete, namely a classical bit string. Eq. (1.6) ensures that when repeating the measurement on the same system, it will not change the outcome—once x has been determined, it remains fixed.

What kind of states can we distinguish with such a measurement? Certainly, the ℓ_2 norm of the state does not matter. At first sight, neither would the argument or *phase* ϑ_x of the complex number $\alpha_x = |\alpha_x| e^{\mathrm{i}\,\vartheta_x}$, $0 \leq \vartheta_x < 2\pi$. However, in case the system undergoes evolution according to the Schrödinger equation, the phase does have measurable relevance. For example, if we have a single qubit which evolves in one discrete time step according to the unitary *Hadamard* matrix

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

the qubit $(|0\rangle + |1\rangle)/\sqrt{2}$ evolves in one time step to

$$\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \rightsquigarrow H\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) = |0\rangle$$

whereas $(|0\rangle - |1\rangle)/\sqrt{2}$, which differs only by the relative phase between $|0\rangle$ and $|1\rangle$, behaves as

$$\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \rightsquigarrow H\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) = |1\rangle \ .$$

Hence, the two states that were initially indistinguishable by our simple measurement became perfectly distinguishable. Applying a unitary operator can be regarded as a basis change from the *orthonormal* basis $\{|x\rangle : x \in \{0,1\}^n\}$ to some other orthonormal basis $\{|\varphi_x\rangle : x \in \{0,1\}^n\}$. Incorporating this basis change in the measurement, we say that measuring $|\psi\rangle$ in the basis $\{|\varphi_x\rangle : x \in \{0,1\}^n\}$ we observe x with probability

$$\Pr[\text{observe } x] = \frac{|\langle\psi|\varphi_x\rangle|^2}{\langle\psi|\psi\rangle} = \frac{\langle\psi|\varphi_x\rangle\langle\varphi_x|\psi\rangle}{\langle\psi|\psi\rangle} \tag{1.7}$$

and when obtaining outcome x, the system is afterwards in state

$$|\psi'\rangle = |\varphi_x\rangle \ . \tag{1.8}$$

In our example, a measurement of $(|0\rangle + |1\rangle)/\sqrt{2}$ in the basis $\{(|0\rangle + |1\rangle)/\sqrt{2}, (|0\rangle - |1\rangle)/\sqrt{2}\}$ would yield 0 with probability 1.

In Eq. (1.7), the probability of outcome x is determined by the length of the projection of $|\psi\rangle$ onto $|\varphi_x\rangle$. We define P_x to be the projection onto the linear subspace spanned by $|\varphi_x\rangle$. Using the Dirac notation we can express P_x succinctly as the matrix product of $|\varphi_x\rangle$ with its dual: $P_x := |\varphi_x\rangle\langle\varphi_x|$. Then we can rewrite (1.7) as

$$\Pr[\text{observe } x] = \frac{\langle\psi|P_x|\psi\rangle}{\langle\psi|\psi\rangle} = \frac{\langle\psi|\varphi_x\rangle\langle\varphi_x|\psi\rangle}{\langle\psi|\psi\rangle} \tag{1.9}$$

and (1.8) as

$$|\psi'\rangle = P_x|\psi\rangle \ , \tag{1.10}$$

respectively. Hence, this kind of measurement amounts to a decomposition of the state space $\mathcal{H}$ into orthogonal subspaces that are labeled by the bit strings x. Expressing this fact in terms of operators, we say the measurement projectors P_x have the property

$$P_xP_y = \delta_{x,y}P_x \quad \text{and} \quad \sum_x P_x = \mathbb{1} \ . \tag{1.11}$$

Projective measurements Our third generalization is to allow as measurement any set of operators $\{P_x\}$ that satisfy (1.11), with the probabilities as defined by Eq. (1.9) and the state after the measurement given by (1.10). In particular, we gain the option of performing *partial measurements*: the subspaces onto which we project may have dimension greater than one. The important fact is that the uncertainty about the state is conserved to the greatest extent that is compatible with the measurement outcome. For example, consider the measurement

$$\begin{aligned} P_0 &= |00\rangle\langle 00| + |11\rangle\langle 11| \\ P_1 &= |01\rangle\langle 01| + |10\rangle\langle 10| \end{aligned}$$

applied to a two-qubit system. Outcome 0 means that the two qubits are in the same state, whereas outcome 1 indicates that the two qubits have opposite value. In either case the concrete values are not determined. Applying this measurement to the state

$$|\psi\rangle = \frac{1}{2}\left(|00\rangle + |01\rangle - \sqrt{2}|10\rangle\right)$$

we obtain outcome 0 with probability 1/4 and outcome 1 with probability 3/4. If the outcome is 0, the state becomes $|\psi'\rangle = |00\rangle/2$; if the outcome is 1, the state becomes $|\psi'\rangle = (|01\rangle - \sqrt{2}|10\rangle)/2$.

Observables If the labels of the measurement operators P_x are real numbers, i.e., $x \in \mathbb{R}$, there is a succinct way to represent the entire measurement by a single operator, called an *observable*:

$$A = \sum_x x P_x \quad . \tag{1.12}$$

An observable A is self-adjoint, its eigenvalues are the labels x and the projectors to its eigenspaces the operators P_x. Moreover,

$$\frac{\langle\psi|A|\psi\rangle}{\langle\psi|\psi\rangle} = \sum_x x \frac{\langle\psi|P_x|\psi\rangle}{\langle\psi|\psi\rangle}$$

is the expected value of the measurement. Since every self-adjoint operator A can be decomposed as in Eq. (1.12), every self-adjoint operator is an observable.

Hamiltonian as observable Postulate 1.2.3 on page 6 stated that the Hamiltonian $H(t)$ in the Schrödinger equation represents the total energy of the system. Indeed, since $H(t)$ is self-adjoint, it is a valid observable and eigenvectors of $H(t)$ evolve by the Schrödinger equation to eigenvectors of the same eigenvalue, thus conserving the eigenvalue.

General measurements The measurements so far can be reduced to partial measurements in a canonical basis if we can apply arbitrary unitary operations. However, more powerful measurements are possible if we can let the quantum system under consideration interact with another quantum system in a known initial state. Such a helper quantum system is often referred to as an *ancilla*. This leads to the most general quantum measurement:

1.2.4. Postulate. A quantum measurement is a family $\{M_x : x \in X\}$ of operators on the Hilbert space $\mathcal{H}$ such that

$$\sum_x M_x^* M_x = \mathbb{1} \ . \tag{1.13}$$

$x \in X$ are the labels that are output by the measurement process. The probability of obtaining outcome x when measuring state $|\psi\rangle$ is

$$\Pr[\text{outcome } x] = \frac{\langle\psi|M_x^* M_x|\psi\rangle}{\langle\psi|\psi\rangle} \tag{1.14}$$

and if the outcome x was obtained, the system after the measurement is in state

$$|\psi'\rangle = M_x|\psi\rangle \ . \tag{1.15}$$

This postulate encompasses our previous measurements: a projective measurement has $M_x = P_x$ and the complete measurement in the canonical basis has $M_x = |x\rangle\langle x|$. Conversely, we can implement a general measurement by a projective measurement on a larger state space: the mapping

$$|0\rangle|\psi\rangle \mapsto \sum_x |x\rangle M_x|\psi\rangle$$

preserves the inner product and therefore can be extended to a unitary mapping U. According to Eq. (1.9), the projective measurement $\{P_x\}$ with $P_x = U^*(|x\rangle\langle x| \otimes \mathbb{1})U$ will yield on $|0\rangle|\psi\rangle$ the same probabilities as $\{M_x\}$ on $|\psi\rangle$ by Eq. (1.14); furthermore, by Eq. (1.10), if outcome x was obtained in the projective measurement, the system is afterwards in the state

$$|\psi''\rangle = U^*(|x\rangle M_x|\psi\rangle)$$

so that an application of U will yield $|x\rangle M_x|\psi\rangle$, from which $M_x|\psi\rangle$ can be recovered by discarding the first quantum subsystem. Thus our simulation also obtains the final state from Eq. (1.15).

POVM The operators $E_x := M_x^* M_x$ are sufficient to compute the probability of outcome x in a general measurement:

$$\Pr[\text{outcome } x] = \frac{\langle\psi|E_x|\psi\rangle}{\langle\psi|\psi\rangle} \ .$$

Assuming there are no redundant $M_x = 0$, the operators E_x are *positive* and by Eq. (1.13), $\sum_x E_x = \mathbb{1}$. Conversely, every family $\{E_x\}$ of positive operators summing to $\mathbb{1}$ gives rise to a general measurement because for every positive T there exists an operator S so that $T = S^* S$. (In fact, in general there are many such "square roots" of T.) Such families are called *positive operator valued measures* (POVM). They are useful because they characterize all possible probability distributions $\Pr[\text{outcome } x \mid \text{state } |\psi\rangle]$ from general quantum measurements. The state after the measurement is "factored out" from this representation.

1.2.4 Entanglement

EPR pairs Consider the following state of two qubits

$$|\psi\rangle = |00\rangle + |11\rangle \ . \tag{1.16}$$

Note that the first 0 and the first 1 form the first qubit and the second 0 and the second 1 form the second qubit. This state is called an *EPR pair* after its inventors Einstein, Podolsky, and Rosen [49]. The purpose of this state was to devise a thought experiment to show the paradoxical implications of quantum mechanics. Imagine that we have this EPR state and that Alice has the first qubit somewhere on Mars and that Bob has the second, say, here on earth. If Alice measures her qubit she will see a 0 or a 1 with equal probability and the state will have collapsed to either $|00\rangle$, if she saw a 0 or $|11\rangle$ in case it was a 1. The same is true for Bob. This leads to the following situation. Suppose that the first qubit, on Mars, was measured first and that Alice saw a 1. This now means that when Bob measures his qubit he will also measure a 1. It appears that some information, i.e., the outcome of Alice's measurement, has somehow traveled to earth *instantaneously*. This appears to be in contradiction to the common belief that nothing can travel faster than the speed of light.

It turns out that EPR pairs cannot be used directly for communication. Hence, they do not violate relativistic causality and the notion that "no information can be transmitted faster than the speed of light." To show this, we introduce some tools from linear algebra that will be of use later on as well.

Density matrices Suppose we are given an *ensemble* of quantum states, i.e., a mixture of quantum states where state $|\psi_j\rangle$ occurs with probability p_j. For notational simplicity, we assume that $\langle\psi_j|\psi_j\rangle = 1$ for all j. For a fixed general measurement $\{M_x\}$, the probability to obtain outcome x on the ensemble is

$$\Pr[\text{outcome } x] = \sum_j p_j \langle\psi_j|M_x^* M_x|\psi_j\rangle = \operatorname{tr}\left(M_x\left(\sum_j p_j|\psi_j\rangle\langle\psi_j|\right)M_x^*\right) \tag{1.17}$$

where $\operatorname{tr} A = \sum_k A_{kk}$ denotes the *trace* of matrix A, which has the property that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ for every $n \times m$ matrix A and every $m \times n$ matrix B. If the measurement outcome is x, the state after the measurement will be $M_x|\psi_j\rangle$ with probability p_j. It is convenient to express the situation before the measurement by the *density matrix* ρ,

$$\rho := \sum_j p_j|\psi_j\rangle\langle\psi_j| \ .$$

Then the probability for outcome x on ensemble ρ is

$$\Pr[\text{outcome } x] = \operatorname{tr}\left(M_x\rho M_x^*\right)$$

and the density matrix after obtaining x is

$$\rho' = \frac{M_x\rho M_x^*}{\operatorname{tr}\left(M_x\rho M_x^*\right)} \ .$$

It is easy to see that density matrices are exactly the self-adjoint matrices that have trace 1. Observe that different mixtures can give rise to the same density matrix: the *completely mixed* state of one qubit,

$$\rho = \frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \ ,$$

is induced by the mixture $|0\rangle$ with probability 1/2 and $|1\rangle$ with probability 1/2, as well as by the mixture $|0\rangle + |1\rangle$ with probability 1/2 and $|0\rangle - |1\rangle$ with probability 1/2. In fact, any orthogonal basis of the state space with the uniform distribution gives rise to this completely mixed state. Remarkably, Eq. (1.17) implies that for a given measurement $\{M_x\}$, the measurement probabilities are independent on the particular way of obtaining ρ and this also holds for any subsequent operation on the ensemble state after the measurement. Therefore we call ρ a *mixed* state of the quantum system; this is in contrast to the *pure* states $|\psi\rangle$, which have density matrix $|\psi\rangle\langle\psi|$.

Reduced density matrices The joint state space of two quantum systems with state spaces $\mathcal{A}$ and $\mathcal{B}$ is $\mathcal{H} = \mathcal{A} \otimes \mathcal{B}$. Suppose Alice has the first part, with state space $\mathcal{A}$, and Bob has the second part, with state space $\mathcal{B}$. What can they locally find out about the global state? A pure state $|\psi\rangle \in \mathcal{H}$ has the general form

$$|\psi\rangle = \sum_j \sum_{j'} \alpha_{j,j'} |j\rangle |j\rangle'$$

where the j range over a basis of $\mathcal{A}$ and the j' range over a basis of $\mathcal{B}$. By the *Schmidt decomposition* theorem, there exist for $|\psi\rangle$ two orthonormal families $\{\psi_{A,j}\} \subset \mathcal{A}$ and $\{\psi_{B,j}\} \subset \mathcal{B}$ so that

$$|\psi\rangle = \sum_j \alpha_j |\psi_{A,j}\rangle |\psi_{B,j}\rangle$$

where $\alpha_j \in \mathbb{R}$ and $\alpha_j \geq 0$ for all j. The corresponding density matrix is

$$|\psi\rangle\langle\psi| = \sum_{j,j'} \alpha_j \alpha^*_{j'} |\psi_{A,j}\rangle\langle\psi_{A,j'}| \otimes |\psi_{B,j}\rangle\langle\psi_{B,j'}| \ . \tag{1.18}$$

We define the *reduced* density matrix of part $\mathcal{A}$ by "tracing out" the subsystem $\mathcal{B}$, i.e., we replace in Eq. (1.18) each operator $|\psi_{B,j}\rangle\langle\psi_{B,j'}|$ by the complex number $\operatorname{tr}(|\psi_{B,j}\rangle\langle\psi_{B,j'}|)$:

$$\rho_A := \operatorname{tr}_B |\psi\rangle\langle\psi| := \sum_{j,j'} \alpha_j \alpha^*_{j'} \operatorname{tr}\left(|\psi_{B,j}\rangle\langle\psi_{B,j'}|\right) |\psi_{A,j}\rangle\langle\psi_{A,j}| \ .$$

tr_B is called the *partial trace* function. ρ_A is a density matrix over $\mathcal{A}$; ρ_A contains all the information that Alice can obtain about the global state without communicating with Bob. More precisely, every measurement $\{M_x\}$ of Alice corresponds to the global measurement $\{M_x \otimes \mathbb{1}_\mathcal{B}\}$, which has probabilities

$$\begin{aligned}
\Pr[\text{outcome } x] &= \operatorname{tr}\left((M_x \otimes \mathbb{1}_\mathcal{B}) |\psi\rangle\langle\psi| (M^*_x \otimes \mathbb{1}_\mathcal{B})\right) \\
&= \operatorname{tr}\left(\sum_{j,j'} \alpha_j \alpha^*_{j'} M_x |\psi_{A,j}\rangle\langle\psi_{A,j'}| M^*_x \otimes |\psi_{B,j}\rangle\langle\psi_{B,j'}|\right) \\
&= \operatorname{tr}\left(\sum_{j,j'} \alpha_j \alpha^*_{j'} \operatorname{tr}\left(|\psi_{B,j}\rangle\langle\psi_{B,j'}|\right) M_x |\psi_{A,j}\rangle\langle\psi_{A,j'}| M^*_x\right) \\
&= \operatorname{tr}\left(M_x \rho_A M^*_x\right) \ .
\end{aligned}$$

The state after the measurement is $(M_x \otimes \mathbb{1}_\mathcal{B})|\psi\rangle$, which has the reduced density matrix $\rho'_A = M_x \rho_A M^*_x / \operatorname{tr}(M_x \rho_A M^*_x)$.

How does the measurement affect Bob's reduced density matrix $\rho_B := \operatorname{tr}_A |\psi\rangle\langle\psi|$ if he does not learn the outcome of the measurement? It becomes

$$\begin{aligned}
\rho'_B &:= \operatorname{tr}_A \left(\sum_x \Pr[\text{outcome } x] \frac{(M_x \otimes \mathbb{1}_{\mathcal{B}})\,|\psi\rangle\langle\psi|\,(M_x^* \otimes \mathbb{1}_{\mathcal{B}})}{\operatorname{tr}\left((M_x \otimes \mathbb{1}_{\mathcal{B}})\,|\psi\rangle\langle\psi|\,(M_x^* \otimes \mathbb{1}_{\mathcal{B}})\right)} \right) \\
&= \operatorname{tr}_A \left(\sum_x (M_x \otimes \mathbb{1}_{\mathcal{B}})\,|\psi\rangle\langle\psi|\,(M_x^* \otimes \mathbb{1}_{\mathcal{B}}) \right) \\
&= \operatorname{tr}_A \left(\sum_{j,j'} \alpha_j \alpha_{j'}^* \left(\sum_x \operatorname{tr}\left(M_x |\psi_{A,j}\rangle\langle\psi_{A,j'}| M_x^*\right) \right) |\psi_{B,j}\rangle\langle\psi_{B,j'}| \right) \\
&= \operatorname{tr}_A \left(\sum_{j,j'} \alpha_j \alpha_{j'}^* \operatorname{tr}\left(\langle\psi_{A,j'}| \sum_x M_x M_x^* |\psi_{A,j}\rangle \right) |\psi_{B,j}\rangle\langle\psi_{B,j'}| \right) \\
&= \rho_B \ .
\end{aligned}$$

Hence, any measurement on Alice's side leaves Bob's reduced density matrix unchanged.

If a unitary operator $U_{\mathcal{A}}$ is applied to Alice's subsystem, the global state $|\psi\rangle$ evolves to $(U_{\mathcal{A}} \otimes \mathbb{1}_{\mathcal{B}})|\psi\rangle$ and the reduced density matrix of Alice evolves from ρ_A to $U_{\mathcal{A}} \rho_A U_{\mathcal{A}}^*$. As for measurements on Alice's side, applying $U_{\mathcal{A}}$ on $\mathcal{A}$ leaves Bob's reduced density matrix unchanged.

By linearity, all these considerations extend to global states that are mixed rather than pure.

Nonlocality Let us now resume the discussion of EPR pairs. We claimed that EPR pairs cannot be used for communication. In the preceding paragraph we established that measurements and unitary operations on Alice's side leave Bob's reduced density matrix unchanged. Furthermore, we showed that all information that Bob can extract from the global state is represented by his reduced density matrix. Hence, no communication is possible. However, EPR pairs do have many strange properties and interesting applications as we will see. Perhaps the most simple instance is when we consider *correlations* between measurement outcomes of Alice and Bob. In other scenarios Alice and Bob share EPR pairs and also can communicate.

The state from Eq. (1.16) is *entangled*: there are no qubit states $|\psi_A\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ and $|\psi_B\rangle = \beta_0|0\rangle + \beta_1|1\rangle$ so that $|\psi\rangle = |\psi_A\rangle \otimes |\psi_B\rangle$. The reason for this is that

$$\begin{aligned}|\psi_A\rangle \otimes |\psi_B\rangle &= (\alpha_0|0\rangle + \alpha_1|1\rangle) \otimes (\beta_0|0\rangle + \beta_1|1\rangle)\\ &= \alpha_0\beta_0|00\rangle + \alpha_0\beta_1|01\rangle + \alpha_1\beta_0|10\rangle + \alpha_1\beta_1|11\rangle\\ &= |00\rangle + |11\rangle\end{aligned}$$

imposes unsatisfiable constraints on the complex amplitudes α_0, α_1, β_0, β_1, because $\alpha_0\beta_0 = \alpha_1\beta_1 = 1$ and hence $\alpha_0\beta_1 \neq 0$ and $\alpha_1\beta_0 \neq 0$.

At first glance, entanglement appears to have no "nonclassical" observable consequences—Einstein, Rosen, and Podolsky's objections to quantum mechanics were largely philosophical in that they questioned the reality of a state that is undetermined between being two zeroes or two ones. However, Bell [17] proved that there can be no classical equivalent of the correlations when Alice and Bob measure an EPR state in bases chosen locally at random from a predetermined set of different measurements. There has been a lot of theoretical and experimental follow-up on Bell's seminal paper, and in Chapter 5 we present our results in improving experiments for detecting nonlocality with imperfect apparatus in the laboratory. This is an important goal since so far there is no nonlocality experiment that simultaneously rules out all plausible classical "loopholes."

1.2.5 Perspective

In our exposition of quantum mechanics, we glossed over several theoretical aspects that are important in concrete physical modelling but which are not essential for the present work. Among these are additional postulates about indistinguishable particles—two photons cannot be told apart and therefore the theory should not attribute to them separate identities. This restricts the state space to subspaces invariant under certain permutations of the particles. The details depend on the concrete *particle statistics*. Another limitation is that in reality, one observes that certain quantities cannot ever exist in superposition—quantum mechanics has *superselection rules* for this. Both particle statistics and superselection rules make quantum mechanics less powerful and therefore will not confer extra computing power to quantum computers. For cryptography and fault tolerance, however, the restrictions may be of interest.

We considered only state spaces of finite dimension. The presented mathematical tools can be extended, sometimes with considerable mathematical effort, to a continuum of dimensions. This is necessary, e.g., for position or momentum observables, which should have arbitrary real values. However, for our applications, the finite dimensionality is sufficient.

Can we expect that new physical theories will allow computers that are even more powerful than quantum computers? While this cannot be ruled out, we note that on one hand, there is to date no convincing experimental data that contradicts quantum mechanics and that indicates directions in which it needs amending. On the other hand, quantum mechanics and general relativity are incompatible, but no single theory that merges them has yet gained widespread acceptance. Hence considering the computing power of such theories is regarded as premature or esoteric.

1.3 Quantum Computation and Information

1.3.1 Quantum circuits

In theoretical computer science, the most common models of computation for computability and complexity analysis are the *Turing machine* and *circuits*. The corresponding models of computation motivated by quantum mechanics are the *quantum Turing machine* and *quantum circuits*. Here we focus on quantum circuits since they have a simple description in terms of small unitary matrices and they are closer to implementation.

Classical circuits A classical Boolean circuit is a directed acyclic graph whose vertices are called *gates* and the edges are the *wires* transmitting bits. A gate has zero or more labeled input bits and zero or more output bits. The logical connectives $\wedge$ (and), $\vee$ (or), and $\oplus$ (exclusive or) are represented by And, Or, and Xor gates, respectively, with $k \geq 2$ inputs and 1 output; the logical not $\neg$ is represented by a gate with one input and one output. Designated gates with one input and no output are output gates; input gates are labeled with a Boolean variable $x_j \in \{0,1\}$, have no input and one output. If there is a single output gate, the circuit computes a Boolean function of the input variables $x_1, \ldots, x_n$; if there are more output gates, the circuit simultaneously computes several Boolean functions. See Figure 1.2 for an example. Further details about classical circuits can be found in textbooks on complexity theory, e.g., in the book by Papadimitriou [96].

Circuit complexity A circuit computes a Boolean function $f : \{0,1\}^n \rightarrow \{0,1\}^m$. When we fix a set of permissible gates, we can ask how many gates are needed to realize a given Boolean function. Thus we obtain a notion of complexity of f under the given gate constraints. Gate families such as $\{\text{And}, \text{Not}\}$ and $\{\text{Nand}\}$ are *universal* in the sense that every Boolean function has a circuit using only gates from those families.

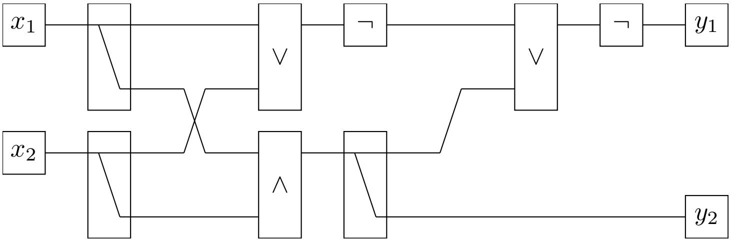


Figure 1.2: A classical circuit. Using the input gates x_1 and x_2, FanOut, And, Or, and Not gates, this circuit computes outputs $y_1 = x_1 \oplus x_2$ and $y_2 = x_1 \wedge x_2$. All edges are assumed to be oriented from the left to right.

x_1	x_2	x_3	y_1	y_2	y_3
0	0	0	0	0	0
0	0	1	0	0	1
0	1	0	0	1	0
0	1	1	0	1	1
1	0	0	1	0	0
1	0	1	1	0	1
1	1	0	1	1	1
1	1	1	1	1	0

Figure 1.3: The CCNot gate maps the inputs x_1, x_2, and x_3 to $y_1 = x_1$, $y_2 = x_2$, and $y_3 = x_3 \oplus (x_1 \wedge x_2)$.

Reversible gates Quantum evolution is unitary and therefore reversible. Our first step towards quantum circuits are classical reversible circuits. Here gates have as many outputs as they have inputs and they are one-to-one functions. This precludes the use of FanOut, And, and Or gates. While this may appear prohibitive, there are well-known constructions to convert each classical circuit into a reversible classical circuit at little overhead. One way is to use the CCNot or *Toffoli* gate [60], which has the three inputs and three outputs with the truth table in Figure 1.3. Taking x_1 and x_2 as control lines, this is a Not operation on the third input conditional on the two first inputs being one, hence the name CCNot for "controlled controlled not." Since $y_3 = x_3 \oplus (x_1 \wedge x_2)$ the CCNot computes the And of x_1 and x_2 if $x_3 = 0$; so each And gate can be simulated using a zero bit and a CCNot gate. Fanout can be implemented similarly: if $x_2 = 1$ and $x_3 = 0$, then for every value of x_1, CCNot outputs $y_1 = y_3 = x_1$ and $y_2 = 1$. In both cases, the constant bits can be "recycled" so that only a constant factor in overhead is incurred.

Quantum gates A natural way to define quantum gates is to let them be unitary transformations. Then every classical reversible gate is a quantum gate. Bounds on fanin and fanout translate to bounds on the number of qubits on which these unitary transformations may act.

Universality In classical circuits, we call a set of gates universal if there are circuits using only those gates for every Boolean function. Similarly, there are sets of quantum gates that approximate every unitary transformation arbitrarily well.

Turing machines are universal: there are Turing machines that take as input a Turing-machine program p and an input x and that simulate with polynomial overhead the operation of p on x. The *strong Church-Turing thesis* states that every "realistic" model of computation is polynomially equivalent to probabilistic Turing machines, i.e., it can be simulated with polynomial overhead. Here "realistic" is a vague term alluding to the possibility to physically implement an arbitrarily long but finite computation in a model of computation in real time proportional to the time complexity of the computation in the model. The vagueness of this formulation leads to the belief that the strong Church-Turing thesis cannot be proved formally.

A circuit only operates on inputs of a fixed length. To compare the computational power of circuits to Turing machines, we have to consider families of circuits that contain one circuit for each input length. Moreover, we need to require that these circuits do not differ too much. *Uniform* circuit families are those for which there exists a Turing machine that on input n produces as output the circuit for input length n in time polynomial in n. It is not hard to see that such uniform families of classical circuits are polynomially equivalent to Turing machines. Since a reversible classical circuit is also a quantum circuit, uniform quantum circuits are polynomially at least as powerful as Turing machines and by the strong Church-Turing thesis Turing complete. However, one of the motivations for quantum computing is the conjecture that the strong Church-Turing thesis does not hold for quantum computers in the sense that quantum computers may be a realistic computational model that cannot be simulated efficiently with classical computers.

1.3.2 Quantum black-box algorithms

The overwhelming majority of results about the complexity of problems on quantum computers are in the *black-box* model, where the input gates are replaced by *oracle* gates giving random access to bits of the input. Instead of time, depth of the circuit, or total number of gates, the complexity measure is the number of queries to bits of the input. A large part of the power of quantum computing is captured by this simple model of query complexity;

the existing quantum algorithms all make far fewer such input queries than classical algorithms for the same problems.

Quantum query For $N = 2^n$, a *quantum query* or *quantum oracle gate* for a Boolean function $f : \{0,1\}^n \to \{0,1\}$ is a unitary operator U_f on $\mathbb{C}^N \otimes \mathbb{C}^2$ that operates on basis states like a reversible classical gate for computing f. In particular,

$$U_f : |j\rangle|0\rangle \mapsto |j\rangle|f(j)\rangle$$

puts the value of $f(j)$ into the second register. By requiring

$$U_f : |j\rangle|1\rangle \mapsto |j\rangle|1 - f(j)\rangle$$

we turn U_f into a reversible classical gate on states in the computational basis. From linear algebra it follows that a linear function is uniquely defined by its values on a basis, so this fixes U_f.

In the computational basis, U_f is a permutation matrix, i.e., it has exactly one 1 in each row and column and 0s elsewhere. This may not appear to be a very exciting operation, but since we can run it on a superposition of indices, say, $\sum_{j=1}^N |j\rangle|0\rangle$, we can actually query all entries of the database at once! Unfortunately, measuring the resulting state

$$U_f\left(\sum_{j=1}^N |j\rangle|0\rangle\right) = \sum_{j=1}^N |j\rangle|f(j)\rangle$$

gives us each $|j\rangle|f(j)\rangle$ with probability $1/N$, hardly an improvement over the classical case. It takes a little more effort to uncover the quantum advantage.

The Deutsch-Jozsa algorithm Consider the following toy problem: given $f : \{1,\ldots,N\} \to \{0,1\}$ where f is either constant or *balanced* in the sense that it takes the value $f(j) = 0$ for exactly as many indices j as it does for $f(j) = 1$; find out whether f is constant or balanced. This problem was thought up by Deutsch and Jozsa in 1992 [48] and they gave an ingenuous solution using a quantum computer, which foreshadowed many future quantum algorithms. The quantum circuit depicted in Figure 1.4 operates for $N = 2^n$ as follows: the initial state $|\psi_0\rangle := |0^n\rangle|1\rangle$ is mapped by Hadamard transformations on each of the $n+1$ qubits to

$$\begin{aligned} |\psi_1\rangle &:= H^{\otimes n+1}|\psi_0\rangle = H^{\otimes n+1}\left(|0^n\rangle|1\rangle\right) \\ &= \left[\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)\right]^{\otimes n} \otimes \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \\ &= \frac{1}{\sqrt{2^{n+1}}} \sum_{x\in\{0,1\}^n} |x\rangle(|0\rangle - |1\rangle) \ . \end{aligned} \tag{1.19}$$

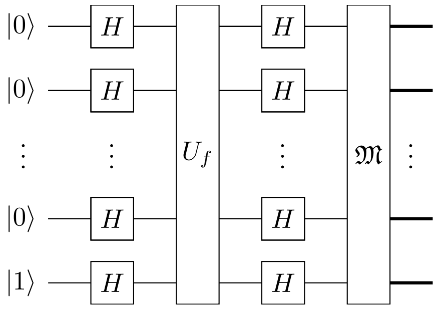


Figure 1.4: The Deutsch-Jozsa algorithm

For the next step, observe that

$$\begin{aligned} U_f\left(|x\rangle\left(|0\rangle-|1\rangle\right)\right) &= U_f\left(|x\rangle|0\rangle\right)-U_f\left(|x\rangle|1\rangle\right) \\ &= |x\rangle|f(x)\rangle-|x\rangle|1-f(x)\rangle \\ &= (-1)^{f(x)}|x\rangle\left(|0\rangle-|1\rangle\right) \quad , \end{aligned}$$

i.e., applying a quantum query to f on state $|x\rangle\left(|0\rangle-|1\rangle\right)$ leaves the state unchanged except for a phase factor $(-1)^{f(x)}$ that depends on $f(x)$. Hence, the next step of the circuit in Figure 1.4 maps $|\psi_1\rangle$ to

$$|\psi_2\rangle := \frac{1}{\sqrt{2^{n+1}}}\sum_{x\in\{0,1\}^n}(-1)^{f(x)}|x\rangle\left(|0\rangle-|1\rangle\right) \quad . \tag{1.20}$$

The final state before the measurement is

$$\begin{aligned} |\psi_3\rangle &:= H^{\otimes n+1}|\psi_2\rangle \\ &= \frac{1}{\sqrt{2^{n+1}}}\sum_{x\in\{0,1\}^n}(-1)^{f(x)}H^{\otimes n+1}\left(|x\rangle\left(|0\rangle-|1\rangle\right)\right) \quad . \end{aligned} \tag{1.21}$$

In order to analyze this expression, note that for $x\in\{0,1\}^n$

$$\begin{aligned} H^{\otimes n}|x\rangle &= H|x_1\rangle\otimes H|x_2\rangle\otimes\cdots\otimes H|x_n\rangle \\ &= \frac{1}{\sqrt{2^n}}\left(|0\rangle+(-1)^{x_1}|1\rangle\right)\otimes\cdots\otimes\left(|0\rangle+(-1)^{x_n}|1\rangle\right) \\ &= \frac{1}{\sqrt{2^n}}\sum_{y\in\{0,1\}^n}(-1)^{x\cdot y}|y\rangle \quad , \end{aligned}$$

where

$$x\cdot y := \sum_{j=1}^{n}x_jy_j \mod 2$$

denotes the *inner product modulo 2* of the binary vectors $x, y \in \mathbb{Z}_2^n$. Substituting this into Eq. (1.21) yields

$$|\psi_3\rangle = \frac{1}{2^n} \sum_{x\in\{0,1\}^n} \sum_{y\in\{0,1\}^n} (-1)^{f(x)+x\cdot y} |y\rangle|1\rangle \ . \tag{1.22}$$

This expression may appear unwieldy, but now we can bring the special structure of f into play. Let us consider the terms of the above sum where $y = 0^n$. Then

$$\begin{aligned}\sum_{x\in\{0,1\}^n} (-1)^{f(x)+x\cdot y} &= \sum_{x\in\{0,1\}^n} (-1)^{f(x)} \\ = |\{x \in \{0,1\}^n : f(x) = 0\}| &- |\{x \in \{0,1\}^n : f(x) = 1\}| \ .\end{aligned}$$

Hence, if f is balanced, measuring $|\psi_3\rangle$ will never yield outcome 0^n1. Conversely, if f is constant, the amplitude of $|0^n\rangle|1\rangle$ in $|\psi_3\rangle$ is 1 or -1 and since our analysis started with a state of norm 1 and we applied only unitary transformations, all $|y\rangle|1\rangle$ with $y \neq 0^n$ must have amplitude 0 so $|\psi_3\rangle = \pm|0\rangle^n|1\rangle$. In other words, if f is constant, then measuring $|\psi_3\rangle$ will yield outcome 0^n1 with certainty, whereas if f is balanced, this outcome has probability 0.

One quantum query thus suffices to distinguish the balanced from the constant case with certainty. Classically, a deterministic algorithm will need $N/2 + 1 = 2^{n-1} + 1$ queries in the worst case: for any sequence of fewer query positions, there exist both constant and balanced functions that are consistent with all queries having answer, say, 0.

Separations Probabilistically, however, the Deutsch-Jozsa problem can be solved classically with great efficiency merely by sampling f in a constant number of places. Stronger separations between classical and quantum query complexity were obtained by Bernstein and Vazirani [22] and Simon [109]. In terms of the domain size N of the input function f, they give separations of $\mathrm{O}(1)$ versus $\Omega(\log N)$ and, as strengthened in [26], $\mathrm{O}(\log N)$ versus $\Omega(\sqrt{N})$, respectively, for classical randomized versus quantum exact query complexity. Since these problems serve us in Chapter 3 as a point of departure for separations in property testing, we will present them in detail there. The best separations to date are 1 versus $\Omega(\sqrt{N})$ [16].

All exponential separations are for partial problems—the input functions f are constrained by a *promise* such as "f is constant or balanced" for the Deutsch-Jozsa problem. This is no accident; Beals, Buhrman, Cleve, Mosca, and de Wolf [15] proved that for *total* problems, the gap between classical and quantum bound-error complexity is at most polynomial. Our considerations about quantum property testing in Chapter 3 can be seen as investigations into what kind of generic promises still yield strong separations between the classical and quantum mechanical models of computation.

1.3.3 Hallmark results

Factoring Quantum computing first got widespread attention with Shor's 1994 discovery of a polynomial-time quantum algorithm for factoring large integers. This was the first arguably useful[2] task where quantum computing appears to beat classical computers. Today's public-key cryptography like RSA [101] relies on the assumption that factoring or related problems such as the discrete logarithm cannot be performed efficiently. This belief is founded on the fact that after many years of intense research, the best published algorithms for these problems have superpolynomial running time in the length n of the input, e.g., $2^{\log n^{\alpha}}$ for some constant α [81, 82].

Shor's approach was to use a classical reduction from factoring to finding the period of a certain class of functions. Using the efficient *quantum Fourier transform* algorithm, he then devised a way to obtain the period. The quantum query complexity of the period-finding subproblem is provably exponentially smaller than the classical query complexity [42]. However, factoring itself is a problem whose time complexity is in the gray zone between NP-hardness and P. Other families of problems that have so far eluded efficient quantum algorithms are the class SZK = "statistical zero-knowledge," notably graph nonisomorphism, and the problem of constructing solutions to problems where each instance is guaranteed to have a solution; this is the class TFNP = "total function NP."

Quantum search Who answers the phone at 736-5000? Telephone directories are ordered alphabetically by name, therefore using a telephone directory to find a number from a name amounts to going through the names one by one. For N entries, looking up a phone number for a given name can be done in $\mathrm{O}(\log N)$ steps using *binary search* whereas search in an unordered list takes $\Omega(N)$ lookups on average, even with randomization.

Surprisingly, one can do much better on a quantum computer. This was shown by Grover [69] who in 1996 gave a quantum algorithm for *unordered search* that finds the solution with high probability using $\mathrm{O}(\sqrt{N})$ quantum queries. Moreover, this algorithm can be generalized to an amplification procedure for quantum algorithms that can be represented as a unitary transformation. In Chapter 2 we will review the basic search algorithm and its

[2] Although one might reason that the existence of any factoring device would lead to instant abolishment of any scheme that assumes that factoring is infeasible. Hence, a single quantum computer would suffice and it would not even be necessary to operate it.

generalization to *amplitude amplification* before we present applications and modifications of the quantum-search paradigm.

The power of quantum computing As introduced in this chapter, quantum computers are a physically plausible computational model with the same notion of computability as classical computers but with potentially greater efficiency. This is in line with the *Church-Turing thesis*—that all powerful but realistic computational models are equivalent in terms of what can be computed—but possibly contradicts the so-called *strong Church-Turing thesis*, namely that even what is *efficiently* computable is the same in all sensible computational models. Here the complexity measure is general time complexity and efficient means within a polynomial time bound. We saw that in restricted models like the black-box model, sharp separations can be proved but those separations lead at best to indirect implications for the general question.

Part I

Quantum Query Complexity

Chapter 2

Quantum Search

In this chapter we present research inspired by Grover's seminal quantum search algorithm [69]. In Section 2.1 we review the basic search algorithm and its generalization to *amplitude amplification*. An application for computing convolution products is proposed in Section 2.2. In Section 2.3 we express the iteration of the search algorithm in terms of density matrices, so that we can analyze its performance in the presence of decoherence. Nonclassical databases are the point of departure for the considerations in Section 2.4, where we derive algorithms to compare the degeneracy of energy levels of a given Hamiltonian. Section 2.4 is based on joint work with Ozhigov [94]; Sections 2.2 and 2.3 are unpublished so far.

2.1 Quantum Amplitude Amplification

2.1.1 Grover's algorithm

Unordered search is the problem of finding a database entry matching the search criteria merely by using queries of the type "does entry j match?" An example is finding a name in a telephone directory given a phone number. The telephone directory is ordered by name and the phone numbers are practically random. It is easy to see that classically, even with randomization, $\Omega(N)$ queries are required on average in an N-entry telephone directory.

Database query The algorithm makes use of the function $f : \{1, \ldots, N\} \to \{0, 1\}$, where $f(j) = 1$ if and only if j is the index we are looking for, i.e., *PhoneNumber*$(j) =$ 736-5000. In the following we assume that $N = 2^n$ for some $n \in \mathbb{N}$ and we identify the domain of f with $\{0, 1\}^n$; since the input is unordered, there is no structure to be respected. Recall that in Subsection 1.3.2

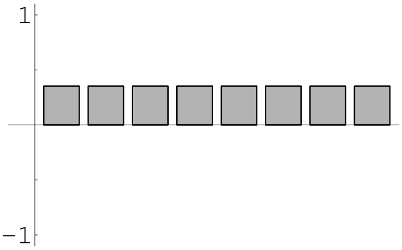


Figure 2.1: The initial state of quantum search for $N = 8$ and $f(j) = \delta_{j,3}$. The bars give the amplitudes α_j of the state $\sum_j \alpha_j |j\rangle|-\rangle$.

we defined a quantum query to f as the unitary transformation

$$U_f : |j\rangle|b\rangle \mapsto |j\rangle|f(j) \oplus b\rangle \ .$$

The idea of quantum search is to start with a *uniform superposition* of indices

$$|\psi_0\rangle := \frac{1}{\sqrt{N}} \sum_{j \in \{0,1\}^n} |j\rangle \ ,$$

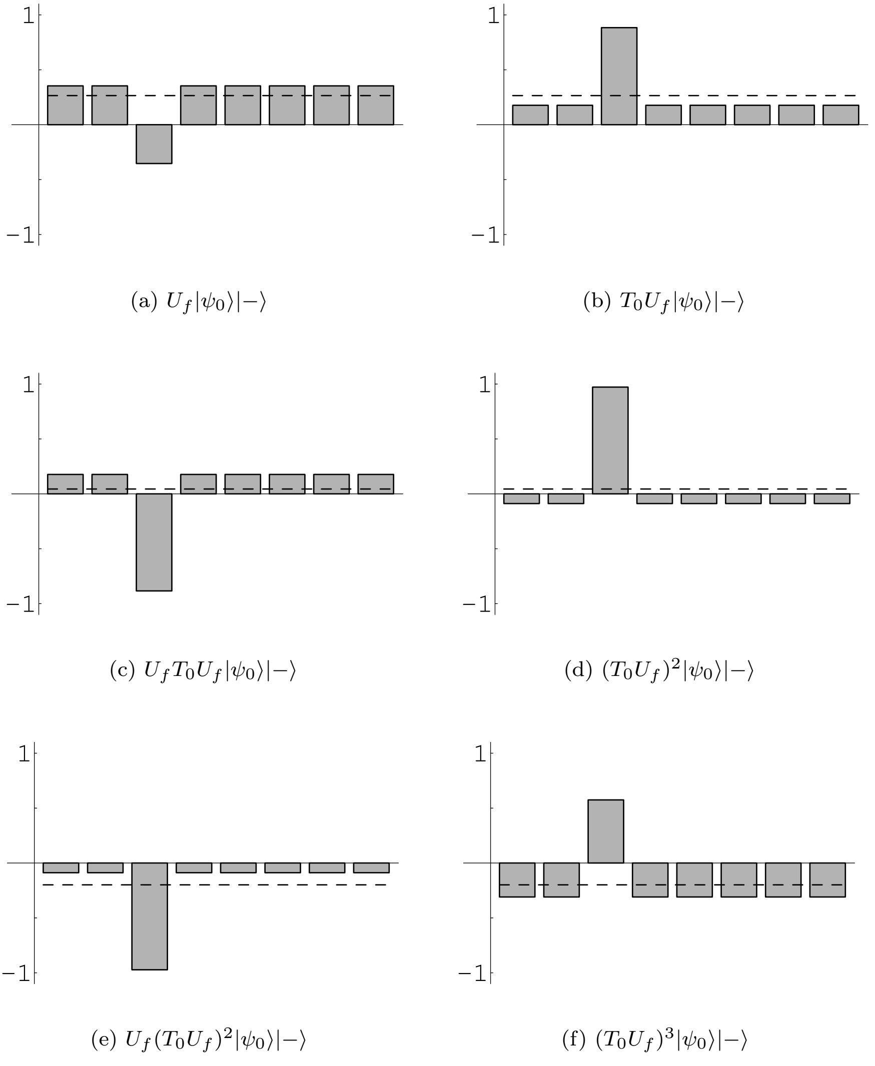
representing the initial knowledge about the j with $f(j) = 1$ and to progressively "transfer" amplitude from basis states $|j'\rangle$ with $f(j') = 0$ to basis states $|j\rangle$ with $f(j) = 1$. The operations have to be unitary and what counts is how often the query gate U_f is invoked. In $|\psi_0\rangle$ and throughout the quantum-search algorithm, the amplitudes of the basis vectors are real and therefore we can represent them as a bar chart like the example in Figure 2.1. We saw in Subsection 1.3.2 that by initializing the last qubit to $|-\rangle := (|0\rangle - |1\rangle)/\sqrt{2}$ we can realize the mapping

$$|j\rangle|-\rangle \mapsto (-1)^{f(j)}|j\rangle|-\rangle$$

using one invocation of U_f. This flips the amplitude of the $|j\rangle$ with $f(j) = 1$ from $1/\sqrt{N}$ to $-1/\sqrt{N}$.

Reflection about the average This substantial change in phase can be translated to a change in absolute value by performing a *reflection about the average* operation as outlined in the step from Figure 2.2(a) to Figure 2.2(b). On input $|\psi\rangle = \sum_j \alpha_j|j\rangle$, it maps each individual amplitude α_j to $\tilde{\alpha} - (\alpha_j - \tilde{\alpha}) = 2\tilde{\alpha} - \alpha_j$ where $\tilde{\alpha} := (1/N)\sum_j \alpha_j$ is the average of the α_j and $(\alpha_j - \tilde{\alpha})$ is the deviation of α_j from the average. It turns out that this operation is unitary and can be implemented efficiently without any U_f gate;

$$T_0 := -WS_0W$$



(a) $U_f|\psi_0\rangle|-\rangle$

(b) $T_0U_f|\psi_0\rangle|-\rangle$

(c) $U_fT_0U_f|\psi_0\rangle|-\rangle$

(d) $(T_0U_f)^2|\psi_0\rangle|-\rangle$

(e) $U_f(T_0U_f)^2|\psi_0\rangle|-\rangle$

(f) $(T_0U_f)^3|\psi_0\rangle|-\rangle$

Figure 2.2: The first iterations of quantum search ($N = 8$ and $f(j) = \delta_{j,3}$). The bars give the amplitudes α_j of the state $\sum_j \alpha_j|j\rangle|-\rangle$; the dashed line indicates the average.

achieves the desired result. Here $W := H^{\otimes n}$ denotes a Hadamard transform on all n qubits individually and

$$S_0 := \mathbb{1} - 2|0\rangle\langle 0| = \begin{pmatrix} -1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}$$

changes the phase of the $|0^n\rangle$ basis state by a factor of -1, leaving all other basis states unchanged. To see that T_0 implements the reflection about the average, note that

$$\begin{aligned} T_0|\psi\rangle &= -\left(\mathbb{1} + 2W|0\rangle\langle 0|W\right)|\psi\rangle \\ &= -\sum_j \alpha_j|j\rangle + 2\left(\frac{1}{\sqrt{N}}\sum_k |k\rangle\right)\left(\frac{1}{\sqrt{N}}\sum_\ell \langle \ell|\right)\sum_m \alpha_m|m\rangle \\ &= \sum_j \left(\frac{2}{N}\left(\sum_m \alpha_m\right) - \alpha_j\right)|j\rangle = \sum_j \left(2\tilde{\alpha} - \alpha_j\right)|j\rangle \end{aligned} \quad . \qquad (2.1)$$

What is the gain in amplitude? For a single j with $f(j) = 1$, the amplitude $\alpha_j = -1/\sqrt{N}$ is mapped to

$$\frac{2}{N}\left((N-1)\frac{1}{\sqrt{N}} - \frac{1}{\sqrt{N}}\right) - \left(-\frac{1}{\sqrt{N}}\right) > \frac{2}{\sqrt{N}} \ .$$

Hence, the amplitude of basis state $|j\rangle$ increased by an additive term of more than $1/\sqrt{N}$.

So far, we prepared the uniform superposition, performed one query and the “reflection about the average” operation; this corresponds to the unitary operator

$$G := (T_0 \otimes \mathbb{1})\, U_f$$

applied to the initial state

$$|\psi_0\rangle := (W \otimes H)\,|0^n 1\rangle = \frac{1}{\sqrt{N}}\sum_j |j\rangle \frac{1}{\sqrt{2}}\left(|0\rangle - |1\rangle\right) \ . \qquad (2.2)$$

One application of G improves our success probability. It is natural to ask whether repeating G is helpful; an oblivious phase-flip followed by the reflection operation should boost the amplitude of the states $|j\rangle$ with $f(j) = 1$. Indeed, these iterations are at the heart of Grover’s algorithm. It remains to determine a judicious number of repetitions r so that when measuring $G^r|\psi_0\rangle$ the probability of observing j with $f(j) = 1$ is large.

Two-dimensional evolution The observation that an iteration of the algorithm treats basis states $|j\rangle$ with the same value of f the same leads to an elegant way to analyze the behavior of the algorithm [24]. Let $M := |\{j : f(j) = 1\}|$ denote the number of solutions and

$$|\chi\rangle := \frac{1}{\sqrt{M}} \sum_{j:f(j)=1} |j\rangle H|1\rangle \quad \text{and} \quad |\chi^{\perp}\rangle := \frac{1}{\sqrt{N-M}} \sum_{j:f(j)=0} |j\rangle H|1\rangle \tag{2.3}$$

the uniform superposition of "good" and "bad" basis states, respectively. The initial state from Equation (2.2) is a superposition of those states:

$$|\psi_0\rangle = \sqrt{\frac{M}{N}}|\chi\rangle + \sqrt{\frac{N-M}{N}}|\chi^{\perp}\rangle \ . \tag{2.4}$$

From Equation (2.1) we obtain

$$G|\chi\rangle = -T_0|\chi\rangle = \left(1 - \frac{2M}{N}\right) \qquad |\chi\rangle - \frac{2\sqrt{M(N-M)}}{N} \qquad |\chi^{\perp}\rangle$$

and

$$G|\chi^{\perp}\rangle = T_0|\chi^{\perp}\rangle = \frac{2\sqrt{M(N-M)}}{N}|\chi\rangle + \left(-1 + 2\frac{N-M}{N}\right)|\chi^{\perp}\rangle \ .$$

Hence, one iteration G can be expressed as a mapping in the two-dimensional subspace spanned by $|\chi\rangle$ and $|\chi^{\perp}\rangle$. For $|\psi\rangle = \alpha|\chi\rangle + \beta|\chi^{\perp}\rangle$, $\alpha, \beta \in \mathbb{C}$, we get

$$G|\psi\rangle = \begin{pmatrix}|\chi\rangle & |\chi^{\perp}\rangle\end{pmatrix} \hat{G} \begin{pmatrix}\alpha \\ \beta\end{pmatrix}$$

where the first matrix product on the right-hand side is to be interpreted formally as $\begin{pmatrix}|\chi\rangle & |\chi^{\perp}\rangle\end{pmatrix} \begin{pmatrix}\alpha' & \beta'\end{pmatrix}^T = \alpha'|\chi\rangle + \beta'|\chi^{\perp}\rangle$ and $\hat{G}$ is the two-dimensional version of G,

$$\hat{G} = \frac{1}{N} \begin{pmatrix} N-2M & 2\sqrt{M(N-M)} \\ -2\sqrt{M(N-M)} & N-2M \end{pmatrix} \ .$$

We are interested in $\hat{G}^r$, which describes the effect of r iterations of G in the two-dimensional subspace spanned by $|\chi\rangle$ and $|\chi^{\perp}\rangle$. $\hat{G}$ is a real unitary matrix, therefore it is a rotation in the real plane, possibly combined with a reflection. Choosing the smallest $\vartheta \geq 0$ such that $\cos\vartheta = (N-2M)/N$,

$$\hat{G} = \begin{pmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{pmatrix} \quad \text{and therefore} \quad \hat{G}^r = \begin{pmatrix} \cos(r\vartheta) & \sin(r\vartheta) \\ -\sin(r\vartheta) & \cos(r\vartheta) \end{pmatrix} \ .$$

Using the same substitution and the observation that $1+\cos^2(\vartheta/2) = 2\cos\vartheta$, the initial state from Equation (2.4) becomes

$$|\psi_0\rangle = \sin(\vartheta/2)|\chi\rangle + \cos(\vartheta/2)|\chi^\perp\rangle = \begin{pmatrix}|\chi\rangle & |\chi^\perp\rangle\end{pmatrix}\begin{pmatrix}\sin(\vartheta/2)\\ \cos(\vartheta/2)\end{pmatrix} \tag{2.5}$$

The probability of obtaining a measurement outcome j with $f(j) = 1$ after r iterations is

$$\begin{aligned}\left\|\sum_{j:f(j)=1}|j\rangle\langle j|G^r|\psi_0\rangle\right\|^2 &= \left|\begin{pmatrix}1 & 0\end{pmatrix}\hat{G}^r\begin{pmatrix}\sin(\vartheta/2)\\ \cos(\vartheta/2)\end{pmatrix}\right|^2\\ &= |\cos(r\vartheta)\sin(\vartheta/2)+\sin(r\vartheta)\cos(\vartheta/2)|^2\\ &= \sin^2\left(\left(r+\frac{1}{2}\right)\vartheta\right) \quad .\end{aligned} \tag{2.6}$$

The last transformation uses the trigonometric identity

$$\sin(\alpha+\beta) = \cos\alpha\sin\beta + \sin\alpha\cos\beta \quad .$$

Success probability From Equation (2.6) it follows that the success probability of quantum search is periodic in r; when $(r+1/2)\vartheta \approx \pi/2$, we have a high probability of obtaining a good measurement outcome. The first maximum is at $r_{\text{opt}} = \pi/(2\vartheta) - 1/2 + \Delta$ for a $\Delta \in \mathbb{R}$ with $|\Delta| \leq 1/2$ that ascertains that r_{opt} is an integer. For $\vartheta \leq \pi/2$, we can bound the success probability as follows:

$$\sin^2\left(\left(r_{\text{opt}}+\frac{1}{2}\right)\vartheta\right) = \sin^2\left(\frac{\pi}{2}+\Delta\vartheta\right) = 1-\sin^2(\Delta\vartheta) \geq 1-\frac{\vartheta^2}{4} \geq \frac{1}{3}$$

whereas $\vartheta > \pi/2$ implies $2M > N$ and $r_{\text{opt}} = 0$. Since in this case, measuring the initial state gives success probability greater than 1/2, we have constant success probability in all cases.

To obtain an asymptotic bound on r in terms of N and M, let $\vartheta' := 2\sqrt{M/N}$. Since $x \geq \sin x$ for $x \geq 0$, we have

$$\frac{\vartheta}{2} \geq \sin\left(\frac{\vartheta}{2}\right) = \sqrt{\frac{M}{N}} = \frac{\vartheta'}{2}$$

where the first equality is as in Equation (2.5). Hence, $\vartheta \geq \vartheta'$ and

$$r_{\text{opt}} \leq \left\lceil \frac{\pi}{4}\sqrt{\frac{N}{M}} \right\rceil \quad .$$

For our telephone-directory example, this implies that using quantum queries we can find the single matching entry with high probability using $\mathrm{O}(\sqrt{N})$ quantum queries.

	Grover's algorithm	amplitude amplification
"mixing" operator	$W \otimes \mathbb{1}$	arbitrary unitary operator A
initial state	$W \otimes \mathbb{1}\|0^n\rangle\|-\rangle$	$A\|\psi_0\rangle$ for arbitrary $\|\psi_0\rangle$
first phase flip	query U_f	$\mathbb{1} - 2\sum_{\|\varphi_x\rangle \text{ good}} \|\varphi_x\rangle\langle\varphi_x\|$
second phase flip	$\mathbb{1} - 2\|0\rangle\langle 0\|$	$\mathbb{1} - 2\|\psi_0\rangle\langle\psi_0\|$

Table 2.1: From Grover's algorithm to amplitude amplification

Tuning So far, we need to know the number of solutions M in order to determine the sufficient number of iterations. For M unknown, there are ways using doubling techniques [24] to find a solution with an expected number of queries $O(\sqrt{N/M})$. If M is known, the success probability of the quantum-search algorithm can be improved to 1, e.g., by changing the ϑ for the last iteration [26, 28]. With regard to lower bounds, Grover's algorithm and its extensions have been shown to be optimal in many respects [20, 120, 30].

2.1.2 Amplitude amplification

The preceding analysis of the quantum-search algorithm hinged on the fact that iterations of the quantum-search algorithm can be expressed as rotations in a plane spanned by "good" and "bad" states. "Amplitude amplification" is a general framework [27, 70] for increasing the amplitude of "good" states when those can be recognized efficiently.

The framework The generalization from Grover's algorithm to amplitude amplification is outlined in Table 2.1. The only operator that is genuinely quantum in Grover's algorithm is the Hadamard transform. Let us investigate what happens if we replace it by an arbitrary unitary operator A, start on an arbitrary quantum state $|\psi_0\rangle$, and use an arbitrary orthonormal family $F := \{|\varphi_x\rangle : x \in X\}$ as the set of "good" states. The iteration of Grover's algorithm began with a database query U_f, which effectively flipped the sign of the good states. So now we just perform an analogous step, namely applying the operator

$$S_F := \mathbb{1} - 2\sum_{x \in X} |\varphi_x\rangle\langle\varphi_x| \ .$$

The next step in the iteration was to reflect the amplitudes about their average, realized by a phase-flip in the W-basis. We mimic the property that the "reflection about the average" flips the phase of the initial state and leaves all orthogonal states invariant by defining the new

$$T_{\psi_0} := -A(\mathbb{1} - 2|\psi_0\rangle\langle\psi_0|)A^{-1} \ .$$

What properties does our new iteration operator $Q := T_{\psi_0} S_F$ have when applied repeatedly to the initial state $A|\psi_0\rangle$? Our definitions are validated insofar as we can repeat the analysis in two-dimensions: in analogy to (2.3) define the "good" and "bad" portions of $A|\psi_0\rangle$ as

$$|\tilde{\chi}\rangle := \sum_{x\in X} |\varphi_x\rangle\langle\varphi_x|A|\psi_0\rangle \qquad \text{and} \qquad |\tilde{\chi}^\perp\rangle := \left(\mathbb{1} - \sum_{x\in X} |\varphi_x\rangle\langle\varphi_x|\right) A|\psi_0\rangle$$

and with $a := \sqrt{\langle\tilde{\chi}|\tilde{\chi}\rangle}$ normalize to

$$|\chi\rangle := \frac{1}{a}|\tilde{\chi}\rangle \qquad \text{and} \qquad |\chi^\perp\rangle := \frac{1}{\sqrt{1-a^2}}|\tilde{\chi}^\perp\rangle \ .$$

Then by simple arithmetic we obtain

$$\begin{aligned} A|\psi_0\rangle &= |\tilde{\chi}\rangle + |\tilde{\chi}^\perp\rangle = a|\chi\rangle + \sqrt{1-a^2}|\chi^\perp\rangle \ , \\ Q|\chi\rangle &= \frac{1}{a}\left(A(\mathbb{1} - 2|\psi_0\rangle\langle\psi_0|)A^{-1} \sum_{x\in X} |\varphi_x\rangle\langle\varphi_x|A|\psi_0\rangle\right) \\ &= |\chi\rangle - 2aA|\psi_0\rangle = (1-2a^2)|\chi\rangle - 2a\sqrt{1-a^2}|\chi^\perp\rangle \ , \end{aligned}$$

and

$$\begin{aligned} Q|\chi^\perp\rangle &= \frac{1}{\sqrt{1-a^2}}\left(-A(\mathbb{1} - 2|\psi_0\rangle\langle\psi_0|)A^{-1}\left(\mathbb{1} - \sum_{x\in X} |\varphi_x\rangle\langle\varphi_x|\right) A|\psi_0\rangle\right) \\ &= -|\chi^\perp\rangle + 2\sqrt{1-a^2}A|\psi_0\rangle = 2a\sqrt{1-a^2}|\chi\rangle + (1-2a^2)|\chi^\perp\rangle \end{aligned}$$

so that we can again define a two-dimensional rotation

$$\hat{Q} := \begin{pmatrix} 1-2a^2 & 2a\sqrt{1-a^2} \\ -2a\sqrt{1-a^2} & 1-2a^2 \end{pmatrix}$$

with

$$Q(\alpha|\chi\rangle + \beta|\chi^\perp\rangle) = \begin{pmatrix} |\chi\rangle & |\chi^\perp\rangle \end{pmatrix} \hat{Q} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \ .$$

Hence, with the smallest $\vartheta \geq 0$ such that $a = \sin(\vartheta/2)$, $\hat{Q}$ is a rotation by ϑ and after at most $\lceil \pi/4a \rceil$ iterations we are close to the good states in the sense that measuring the observable $\sum_{x\in X} |\varphi_x\rangle\langle\varphi_x|$ will yield outcome 1 with constant probability and in this case project the state into the subspace spanned by the $\{|\varphi_x\rangle : x \in X\}$.

Applications So what is amplitude amplification good for? Clearly, it generalizes quantum search. Furthermore, we can amplify the success probability of an arbitrary quantum algorithm if the following conditions are met:

1. the initial state of the algorithm is a pure state $|\psi_0\rangle$ and we have a transformation $S_{\psi_0} = 1 - 2|\psi_0\rangle\langle\psi_0|$;
2. the algorithm only uses unitary gates and in particular does not make any measurements;
3. there is a projective measurement $\{P_{\text{success}}, \mathbb{1} - P_{\text{success}}\}$ that determines for an output whether the run was successful or not, and we have the corresponding unitary transform $S_F = 1 - 2P_{\text{success}}$.

By Condition 2, the algorithm corresponds to an overall unitary operator A. On initial state $|\psi_0\rangle$, the success probability is $p_{\text{success}} := \|P_{\text{success}}A|\psi_0\rangle\|^2$ according to Condition 3. We fit this into the amplitude-amplification framework by letting the set of good states $F = \{|\varphi_x\rangle : x \in X\}$ be a basis of the range of P_{success}. Then $P_{\text{success}} = \sum_{x\in X} |\varphi_x\rangle\langle\varphi_x|$, $S_F = 1 - 2\sum_{x\in X} |\varphi_x\rangle\langle\varphi_x|$, and $a = \|P_{\text{success}}A|\psi_0\rangle\| = \sqrt{p_{\text{success}}}$. Hence, applying amplitude amplification we can boost a small p_{success} to constant in $\mathrm{O}(1/\sqrt{p_{\text{success}}})$ iterations, whereas classically, boosting the success probability of algorithms that indicate whether they were successful takes

$$1 - (1 - p_{\text{success}})^r \geq c \quad \Rightarrow \quad r = \Omega\left(\frac{1}{p}\right)$$

repetitions.

For a concrete example, consider the following instance of the *claw-finding* problem, derived as a special case from [32]: given two functions f and g with domain $[N] = \{1, \ldots, N\}$, find x and $y \in [N]$ with $f(x) = g(y)$. Our quantum algorithm A selects uniformly at random a set $I \subseteq [N]$ of size $|I| = \sqrt{N}$. It queries f on all $x \in I$ and uses this to construct an oracle $h : [N] \to \{0, 1\}$ for quantum search on g by defining for $h(y) = 1 \Leftrightarrow \exists x \in I : f(x) = g(y)$. Evaluating h takes one query to g and no query to f. A then performs quantum search for $h(y) = 1$. This takes $|I| = \sqrt{N}$ queries to f and $\mathrm{O}(\sqrt{N})$ queries to h and thus to g. A finds a claw $f(x) = f(y)$ if it chose I such that $x \in I$ and if the quantum search was successful. This happens with probability $(|I|/N) \cdot \text{const} = \Omega(1/\sqrt{N})$. Now we use amplitude amplification on A to boost the success probability to constant in $\mathrm{O}(N^{1/4})$ iterations, performing in total $\mathrm{O}(N^{1/4+1/2}) = \mathrm{O}(N^{3/4})$ queries. Since this is a special case of quantum search, classically $\Omega(N)$ queries are necessary in the worst case. The best known quantum upper bound to date is $\mathrm{O}(N^{2/3})$ [9].

2.2 Convolution Products

Can a quantum computer speed up multiplication or applications relying on multiplication? This question, the efficient quantum Fourier transform [46, 108, 50, 44], and the utility of convolution products, e.g., for pattern matching, were our motivations for examining computing convolution products on a quantum computer.

Convolution and the discrete Fourier transform For two vectors $\mathbf{a} = \begin{pmatrix} a_0 & \cdots & a_{N-1} \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} b_0 & \cdots & b_{N-1} \end{pmatrix}$, the *convolution product* is

$$\mathbf{a} * \mathbf{b} = \begin{pmatrix} c_0 & \cdots & c_{N-1} \end{pmatrix}$$

with

$$c_j = \sum_{k,k':k+k'=j} a_k b_{k'} \ . \tag{2.7}$$

Evidently, $\mathbf{c}$ is just the vector of coefficients of the polynomial

$$\left(\sum_{k=0}^{N-1} a_k x^k \right) \left(\sum_{k'=0}^{N-1} b_{k'} x^{k'} \right) \ .$$

Computing $\mathbf{c}$ directly via Equation (2.7) requires $\Omega(N^2)$ arithmetic operations. This can be reduced to $\mathrm{O}(N \log N)$ operations using the *discrete Fourier transform* and its inverse. Let ω denote the Nth root of unity $\omega = e^{2\pi\,\mathrm{i}/N}$. The discrete Fourier transform is the mapping

$$\mathrm{DFT} : \mathbf{a} \mapsto \hat{\mathbf{a}} := \begin{pmatrix} \hat{a}_0 & \cdots & \hat{a}_{N-1} \end{pmatrix} \quad \text{with} \quad \hat{a}_\ell = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} a_k \omega^{k\ell} \tag{2.8}$$

and its inverse is

$$\mathrm{DFT}^{-1} : \hat{\mathbf{a}} \mapsto \mathbf{a} := \begin{pmatrix} a_0 & \cdots & a_{N-1} \end{pmatrix} \quad \text{with} \quad a_\ell = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \hat{a}_k \omega^{-k\ell} \ . \tag{2.9}$$

First, we compute $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$. Then we compute the product of $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ component by component, i.e.,

$$\hat{\mathbf{c}} = \begin{pmatrix} \hat{a}_0 \hat{b}_0 & \cdots & \hat{a}_{N-1} \hat{b}_{N-1} \end{pmatrix}$$

and use the inverse Fourier transform to obtain $\mathbf{c}$ with

$$c_\ell = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \hat{a}_k \hat{b}_k \omega^{-k\ell}$$

$$
\begin{aligned}
&= \frac{1}{N^{3/2}} \sum_{k',k''=0}^{N-1} a_{k'} b_{k''} \sum_{k=0}^{N-1} \omega^{k(k'+k''-\ell)} \\
&= \frac{1}{\sqrt{N}} \sum_{(k',k'')\in\Sigma(\ell)} a_{k'} b_{k''} \quad ,
\end{aligned}
$$

with $\Sigma(\ell) = \Sigma_2(\ell)$ and $\Sigma_t(\ell) := \{(i_1, \ldots, i_t) : 0 \le i_j < N$ for all j and $\sum i_j \equiv \ell \mod N\}$; we drop the subscript from $\Sigma_t(\ell)$ whenever t is evident from the context.

Hence, if $a_k = b_k = 0$ for $k > N/2$, then $\sqrt{N}\mathbf{c} = \mathbf{a} * \mathbf{b}$. The *fast Fourier transform* algorithm [45] computes the discrete Fourier transform or its inverse in $\mathrm{O}(N \log N)$ steps, therefore we can compute the convolution product with $\mathrm{O}(N \log N)$ steps as well.

Quantum Fourier transform In the setting of quantum circuits, the vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, etc. from the preceding paragraphs map in a natural way to quantum states, e.g.,

$$|\psi_{\mathbf{a}}\rangle := \sum_{k=0}^{N-1} a_k |k\rangle \quad .$$

Moreover, as defined in Equation (2.8) the discrete Fourier transform is a unitary transformation. The corresponding quantum operation

$$\mathrm{QFT} : |j\rangle \mapsto \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \omega^{kj} |k\rangle$$

is called the *quantum Fourier transform*. That it can be approximated efficiently with $\mathrm{O}(\log N \log\log N)$ operations [46, 108, 50] is the foundation of many quantum algorithms.

Keeping this in mind, it is straightforward to compute convolution products on a quantum computer by transforming the input state

$$|\psi_{\text{input}}\rangle := \left(\sum_{k=0}^{N-1} a_k |k\rangle\right) \otimes \left(\sum_{k'=0}^{N-1} b_{k'} |k'\rangle\right)$$

into

$$|\psi_{\text{output}}\rangle := \left(\sum_{j=0}^{N-1} \left(\sum_{(k,k')\in\Sigma(j)} a_k b_{k'}\right) |j\rangle\right) \otimes |\text{rest}\rangle \quad ,$$

where $N = 2^n$ and $|k\rangle$, $|k'\rangle$, $|j\rangle$, and $|\text{rest}\rangle$ are n-bit quantum registers. These vectors are not necessarily normalized; note, however, that we need to require that $\mathbf{a} \neq 0$ and $\mathbf{b} \neq 0$. A straightforward approach is to perform a QFT gate on the first N qubits and the last N qubits, leading to state

$$|\psi_1\rangle = \left(\sum_{\ell=0}^{N-1} \hat{a}_\ell |\ell\rangle\right) \otimes \left(\sum_{\ell'=0}^{N-1} \hat{b}_{\ell'} |\ell'\rangle\right) = \sum_{\ell,\ell'=0}^{N-1} \hat{a}_\ell \hat{b}_{\ell'} |\ell\rangle|\ell'\rangle \ .$$

We then permute the basis states to map $|\ell\rangle|\ell\rangle$ to $|\ell\rangle|0\rangle$ for each ℓ; call this state $|\psi_2\rangle$. Now we measure the second register. If the outcome is not $|0\rangle$, then the algorithm fails, otherwise the system is projected to

$$|\psi_3\rangle = \sum_{\ell=0}^{N-1} \hat{a}_\ell \hat{b}_\ell |\ell\rangle|0\rangle = \frac{1}{N} \sum_{\ell=0}^{N-1} \left(\sum_{k,k'=0}^{N-1} a_k b_{k'} \omega^{(k+k')\ell} \right) |\ell\rangle|0\rangle$$

and we apply the inverse QFT^{-1} on the first register. Hence, we obtain

$$\frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} \left(\sum_{\ell=0}^{N-1} \hat{a}_\ell \hat{b}_\ell \omega^{-\ell j} \right) |j\rangle|0\rangle = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} \sum_{(k,k')\in\Sigma(j)} a_k b_{k'} |j\rangle|0\rangle \ ,$$

which is the desired output state $|\psi_{\mathrm{output}}\rangle$.

Amplification Unfortunately, the success probability of this algorithm is not constant and, moreover, is dependent on the inputs $\mathbf{a}$ and $\mathbf{b}$. Therefore we resort to amplitude amplification. The price to pay will be the need to repeatedly execute the steps we outlined before and the input $|\psi_{\mathrm{input}}\rangle$ must be given by means of an operator V preparing $|\psi_{\mathrm{input}}\rangle$ from the initial state of amplitude amplification $|0\rangle$, i.e., $V|0\rangle = |\psi_{\mathrm{input}}\rangle$.

To define the algorithm A formally, let R be the permutation that maps $|\ell_0 \dots \ell_{n-1}\ell'_0 \dots \ell'_{n-1}\rangle$ to $|\ell_0\ell'_0 \dots \ell_{n-1}\ell'_{n-1}\rangle$. Since $R^{-1}\mathrm{CNOT}^{\otimes n}R$ maps $|\ell\rangle|\ell\rangle$ to $|\ell\rangle|0\rangle$, we can express the operations up to the measurement by

$$A := R^{-1}\ \mathrm{CNOT}^{\otimes n}\ R\ (\mathrm{QFT} \otimes \mathrm{QFT})\ V \ .$$

Thus $A|0\rangle = |\psi_2\rangle$. We would like to amplify the basis states $F := \{|\ell\rangle|0\rangle : 0 \leq \ell < N\}$. Let $a := \||\psi_3\rangle\| / \||\psi_2\rangle\|$, i.e., a^2 is the success probability when A is applied exactly once. This is the initial success probability that we boost by amplitude amplification using $\Theta(1/a)$ applications of A. We derive a lower bound on a^2 under the additional assumption that all a_k and b_k are nonnegative:

$$a^2 = \frac{1}{\||\psi_2\rangle\|^2} \||\psi_3\rangle\|^2 = \frac{1}{\||\psi_{\mathrm{input}}\rangle\|^2} \||\psi_3\rangle\|^2 = \frac{1}{\|\mathbf{a}\|^2 \|\mathbf{b}\|^2} \sum_{\ell=0}^{N-1} |\hat{a}_\ell \hat{b}_\ell|^2$$

$$
\begin{aligned}
&= \frac{1}{\|\mathbf{a}\|^2\|\mathbf{b}\|^2 N^2} \sum_{\ell=0}^{N-1} \left| \sum_{k,k'=0}^{N-1} a_k b_{k'} \omega^{(k+k')\ell} \right|^2 \\
&= \frac{1}{\|\mathbf{a}\|^2\|\mathbf{b}\|^2 N^2} \sum_{\ell=0}^{N-1} \sum_{k,k',k'',k'''=0}^{N-1} a_k b_{k'} \overline{a_{k''} b_{k'''}} \omega^{(k+k'-k''-k''')\ell} \\
&= \frac{1}{\|\mathbf{a}\|^2\|\mathbf{b}\|^2 N^2} \sum_{(k,k',N-k'',N-k''')\in\Sigma(0)} a_k b_{k'} \overline{a_{k''} b_{k'''}} N \\
&\geq \frac{1}{\|\mathbf{a}\|^2\|\mathbf{b}\|^2 N} \sum_{k,k'=0}^{N-1} |a_k|^2 |b_{k'}|^2 = \frac{1}{N}
\end{aligned}
$$

The inequality holds because we impose the restriction of summing only over those $(k, k', N-k'', N-k''') \in \Sigma(0)$ where $k = k''$ and $k' = k'''$. It follows that $\mathrm{O}(\sqrt{N})$ repetitions of the amplitude-amplification procedure are sufficient in all cases.

How expensive is one iteration? The iteration operator is

$$Q = -AS_0 A^{-1} S_F \ ,$$

where A is as defined above, S_0 is the phase rotation by -1 conditional on the input being $|0\rangle$, i.e., $S_0 = \mathbb{1} - 2|0\rangle\langle 0|$, and

$$S_F = \mathbb{1} - 2 \sum_{\ell=0}^{N-1} |\ell\rangle\langle\ell| \otimes |0\rangle\langle 0|$$

rotates the phase of basis vectors $|\ell\rangle|0\rangle$ by -1 and leaves all other basis vectors invariant. Preskill [99] shows that S_0 can be implemented with $\mathrm{O}(\log N)$ gates; similarly, S_F can be realized using $\mathrm{O}(\log N)$ gates and three auxiliary qubits. QFT takes $\mathrm{O}(\log N \log\log N)$ operations [50] and R can be implemented by $\mathrm{O}(\log N)$ swaps of adjacent qubits, which in turn can be constructed from three CNOT gates. If v is the number of gates needed for implementing V, we get a bound of $\mathrm{O}(v + (\log N)^2)$. Thus, if we measure the observable $\sum_{\ell=0}^{N-1} |\ell\rangle\langle\ell| \otimes |0\rangle\langle 0|$ after

$$\mathrm{O}\left(\sqrt{N}\left(v + (\log N)^2\right)\right)$$

operations, we have constant success probability for projecting the system into state $|\psi_3\rangle$. Since we only have a lower bound on the success probability, it will in general be necessary to apply the techniques of quantum search with unknown number of solutions. Finally, we can convert $|\psi_3\rangle$ by an inverse QFT on the first register with $\mathrm{O}(\log N \log\log N)$ operations into $|\psi_{\text{output}}\rangle$.

An application Suppose we want to implement the preparation V using an oracle for the amplitudes $\begin{pmatrix} a_0 & \cdots & a_{N-1} \end{pmatrix}$ and $\begin{pmatrix} b_0 & \cdots & b_{N-1} \end{pmatrix}$ by using the technique from [72]. We assume that $N = 2^n$, that the components of $\mathbf{a}$ and $\mathbf{b}$ are nonnegative multiples of 2^{-m} for some $m \in \mathbb{N}$, and that $\sum_{0 \le k < N} a_k^2 = \sum_{0 \le k < N} b_k^2 = 1$. Using oracles for the m bits of precision of the components, we can implement transformations for putting the components in the amplitudes,

$$\begin{aligned} U_{\mathbf{a}} &: |k\rangle|b\rangle \mapsto a_k|k\rangle|b\rangle + (-1)^b\sqrt{1-|a_k|^2}|k\rangle|b\oplus 1\rangle \\ U_{\mathbf{b}} &: |k\rangle|b\rangle \mapsto b_k|k\rangle|b\rangle + (-1)^b\sqrt{1-|b_k|^2}|k\rangle|b\oplus 1\rangle \end{aligned}$$

and their inverses

$$\begin{aligned} U_{\mathbf{a}}^{-1} &: |k\rangle|b\rangle \mapsto \bar{a}_k|k\rangle|b\rangle - (-1)^b\sqrt{1-|a_k|^2}|k\rangle|b\oplus 1\rangle \\ U_{\mathbf{b}}^{-1} &: |k\rangle|b\rangle \mapsto \bar{b}_k|k\rangle|b\rangle - (-1)^b\sqrt{1-|b_k|^2}|k\rangle|b\oplus 1\rangle \end{aligned}$$

where $0 \le k < N$ and $b \in \{0, 1\}$. These operators are weak in the sense that, e.g., for most $\mathbf{a}$, the a_k are going to be small and therefore the states $U_{\mathbf{a}}|k\rangle|b\rangle$ close to $|k\rangle|b\rangle$; however, this construction has the advantage of uniformly operating on the table of amplitudes without preprocessing.

Amplitude amplification on $U_{\mathbf{a}}H^{\otimes n}$ lets us map $|0^{n+1}\rangle$ exactly to

$$\sum a_k|k\rangle|0\rangle$$

in $\Theta(\sqrt{N})$ iterations; from this we get the input preparation operator V. Applying the result of the previous paragraph we can thus produce

$$\sum_{j=0}^{N-1} \sum_{(k,k')\in\Sigma(j)} a_k b_{k'}|j\rangle \tag{2.10}$$

with high probability using $\mathrm{O}(\sqrt{N}(m\sqrt{N}+(\log N)^2)) = \mathrm{O}(mN)$ oracle queries and operations.

An efficient method to produce the state (2.10) may be of interest, e.g., for approximate pattern matching. However, our result is disappointing in this respect; the present algorithm requires reading a constant fraction of the input. Moreover, a very similar classical problem has an efficient solution: reading the entire input allows us to sample efficiently from the distribution $\Pr[j] = \sum_{(k,k')\in\Sigma(j)} a_k^2 b_{k'}^2$ simply by choosing k with probability a_k^2 and k' with probability $b_{k'}^2$ and outputting $k + k'$. One way to improve the quantum complexity would be to realize the reflection about the input state, $1 - 2|\psi_{\text{input}}\rangle\langle\psi_{\text{input}}|$ directly using $U_{\mathbf{a}}$ and $U_{\mathbf{b}}$ instead of relying on

$1 - 2|\psi_{\text{input}}\rangle\langle\psi_{\text{input}}| = V(1 - 2|0\rangle\langle 0|)V^*$. This operation requires $\Theta(\sqrt{N})$ invocations of $U_{\mathbf{a}}$ and $U_{\mathbf{b}}$, since obtaining V from $U_{\mathbf{a}}$ and $U_{\mathbf{b}}$ is a generalization of quantum search, which has a lower bound of $\Omega(\sqrt{N})$ queries [20, 71, 15]. Similarly, implementing V using the reflection operator $1 - 2|\psi_{\text{input}}\rangle\langle\psi_{\text{input}}|$ requires in general $\Omega(\sqrt{N})$ iterations of amplitude amplification, hence, one might hope that implementing the "weak" reflection operator by means of the "weak" amplitude queries $U_{\mathbf{a}}$ and $U_{\mathbf{b}}$ should be efficiently feasible, but alas we did not to find a way to achieve this.

2.3 Search in the Density-Matrix Formalism

In the real world, a quantum computer will be subject to noise and imperfections. For instance, it is hard to implement quantum gates exactly and the approximation error will accumulate over the course of a computation. An altogether different error source arises from the difficulty of isolating a quantum mechanical system from its environment; unintended interaction with the environment is called *decoherence* and manifests itself in uncontrolled measurements that "collapse" the current quantum state. These problems have attracted much attention and were in part solved by *quantum error correction*: by computing on encoded states, interleaving the computation with error-correction stages, and recursively applying these techniques, fault-tolerant quantum computing was shown to be possible whenever the errors are sufficiently local and uncorrelated, there is a supply of "fresh" qubits or sufficient parallelism, and the individual error probability is below a model-specific threshold [106, 112, 107, 2, 78].

However, the generic transformations for making a quantum circuit fault-tolerant are quite expensive and may be prohibitive for simple quantum computers. Therefore it is of interest to study the behavior of fundamental quantum algorithms when subjected to typical errors—with or without minimal fault detection and correction. In this section, we generalize the elegant analysis of Grover's algorithm as a rotation in a two-dimensional vector space spanned by two pure quantum states: now the current state of the algorithm is a mixed state, and to accommodate the decoherence operator, we have to analyze the algorithm as a linear transformation in a *four*-dimensional space spanned by four density matrices.

Evolution in density matrices Consider Grover's algorithm for database search [69] with one target state. Let $N = 2^n$ be the size of the database, $|t\rangle \in \mathcal{H}_N$ the target state, $S_k = 1 - 2|k\rangle\langle k|$ the reflection conditional on k, and W the N-dimensional Hadamard transform. As before, one iteration WS_0WS_t of the algorithm can be seen as a unitary mapping in the two-dimensional subspace spanned by $|t\rangle$ and $|t^\perp\rangle := \sum_{k\neq t} |k\rangle = \sqrt{N}(1 - |t\rangle\langle t|)W|0\rangle$:

$$WS_0WS_t(\alpha|t\rangle + \beta|t^\perp\rangle) = \begin{pmatrix}|t\rangle & |t^\perp\rangle\end{pmatrix} \frac{1}{N}\begin{pmatrix} N-2 & 2(N-1)\\ -2 & N-2\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}$$

In order to investigate the effects of decoherence on the algorithm, we express this evolution in the language of density matrices. For $\alpha, \beta \in \mathbb{R}$, we write the undisturbed iteration

$$\begin{aligned} &WS_0WS_t\left(\alpha|t\rangle + \beta|t^\perp\rangle\right)\left(\bar{\alpha}\langle t| + \bar{\beta}\langle t^\perp|\right) S_t^* W^* S_0^* W^* \\ &= WS_0WS_t\left(\alpha^2|t\rangle\langle t| + \beta^2|t^\perp\rangle\langle t^\perp| + \alpha\beta\left(|t^\perp\rangle\langle t| + |t\rangle\langle t^\perp|\right)\right) S_tWS_0W \end{aligned}$$

as a linear mapping in the subspace of matrices spanned by

$$\rho_t := |t\rangle\langle t| = \begin{array}{cc} \begin{matrix} & t & \\ & \downarrow & \end{matrix} & \\ \begin{pmatrix} 0 & & 0 \\ & 1 & \\ 0 & & 0 \end{pmatrix} & \leftarrow t \end{array} \ ,$$

$$\rho_{t^\perp} := |t^\perp\rangle\langle t^\perp| = \textstyle\sum_{j,k\neq t} |j\rangle\langle k| = \begin{array}{cc} \begin{matrix} & t & \\ & \downarrow & \end{matrix} & \\ \begin{pmatrix} 1\cdots 1 & 0 & 1\cdots 1 \\ \vdots\ \ \ \vdots & \vdots & \vdots\ \ \ \vdots \\ 1\cdots 1 & 0 & 1\cdots 1 \\ 0\cdots 0 & 0 & 0\cdots 0 \\ 1\cdots 1 & 0 & 1\cdots 1 \\ \vdots\ \ \ \vdots & \vdots & \vdots\ \ \ \vdots \\ 1\cdots 1 & 0 & 1\cdots 1 \end{pmatrix} & \leftarrow t \end{array} \ , \text{ and}$$

$$\rho_\times := |t^\perp\rangle\langle t| + |t\rangle\langle t^\perp| = \begin{array}{cc} \begin{matrix} & t & \\ & \downarrow & \end{matrix} & \\ \begin{pmatrix} & 1 & \\ 0 & \vdots & 0 \\ & 1 & \\ 1\cdots 1 & 0 & 1\cdots 1 \\ & 1 & \\ 0 & \vdots & 0 \\ & 1 & \end{pmatrix} & \leftarrow t \end{array} \ .$$

For ρ_t,

$$\begin{aligned}
&WS_0WS_t\rho_tS_tWS_0W\\
&= (WS_0WS_t|t\rangle)\,(\langle t|S_tWS_0W)\\
&= \left(\frac{N-2}{N}|t\rangle - \frac{2}{N}|t^\perp\rangle\right)\left(\frac{N-2}{N}\langle t| - \frac{2}{N}\langle t^\perp|\right)\\
&= \left(\frac{N-2}{N}\right)^2\rho_t + \frac{4}{N^2}\rho_{t^\perp} - \frac{2(N-2)}{N^2}\rho_\times \quad .
\end{aligned}$$

For $\rho_{t^\perp}$,

$$\begin{aligned}
&WS_0WS_t\rho_{t^\perp}S_tWS_0W\\
&= \left(WS_0WS_t|t^\perp\rangle\right)\,\left(\langle t^\perp|WS_tWS_0W\right)\\
&= \left(\frac{2(N-1)}{N}|t\rangle + \frac{N-2}{N}|t^\perp\rangle\right)\left(\frac{2(N-1)}{N}\langle t| + \frac{N-2}{N}\langle t^\perp|\right)\\
&= 4\left(\frac{N-1}{N}\right)^2\rho_t + \left(\frac{N-2}{N}\right)^2\rho_{t^\perp} + 2\frac{(N-1)(N-2)}{N^2}\rho_\times \quad .
\end{aligned}$$

For $\rho_\times$,

$$\begin{aligned}
&WS_0WS_t\rho_\times S_tWS_0W\\
&= \left(WS_0WS_t|t^\perp\rangle\right)\,(\langle t|WS_tWS_0W) + (WS_0WS_t|t\rangle)\,\left(\langle t^\perp|WS_tWS_0W\right)\\
&= \left(\frac{2(N-1)}{N}|t\rangle + \frac{N-2}{N}|t^\perp\rangle\right)\left(\frac{N-2}{N}\langle t| - \frac{2}{N}\langle t^\perp|\right) +\\
&\quad \left(\frac{N-2}{N}|t\rangle - \frac{2}{N}|t^\perp\rangle\right)\left(\frac{2(N-1)}{N}\langle t| + \frac{N-2}{N}\langle t^\perp|\right)\\
&= 4\frac{(N-1)(N-2)}{N^2}\rho_t - 4\frac{N-2}{N^2}\rho_{t^\perp} + \frac{N^2-8N+8}{N^2}\rho_\times \quad .
\end{aligned}$$

Thus, one iteration of database search acts as

$$a\,\rho_t + b\,\rho_{t^\perp} + c\,\rho_\times \mapsto \begin{pmatrix}\rho_t & \rho_{t^\perp} & \rho_\times\end{pmatrix} R \begin{pmatrix}a\\ b\\ c\end{pmatrix}$$

where

$$R = \frac{1}{N^2}\begin{pmatrix} (N-2)^2 & 4(N-1)^2 & 4(N-1)(N-2)\\ 4 & (N-2)^2 & -4(N-2)\\ -2(N-2) & 2(N-1)(N-2) & N^2-8N+8 \end{pmatrix} \quad .$$

The initial state $W|0\rangle$ has density matrix

$$W|0\rangle\langle 0|W = \frac{1}{N}\left(\rho_t + \rho_{t^\perp} + \rho_\times\right)$$

and is represented by the 3-vector

$$\frac{1}{N}\begin{pmatrix}1\\1\\1\end{pmatrix} \ .$$

Decoherence processes We considered two decoherence processes that are motivated by NMR [39]:

$$D_1 : \rho \mapsto (1-\lambda_1)\rho + \lambda_1 \sum_k |k\rangle\langle k|\rho|k\rangle\langle k|$$

corresponds to performing a measurement in the basis $\{|k\rangle\}$ with probability λ_1. Note that D_1 pushes the system towards a "preferred" basis, which we assume to coincide with the computational basis. On the other hand, a usually weaker but more devastating decoherence effect is

$$D_2 : \rho \mapsto (1-\lambda_2)\rho + \lambda_2 \frac{1}{N}\mathbb{1} \ ,$$

which models relaxation to the totally mixed state $\mathbb{1}/N$ with probability λ_2. Since the action of D_2 commutes with every other linear transformation, we restrict our attention to D_1. We compute how D_1 acts on the four-dimensional subspace spanned by ρ_t, $\rho_{t^\perp}$, $\rho_\times$, and $\mathbb{1}$:

$$\begin{aligned} D_1\rho_t &= \rho_t \\ D_1\rho_{t^\perp} &= (1-\lambda_1)\rho_{t^\perp} + \lambda_1\mathbb{1} - \lambda_1\rho_t \\ D_1\rho_\times &= (1-\lambda_1)\rho_\times \\ D_1\mathbb{1} &= \mathbb{1} \ . \end{aligned}$$

Thus D_1 acts as

$$a\,\rho_t + b\,\rho_{t^\perp} + c\,\rho_\times + d\,\mathbb{1} \mapsto \begin{pmatrix}\rho_t & \rho_{t^\perp} & \rho_\times & \mathbb{1}\end{pmatrix} D_1 \begin{pmatrix}a\\b\\c\\d\end{pmatrix}$$

with

$$D_1 = \begin{pmatrix} 1 & -\lambda_1 & 0 & 0 \\ 0 & 1-\lambda_1 & 0 & 0 \\ 0 & 0 & 1-\lambda_1 & 0 \\ 0 & \lambda_1 & 0 & 1 \end{pmatrix} \ .$$

We investigate the behavior of search when the state of the system is disturbed by D_1 before each rotation. One iteration then corresponds to the linear mapping

$$a\,\rho_t + b\,\rho_{t^\perp} + c\,\rho_\times + d\,\mathbb{1} \mapsto \begin{pmatrix}\rho_t & \rho_{t^\perp} & \rho_\times & \mathbb{1}\end{pmatrix} R_1 D_1 \begin{pmatrix}a\\ b\\ c\\ d\end{pmatrix}$$

with

$$R_1 = \begin{pmatrix}R & 0\\ 0 & 1\end{pmatrix} \;.$$

The initial state is represented by the 4-vector

$$s = \frac{1}{N}\begin{pmatrix}1\\ 1\\ 1\\ 0\end{pmatrix}.$$

The probability of successfully measuring $|t\rangle$ after ℓ iterations is

$$\begin{aligned} p_{N,\lambda_1,\ell} &= \langle t|(WS_0WS_tD_1)^\ell W|0\rangle\langle 0|W(D_1^* S_t W S_0 W)^\ell |t\rangle \\ &= \begin{pmatrix}1 & 0 & 0 & 1\end{pmatrix}(R_1D_1)^\ell s. \end{aligned}$$

Numerical simulations Figure 2.3 gives an example of the undisturbed evolution in the four-dimensional subspace; Figure 2.4 shows the same evolution when subjected to decoherence via D_1. Figure 2.5 indicates that for constant success probability, smaller and smaller λ_1 can be tolerated with growing N and it suggests that constant success probability can be achieved with $\lambda_1 = \omega(1/\sqrt{N})$—this would mean that the decoherence process D_1 is somewhat less destructive to quantum search than D_2, which in each iteration replaces the state of the computation with the completely mixed state with probability λ_2 and which clearly can tolerate error probability $\lambda_2 = \mathrm{O}(1/\sqrt{N})$ only. The susceptibility of quantum search to other kinds of errors has been studied before both numerically and analytically [85, 95, 105].

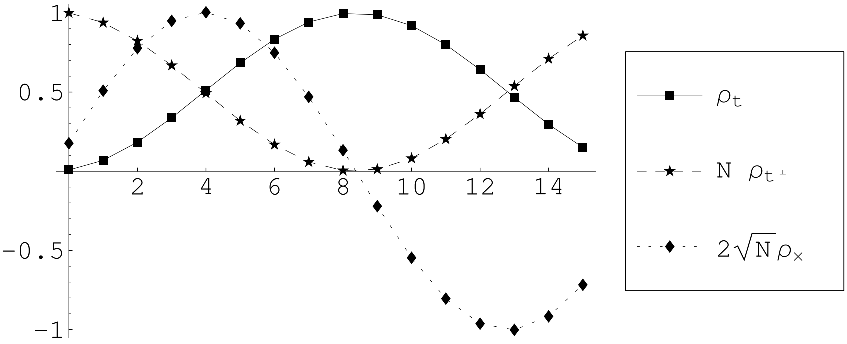


Figure 2.3: Example of undisturbed database search ($N = 128$)

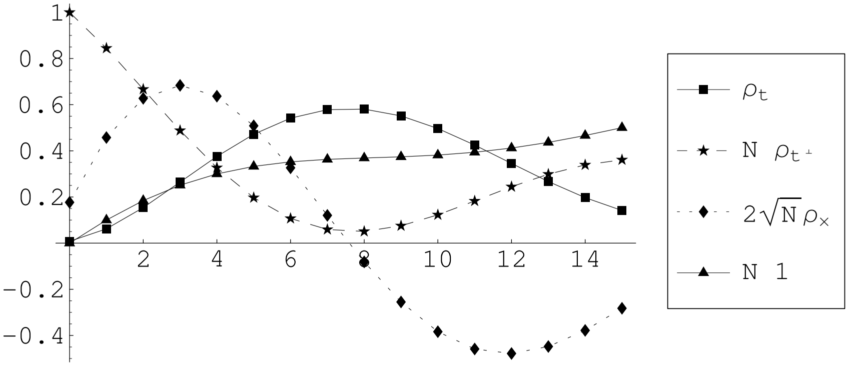


Figure 2.4: Example of database search disturbed by D_1 before each iteration ($N = 128$, $\lambda_1 = 0.1$)

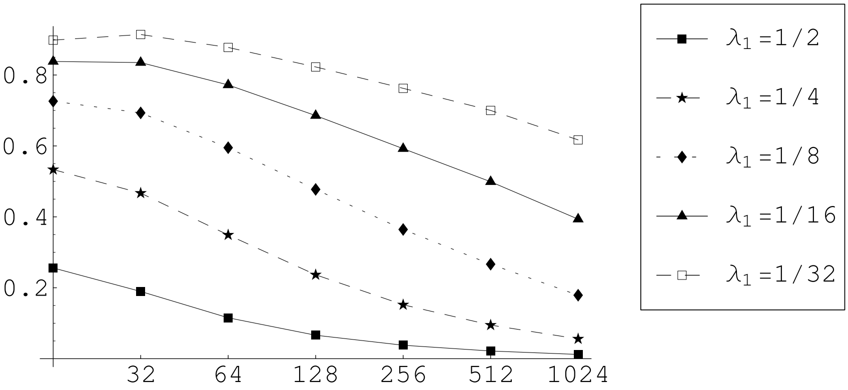


Figure 2.5: For the case of database search disturbed by D_1 before each iteration, plot of success probability after $(\pi/4)\sqrt{N}$ iterations against database size for several λ_1.

2.4 Energy Levels of a Hamiltonian

In this section we study an application of quantum search to physics. Recall from Subsection 1.2.2 that a time-independent Hamiltonian is a self-adjoint operator H that describes the evolution of a quantum system via $|\psi_t\rangle = e^{\mathrm{i}\,Ht}|\psi_0\rangle$. The eigenvalues of H are real and describe the "energy levels" of the system; the energy is a preserved quantity for every state. If an eigenspace of H has dimension greater than 1, we call it degenerate; the dimension of the eigenspace we call the degeneracy degree of the given energy level.

We study the following problem: suppose in an N-dimensional Hilbert space $\mathcal{H}$ we are given a Hamiltonian H with three energy levels, 0, E, and $E+d$. By k and ℓ we denote the degeneracy degree of E and $E+d$, respectively; we assume that they are much less than N, i.e., $k+\ell = \mathrm{o}(N)$. The goal is to sample states from level E (or from level $E+d$) and to determine which degree is larger as efficiently as possible for large N, $E=\pi$ fixed, and d fixed or a decreasing function of N. In unit time, an eigenstate of the Hamiltonian is unchanged at energy 0, acquires a phase of -1 at energy $E=\pi$, and a phase of $-e^{\mathrm{i}\,d}$ at energy $E=\pi+d$. Taking this evolution as a query operator, sampling amounts to quantum search with a phase error. Thus we extend quantum search beyond perfect phase flips; our case study is of different scope to the robustness construction of Høyer, Mosca, and de Wolf [75] since here our goal is to distinguish between the different phases.

	sampling	comparing
Method 1	$d = \mathrm{O}\left(\sqrt{\frac{k}{N}}\right)$	$d = \mathrm{O}\left(\frac{\lvert k-\ell\rvert}{\sqrt{(k+\ell)N}}\right)$
	time $\mathrm{O}\left(\frac{1}{d}\sqrt{\frac{N}{k}}\right)$	time $\mathrm{O}\left(\frac{\sqrt{(k+\ell)N}}{\lvert k-\ell\rvert}\right)$
Method 2	$d = \Omega\left(\sqrt{\frac{k+\ell}{N}}\right)$	$d = \Omega\left(\sqrt{\frac{k+\ell}{N}}\right)$
	time $\mathrm{O}\left(\sqrt{\frac{N}{k}}\right)$	time $\approx \mathrm{O}\left(\sqrt{\frac{N}{\min(k,\ell)}}\right)$

Table 2.2: Summary of results for Hamiltonian energy levels; the conditions on d indicate in what regime the corresponding time bounds hold.

We derive quantum algorithms for this problem based on Grover's search technique. Our results are summarized in Table 2.2; we used different approaches depending on whether d is small or large; for $\ell = \mathrm{o}(k)$, the full range of possible d is covered.

We use the following notation. Let $U := e^{-\mathrm{i}H}$ be the evolution operator of the system in unit time and let $U' := e^{-\mathrm{i}H\pi/(\pi+d)}$ be the evolution in time $\pi/(\pi+d)$. Let $\{|m\rangle : 0 \leq m < N\}$ be an orthonormal basis of the eigenstates of H, U, and U'. Let M_ϑ be the indices of the eigenstates with energy ϑ. Thus U multiplies $|m\rangle$ by $e^{\mathrm{i}\pi} = -1$ if $m \in M_\pi$, by $e^{\mathrm{i}(\pi+d)}$ if $m \in M_{\pi+d}$, and leaves it unchanged if $m \in M_0$. With $d' = -d + d^2/(\pi+d) = -d + \mathrm{O}(d^2)$, U' phase-shifts $|m\rangle$ by $e^{\mathrm{i}(\pi+d')}$ if $m \in M_\pi$, by -1 if $m \in M_{\pi+d}$, and leaves it unchanged otherwise.

2.4.1 Sampling from the energy levels

Sampling using Grover's search technique Building on Grover's search technique, we discuss in this subsection how to approximately generate a uniform superposition of the states $|m\rangle$ with $m \in M_\pi$ if k and ℓ are known. Uniformly sampling an $|m\rangle$ with $m \in M_\pi$ then amounts to measuring this superposition in the basis $\{|m\rangle : 0 \leq m < N\}$.

Grover's search algorithm consists of a number of repeated applications of the operator $G = T_0 S_t$ to the start state $W|0\rangle$. Here $W := H^{\otimes n}$ denotes again the Hadamard transformation applied to all $\log N$ qubits; S_t denotes the conditional phase-shift operator that acts on the computational basis by multiplying the phase of certain "marked" basis states by -1 and leaving the remaining basis states unchanged; T_0 is again the reflection about the average in the computational basis. In Section 2.1 we saw that with $\mathrm{O}(\sqrt{N})$ applications of G to $W|0\rangle$ it is possible to approximate the uniform superposition of the marked basis states.

In our setting, we do not have an operator S_t; U acts like S_t on the M_π and M_0 states but deviates on the $M_{\pi+d}$ states. We present ways to recover the properties of Grover's search algorithm when S_t is replaced by U. Note that using U' in place of U will yield essentially the same results with the role of M_π and $M_{\pi+d}$ interchanged.

Small energy difference First, we quickly discuss the case

$$d \leq \frac{\pi}{40}\sqrt{\frac{k}{N}} \ .$$

For $q \in \mathbb{N}$, the operator U^{2q+1} phase-shifts the M_π-states by -1, the $M_{\pi+d}$ states by $e^{\mathrm{i}(\pi+(2q+1)d)}$, and leaves the remaining states unchanged. So we can select a q that minimizes the impact of the $M_{\pi+d}$ states: with q the integer closest to $(\pi/d-1)/2$, the $M_{\pi+d}$ states get phase-shifted by $e^{\mathrm{i}d_0}$ with $|d_0| \leq d$. Hence in the operator norm, $\| U^{2q+1} - S_t \| \leq d$. Applying Grover search for $(\pi/4)\sqrt{N/k}$ steps with U^{2q+1} in place of S_t thus causes a total deviation from the ideal evolution of quantum search of $\mathrm{O}(d\sqrt{N/k}) < 1/3$ with high probability.

We now turn our attention to the case $d = \Omega(\sqrt{(k+\ell)/N})$, for which a much more efficient algorithm can be derived by showing that the $M_{\pi+d}$-states do not cause any noticeable disturbance.

Evolution in a three-dimensional subspace In Section 2.1 we derived that the evolution of the system under subsequent applications of G is confined to the two-dimensional subspace spanned by the uniform superposition of marked and unmarked states. In our setting, $G = T_0 U$; we derive the evolution of $W|0\rangle$ under repeated applications of G as a transformation in a *three*-dimensional subspace. Let

$$\begin{aligned} |\bar{0}\rangle &:= \frac{1}{\sqrt{N-k-\ell}} \sum_{m \in M_0} |m\rangle \ , \\ |\bar{k}\rangle &:= \frac{1}{\sqrt{k}} \sum_{m \in M_\pi} |m\rangle \ , \text{ and} \\ |\bar{\ell}\rangle &:= \frac{1}{\sqrt{\ell}} \sum_{m \in M_{\pi+d}} |m\rangle \ . \end{aligned}$$

For every $M \subseteq \{0, \ldots, N-1\}$, the reflection about the average T_0 acts as

$$T_0 : \sum_{m \in M} |m\rangle \mapsto \left(\frac{2|M|}{N} - 1 \right) \sum_{m \in M} |m\rangle + \frac{2|M|}{N} \sum_{m \notin M} |m\rangle \ ,$$

hence,

$$\begin{aligned} G|\bar{\ell}\rangle &= T_0 U |\bar{\ell}\rangle = -e^{\mathrm{i}\,d} T_0 |\bar{\ell}\rangle \\ &= e^{\mathrm{i}\,d} \left(\left(1 - \frac{2\ell}{N} \right) |\bar{\ell}\rangle - \frac{2\ell}{N} \left(\frac{\sqrt{N-k-\ell}}{\sqrt{\ell}} |\bar{0}\rangle + \frac{\sqrt{k}}{\sqrt{\ell}} |\bar{k}\rangle \right) \right) \\ &= -2e^{\mathrm{i}\,d} \frac{\sqrt{\ell(N-k-\ell)}}{N} |\bar{0}\rangle - 2e^{\mathrm{i}\,d} \frac{\sqrt{k\ell}}{N} |\bar{k}\rangle + e^{\mathrm{i}\,d} \left(1 - \frac{2\ell}{N} \right) |\bar{\ell}\rangle \ . \end{aligned}$$

Similar calculations give rise to a matrix

$$R = \begin{pmatrix} 1 - \frac{2(k+\ell)}{N} & -2\frac{\sqrt{k(N-k-\ell)}}{N} & -2e^{\mathrm{i}\,d}\frac{\sqrt{\ell(N-k-\ell)}}{N} \\ 2\frac{\sqrt{k(N-k-\ell)}}{N} & 1 - \frac{2k}{N} & -2e^{\mathrm{i}\,d}\frac{\sqrt{k\ell}}{N} \\ 2\frac{\sqrt{\ell(N-k-\ell)}}{N} & -\frac{2\sqrt{k\ell}}{N} & e^{\mathrm{i}\,d}\left(1 - \frac{2\ell}{N}\right) \end{pmatrix}$$

so that the evolution of the system starting in state

$$W|0\rangle = \sqrt{\frac{N-k-\ell}{N}}|\bar{0}\rangle + \sqrt{\frac{k}{N}}|\bar{k}\rangle + \sqrt{\frac{\ell}{N}}|\bar{\ell}\rangle$$

to which G is applied repeatedly can be expressed as a transformation in the 3-dimensional subspace of $\mathcal{H}_N$ spanned by $|\bar{0}\rangle$, $|\bar{k}\rangle$, and $|\bar{\ell}\rangle$:

$$G(a|\bar{0}\rangle + b|\bar{k}\rangle + c|\bar{\ell}\rangle) = \begin{pmatrix}|\bar{0}\rangle & |\bar{k}\rangle & |\bar{\ell}\rangle\end{pmatrix} R \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

Discarding in R terms that are $\mathrm{O}((k+\ell)/N)$ and substituting $x := 2\sqrt{k/N}$, $y := 2\sqrt{\ell/N}$, and $v := e^{\mathrm{i}\,d}$, we get $R = \tilde{R} + \mathrm{O}((k+\ell)/N)$ with

$$\tilde{R} = \begin{pmatrix} 1 & -x & -vy \\ x & 1 & 0 \\ y & 0 & v \end{pmatrix} \ .$$

Here $R = \tilde{R} + \mathrm{O}((k+\ell)/N)$ is shorthand for $\|R - \tilde{R}\| = \mathrm{O}((k+\ell)/N)$ in the operator norm.

Finding the eigenvalues To find the eigenvalues of $\tilde{R}$ we consider its characteristic polynomial $p(\lambda) = \det(\tilde{R} - \lambda\mathbb{1})$. It has the form

$$p(\lambda) = (\lambda - 1 + \mathrm{i}\,x)(\lambda - 1 - \mathrm{i}\,x)(\lambda - v) + vy^2(\lambda - 1) \ . \tag{2.11}$$

We show that $\tilde{\lambda}_1 = 1 - \mathrm{i}\,x$, $\tilde{\lambda}_2 = 1 + \mathrm{i}\,x$ and $\tilde{\lambda}_3 = v$ are the zeroes of $p(\lambda)$ up to order $1/(dN)$, i.e., there exist roots λ_1, λ_2, λ_3 of $p(\lambda)$ such that $\lambda_k = \tilde{\lambda}_k + \mathrm{O}(1/(dN))$.

By the definition of the derivative and the inverse-function theorem from elementary calculus,

$$\begin{aligned} p^{-1}(0) &= p^{-1}(h) - (p^{-1})'(h) \cdot h + \mathrm{o}(h) \\ &= p^{-1}(h) - \frac{h}{p'(p^{-1}(h))} + \mathrm{o}(h) \ , \end{aligned}$$

that is, for $h = p(\tilde{\lambda}_k)$,

$$\lambda_k = \tilde{\lambda}_k - \frac{p(\tilde{\lambda}_k)}{p'(\tilde{\lambda}_k)} + \mathrm{o}(p(\tilde{\lambda}_k)) \ .$$

From Eq. (2.11) we have $p(\tilde{\lambda}_k) = vy^2(\tilde{\lambda}_k - 1)$, thus $p(\tilde{\lambda}_{1,2}) = \mathrm{O}(1/N^{3/2})$ and $p(\tilde{\lambda}_3) = \mathrm{O}(d/N)$. Moreover,

$$p'(\tilde{\lambda}_{1,2}) = \pm 2\,\mathrm{i}(v-1)x - 2x^2 + vy^2 = \Omega\left(\frac{d}{\sqrt{N}}\right) \qquad \text{and}$$
$$p'(\tilde{\lambda}_3) = 1 + v^2 + x^2 + v(y^2 - 2) = \Omega(1) \ .$$

Altogether, $\lambda_{1,2} = \tilde{\lambda}_{1,2} + \mathrm{O}(1/(dN))$ and $\lambda_3 = \tilde{\lambda}_3 + \mathrm{O}(d/N) = \tilde{\lambda}_3 + \mathrm{O}(1/(dN))$.

Finding the eigenvectors Let $\gamma := (k+\ell)/N$ and denote eigenvectors of $\tilde{R}$ by $\bar{a} = (a, b, w)$. We assume that they are of unit length: $a^2+b^2+w^2 = 1$. The system of linear equations $\bar{a}(\tilde{R} - \lambda \mathbb{1}) = 0$ for finding approximate eigenvectors up to $\mathrm{O}(\gamma)$ has for $\tilde{\lambda}_1$ the form

$$\begin{cases} \mathrm{i}\,xa & +xb & & +yw & = \mathrm{O}(\gamma) \\ -xa & +\mathrm{i}\,xb & & & = \mathrm{O}(\gamma) \\ -vya & & & +(v-1+\mathrm{i}\,x)w & = \mathrm{O}(\gamma) \end{cases}$$

It has the solution $a = -1 + \mathrm{o}(1)$, $b = \mathrm{i} + \mathrm{o}(1)$, $w = \mathrm{o}(1)$. For the second root, $\tilde{\lambda}_2$, the corresponding equations yield $a = 1 + \mathrm{o}(1)$, $b = \mathrm{i} + \mathrm{o}(1)$, $w = \mathrm{o}(1)$. For the third root, $\tilde{\lambda}_3$, we obtain $a = \mathrm{o}(1)$, $b = \mathrm{o}(1)$, $w = 1 + \mathrm{o}(1)$.

Comparing this with the two-dimensional quantum-search iteration,

$$G = \frac{1}{N}\begin{pmatrix} N-2M & 2\sqrt{M(N-M)} \\ -2\sqrt{M(N-M)} & N-2M \end{pmatrix} = \begin{pmatrix} \mathrm{i} & -\mathrm{i} \\ 1 & 1 \end{pmatrix} D \begin{pmatrix} -\mathrm{i} & 1 \\ \mathrm{i} & 1 \end{pmatrix} \text{ with}$$
$$D = \begin{pmatrix} 1 - 2\frac{M}{N} - 2\,\mathrm{i}\sqrt{\frac{M}{N}\left(1 - \frac{M}{N}\right)} & 0 \\ 0 & 1 - 2\frac{M}{N} + 2\,\mathrm{i}\sqrt{\frac{M}{N}\left(1 - \frac{M}{N}\right)} \end{pmatrix} ,$$

we see that the eigenvalues are up to $\mathrm{O}(\gamma)$ the same and the eigenvectors coincide up to terms of $\mathrm{o}(1)$. This means that for up to $\mathrm{o}(1/\gamma)$ iterations, the behavior of our algorithm can be approximated by the behavior of Grover's algorithm.

2.4.2 Comparing degeneracy degrees

In this subsection we apply the quantum approximate counting technique by Brassard, Høyer, and Tapp [28] to our setting:

2.4.1. Lemma (Theorem 5 of [28]). *Let $F : [N] \to \{0,1\}$ be a Boolean function, $t = |F^{-1}(1)| < N/2$, and $P \in \mathbb{N}$ with $0 < P \le N$. There is a quantum algorithm* **Count**(F, P) *whose output $\tilde{t}$ satisfies*

$$|t - \tilde{t}| \le \frac{2\pi}{P}\sqrt{tN} + \frac{\pi^2}{P^2}N \ .$$

Furthermore, **Count**(F, P) *makes* P *quantum queries to* F.

As before, we study two cases, namely that d is small enough to construct a good approximation of S_t for the M_π states and that d is so large that it does not influence

Small energy difference The same construction as for sampling gives us an approximation U^{2q+1} of S_t with $\| U^{2q+1} - S_t \| \leq d$. Hence, with $P = \mathrm{O}(1/d)$ the algorithm **Count**(F, P) with U^{2q+1} in place of S_t will still work with constant probability. To compare the degeneracy degrees $\ell = |M_\pi|$ and $k = |M_{\pi+d}|$, we obtain an approximate count $\tilde{\ell}$ and, with U' in place of U, an approximation $\tilde{k}$. Sufficient conditions for the comparison to succeed with constant probability are

$$|\ell - \tilde{\ell}| < \frac{1}{2}|k - \ell| \qquad \text{and}$$
$$|k - \tilde{k}| < \frac{1}{2}|k - \ell| \ .$$

These are satisfied if we choose P so that

$$\frac{2\pi}{P}\sqrt{(k+\ell)N} + \frac{\pi^2}{P^2}N < \frac{1}{2}|k - \ell| \ ,$$

or

$$P = \Theta\left(\frac{\sqrt{(k+\ell)N}}{|k - \ell|}\right) \ .$$

Large energy difference If d is large enough to allow P iterations of search with only constant total deviation, then we can just use U and U' in place of S_t. In the previous subsection we showed that for $d = \Omega(\sqrt{(k+\ell)/N})$ we can execute as many as $\mathrm{o}(N/(k+\ell))$ iterations of search with U and U'. With $P_k = \omega(\sqrt{N/k})$ and $P_\ell = \omega(\sqrt{N/\ell})$, respectively, we obtain approximations $\tilde{k}$ and $\tilde{\ell}$ with

$$|k - \tilde{k}| + |\ell - \tilde{\ell}| = \mathrm{o}(1)$$

so that asymptotically we can detect any difference between k and ℓ. This takes time $\mathrm{O}(\sqrt{N/\min(k,\ell)}r(N))$ where $r(N) = \omega(1)$ is an unbounded and arbitrarily slow growing function.

2.4.3 Numerical simulations

To complement our theoretical results, we simulated the sampling algorithms from Subsection 2.4.1 with one state that is rotated by $e^{\mathrm{i}\,\pi}$ and one state that is rotated by $e^{\mathrm{i}(\pi+d)}$. Figures 2.6 and 2.7 show the probability of finding the state rotated by $e^{\mathrm{i}\,\pi}$ as a function of the dimension of the state space. In Figure 2.6, d is chosen independently of N. Observe that for small d and small N, the success probability is about 1/2. This is because in this regime, the system evolves as search with two target states that are rotated by $e^{\mathrm{i}\,\pi}$: the probability that we hit the desired of the two target states is 1/2. With growing N, the success probability converges to 1, as theoretically predicted—the state rotated by $e^{\mathrm{i}(\pi+d)}$ causes negligible distortion. The graph suggests that the "speed" of convergence depends linearly on d: given $\varepsilon > 0$, the smallest N for which the success probability is greater than $1 - \varepsilon$ appears to be a linear function of d. Figure 2.7 illustrates the case that d is a function of N. Our analysis that for $d = \omega(1/\sqrt{N})$ the success probability will converge to 1, is mirrored by the curves for $a < 1/2$ appearing to converge to 1. We do not have an analytical result for $d = \Theta(1/\sqrt{N})$ or $a = 1/2$, but the graph suggests that the success probability does not rise above 1/2.

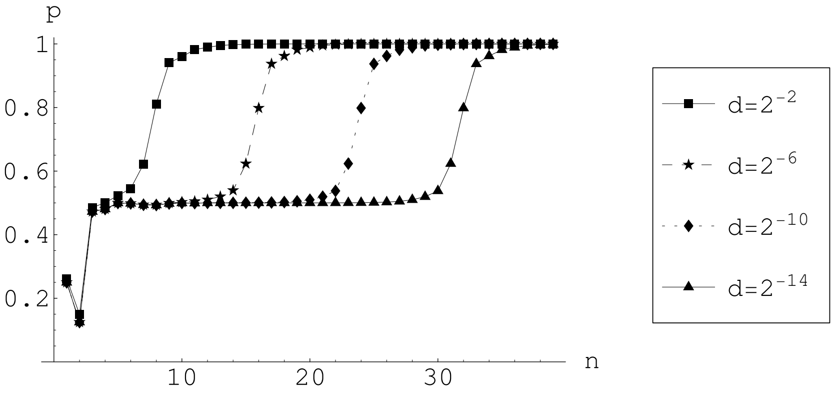


Figure 2.6: Success probability for finding one of the $|\bar{k}\rangle$ states for d constant: Plot of success probability against dimension $N = 2^n$ for $n = 2, \ldots, 40$ and $d = 2^{-2}$, 2^{-6}, 2^{-10}, 2^{-14}.

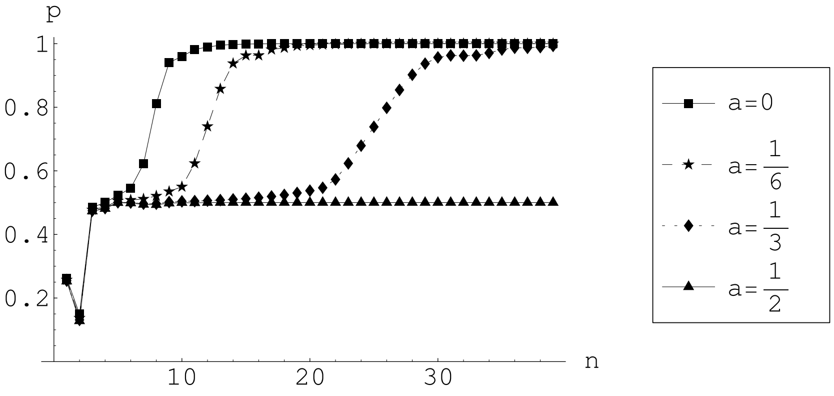


Figure 2.7: Success probability for finding one of the $|\bar{k}\rangle$ states for d a function of N: Plot of success probability against dimension $N = 2^n$ for $n = 2, \ldots, 40$ and $d = 2^{-2-an}$ where $a = 0$, $1/6$, $1/3$, and $1/2$.

Chapter 3

Property Testing

This chapter is based on joint work with Buhrman, Fortnow, and Newman [33].

3.1 Introduction

Suppose we have a large data set, for example, a large chunk of the worldwide web or a genomic sequence. We would like to test whether the data has a certain property, but we may not have the time to look at the entire data set or even a large portion of it.

To handle these types of problems, Rubinfeld and Sudan [103] and Goldreich, Goldwasser and Ron [65] have developed the notion of *property testing*. Testable properties come in many varieties including graph properties, e.g., [65, 7, 57, 58, 5, 66], algebraic properties of functions [23, 103, 51], and regular languages [8]. Nice surveys of this area can be found in [102] [56].

In this model, the property tester has random access to the n input bits similar to the black-box oracle model. The tester can query only a small number of input bits; the set of indices is usually of constant size and chosen probabilistically. Clearly we cannot determine from this small number of bits whether the input sits in some language L. However, for many languages we can distinguish the case that the input is in L from the case that the input differs from all inputs in L of the same length by some constant fraction of input bits.

Since there are many examples where quantum computation gives us an advantage over classical computation [22, 109, 108, 69] one may naturally ask whether using quantum computation may lead to better property testers. By using the quantum oracle-query model we can easily extend the definitions of property testing to the quantum setting.

Beals, Buhrman, Cleve, Mosca, and de Wolf [15] have shown that for all total functions we have a polynomial relationship between the number of queries required by quantum machine and that needed by a deterministic machine. For greater separations one needs to impose a promise on the input. The known examples, such as those due to Simon [109] and Bernstein and Vazirani [22], require considerable structure in the promise. Property testing amounts to the natural promise of either being in the language or far from each input in the language. This promise would seem to have too little structure to give a separation but in fact we can prove that quantum property testing can greatly improve on classical testing.

We show that every subset of Hadamard codes has a quantum property tester with $\mathrm{O}(1)$ queries and that most subsets would require $\Theta(\log n)$ queries to test with a probabilistic tester. This shows that indeed quantum property testers are more powerful than classical testers. Moreover, we also give an example of a language where the quantum tester is exponentially more efficient.

Beals, Buhrman, Cleve, Mosca, and de Wolf [15] observed that every k-query quantum algorithm gives rise to a degree-$2k$ polynomial in the input bits, which gives the acceptance probability of the algorithm; thus, a quantum property tester for P gives rise to a polynomial that is on all binary inputs between 0 and 1, that is at least $2/3$ on inputs with the property P and at most $1/3$ on inputs far from having the property P. Szegedy [114] suggested to algebraically characterize the complexity of classical testing by the minimum degree of such polynomials; however, our separation results imply that there are for example properties, for which such polynomials have constant degree, but for which the best classical tester needs $\Omega(\log n)$ queries. Hence, the minimum degree is only a lower bound, which sometimes is not tight.

A priori it is conceivable that every language has a quantum property tester with a small number of queries. We show that this is not the case. We prove that for most properties of a certain size, every quantum algorithm requires $\Omega(n)$ queries. We then show that a natural explicit property, namely, the range of a d-wise independent pseudorandom generator cannot be quantumly tested with less than $(d+1)/2$ queries for every odd $d \leq n/\log n - 1$.

3.2 Preliminaries

We will use the following formal definition of property testing from Goldreich [64]:

3.2.1. Definition. Let S be a finite set, and P a set of functions mapping S to $\{0,1\}$. A *property tester* for P is a probabilistic oracle machine M, which given a distance parameter $\varepsilon > 0$ and oracle access to a function $f : S \to \{0,1\}$, satisfies the following conditions:

1. the tester accepts f if it is in P: if $f \in P$ then $\Pr(M^f(\varepsilon) = 1) \geq 2/3$
2. the tester rejects f if it is far from P: if $|\{x \in S : f(x) \neq g(x)\}| > \varepsilon \cdot |S|$, for every $g \in P$, then $\Pr(M^f(\varepsilon) = 1) \leq 1/3$.

Here M^f denotes that the machine M is provided with the oracle for f.

3.2.2. Definition. The complexity of the tester is the number of oracle queries it makes: A property P has an (ε, q)-*tester* if there is a tester for P that makes at most q oracle queries for distance parameter ε.

We often consider a language $L \subseteq \{0,1\}^*$ as the family of properties $\{P_n\}$ with P_n the characteristic functions of the length-n strings from L, and analyze the query complexity $q = q(\varepsilon, n)$ asymptotically for large n. We say L is ε-*testable* with $q(n)$ queries, if for each n, P_n has a $(\varepsilon, q(n))$ tester.

To define quantum property testing we simply modify Definition 3.2.1 by allowing M to be a quantum oracle machine.

3.3 Separating Quantum and Classical Property Testing

We show that there exist languages with $(\varepsilon, \mathrm{O}(1))$ quantum property testers that do not have $(\varepsilon, \mathrm{O}(1))$ classical testers.

3.3.1. Theorem. *There is a language L that is ε-testable by a quantum test with $\mathrm{O}(1/\varepsilon)$ number of queries but for which every probabilistic $1/3$-test requires $\Omega(\log n)$ queries.*

We use Hadamard codes to provide examples for Theorem 3.3.1:

3.3.2. Definition. The *Hadamard code* of $y \in \{0,1\}^{\log n}$ is $x = h(y) \in \{0,1\}^n$ such that $x_i = y \cdot i$ where $y \cdot i$ denotes the inner product of two vectors $y, i \in \mathbb{F}_2^{\log n}$.

Note: the Hadamard mapping $h : \{0,1\}^{\log n} \to \{0,1\}^n$ is one-to-one. Bernstein and Vazirani [22] showed that a quantum computer can extract y with one query to an oracle for the bits of x, whereas a classical probabilistic procedure needs $\Omega(\log n)$ queries. Based on this separation for a decision problem we construct for $A \subseteq \{0,1\}^{\log n}$ the property $P_A \subseteq \{0,1\}^n$,

$$P_A := \{x : \exists y \in A \text{ s.t. } x = h(y)\}.$$

Theorem 3.3.1 follows from the following two lemmas.

3.3.3. Lemma. *For every A, P_A has an $(\varepsilon, \mathrm{O}(1/\varepsilon))$ quantum tester. Furthermore, the test has one-sided error.*

3.3.4. Lemma. *For most A of size $|A| = n/2$, P_A requires $\Omega(\log n)$ queries for a probabilistic $1/3$-test, even for testers with two-sided error.*

Before we prove Lemma 3.3.3 we note that for every A, P_A can be tested by a one-sided algorithm with $\mathrm{O}(1/\varepsilon + \log n)$ queries even nonadaptively; hence, the result of Lemma 3.3.4 is tight. An ε-test with $\mathrm{O}((\log n)/\varepsilon)$ queries follows from Theorem 3.3.5 below. The slightly more efficient test is the following: First we query x_{2^i}, $i = 1, \dots, \log n$. Note that if $x = h(y)$ then $y_i = x_{2^i}$ for $i = 1, \dots, \log n$. Thus a candidate y for $x = h(y)$ is found. If $y \notin A$ then x is rejected. Then $k := \mathrm{O}(1/\varepsilon)$ times the following check is performed: a random index $i \in \{1, \dots, n\}$ is chosen independently at random and if $x_i \neq y \cdot i$, then x is rejected. Otherwise, x is accepted. Clearly if x is rejected then $x \notin P_A$. It is easily verified that if x has Hamming distance more than εn from every z in P_A then with constant probability x is rejected.

Proof of Lemma 3.3.3. P_A can be checked with $\mathrm{O}(1/\varepsilon)$ queries on a quantum computer: The test is similar to the test above except that y can be found in $\mathrm{O}(1)$ queries: k times query for random i, j values x_i, x_j, and $x_{i \oplus j}$. If $x_i \oplus x_j \neq x_{i \oplus j}$ reject. $k = \mathrm{O}(1/\varepsilon)$ is sufficient to detect an input x that is εn-far from being a Hadamard codeword with high probability. Now run the Bernstein-Vazirani algorithm to obtain y. Accept if and only if $y \in A$. Obviously, if $x \in P_A$, the given procedure accepts, and if x is far from each $x' \in P_A$, then it is either far from being a Hadamard codeword or it is close to a Hadamard codeword $h(y')$ for a $y' \notin A$; note that in this case x is far from every $h(y)$, $y \in A$ as two distinct Hadamard codewords are of Hamming distance $n/2$. Thus, in this case the second part of the tester succeeds with high probability in finding y' and rejects because $y' \notin A$. We note also that this algorithm has one-sided error. □

Proof of Lemma 3.3.4. The lower bound makes use of the Yao principle [118]: let D be an arbitrary probability distribution on *positive* and *negative* inputs, i.e., on inputs that either belong to P_A or are εn-far from P_A. Then if every deterministic algorithm that makes at most q queries, errs with probability at least $1/8$ with respect to input chosen according to D, then q is a lower bound on the number of queries of any randomized algorithm for testing P_A with error probability bounded by $1/8$.

D will be the uniform distribution over Hadamard codewords of length n, namely, generated by choosing $y \in \{0,1\}^{\log n}$ uniformly at random and setting $x = h(y)$. Note that for any $A \subset \{0,1\}^{\log n}$, D is concentrated on positive and negative inputs as required, as two Hadamard codewords are of Hamming distance $n/2$ apart.

The lower bound will be established by a counting argument. We show that for a fixed tester that makes $q \leq (\log n)/2$ queries, the probability over random choices of A that the algorithm errs on at most $1/8$ of the inputs is bounded from above by $1/(10T)$ where T is the number of such algorithms. By the union bound it follows that for most properties there is no such algorithm.

Indeed, let $A \subseteq \{0,1\}^{\log n}$ be chosen by picking independently each $i \in \{0,1\}^{\log n}$ to be in A with probability $1/2$; this will not necessarily result in a set A of size $n/2$ but we can condition on the event that $|A| = n/2$ and will not lose much. Let $\mathcal{T}$ be any fixed deterministic decision tree performing at most q queries in every branch. Then let $c(\mathcal{T}) := \{y|\mathcal{T}(h(y)) = \text{accept}\}$ and let $\mu(\mathcal{T}) := |c(\mathcal{T})|/n$, i.e., $\mu(\mathcal{T})$ is the fraction of inputs that $\mathcal{T}$ accepts. Assume first that $\mu(\mathcal{T}) \leq 1/2$. Since for a random y we have $\Pr_y[\mathcal{T}(h(y)) = \text{accept}] = \mu(\mathcal{T}) \leq 1/2$, it follows by a Chernoff-type bound that $\Pr_A[|A \cap c(\mathcal{T})| \geq (3/4)|A|] \leq 2^{-n/8}$. However, if $|A \cap c(\mathcal{T})| < (3/4)|A|$ then $\mathcal{T}$ will be wrong on at least $1/4$ of the positive inputs which is at least $n/8$ of all inputs. Hence, with probability at most $2^{-n/8}$, $\mathcal{T}$ will be correct on at least $7/8$ of the inputs. If $\mu(\mathcal{T}) > 1/2$ the same reasoning shows that with probability of at most $1 - 2^{-n/8}$ it will err on at least a $1/4$-fraction of the negative inputs. Hence, in total, for every fixed $\mathcal{T}$, $\Pr_A[\mathcal{T}$ is correct on at least $7/8$ of the inputs$] \leq 2^{-n/8}$.

Now, let us bound from above the number of algorithms that make at most q queries. As an algorithm may be adaptive, it can be defined by $2^q - 1$ query positions for all queries on all branches and a Boolean function $f : \{0,1\}^q \to \{\text{accept}, \text{reject}\}$ of the decision made by the algorithm for the possible answers. Hence, there are at most $T \leq (2n)^{2^q}$ such algorithms. However, for $q < (\log n)/2$, we have $T \cdot 2^{-n/8} = \mathrm{o}(1)$, which shows that for most A as above, every ε-test that queries at most $(\log n)/2$ many queries has error probability of at least $1/8$. Standard amplification techniques then imply that for some constant c every algorithm that performs $c \log n$ many queries has error at least $1/3$. □

3.3.5. Theorem. *Let $P \subseteq \{0,1\}^n$ be a property with $|P| = s > 0$. For any $\varepsilon > 0$, P can be ε-tested by a one-sided classical algorithm using* $\mathrm{O}((\log s)/\varepsilon)$ *many queries.*

Proof. Denote the input by $y \in \{0,1\}^n$. Consider the following algorithm: query the input y in $k := \ln(3s^2)/\varepsilon$ random places; accept if there is at least one $x \in P$ consistent with the bits from the input and reject otherwise. Clearly, if $y \in P$, this algorithm works correctly.

If y is ε-far from each $x \in P$, then for every specific $x \in P$, $\Pr[x_i = y_i] \leq 1 - \varepsilon$ when choosing an $i \in [n]$ uniformly at random. With k indices chosen independently and uniformly at random, the probability for no disagreement with x becomes $(1-\varepsilon)^k \leq 1/(3s^2)$. Therefore, the probability that there is no disagreement for at least one of the s members of P is at most $1/(3s)$, so with probability $2/3$ for a y that is far from P, we will rule out every $x \in P$ as being consistent with y. □

3.4 An Exponential Separation

In this section, we show that a quantum computer can be exponentially more efficient in testing certain properties than a classical computer.

3.4.1. Theorem. *There exists a language L that for every $\varepsilon = \Omega(1)$ is $(\varepsilon, \log n \, \log\log n)$ quantumly testable but every probabilistic $1/8$-test for L requires $n^{\Omega(1)}$ queries.*

The language that we provide is inspired by Simon's problem [109] and our quantum testing algorithm makes use of Brassard and Høyer's algorithm for Simon's problem [26]. Simon's problem is to find $s \in \{0,1\}^n \setminus \{0^n\}$ from a function-query oracle for some $f : \{0,1\}^n \to \{0,1\}^n$, such that $f(x) = f(y) \Leftrightarrow x = y \oplus s$. Simon proved that classically, $\Omega(2^{n/2})$ queries are required on average to find s, and gave a quantum algorithm for determining s with an expected number of queries that is polynomial in n; Brassard and Høyer improved the algorithm to worst-case polynomial time. Their algorithm produces in each run a z with $z \cdot s = 0$ that is linearly independent to all previously computed such zs. Essentially, our quantum tester uses this subroutine to try to extract information about s until it fails repeatedly. Høyer [74] and also Friedl et al. [61] analyzed this approach in group-theoretic terms, obtaining an alternative proof to Theorem 3.4.3.

In the following, let $N = 2^n$ denote the length of the binary string encoding a function $f : \{0,1\}^n \to \{0,1\}$. For $x \in \{0,1\}^n$ let $x[j]$ be the jth bit of x, i.e., $x = x[1] \ldots x[n]$. We define

$$L := \{f \in \{0,1\}^N : \exists s \in \{0,1\}^n \setminus \{0^n\} \; \forall x \in \{0,1\}^n \; f(x) = f(x \oplus s)\}$$

Theorem 3.4.1 follows from the following two theorems.

3.4.2. Theorem. *Every classical $1/8$-tester for L must make $\Omega(\sqrt{N})$ queries, even when allowing two-sided error.*

3.4.3. Theorem. *There is a quantum property tester for L making $\mathrm{O}(\log N \, \log\log N)$ queries. Moreover, this quantum property tester makes all its queries nonadaptively.*

Proof of Theorem 3.4.2. We again apply the Yao principle [118] as in the proof of Lemma 3.3.4: we construct two distributions, P and U, on positive and at least $N/8$-far negative inputs, respectively, such that every deterministic adaptive decision tree $\mathcal{T}$ with few queries has error $1/2 - \mathrm{o}(1)$ when trying to distinguish whether an input is chosen from U or P. Indeed, we will show a stronger statement: Let $\mathcal{T}$ be any deterministic decision tree. Let v be a vertex of $\mathcal{T}$. Let $\Pr_P(v)$ and $\Pr_U(v)$ be the probability that an input chosen according to P and U, respectively, is consistent with v. We will show that for every vertex v of $\mathcal{T}$ we have $|\Pr_P(v) - \Pr_U(v)| = \mathrm{o}(1)$; hence, $\mathcal{T}$ has error $1/2 - \mathrm{o}(1)$ if with probability $1/2$ we choose v according to P and with probability $1/2$ from U.

The distribution P is defined as follows: We first choose $s \in \{0,1\}^n$ at random. This defines a matching M_s of $\{0,1\}^n$ by matching x with $x \oplus s$. Now a function f_s is defined by choosing for each matched pair independently $f_s(x) = f_s(x \oplus s) = 1$ with probability $1/2$ and $f_s(x) = f_s(x \oplus s) = 0$ with probability $1/2$. Clearly, this defines a distribution that is concentrated on positive inputs. Note that it might be that by choosing different s's we end up choosing the same function, however, these functions will be considered different events in the probability space. Namely, the atomic events in P really are the pairs (s, f_s) as described above.

Now let U be the uniform distribution over all functions, namely, we select the function by choosing for each x independently $f(x) = 1$ with probability $1/2$ and 0 with probability $1/2$. Since every function has a nonzero probability, U is not supported exclusively on the negative instances. However, as we proceed to show, a function chosen according to U is $N/8$-far from having the property with very high probability, and hence U will be a good approximation to the desired distribution:

3.4.4. Definition. For $f : \{0,1\}^n \to \{0,1\}$ and $s \in \{0,1\}^n$ we define $n_s := |\{x : f(x) = f(x \oplus s)\}|$.

3.4.5. Lemma. *Let f be chosen according to U. Then* $\Pr_U[\exists s \in \{0,1\}^n : n_s \geq N/8] \leq e^{-\Omega(N)}$.

Proof. Let f be chosen according to U and $s \in \{0,1\}^n$. By a Chernoff bound we obtain $\Pr_U[n_s \geq N/8] \leq e^{-\Omega(N)}$. Together with the union bound over all s's this yields $\Pr_U[\exists s \in \{0,1\}^n : n_s \geq N/8] \leq 2^n \cdot e^{-\Omega(N)} \leq e^{-\Omega(N)}$. □

In particular, a direct consequence of Lemma 3.4.5 is that with probability $1 - e^{-\Omega(N)}$ an input chosen according to U will be $N/8$-far from having the property.

From the definition of U, we immediately obtain the following:

3.4.6. Lemma. *Let $\mathcal{T}$ be any fixed deterministic decision tree and let v be a vertex of depth d in $\mathcal{T}$. Then* $\Pr_U[f$ *is consistent with the path to* $v] = 2^{-d}$.

We now want to derive a similar bound as in the lemma for functions chosen according to P. For this we need the following definition for the event that after d queries, nothing has been learned about the hidden s:

3.4.7. Definition. Let $\mathcal{T}$ be a deterministic decision tree and u a vertex in $\mathcal{T}$ at depth d. We denote the path from the root of $\mathcal{T}$ to u by path(u). Every vertex v in $\mathcal{T}$ defines a query position $x_v \in \{0,1\}^n$. For $f = f_s$ chosen according to P, we denote by B_u the event $B_u := \{(s, f_s) : s \neq x_v \oplus x_w \text{ for all } v, w \in \text{path}(u)\}$.

3.4.8. Lemma. *Let v be a vertex of depth d in a decision tree $\mathcal{T}$. Then* $\Pr_P[B_v] \geq 1 - \binom{d-1}{2}/N$

Proof. B_v does not occur if for some v, w on the path to v we have $s = x_v \oplus x_w$. As there are $d-1$ such vertices, there are at most $\binom{d-1}{2}$ pairs. Each of these pairs excludes exactly one s and there are N possible s's. □

3.4.9. Lemma. *Let v be a vertex of depth d in a decision tree $\mathcal{T}$ and let f be chosen according to P. Then* $\Pr_P[f$ *is consistent with* $v|B_v] = 2^{-d}$.

Proof. By the definition of P, f gets independently random values on vertices that are not matched. But if B_v occurs, then no two vertices along the path to v are matched and hence the claim follows. □

Now we can complete the proof of the theorem: assume that $\mathcal{T}$ is a deterministic decision tree of depth $d = \mathrm{o}(\sqrt{N})$ and let v be any leaf of $\mathcal{T}$. Then by Lemmas 3.4.8 and 3.4.9, we get that $\Pr_P[f$ is consistent with $v] = (1-\mathrm{o}(1))2^{-d}$. On the other hand, let U' be the distribution on negative inputs defined by U conditioned on the event that the input is at least $N/8$-far from the property. Then by Lemmas 3.4.5 and 3.4.6 we get that $\Pr_{U'}[f$ is consistent with $v] = (1-\mathrm{o}(1))2^{-d}$ and hence $\mathcal{T}$ has only o(1) bias of being right on *every* leaf. This implies that its error probability is $1/2 - \mathrm{o}(1)$. □

Proof of Theorem 3.4.3. We give a quantum algorithm making $\mathrm{O}(\log N \log\log N)$ queries to the quantum oracle for input $f \in \{0,1\}^N$. We will show that it accepts with probability 1 if $f \in L$ and rejects with high probability if the Hamming distance between f and every $g \in L$ is at least εN. Pseudo code for our algorithm is given on page 65; it consists of a classical main program SimonTester and a quantum subroutine SimonSampler adapted from Brassard and Høyer's algorithm for Simon's problem [26, Section 4]. The

Procedure SimonTester

```
for $k = 0$ to $n - 1$ do
  $l \leftarrow 0$
  repeat
    $z \leftarrow$ SimonSampler($z_1, \dots, z_k$)
    $l \leftarrow l + 1$
  until $z \neq 0$ or $l > 2(\log n)/\varepsilon^2$
  if $z = 0$ then
    accept
  else
    $z_{k+1} \leftarrow z$
reject
```

Procedure SimonSampler($z_1, \dots, z_k$)

```
input: $z_1, \dots, z_k \in \{0,1\}^n$
output: $z \in \{0,1\}^n$
quantum workspace: $\mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{Z}$ where
$\mathcal{X}$ is $n$ qubits $\mathcal{X} = \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$, $\mathcal{X}_i = \mathbb{C}^2$,
$\mathcal{Y} = \mathbb{C}^2$ is one qubit, and
$\mathcal{Z}$ is $k$ qubits $\mathcal{Z} = \mathcal{Z}_1 \otimes \cdots \otimes \mathcal{Z}_k$, $\mathcal{Z}_j = \mathbb{C}^2$
initialize the workspace to $|0^n\rangle|0\rangle|0^k\rangle$
apply $H_{2^n}$ to $\mathcal{X}$
apply $U_f$ to $\mathcal{X} \otimes \mathcal{Y}$
apply $H_{2^n}$ to $\mathcal{X}$
for $j = 1$ to $k$ do
  $i \leftarrow \min\{i : z_j[i] = 1\}$
  apply CNOT with control $\mathcal{X}_i$ and target $\mathcal{Z}_j$
  apply $|x\rangle \mapsto |x \oplus z_j\rangle$ to $\mathcal{X}$ conditional on $\mathcal{Z}_j$
  apply $H_2$ to $\mathcal{Z}_j$
return measurement of $\mathcal{X}$
```

quantum gates used are the 2^n-dimensional Hadamard transform H_{2^n}, which applies

$$\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}$$

individually to each of n qubits, the quantum oracle query U_f, and classical reversible operations run in quantum superposition.

The following technical lemma captures the operation of the quantum subroutine SimonSampler. For $i_1, \ldots, i_J$ fixed, let $Y_J := \{y \in \{0,1\}^n : \forall j \leq J\ y[i_j] = 0\}$ denote the length-n binary strings that are 0 at positions $i_1, \ldots, i_J$.

3.4.10. Lemma. *When* SimonSampler *is passed k vectors $z_1, \ldots, z_k$ so that all $i_j := \min\{i : z_j[i] = 1\}$ are distinct for $1 \leq j \leq k$, then the state $|\psi\rangle$ before the measurement is*

$$\frac{\sqrt{2^k}}{N} \sum_{x\in\{0,1\}^n} \sum_{y\in Y_k} (-1)^{x\cdot y} |y\rangle |f(x)\rangle |x\cdot z_1\rangle \cdots |x \cdot z_k\rangle \ .$$

Proof. We follow the steps of subroutine SimonSampler.

$$\begin{aligned} |0^n\rangle|0\rangle|0^k\rangle \mapsto \frac{1}{\sqrt{N}} \sum_{x\in\{0,1\}^n} |x\rangle|0\rangle|0^k\rangle &\mapsto \frac{1}{\sqrt{N}} \sum_{x\in\{0,1\}^n} |x\rangle|f(x)\rangle|0^k\rangle \\ &\mapsto \frac{1}{N} \sum_{x,y\in\{0,1\}^n} (-1)^{x\cdot y}|y\rangle|f(x)\rangle|0^k\rangle \end{aligned}$$

This is the state before the **for** loop is entered. We claim and proceed to show by induction that after the Jth execution of the loop body, the state is

$$\frac{\sqrt{2^J}}{N} \sum_{x\in\{0,1\}^n} \sum_{y\in Y_J} (-1)^{x\cdot y} |y\rangle |f(x)\rangle |x\cdot z_1\rangle \cdots |x \cdot z_J\rangle |0^{k-J}\rangle \ .$$

Executing the body of the loop for $j = J + 1$,

$$\begin{aligned} &\frac{\sqrt{2^J}}{N} \sum_{x\in\{0,1\}^n} \sum_{y\in Y_J} (-1)^{x\cdot y} |y\rangle |f(x)\rangle |x\cdot z_1\rangle \cdots |x \cdot z_J\rangle |0\rangle |0^{k-J-1}\rangle \\ \mapsto &\frac{\sqrt{2^J}}{N} \sum_{x\in\{0,1\}^n} \sum_{y\in Y_J} (-1)^{x\cdot y} |y\rangle |f(x)\rangle |x\cdot z_1\rangle \cdots |x \cdot z_J\rangle |y[i_{j+1}]\rangle |0^{k-J-1}\rangle \\ = &\frac{\sqrt{2^J}}{N} \sum_{\substack{x\in\{0,1\}^n \\ y\in Y_{J+1} \\ b\in\{0,1\}}} (-1)^{x\cdot(y\oplus bz_{J+1})} |y \oplus bz_{J+1}\rangle |f(x)\rangle |x\cdot z_1\rangle \cdots |x \cdot z_J\rangle |b\rangle |0^{k-J-1}\rangle \end{aligned}$$

(Here, we used the fact that $Y_J = Y_{J+1} \dot{\cup} (z_{J+1} \oplus Y_{J+1})$.)

$$
\begin{aligned}
&\mapsto \frac{\sqrt{2^J}}{N} \sum_{\substack{x\in\{0,1\}^n\\ y\in Y_{J+1}\\ b\in\{0,1\}}} (-1)^{x\cdot(y\oplus bz_{J+1})} |y\rangle|f(x)\rangle|x\cdot z_1\rangle\cdots|x\cdot z_J\rangle|b\rangle|0^{k-J-1}\rangle\\
&= \frac{\sqrt{2^{J+1}}}{N} \sum_{x\in\{0,1\}^n} \sum_{y\in Y_{J+1}} (-1)^{x\cdot y} |y\rangle|f(x)\rangle|x\cdot z_1\rangle\cdots|x\cdot z_J\rangle\\
&\qquad\qquad \frac{1}{\sqrt{2}} \sum_{b\in\{0,1\}} (-1)^{x\cdot(bz_{J+1})} |b\rangle|0^{k-J-1}\rangle\\
&\mapsto \frac{\sqrt{2^{J+1}}}{N} \sum_{x\in\{0,1\}^n} \sum_{y\in Y_{J+1}} (-1)^{x\cdot y} |y\rangle|f(x)\rangle|x\cdot z_1\rangle\cdots|x\cdot z_{J+1}\rangle|0^{k-J-1}\rangle\ .
\end{aligned}
$$

□

This establishes the following invariants for SimonTester:

3.4.11. LEMMA. *If measuring the first register, $\mathcal{X}$, yields a nonzero value z, then*

1. *$\{z_1, \dots, z_k, z\}$ is linearly independent,*
2. *$\min\{i : z[i] = 1\}$ is distinct from i_j for $1 \le j \le k$, and*
3. *if $f \in L$, then $z \cdot s = 0$ for every $s \neq 0$ such that $f(x) = f(x \oplus s)$ for all x.*

Proof. If we measure the state from Lemma 3.4.10, then for the value z of the first register holds $z \in Y_k$. This implies 2, from which follows 1. For 3: as in Simon's original algorithm, if there is a $s \neq 0$ so that for all x, $f(x) = f(x \oplus s)$, then we can rewrite the state from Lemma 3.4.10 as

$$
\begin{aligned}
&\frac{\sqrt{2^k}}{N} \sum_{\substack{x:x<x\oplus s\\ y\in Y_k}} |y\rangle \Big((-1)^{x\cdot y}|f(x)\rangle + (-1)^{(x\oplus s)\cdot y}|f(x\oplus s)\rangle\Big) |x\cdot z_1\rangle\cdots|x\cdot z_k\rangle\\
&\quad= \frac{\sqrt{2^k}}{N} \sum_{x:x<x\oplus s} \sum_{y\in Y_k} |y\rangle(-1)^{x\cdot y}\left(1+(-1)^{s\cdot y}\right)|f(x)\rangle|x\cdot z_1\rangle\cdots|x\cdot z_k\rangle\ .
\end{aligned}
$$

Hence, only y with $s \cdot y = 0$ will have nonzero amplitude. □

Next, we want to assess the probability of obtaining $z = 0$ in SimonTester Line 4. We let P_0 denote the projection operator mapping $|0\rangle|y\rangle|z\rangle \mapsto |0\rangle|y\rangle|z\rangle$ and $|x\rangle|y\rangle|z\rangle \mapsto 0$ for $x \neq 0$; hence, $\|P_0|\psi\rangle\|^2$ is the probability of obtaining 0 when measuring subspace $\mathcal{X}$ of the quantum register in state $|\psi\rangle$. We can characterize the probability for outcome $z = 0$ in terms of the following definition and lemma:

3.4.12. Definition. For $c \in \{0,1\}^k$ and $z_1, \ldots, z_k \in \{0,1\}^n$ we define $D_c := \{x \in \{0,1\}^n : x \cdot z_1 = c[1], \ldots, x \cdot z_k = c[k]\}$.

3.4.13. Lemma. *Let $|\psi\rangle$ be the state before the measurement in* SimonSampler*, when* SimonSampler *is passed k linearly independent vectors $z_1, \ldots, z_k$ so that all $i_j := \min\{i : z_j[i] = 1\}$ are distinct for $1 \leq j \leq k$.*

1. *$\|P_0|\psi\rangle\|^2 = 1$ if and only if for every $c \in \{0,1\}^k$, f is constant when restricted to D_c.*

2. *If $\|P_0|\psi\rangle\|^2 \geq 1 - \varepsilon^2/2$, then f differs in at most εN points from some function g that is constant when restricted to D_c for every $c \in \{0,1\}^k$.*

Proof. For $b \in \{0,1\}$ let $D_{b,c} := D_c \cap f^{-1}\{b\} = \{x : f(x) = b \text{ and } x \cdot z_1 = c[1], \ldots, x \cdot z_k = c[k]\}$. Note that the $D_{b,c}$ and D_c also depend on $z_1, \ldots, z_k$ and the $D_{b,c}$ depend on f. Let

$$\begin{aligned} |\psi_0\rangle &:= \frac{\sqrt{2^k}}{N} \sum_{x \in \{0,1\}^n} |0\rangle|f(x)\rangle|x \cdot z_1\rangle \cdots |x \cdot z_k\rangle \\ &= \frac{\sqrt{2^k}}{N} \sum_{b \in \{0,1\}} \sum_{c \in \{0,1\}^k} |D_{b,c}|\, |0\rangle|b\rangle|c[1]\rangle \cdots |c[k]\rangle \ . \end{aligned}$$

By Lemma 3.4.10, at the end of SimonSampler the system is in state $|\psi\rangle = |\psi_0\rangle + |\psi_0^\perp\rangle$ for some $|\psi_0^\perp\rangle$ orthogonal to $|\psi_0\rangle$. We consider the case $\|P_0|\psi\rangle\|^2 = 1$. Then the register $\mathcal{X}$ must be in state $|0\rangle$ and thus $|\psi\rangle = |\psi_0\rangle$. Since the state has norm 1, we know that

$$\sum_{b \in \{0,1\}} \sum_{c \in \{0,1\}^k} |D_{b,c}|^2 = \frac{N^2}{2^k} \ . \tag{3.1}$$

The $D_{b,c}$ partition $\{0,1\}^n$ and the $D_c = D_{0,c} \cup D_{1,c}$ have the same size for all $c \in \{0,1\}^k$ because they are cosets of D_0. Therefore,

$$\sum_{b \in \{0,1\}} \sum_{c \in \{0,1\}^k} |D_{b,c}| = N \text{ and } |D_{0,c}| + |D_{1,c}| = \frac{N}{2^k} \text{ for all } c \in \{0,1\}^k \ . \tag{3.2}$$

$|D_{0,c}|^2 + |D_{1,c}|^2 \leq N^2/2^{2k}$, but in order for equation (3.1) to hold, $|D_{0,c}|^2 + |D_{1,c}|^2$ must be exactly $N^2/2^{2k}$. This can only be achieved if either $D_{0,c}$ or $D_{1,c}$ is empty. So f must be constant when restricted to D_c for any $c \in \{0,1\}^k$. Conversely, if f is constant when restricted to D_c for any $c \in \{0,1\}^k$, then equation (3.1) holds, therefore $\||\psi_0\rangle\| = 1$ and $|\psi\rangle = |\psi_0\rangle$. This concludes the proof of case 1 of the lemma.

If $\|P_0|\psi\rangle\|^2 = \||\psi_0\rangle\|^2 \geq 1 - \delta$, then

$$\sum_{b\in\{0,1\}} \sum_{c\in\{0,1\}^k} |D_{b,c}|^2 \geq (1-\delta)\,\frac{N^2}{2^k} \ . \tag{3.3}$$

Still, the constraints (3.2) hold; let $r2^k$ be the number of $c \in \{0,1\}^k$ so that $\min\{|D_{0,c}|, |D_{1,c}|\} \geq \gamma N/2^k$. Then

$$\sum_{b\in\{0,1\}} \sum_{c\in\{0,1\}^k} |D_{b,c}|^2 \leq r2^k(\gamma^2 + (1-\gamma)^2)\frac{N^2}{2^{2k}} + (1-r)2^k\frac{N^2}{2^{2k}} \ ,$$

and using (3.3), we obtain $r \leq \delta/(1 - \gamma^2 - (1-\gamma)^2)$. With $\delta = \varepsilon^2/2$ and $\gamma = \varepsilon/2$, this implies $r \leq \varepsilon$. But then

$$\sum_{c\in\{0,1\}^k} \min\{|D_{0,c}|, |D_{1,c}|\} \leq r2^k\frac{N}{2^{k+1}} + (1-r)2^k\gamma\frac{N}{2^k} \leq \varepsilon N \ .$$

□

We need to relate these two cases to membership in L and bound the number of repetitions needed to distinguish between the two cases. This is achieved by the following two lemmas.

3.4.14. Lemma. *Let k be the minimum number of linearly independent vectors $z_1, \ldots, z_k$ so that for each $c \in \{0,1\}^k$, f is constant when restricted to D_c. Then $f \in L$ if and only if $k < n$.*

Proof. If $k < n$, then there exists an $s \neq 0$ with $s \cdot z_1 = 0, \ldots, s \cdot z_k = 0$. For each such s and all x, we have $x \cdot z_1 = (x \oplus s) \cdot z_1, \ldots, x \cdot z_k = (x \oplus s) \cdot z_k$ and $x \in D_{f(x), x\cdot z_1, \ldots, x\cdot z_k}$ and $x \oplus s \in D_{f(x\oplus s), x\cdot z_1, \ldots, x\cdot z_k}$, therefore $f(x) = f(x \oplus s)$. Conversely, for $f \in L$, $S := \{s : \forall x f(x) = f(x \oplus s)\}$ is a nontrivial subspace of $\{0,1\}^n$, therefore $S^\perp = \{z : z \cdot s = 0 \forall s \in S\}$ is a proper subspace of $\{0,1\}^n$. Let $z_1, \ldots, z_k$ be an arbitrary basis of $S^\perp$. □

3.4.15. Lemma. *Let $0 < q < 1$, and $|\varphi_1\rangle, \ldots, |\varphi_m\rangle$ be quantum states satisfying $\|P_0|\varphi_j\rangle\|^2 < 1 - \delta$ for $1 \leq j \leq m$. If $m = \log q/\log(1-\delta) = \Theta(-(\log q)/\delta)$, then with probability at most q measuring the $\mathcal{X}$ register of $|\varphi_1\rangle, \ldots, |\varphi_m\rangle$ will yield m times outcome 0.*

Proof.

$$\Pr\left[m \text{ times } 0 \,\middle|\, \forall j : \|P_0|\varphi_j\rangle\|^2 < 1-\delta\right] < (1-\delta)^m = (1-\delta)^{\log q/\log(1-\delta)} = q \ .$$

□

Now all the ingredients for wrapping up the argument are at hand; first consider $f \in L$. Let $S := \{s : f(x) = f(x \oplus s)\ \forall x\}$ be the set of all "Simon promises" of f, and $S^\perp := \{z : z \cdot s = 0\ \forall s \in S\}$ the vectors that are orthogonal to all such promises. By Lemma 3.4.11 the nonzero z computed by the algorithm lie in $S^\perp$ and are linearly independent, therefore after $\dim S^\perp$ rounds of **for** loop in SimonTester, we measure $z = 0$ with certainty. Since $f \in L$, $\dim S > 0$ and thus $\dim S^\perp < n$.

If f is εn-far from being in L, then by Lemma 3.4.14 f is εn-far from being close to a function for which a $k < n$ and $z_1, \ldots, z_k$ exist so that f is constant when restricted to D_c for any of the $c \in \{0,1\}^k$. Therefore, by Lemma 3.4.13 case 2, for all $k < n$, $\|P_0|\psi\rangle\|^2 < 1-\varepsilon^2/2$. Thus, Lemma 3.4.15 guarantees that we accept with probability at most 1/3 if we let $q = 1/(3n)$ and thus $m = \mathrm{O}((\log n)/\varepsilon^2)$.

This concludes the proof of Theorem 3.4.3. □

3.5 Quantum Lower Bounds

In this section we prove that not every language has a fast quantum property tester.

3.5.1. Theorem. *Most properties containing $2^{n/20}$ elements of $\{0,1\}^n$ require quantum property testers using $\Omega(n)$ queries.*

Proof. Fix n, a small ε, and a quantum algorithm A making $q := n/400$ queries. Pick a property P as a random subset of $\{0,1\}^n$ of size $2^{n/20}$. Let

$$P_\varepsilon := \{y : d(x,y) < \varepsilon n \text{ for some } x \in P\} \ ;$$

using $\sum_{k=0}^{\varepsilon n} \binom{n}{k} \le 2^{H(\varepsilon)n}$, where

$$H(\varepsilon) := -\varepsilon \log \varepsilon - (1-\varepsilon)\log(1-\varepsilon) \ ,$$

we obtain $|P_\varepsilon| \le 2^{(1/20+H(\varepsilon))n}$. In order for A to test properties of size $2^{n/20}$, it needs to reject with high probability on at least $2^n - 2^{(1/20+H(\varepsilon))n}$ inputs; but then, the probability that A accepts with high probability on a random $x \in \{0,1\}^n$ is bounded by $2^{(1/20+H(\varepsilon))n}/2^n$ and therefore the probability that A accepts with high probability on $|P|$ random inputs is bounded by

$$2^{-(1-1/20-H(\varepsilon))n|P|} = 2^{-2^{n/20+\Theta(\log n)}} \ .$$

We would like to sum this success probability over all algorithms using the union bound to argue that for most properties no algorithm can succeed. However, there is an uncountable number of possible quantum algorithms with arbitrary quantum transitions. But by Beals, Buhrman, Cleve, Mosca, and de Wolf [15], the acceptance probability of A can be written as a multilinear polynomial of degree at most $2q$ where the n variables are the bits of the input; using results of Bennett, Bernstein, Brassard, and Vazirani [20] and Solovay and Yao [110], every quantum algorithm can be approximated by another algorithm such that the coefficients of the polynomials describing the accepting probability are integers of absolute value less than $2^{n^{\mathrm{O}(1)}}$ over some fixed denominator. There are less than $2^{nH(2q/n)}$ degree-$2q$ monomials in n variables, thus we can limit ourselves to $2^{n^{\mathrm{O}(1)}2^{nH(2q/n)}} \leq 2^{2^{(n/20)\cdot(91/100)+\Theta(\log n)}}$ algorithms.

Thus, by the union bound, for most properties of size $2^{n/20}$, no quantum algorithm with q queries will be a tester for it. □

We also give an explicit natural property that requires a large number of quantum queries to test. For $m \ll n$, a pseudorandom number generator is a function $f : \{0,1\}^m \to \{0,1\}^n$ that maps a small seed $s \in \{0,1\}^m$ to a large binary string $f(s) \in \{0,1\}^n$; if s is chosen uniformly at random, the distribution $f(s)$ of n-bit strings should have certain properties of the uniform distribution over n-bit strings. One such property is independence: if $x \in \{0,1\}^n$ is chosen uniformly at random, the values of its bits are independent, i.e., $x[i]$ and $x[j]$ are independent random variables for $i \neq j$. Accordingly, random s, $f(s)[i]$ and $f(s)[j]$ should be independent, i.e., for fixed seed s and index i and each index $j \neq i$, the sets of seeds

$$S_{s,i,j,0} := \{s' : f(s')[j] = 0 \text{ and } f(s')[i] = f(s)[i]\}$$
$$S_{s,i,j,1} := \{s' : f(s')[j] = 1 \text{ and } f(s')[i] = f(s)[i]\}$$

should have the same size. This independence requirement readily extends to fixing up to d bit positions and requiring that for each of the remaining bit positions j, there are as many strings in the image $f(\{0,1\}^m)$ with the jth bit 0 as there are with the jth bit 1. This corresponds to the $(d+1)$-wise independence of the pseudorandom values $f(\{0,1\}^m)$. Of course, choosing $x \in \{0,1\}^n$ uniformly at random gives n-wise independence, but for many applications d-wise independence with $d < n$ is sufficient and permits small seed sizes m.

What we show is that for an arbitrary fixed $f : \{0,1\}^m \to \{0,1\}^n$ that is a d-wise independent pseudorandom number generator, testing whether some $x \in \{0,1\}^n$ is close to satisfying $x \in f(\{0,1\}^m)$ requires many queries on a quantum computer. Intuitively, this means that such pseudorandom numbers look in a certain way random even to a quantum computer.

3.5.2. Theorem. *The range of a d-wise independent pseudorandom number generator requires* $(d+1)/2$ *quantum queries to test for any odd* $d \leq n/\log n - 1$.

We will make use of the following lemma:

3.5.3. Lemma (see [6]). *Suppose* $n = 2^k - 1$ *and* $d = 2t+1 \leq n$. *Then there exists a uniform probability space* Ω *of size* $2(n+1)^t$ *and d-wise independent random variables* $\xi_1, \ldots, \xi_n$ *over* Ω, *each of which takes the values 0 and 1 with probability* $1/2$.

The proof of Lemma 3.5.3 is constructive and the construction uniform in n. For given n and d, consider the language P of bit strings $\xi(z) := \xi_1(z)\ldots\xi_n(z)$ for all events $z \in \Omega = \{1, \ldots, 2(n+1)^t\}$. As a warmup, observe that classically deciding membership in P takes more than d queries: for all d positions $i_1, \ldots, i_d$ and all strings $v_1 \ldots v_d \in \{0,1\}^d$ there is a z such that $\xi_{i_1}(z)\ldots\xi_{i_d}(z) = v_1 \ldots v_d$. On the other hand, $\lfloor \log |\Omega| \rfloor + 1 = \mathrm{O}(d \log n)$ queries are always sufficient.

Proof of Theorem 3.5.2. We first consider the decision problem and then extend the lower bound to testing. A quantum computer *deciding* membership for $x \in \{0,1\}^n$ in $P := \{\xi(z) : z \in \Omega\}$ with T queries gives rise to a degree $2T$ multilinear n-variable approximating polynomial $p(x) = p(x_1, \ldots, x_n)$ [15]. We show that there must be high-degree monomials in p by comparing the expectation of $p(x)$ for randomly chosen $x \in \{0,1\}^n$ with the expectation of $p(x)$ for randomly chosen $x \in P$.

For uniformly distributed $x \in \{0,1\}^n$, we have $\mathrm{E}[p(x)|x \in P] \geq 2/3$ and $\mathrm{E}[p(x)|x \notin P] \leq 1/3$. Since $|P| = \mathrm{o}(2^n)$, $\mathrm{E}[p(x)] \leq 1/3 + \mathrm{o}(1)$ and thus $\Delta := \mathrm{E}[p(x)|x \in P] - \mathrm{E}[p(x)] \geq 1/3 - \mathrm{o}(1)$. Considering $p(x) = \sum_i \alpha_i m_i(x)$ as a linear combination of n-variable multilinear monomials m_i, we have by the linearity of expectation $\mathrm{E}[p(x_1, \ldots, x_n)] = \sum_i \alpha_i \mathrm{E}[m_i(x_1, \ldots, x_n)]$. Because of the d-wise independence of the bits of each $x \in P$, for every m_i of degree at most d holds $\mathrm{E}[m_i(x)] = \mathrm{E}[m_i(x)|x \in P]$. Since $\Delta > 0$, p must comprise monomials of degree greater than d. Hence, the number of queries T is greater than $d/2$.

This proof extends in a straightforward manner to the case of testing the property P: let again $P_\varepsilon := \{y : d(x,y) < \varepsilon n \text{ for some } x \in P\}$. Then

$$|P_\varepsilon| \leq 2^{H(\varepsilon)n}|P| = \mathrm{O}(2^{H(\varepsilon)n + d\log n}) \ ,$$

so

$$\mathrm{E}[p(x)] = \frac{|P_\varepsilon|}{2^n}\,\mathrm{E}[p(x)|x \in P_\varepsilon] + \left(1 - \frac{|P_\varepsilon|}{2^n}\right)\mathrm{E}[p(x)|x \notin P_\varepsilon] \leq \frac{1}{3} + \mathrm{o}(1)$$

for every $d = n/\log n - \omega(1/\log n)$ and every ε with $H(\varepsilon) = 1 - \omega(1/n)$. Again, we have $\Delta > 1/3 - \mathrm{o}(1)$ and we need monomials of degree greater than d. □

3.6 Further Research

The research presented in this chapter initiated the study of quantum property testing. Several interesting problems remain including

- Can one get the greatest possible separation of quantum and classical property testing, i.e., is there a language that requires $\Omega(n)$ classical queries but only $\mathrm{O}(1)$ quantum queries to test?

- Are there other natural problems that do not have quantum property testers? The language $\{uuvv : u, v \in \Sigma^*\}$ appears to be a good candidate for not having a quantum property tester.

- Beals, Buhrman, Cleve, Mosca, and de Wolf [15] observed that every k-query quantum algorithm gives rise to a degree-$2k$ polynomial in the input bits, which gives the acceptance probability of the algorithm; thus, a quantum property tester for P gives rise to a polynomial that is on all binary inputs between 0 and 1, that is at least 2/3 on inputs with the property P and at most 1/3 on inputs far from having the property P. Szegedy [114] suggested to algebraically characterize the complexity of classical testing by the minimum degree of such polynomials; as mentioned in the introduction, our results imply that this cannot be the case for classical testers. However, it is an open question whether quantum property testing can be algebraically characterized in this way.

- Høyer [74] and Friedl et al. [61] put quantum property testing into a group theoretic context. Is a characterization of quantum property testing possible in group-theoretic terms?

Chapter 4

Robustness

In this chapter we study the effect of noisy input on problems in the blackbox setting. It is based on work with Buhrman, Newman, and de Wolf [36].

4.1 Introduction

Consider the following setting: we would like to compute some function $f : \{0,1\}^n \to \{0,1\}$, but our access to the input $x \in \{0,1\}^n$ has to deal with noise: when looking up the bit x_i we get the wrong value $1 - x_i$ with probability ε_i. The precise error probability is unknown to us, but we are given an upper bound $\varepsilon < 1/2$ so that for all bit positions i holds $\varepsilon_i \leq \varepsilon$. Many algorithms designed for noiseless input will fail when given such noisy input. For example, the trivial algorithm for computing OR, "query every bit and output 1 if $x_i = 1$ for at least one i," will fail with high probability on the all-zero input for $\varepsilon > 1/n$.

Feige et al. [53] studied the overhead it takes to make an algorithm *robust*, i.e., resistant against noisy inputs. In general, one can query a variable x_i $\mathrm{O}(\log n)$ times instead of once and take the majority value as the value of x_i. This reduces the uniform bound on the error probability to much less than $1/n$; then the union bound implies that with high probability *all* queries will be given the correct value, so a non-robust algorithm will work. Accordingly, every non-robust algorithm in the decision-tree or query-complexity model can be made robust at the cost of a factor $\mathrm{O}(\log n)$ overhead (in fact, $\mathrm{O}(\log T)$ would suffice for a T-query algorithm). Sometimes this factor of $\mathrm{O}(\log n)$ is necessary: Feige et al. proved that every robust algorithm for the PARITY function needs to make $\Omega(n \log n)$ queries, for fixed ε. On the other hand, for some functions the $\mathrm{O}(\log n)$ can be dispensed with: Feige et al. also designed a non-trivial robust algorithm that computes the OR with $\mathrm{O}(n)$ queries, only a constant factor worse than the noiseless case.

Here we study this model for quantum algorithms. There is an issue as to what a "noisy query" means in this case, since one application of a quantum query can address many different x_i's in superposition:

1. One possibility is that for each quantum query, each of the bits is flipped with probability ε. However, now each quantum query introduces a lot of randomness, and the algorithm's state after the query would no longer be a pure quantum state.

2. Alternatively, we can assume that we have n quantum procedures, A_1, $\ldots$, A_n, such that A_i outputs x_i with probability at least $1 - \varepsilon$. Such algorithms can always be made coherent by pushing measurements to the end, which means that we can apply and reverse them at will. To enable us to apply the A_is in superposition, we assume we have a black box
$$\mathcal{A} : |i\rangle|0\rangle \mapsto |i\rangle A_i|0\rangle \ .$$
One application of this will count as one query.

3. The multiple-faulty-copies model was studied by Szegedy and Chen [115]; here, instead of x_i, the algorithm can only query "perturbed" copies $y_{i,1}$, $\ldots$, $y_{i,m}$ of x_i. The $y_{i,j}$ are independent Boolean random variables with $\Pr[x_i = y_{i,j}] \geq 1 - \varepsilon$ for each $i = 1, \ldots, n$, $j = 1, \ldots, m$. In contrast to the first proposal, this model leaves the queries perfectly reversible, since the perturbed copies are fixed at the start of the algorithm and the same $y_{i,j}$ can be queried more than once. The assumption of this model is also stronger than the second model, since we can construct a 1-query A_i that just outputs a superposition of all $y_{i,j}$. If m is sufficiently large, A_i will compute x_i with high success probability, satisfying the assumption of the second model (see Section 4.3 for details).

Assuming the second model and some fixed ε, we call a quantum algorithm *robust* if it computes f with bounded error probability when its inputs are given by algorithms A_1, $\ldots$, A_n. A first observation is that every T-query non-robust algorithm can be made robust at a multiplicative cost of $\mathrm{O}(\log T)$. With $\mathrm{O}(\log T)$ queries, a majority gate, and an uncomputation step, we can construct a unitary $\tilde{U}_x$ that approximates an exact quantum query $U_x : |i\rangle|b\rangle \mapsto |i\rangle|b \oplus x_i\rangle$ very well: $\|U_x - \tilde{U}_x\| \leq 1/(100T)$. Since errors add linearly in a quantum algorithm, replacing U_x by $\tilde{U}_x$ in a non-robust algorithm gives a robust algorithm with almost the same final state. In some cases better constructions are possible. For instance, a recent result by Høyer et al. [75] immediately implies a quantum algorithm that robustly computes OR with $\mathrm{O}(\sqrt{n})$ queries. This is only a constant factor worse than the noiseless case, which is Grover's algorithm [69]. In fact, we do not know of any function where the robust degree is more than a constant factor larger than the non-robust approximate degree.

Our main result (made precise in Theorem 4.2.1) is the following:

> There exists a quantum algorithm that outputs x with high probability, using $\mathrm{O}(n)$ invocations of the A_i algorithms (i.e., queries).

This result implies that *every* n-bit function f can be robustly quantum computed with $\mathrm{O}(n)$ queries. This contrasts with the classical $\Omega(n \log n)$ lower bound for PARITY. It is quite interesting to note that quantum computers, which usually are more fragile than classical computers, are actually more robust in the case of computing PARITY with noisy inputs. The results for OR and PARITY can be extended to every symmetric function f: for every such function, the optimal quantum algorithm can be made robust with only a constant factor overhead.

Our main result has a direct bearing on the *direct-sum problem*, which is the question how the complexity of computing n independent instances of a function scales with the complexity of one instance. One would expect that computing n instances with bounded-error takes no more than n times the complexity of one instance. However, since we want all n instances to be computed correctly *simultaneously* with high probability, the only known general method is to compute each instance with error probability reduced to $\mathrm{O}(1/n)$, which costs another factor of $\mathrm{O}(\log n)$. In fact, it follows from the $\Omega(n \log n)$ bound for PARITY that this factor of $n \log n$ is optimal when we can only run algorithms for individual instances in a black-box fashion. In contrast, our result implies that in the quantum world, the bounded-error complexity of n instances is at most $\mathrm{O}(n)$ times the bounded-error complexity of one instance. This is a very general result. For example, it also applies to communication complexity [80, Section 4.1.1]. If Alice and Bob have a bounded-error protocol for a distributed function f, using c bits (or qubits) of communication, then there is a bounded-error quantum protocol for n instances of f, using $\mathrm{O}(n(c + \log n))$ qubits of communication. The additive $\log n$ is because Alice and Bob need to communicate (possibly in superposition) the index of the instance that they are computing. In contrast, the best known general classical solution uses $\Theta(cn \log n)$ bits of communication.

In addition to robust quantum algorithms, we also consider robustness for multivariate *polynomials* approximating Boolean functions. In general, there are many connections between the (quantum or classical) query complexity of an n-bit function and the degrees of n-variate polynomials that approximate it [38]. We consider two complementary definitions of robust polynomials. First, in analogy to the multiple-faulty-copies model, we can consider the usual approximating polynomial but on nm instead of just n binary variables, and require that if $\Pr[x_i = y_{i,j}] \geq 1 - \varepsilon$ for each $i = 1, \ldots, n$, $j = 1, \ldots, m$

then the polynomial p satisfies $\Pr[|p(y) - f(x)| \geq 1/3] \leq 1/3$. Secondly, we can define a robust polynomial for a Boolean function f to operate on n variables $z \in \mathbb{R}^n$, so that $|q(z_1, \ldots, z_n) - f(x_1, \ldots, x_n)| \leq 1/3$ whenever $x \in \{0,1\}^n$ and $|z_i - x_i| \leq \varepsilon$ for all i. In Section 4.4 we show that the two types of robust polynomials are essentially equivalent, and that every non-robust approximating polynomial of degree d can be made robust at the cost of increasing its degree by a factor $O(\log d)$. Beals, Buhrman, Cleve, Mosca, and de Wolf [15] showed that every T-query quantum algorithm for f gives rise to a degree-$2T$ approximating polynomial for f, and similarly one can show that every T-query robust quantum algorithm for f induces a degree-$2T$ polynomial that approximates f robustly. This implies, for instance, that the robust degree of OR is $\Theta(\sqrt{n})$, and that every n-bit function has robust degree $O(n)$.

4.2 Robustly Recovering All n Bits

In this section we prove our main result, that we can recover an n-bit string x using $O(n)$ invocations of algorithms $A_1, \ldots, A_n$ where A_i computes x_i with bounded error.

4.2.1. Theorem. *Given ε-error algorithms $A_1, \ldots, A_n$ for the bits $x_1, \ldots, x_n$, there is a quantum algorithm that recovers $x = x_1 \ldots x_n$ with probability $2/3$ using $O(n/(1/2 - \varepsilon)^2)$ queries (invocations of the A_i).*

We assume A_i is a unitary transformation

$$A_i : |0^t\rangle \mapsto \alpha_i|0\rangle|\psi_i^0\rangle + \sqrt{1-\alpha_i^2}|1\rangle|\psi_i^1\rangle$$

for some $\alpha_i \geq 0$ such that $|\alpha_i|^2 \leq \varepsilon$ if $x_i = 1$ and $|\alpha_i|^2 \geq 1 - \varepsilon$ if $x_i = 0$; $|\psi_i^0\rangle$ and $|\psi_i^1\rangle$ are arbitrary $(t-1)$-qubit norm-1 quantum states. Every quantum algorithm can be expressed in this form by postponing measurements; every classical randomized algorithm can be converted into this form by making it reversible and replacing random bits by states $(|0\rangle + |1\rangle)/\sqrt{2}$. By applying a NOT to the first qubit after the execution of A_i, we can easily implement

$$\bar{A}_i : |0^t\rangle \mapsto \alpha_i|1\rangle|\psi_i^0\rangle + \sqrt{1-\alpha_i^2}|0\rangle|\psi_i^1\rangle \ ,$$

which operates like A_i but outputs 1 when A_i would have output 0 and vice versa. Let

$$A_i(b) := \begin{cases} A_i & \text{if } b = 0 \\ \bar{A}_i & \text{if } b = 1 \end{cases}$$

Procedure RobustFind(n, $\mathcal{A}$, ε, β, γ, δ)

$n \in \mathbb{N}$, $\mathcal{A}$: n quantum algorithms, $\varepsilon, \beta, \gamma, \delta > 0$

Output: $i \in [n] \cup \{\perp\}$ with the following properties:

1. if $\mathcal{A}$ is ε-close to $x \in \{0,1\}^n$ and $|x| \geq \beta n$, then $i \neq \perp$ with probability at least $1 - \delta$
2. if $\mathcal{A}$ is ε-close to $x \in \{0,1\}^n$ and if $i \neq \perp$, then $x_i = 1$ with probability at least $1 - \gamma$

Complexity:

$$\mathrm{O}\left(\frac{1}{\left(\frac{1}{2} - \varepsilon\right)^2} \cdot \sqrt{\frac{1}{\beta}} \cdot \log \frac{1}{\gamma\delta}\right) \text{ invocations of the } A_i$$

If we plug the right bit x_i into A_i, then for all A_i we expect output 0: for the unique good $x \in \{0,1\}^n$, $\mathcal{A}(x) := (A_1(x_1), \ldots, A_n(x_n))$ is ε-close to 0^n by the following notion of closeness:

4.2.2. Definition. For $\varepsilon < 1/2$ and decision algorithms $\mathcal{A} = (A_1, \ldots, A_n)$, we say $\mathcal{A}$ is *ε-close* to $x \in \{0,1\}^n$ if $\Pr[A_i \text{ outputs } x_i] \geq 1 - \varepsilon$ for all $i \in [n]$.

Our algorithm builds on a robust quantum search algorithm by Høyer, Mosca, and de Wolf [75]: the RobustFind subroutine above takes a vector $\mathcal{A}$ of n quantum algorithms and in the good case returns an index i so that the "high probability" output of A_i is 1. This allows us to verify a purported solution $\tilde{x} \in \{0,1\}^n$ by running RobustFind on $\mathcal{A}_{\tilde{x}}$ to find differences with the real input x. In fact, adjusting the parameters to RobustFind as we move closer and closer to a good solution, our main program AllOutputs (as defined by the pseudo code on page 80) manages to construct the unique x with high probability. Note that RobustFind is the only quantum component of our otherwise classical algorithm.

Success probability The first step of our algorithm (Line 1 in AllOutputs) is to classically sample each i once and to store this initial approximation into a variable $\tilde{x}$. The following rounds of the algorithm refine $\tilde{x}$ until with high probability it is correct (i.e., equal to x).

We call i a *bad* index if $i \in [n]$ and $\Pr[A_i \text{ outputs } x_i] \leq \varepsilon$. Let B_0 denote the random variable counting the number of bad indices after Line 1 in AllOutputs and let B_k denote the random variable of the number of bad

Procedure InitialGuess(n, $\mathcal{A}$)

$n \in \mathbb{N}$, $\mathcal{A}$: n algorithms

```
for $i \leftarrow$ to $n$ do
    run $A_i$
    $\tilde{x}_i \leftarrow$ result of $A_i$
return $\tilde{x}$
```

Procedure SampleBad(n, $\mathcal{A}$, $\tilde{x}$, r, ε, β, γ, δ)

$n \in \mathbb{N}$, $\mathcal{A}$: n algorithms, $\tilde{x} \in \{0,1\}^n$, $r \in \mathbb{N}$, $\varepsilon, \beta, \gamma, \delta > 0$

```
for $\ell \leftarrow 1$ to $r$ do
    $i \leftarrow$ RobustFind$(n, \mathcal{A}(\tilde{x}), \varepsilon, \beta, \gamma, \delta)$
    if $i \neq \perp$ then
        $\tilde{x}_i \leftarrow 1 - \tilde{x}_i$
return $\tilde{x}$
```

Procedure FindAllBad(n, $\mathcal{A}$, $\tilde{x}$, ε, β, γ, δ)

$n \in \mathbb{N}$, $\mathcal{A}$: n algorithms, $\tilde{x} \in \{0,1\}^n$, $\varepsilon, \beta, \gamma, \delta > 0$

```
repeat
    $i \leftarrow$ RobustFind$(n, \mathcal{A}(\tilde{x}), \varepsilon, \beta, \gamma, \delta)$
    if $i \neq \perp$ then
        $\tilde{x}_i \leftarrow 1 - \tilde{x}_i$
until $i = \perp$
return $\tilde{x}$
```

Procedure AllOutputs(n, $\mathcal{A}$, ε)

$n \in \mathbb{N}$, $\mathcal{A}$: n algorithms, $\varepsilon > 0$

```
$\tilde{x} \leftarrow$ InitialGuess$(n, \mathcal{A})$
for $k \leftarrow 1$ to $\log(\varepsilon(\log n)^2)$ do
    $\varepsilon' \leftarrow \varepsilon / 2^{k-1}$
    $\tilde{x} \leftarrow$ SampleBad$\left(n, \mathcal{A}, \tilde{x}, 1.7\varepsilon' n, \varepsilon, 0.3\varepsilon', \frac{1}{8}, \frac{1}{8}\right)$
$\tilde{x} \leftarrow$ FindAllBad$\left(n, \mathcal{A}, \tilde{x}, \varepsilon, \frac{1}{10n}, \frac{1}{10n}, \frac{1}{10n}\right)$
return $\tilde{x}$
```

indices after Line 4 in AllOutputs. By G_k we denote the event $B_k \leq n\varepsilon/2^{k-1}$. We have

$$\Pr\left[G_{k_{\max}}\right] \geq \Pr\left[G_0\right] \prod_{k=1}^{k_{\max}} \Pr\left[G_k | G_{k-1}\right] \quad .$$

We now show that $\Pr[G_k|G_{k-1}]$ is large. For $k = 0$, we know that $\mathrm{E}[B_0] \leq \varepsilon n$ and $\Pr[B_0 \leq 2\varepsilon n] \geq 9/10$ by a Chernoff bound. In round k, we want to reduce the upper bound on the number of bad indices from $2n\varepsilon/2^{k-1}$ to $n\varepsilon/2^{k-1}$. If we have the maximum number of bad indices so that still G_k holds, we expect r repetitions of RobustFind to reduce the number of bad indices to

$$2\frac{n\varepsilon}{2^{k-1}} - (1-\delta)\left((1-\gamma)r - \gamma r\right) \leq \frac{9}{10}\frac{n\varepsilon}{2^{k-1}}$$

therefore we choose

$$r := \frac{11}{10}\frac{1}{(1-\delta)(1-2\gamma)}\frac{n\varepsilon}{2^{k-1}} \approx 1.7\frac{n\varepsilon}{2^{k-1}} \quad .$$

On the other hand, if we have only a small number b of bad indices, it is likely that we will make many errors, so we would like

$$b + \gamma r \leq \frac{9}{10}\frac{n\varepsilon}{2^{k-1}} \quad .$$

This is satisfied by choosing $b := 0.3n\varepsilon/2^{k-1}$; this choice of b also ensures that we never get as few as b bad indices if we start the round with $2n\varepsilon/2^{k-1}$ bad indices.

We tune RobustFind to find bad indices with probabilities δ and γ if there are at least b bad indices. Hence, in the extreme cases of either having exactly $2n\varepsilon/2^{k-1}$ or less than b bad indices, we expect to arrive at at most $(9/10) \cdot n\varepsilon/2^{k-1}$ bad indices, and this holds for the intermediary cases as well. By a Chernoff-type argument, the probability that we are a constant factor $10/9$ away from the expectation is exponentially small in the number r of samples, therefore, with $k_{\max} = \log(\varepsilon(\log n)^2)$, we have

$$\Pr[G_k|G_{k-1}] \geq 1 - e^{-\Omega(n/\log n)}$$

and

$$\begin{aligned}\Pr[G_{k_{\max}}] &\geq \Pr[G_0]\left(1 - e^{-\Omega(n/\log n)}\right)^{k_{\max}} \\ &\geq \frac{9}{10}\left(1 - \frac{k_{\max}}{e^{\Omega(n/\log n)}}\right) = \frac{9}{10} - \mathrm{o}(1) \quad .\end{aligned}$$

Hence, for large n with probability $8/10$ we have at most $n/(\log n)^2$ bad indices at Line 6 in AllOutputs. In this case, we will find with constant probability all bad indices by making the individual error probability in RobustFind so small that we can use a union bound: we determine each of the remaining bad indices with error probability $1/(10n)$. This implies an overall success probability $\geq (8/10) \cdot (9/10) > 2/3$.

Complexity We bound the number of queries to f in SampleBad as follows:

$$\sum_{k=1}^{k_{\max}} \sum_{\ell=1}^{n\varepsilon/2^{k-1}} C \frac{1}{\left(\frac{1}{2}-\varepsilon\right)^2} \sqrt{\frac{1}{\varepsilon/2^k}} \leq C' \frac{\sqrt{\varepsilon}}{\left(\frac{1}{2}-\varepsilon\right)^2} n \sum_{k=1}^{\infty} \frac{k}{2^{k/2}} = \mathrm{O}\left(\frac{n}{\left(\frac{1}{2}-\varepsilon\right)^2}\right)$$

for some constants C, C'. The call to FindAllBad results in

$$\mathrm{O}\left(\frac{1}{\left(\frac{1}{2}-\varepsilon\right)^2}\sqrt{(\log n)^2}\log n \cdot \left(\frac{n}{(\log n)^2}\right)\right) = \mathrm{O}\left(\frac{n}{\left(\frac{1}{2}-\varepsilon\right)^2}\right)$$

many queries. Therefore, the total query complexity of AllOutputs also is $\mathrm{O}(n/(1/2-\varepsilon)^2)$.

Consequences Once we have recovered the input x, we can compute an arbitrary function of x without further queries.

4.2.3. Corollary. *For every $f : \{0,1\}^n \to \{0,1\}$, there is a robust quantum algorithm that computes f using $\mathrm{O}(n)$ queries.*

In particular, PARITY can be robustly quantum computed with $\mathrm{O}(n)$ queries while it takes $\Omega(n \log n)$ queries classically [53].

In the context of the direct-sum problem, the complexity of quantum computing a vector of instances of a function scales linearly with the complexity of one instance.

4.2.4. Corollary (Direct Sum). *If there exists a T-query bounded-error quantum algorithm for f, then there is an $\mathrm{O}(Tn)$-query bounded-error quantum algorithm for n independent instances of f.*

As mentioned, the best classical upper bound has an additional factor of $\log n$, and this is optimal in a classical black-box setting.

Finally, all *symmetric* functions can be computed robustly on a quantum computer with the same asymptotic complexity as non-robustly. A function is symmetric if its value only depends on the hamming weight of the input. Let $\Gamma(f) := \min\{|2k-n+1| : f$ flips value if the Hamming weight of the input changes from k to $k+1\}$. The non-robust algorithm for computing f with $\mathrm{O}(\sqrt{n(n-\Gamma(f))})$ queries [15, Theorem 4.10] can be made robust by a similar algorithm as the one used in the proof of Theorem 4.2.1, giving:

4.2.5. THEOREM. *For every symmetric function f, there is a robust quantum algorithm that computes f using $O(\sqrt{n(n-\Gamma(f))})$ quantum queries.*

4.3 The Multiple-Faulty-Copies Model

As mentioned in the introduction, the assumption that we have a bounded-error algorithm A_i for each of the input bits x_i also covers the model of [115] where we have a sequence $y_{i,1}, \ldots, y_{i,m}$ of faulty copies of x_i. These we can query by means of a mapping

$$|i\rangle|j\rangle|0\rangle \mapsto |i\rangle|j\rangle|y_{i,j}\rangle \ .$$

Here we spell out this connection in some more detail. First, by a Chernoff bound, choosing $m := O((\log n)/\varepsilon^2)$ implies that the average $\overline{y}_i := \sum_{j=1}^{m} y_{i,j}/m$ is close to x_i with very high probability:

$$\Pr[|\overline{y}_i - x_i| \geq 2\varepsilon] \leq \frac{1}{100n} \ .$$

By the union bound, with probability 99/100 this closeness will hold for all $i \in [n]$ simultaneously. Assuming this is the case, we implement the following unitary mapping using one query to the $y_{i,j}$:

$$A_i : |0^{\log(m)+1}\rangle \mapsto \frac{1}{\sqrt{m}} \sum_{j=1}^{m} |j\rangle|y_{i,j}\rangle \ .$$

Measuring the last qubit of the resulting state gives x_i with probability at least $1 - 2\varepsilon$. Hence, we can run our algorithm from Section 4.2 and recover x using $O(n)$ queries to the $y_{i,j}$. Similarly, all consequences mentioned in the last section hold for this multiple-faulty-copies model as well.

4.4 Robust Polynomials

In this section we study robust polynomials, of two different but essentially equivalent types. The first type follows the many-faulty-copies model.

4.4.1. DEFINITION. An (ε, m) *perturbation* of $x \in \{0,1\}^n$ is a matrix y of $n \times m$ independent binary random variables $y_{i,j}$ so that $\Pr[y_{i,j} = x_i] \geq 1 - \varepsilon$ for each $1 \leq j \leq m$.

4.4.2. DEFINITION. A *type-1* (ε, m)*-robust polynomial* for the Boolean function $f(x_1, \ldots, x_n)$ is a real polynomial p in nm variables $y_{i,j}$ (with $1 \leq i \leq n$

and $1 \le j \le m$) so that for every $x \in \{0,1\}^n$ and y an (ε, m) perturbation of x, $\Pr[|p(y) - f(x)| \ge 1/3] \le 1/3$. Moreover, for every $v \in \{0,1\}^{nm}$, we require $-1/3 \le p(v) \le 4/3$.

The approximation "quality" of a type-1 robust polynomial can be boosted at constant multiplicative cost in the degree. Analogously we can improve the parameters to any other constant.

4.4.3. Lemma. *If there is a* type-1 (ε, m)-robust polynomial *of degree d for f, then for some $m' = \mathrm{O}(m)$ there exists a type-1 (ε, m')-robust polynomial p of degree $\mathrm{O}(d)$ so that $x \in \{0,1\}^n$ and y an (ε, m') perturbation of x, $\Pr[|p(y) - f(x)| \ge 1/9] \le 1/9$. Moreover, for every $v \in \{0,1\}^{nm'}$, $-1/9 \le p(v) \le 10/9$.*

Proof. Let p_0 denote the type-1 (ε, m)-robust polynomial that we start with. The single-variate polynomial $g(a) := (2a-1)(1+a(2+(a/22)(1+45a(a-2))))$ has the property that $-1/9 \le g^3(a) \le 1/9$ for $-1/3 \le a \le 1/3$ and $8/9 \le g^3(a) \le 10/9$ for $2/3 \le a \le 4/3$; here $g^t(a)$ denotes the t-fold application of g.

$$g^t(a) := \underbrace{g(g(\cdots g}_{t}(a))) \ .$$

Therefore $p_1(y) := g^3(p_0(y))$ satisfies $|p_1(y) - f(x)| \le 1/9$ whenever $|p_0(y) - f(x)| \le 1/3$.

For some r to be determined later and an arbitrary $x \in \{0,1\}^n$, we use r independent (ε, m) perturbations y_k of x, $1 \le k \le r$. Let B denote the random variable counting the number of indices k so that $|p_1(y_k) - f(x)| \ge 1/9$. Choosing r sufficiently large, a Chernoff bound implies $\Pr[B \ge 13r/36] \le 1/9$. Therefore

$$\begin{aligned}
\Pr\left[\left|\frac{1}{r}\sum_{k=1}^{r} p_1(y_k) - f(x)\right| \ge \frac{17}{36}\right] &\le \Pr\left[\sum_{k=1}^{r} |p_1(y_k) - f(x)| \ge \frac{17}{36}r\right] \\
&\le \Pr\left[\frac{10}{9}B + (r-B)\frac{1}{9} \ge \frac{17}{36}r\right] \\
&= \Pr\left[B \ge \frac{13}{36}r\right] \le \frac{1}{9}
\end{aligned}$$

Let $p_2(y_1, \ldots, y_r) := \frac{1}{r}\sum_{k=1}^{r} p_1(y_k)$. We move closer to f in the "good" case: $p(y_1, \ldots, y_r) := g^6(p_2(y_1, \ldots, y_r))$ satisfies

$$\Pr\left[|p(y_1, \ldots, y_r) - f(x)| \ge \frac{1}{9}\right] \le \frac{1}{9} \qquad \text{and} \qquad -\frac{1}{9} \le p(v) \le \frac{10}{9}$$

for all $v \in \{0,1\}^{nm}$. Now we are done: with $m' := rm$ we have that $y_1, \ldots, y_r$ is an (ε, m') perturbation of x and $\deg(p) = \mathrm{O}(\deg(p_0))$. □

The second kind of robust polynomial is the following:

4.4.4. Definition. A *type-2 ε-robust polynomial* for the Boolean function $f : \{0,1\}^n \to \{0,1\}$ is a real polynomial q in n variables $z_1, \ldots, z_n \in \mathbb{R}$ so that for every $x \in \{0,1\}^n$ and $z \in \mathbb{R}^n$ we have $|q(z) - f(x)| \leq 1/3$ if $|z_i - x_i| \leq \varepsilon$ for all $i \in [n]$. If $\varepsilon = 0$, then q is called an *approximating polynomial* for f.

4.4.5. Theorem. *For every type-2 ε-robust polynomial of degree d for f there is a type-1 $(\varepsilon/2, \mathrm{O}(\log(n)/(1/2-\varepsilon)^2))$-robust polynomial of degree d for f. Conversely, for every type-1 (ε, m)-robust polynomial of degree d for f there is a type-2 ε-robust polynomial of degree $\mathrm{O}(d)$ for f.*

Proof. Let p be a type-2 ε-robust polynomial of degree d for f. As in Section 4.3, we choose $m = \mathrm{O}(\log(n)/(1/2-\varepsilon)^2)$. If each $y_{i,j}$ is wrong with probability $\leq \varepsilon/2$, then with probability at least $2/3$, the averages $\overline{y}_i$ will satisfy $|\overline{y}_i - x_i| \leq \varepsilon$ for all $i \in [n]$. Hence the polynomial $p(\overline{y}_1, \ldots, \overline{y}_n)$ will be a type 1 $(\varepsilon/2, \mathrm{O}(\log(n)/(1/2-\varepsilon)^2))$-robust polynomial of degree d for f.

For the other direction, consider a type-1 (ε, m)-robust polynomial of degree d for f. Using Lemma 4.4.3, we boost the approximation parameters to obtain a type-1 (ε, m')-robust polynomial p of degree $\mathrm{O}(d)$, with $m' = \mathrm{O}(m)$, so that for every $x \in \{0,1\}^n$ and (ε, m') perturbation y of x, $\Pr[|p(y) - f(x)| \geq 1/9] \leq 1/9$. For $z \in \mathbb{R}^n$ with $0 \leq z_i \leq 1$ for all i, let $y_{i,j}$ ($i \in [n]$, $j \in [m']$) be independent random variables, where $y_{i,j} = 1$ with probability z_i. Define $q(z) := \mathrm{E}[p(y)]$. This q is a polynomial in z, because $\mathrm{E}[p(y)] = p(\mathrm{E}[y])$ and $\mathrm{E}[y_{i,j}] = z_i$. Moreover, if for z there exists $x \in \{0,1\}^n$ with $|z_i - x_i| \leq \varepsilon$ for all i, then y is an (ε, m') perturbation of x. Therefore $V := \{v : |p(v) - f(x)| \leq 1/9\}$ has probability $\Pr[y \in V] \geq 8/9$ and

$$\begin{aligned} |f(x) - q(z)| &= \left| \sum_{v \in \{0,1\}^{nm}} \Pr[y = v]\,(f(x) - p(v)) \right| \\ &\leq \left| \sum_{v \in V} \Pr[y = v]\,(f(x) - p(v)) \right| + \left| \sum_{v \notin V} \Pr[y = v] \left(1 + \frac{1}{9}\right) \right| \\ &\leq \frac{8}{9} \cdot \frac{1}{9} + \frac{1}{9} \cdot \frac{10}{9} < \frac{1}{3} \ . \end{aligned}$$

This means that $q(z)$ is a type-2 ε-robust polynomial for f of degree $\mathrm{O}(d)$. □

4.4.6. Definition. For $f : \{0,1\}^n \rightarrow \{0,1\}$, let $\mathrm{rdeg}_1(f)$ denote the minimum degree of the type-1 $(1/3, 5\log n)$ polynomials for f, $\mathrm{rdeg}_2(f)$ be the minimum degree of the type-2 1/3-robust polynomials approximating f, and $\widetilde{\deg}(f)$ be the minimum degree among all approximating polynomials for f.

Note that in Definition 4.4.2 we require for type-1 polynomials p that for each Boolean assignment $v \in \{0,1\}^{nm}$ to the (possibly real) variables, the polynomial value $p(v)$ between $-1/3$ and $4/3$. Because of this totality requirement, the following corollaries are given for total Boolean functions.

4.4.7. Corollary. $\mathrm{rdeg}_1(f) = \Theta(\mathrm{rdeg}_2(f))$ *for every (total) Boolean function* $f : \{0,1\}^n \rightarrow \{0,1\}$.

4.4.8. Corollary. $\mathrm{rdeg}_{1,2}(f) = \mathrm{O}(\widetilde{\deg}(f)\log n)$ *for every (total) Boolean function* $f : \{0,1\}^n \rightarrow \{0,1\}$.

Using the notion of *certificate complexity* $C(f)$ and its polynomial relation to $\widetilde{\deg}(f)$, one can strengthen Corollary 4.4.8 to the following theorem [36].

4.4.9. Theorem. $\mathrm{rdeg}_{1,2}(f) = \mathrm{O}(\widetilde{\deg}(f) \cdot \log \widetilde{\deg}(f))$.

4.5 Discussion and Open Problems

In contrast to the classical case, we do not know of any function where making a quantum algorithm or polynomial robust costs more than a constant factor. In the case of symmetric functions, such a constant overhead suffices. It is conceivable that quantum algorithms and polynomials can *always* be made robust at a constant factor overhead. Proving or disproving this would be very interesting.

We have chosen our model of a noisy query so that we can coherently make a query and reverse it. An open question is whether the advantage of quantum algorithms can be maintained for "decohering" queries, like the first model proposed in the introduction. It is not clear to what extent non-robust quantum algorithms can be made resilient against such random noise, since the usual transformations to achieve fault-tolerant quantum computation do not immediately apply to the query gate, which acts on a non-constant number of quantum bits simultaneously.

Part II

Distributed Quantum Computing

Chapter 5

Nonlocality

This chapter is based on research with Buhrman, Høyer, and Massar [34, 35]. Portions of the introduction are culled from a survey compiled with Buhrman [37].

5.1 Introduction

In Subsection 1.2.4 we outlined Einstein, Podolsky, and Rosen's objection about instantaneous "action" at spatially separated parts of a quantum system. The predictions of quantum mechanics of nonlocal effects were given an operational meaning by Bell [17], who came up with an experimental way of testing the nonlocal behavior of quantum mechanics. These tests and the so-called Bell inequalities lead to experiments, first executed by Aspect et al. [13], that appear to demonstrate the nonlocality of quantum mechanics.

Bell showed that the correlations between the outcomes of measurements carried out on entangled quantum systems cannot be reproduced by a local classical theory, often called a local hidden variable model. Since then extensive work has been carried out on quantum nonlocality, both on the experimental and theoretical aspects. On the theory side, research on quantum nonlocality has branched out into many different and complementary directions. One important direction of investigation is the search for qualitatively different types of quantum nonlocality. Of particular interest was the discovery of the Greenberger-Horne-Zeilinger (GHZ) "paradox" [68, 93]. In this and related examples, correlations are characterized as nonlocal by the pattern of zero and nonzero joint probabilities. This property has been called "pseudo telepathy," because in every run of the experiment, the parties appear to agree clandestinely on a subset of admissible outputs. It should be contrasted with other examples where it is the values of these joint probabilities that imply nonlocality.

Another important advance was to show that quantum nonlocality subsists even in the presence of noise as first demonstrated by Clauser, Horne, Shimony, and Holt [41]. This is essential since every experimental test will necessarily be affected by imperfections; the best experiments to date have error rates of the order of a few percent. Much additional work has been devoted to understanding the resistance of quantum nonlocality to imperfections.

Detector efficiency In experiments involving entangled photons, there is one particular kind of imperfection that plays a central role, namely the small efficiency of single-photon detectors. A single-photon detector will register the presence of a photon with probability η, and will not register the presence of the photon with probability $1 - \eta$. For instance, as one goes from visible to infrared wavelengths, η decreases from more than 50% to 10%. Detector inefficiency can be thought of as a specific type of noise. This imperfection was first discussed by Pearle [97] and remains to this day one of the major hurdles to overcome in order to carry out a loophole-free test of quantum nonlocality. Examples show that there are quantum correlations that are highly insensitive to detector inefficiency, but are much more sensitive to other kinds of noise, see Massar [86], and therefore this kind of imperfection should be studied independently of other kinds of noise.

Note that we disregard here the complementary error, namely detectors clicking when they should not. In general, we do not know the precise time when Charlie is sending the particles and so distinguishing false positives from false negatives can only be done by some kind of voting procedure. For this reason we consider this error only as general noise.

Remarkably, the amount of classical communication required to reproduce the quantum correlations and the minimum detector efficiency required to close the detection loophole are closely related quantities as demonstrated by Gisin and Gisin [62], and Massar [86]. In many cases, quantum correlations that require a lot of communication to reproduce classically cannot be simulated classically without communication, even when the actual detectors are very inefficient, see Steiner [113] for examples.

Asymptotics Another question that has been raised in the context of quantum information theory concerns the asymptotic limit when the size of the entangled system grows. Does the gap between classical and quantum correlations grow, and if so, at what rate? Brassard et al. [25] showed that in the bipartite case the amount of communication required to classically reproduce the quantum correlations can increase exponentially with the number of entangled bits shared by the parties. And it follows from the results in

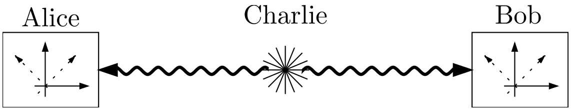


Figure 5.1: Schema of a nonlocality experiment: Charlie sends particles to Alice and Bob who randomly perform one of several possible measurements. We are interested in the probability distribution $\Pr[a, b|x, y]$ where x and y designate Alice's and Bob's measurement, respectively, and a and b their respective measurement outcomes.

Buhrman et al. [31] that there are quantum correlations for n parties each holding a two-dimensional subsystem, so that the amount of communication that must be broadcast in a classical simulation increases logarithmically with the number of parties. Unfortunately these asymptotic results have only been proved in the total absence of noise.

5.1.1 Bell inequalities

Nonlocality experiments Probing the "quantumness" of nature means devising and performing experiments that give different outcomes depending on whether the world is governed by classical physics or quantum mechanics. Under reasonable boundary conditions this should corroborate the validity of quantum mechanics. Nonlocality experiments usually work as sketched in Figure 5.1: two parties, Alice and Bob, receive from a third party, Charlie, each one particle, e.g., a photon. Randomly, they select one of several possible measurements, e.g., measuring the polarization in the vertical-horizontal or diagonal basis, and output the measurement outcome. Denoting Alice's output by a and Bob's by b, and Alice's measurement choice x and Bob's y, each run of such an experiment results in a tuple (a, b, x, y). Repeating the experiment many times allows us to estimate the probability distribution $\Pr[a, b|x, y]$.

The crucial point of the experiment is that Alice and Bob are separated while they make their random choice and perform the measurement—they do not know the other's measurement choice and they do not learn the other's measurement outcome until after they have produced their own output and, hence, are committed to a definite value. That such a separation is possible, is an implication of special relativity, which is assumed to hold both in the quantum mechanical and the classical hypotheses. Two events in *space-time*, i.e., events that occur at a given point x in Euclidean space at a unique moment in time t, are said to be *timelike* separated if a particle emanating from x_1 at time t_1 cannot reach x_2 at time t_2. Since no particle can travel faster than light, there indeed exist events that are timelike separated; no information can be transmitted between these points at the given times. When we want to quantify the information that would have needed to travel faster than light to give a classical explanation for some quantum phenomenon, we speak about *superluminal* or faster-than-light communication without attributing physical reality to it.

We assume that Alice and Bob are timelike separated while they decide on their respective measurement, receive the particle from Charlie, and produce the macroscopic measurement result. For instance, the time difference between Alice's and Bob's measurement needs to be much smaller than the time it takes for a photon to travel from Alice to Bob at the speed of light. Then a *nonlocal* effect is a probability distribution $\Pr[a, b|x, y]$, which is not induced by any local classical theory. Note that we assume that the decision about which measurements to execute is imposed on the detectors by external trusted random number generators; this is of minor importance since we can go over experimental records and perform checks on the generated settings.

Local hidden variable models The classical contenders for explanations of nonlocality experiments are so-called *local hidden variable* theories. We give a formal definition in Section 5.2; the idea is that these theories are *local* in the sense that all properties of a particle are contained within the particle alone and thus manipulations or measurements of Bob's particle do not affect Alice's and vice versa. Perceived nonlocality may be caused by some "hidden" property of objects, which we cannot or do not know how to measure directly, but which is passed along by Charlie at the inception of the particles and can synchronize the classical behavior of the particles. "Variable" refers to the fact that the hidden shared property may be set randomly by Charlie and thus is a random variable.

Bell inequalities Consider the following nonlocality experiments: Alice and Bob choose from two possible measurements A, A' and B, B', all four of which have two outcomes each, which we label by -1 and 1. A, A', B, B' are random variables, hence we have for the expectations of their pairwise products:

$$\mathrm{E}[AB] + \mathrm{E}[A'B] + \mathrm{E}[AB'] - \mathrm{E}[A'B'] = \mathrm{E}[A(B + B') + A'(B - B')]$$

Furthermore,

$$\begin{aligned}\mathrm{E}[A(B + B') + A'(B - B')] &\leq \mathrm{E}\big[|A(B + B') + A'(B - B')|\big] \\ &\leq \mathrm{E}\big[|B + B'| + |B - B'|\big] \ .\end{aligned}$$

In a local hidden variable model, in each run all four random variables take definite values from $\{-1, 1\}$, for each choice of measurements. Therefore, for each fundamental event, either $|B+B'| = 2$ and $|B-B'| = 0$, or $|B+B'| = 0$ and $|B - B'| = 2$, so

$$\mathrm{E}\big[|B + B'| + |B - B'|\big] = 2 \ .$$

The resulting inequality,

$$\mathrm{E}[AB] + \mathrm{E}[A'B] + \mathrm{E}[AB'] - \mathrm{E}[A'B'] \leq 2 \ , \tag{5.1}$$

imposes a linear constraint on the probabilities $\Pr[a, b|x, y]$ by expanding the expectations. This constraint holds for any local hidden variable model in this setting of two parties, two possible measurements each with two outcomes each. It is called a *Bell inequality*. This particular constraint is called the *CHSH inequality* after its inventors Clauser, Horne, Shimony, and Holt [41].

Quantum mechanics allows us to violate Ineq. (5.1). For example, let Charlie create an EPR pair $|\psi\rangle := (|01\rangle - |10\rangle)/\sqrt{2}$ and send the first qubit to Alice and the second to Bob. Alice chooses uniformly at random measurement $A := \sigma_x = |0\rangle\langle 1| + |1\rangle\langle 0|$ or $A' := \sigma_z = |0\rangle\langle 0| - |1\rangle\langle 1|$; Bob measures in the Hadamard basis and flips the outcome at random, i.e.,

$$B := \frac{1}{\sqrt{2}}(-\sigma_z - \sigma_x) = -\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}$$

and

$$B' := \frac{1}{\sqrt{2}}(\sigma_z - \sigma_x) = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix} \ .$$

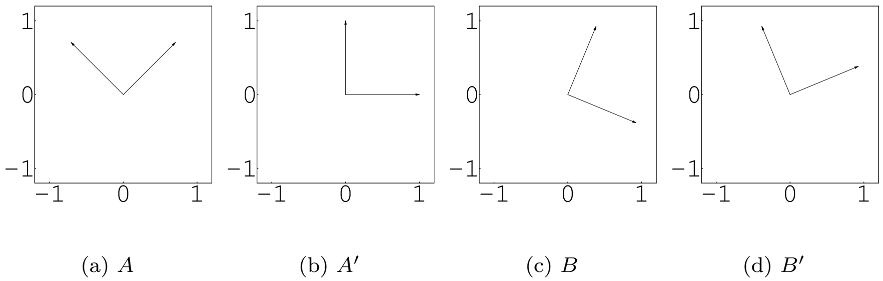
These observables correspond to the measurement bases depicted in Figure 5.2.

By the laws of quantum mechanics, $\mathrm{E}[AB] = \langle\psi|(A \otimes B)|\psi\rangle$ and correspondingly for the other three combinations of measurements. Carrying out the linear algebra gives $\langle\psi|(A \otimes B)|\psi\rangle = 1/\sqrt{2}$ and further

$$\mathrm{E}[AB] + \mathrm{E}[A'B] + \mathrm{E}[AB'] - \mathrm{E}[A'B'] = 2\sqrt{2} \ . \tag{5.2}$$

What is the essential ingredient that permits quantum mechanics to bypass Ineq. (5.1)? Contrary to classical models, it is not possible to assign a value to the measurements that are not carried out. This is because the measurement bases are not mutually orthogonal.

It turns out that in the setting of 2 parties, 2 settings, and 2 detectors, this is the greatest violation achievable by quantum mechanics. This was shown by Cirel'son [40]. The maximum possible value for the left-hand side of (5.1) is 4 and can be achieved if Alice and Bob communicate.



(a) A (b) A' (c) B (d) B'

Figure 5.2: Measurement bases; the measurement outcome 1 corresponds in the first three cases to the vector extended the furthest right and in the last case to the one leaning left.

5.1.2 Imperfections

The detection loophole Experimental realizations of nonlocality tests are hampered by noise and imperfections in the physical apparatus. In particular, measurement devices for individual quantum systems, e.g., single-photon detectors, tend to fail on most runs of the experiment, allowing local classical explanations of the data by means of local classical theories that are allowed to make the same kind of errors and this opens the so-called "detection loophole."

We consider again the nonlocality experiment from the preceding subsection. Let U_A denote the basis transformation from the computational basis to the measurement basis of observable A so that $A = U_A^*(|0\rangle\langle 0| - |1\rangle\langle 1|)U_A$; we define $U_{A'}$, U_B, and $U_{B'}$ in the same way. Then the measurement of A by Alice and B by Bob corresponds to a measurement in the computational basis of the state

$$\begin{aligned}(U_A \otimes U_B)|\psi\rangle &= \frac{1}{\sqrt{2}} \sum_{i\in\{0,1\}} (-1)^i U_A|i\rangle U_B|1-i\rangle \\ &= \frac{1}{\sqrt{2}} \sum_{i\in\{0,1\}} (-1)^i |i\rangle U_B U_A^T|1-i\rangle\end{aligned}$$

Measuring this state in the computational basis will yield outcome i for Alice and j for Bob with probability

$$\frac{1}{2}\left|\langle j|U_B U_A^T|i\rangle\right|^2 \quad .$$

When Alice and Bob want to reproduce these correlations classically, Alice does not know the measurement basis of Bob and vice versa, therefore they cannot easily compute the probabilities locally. So what they do is, they separate the expression in a part that can be evaluated by Alice and a part that can be evaluated by Bob by sandwiching a projector $|\varphi\rangle\langle\varphi|$ between the parts that Alice has knowledge of and the parts Bob knows:

$$\begin{aligned}\frac{1}{2}\left|\langle j|U_B U_A^T|i\rangle\right|^2 &\geq \frac{1}{2}\left|\langle j|U_B|\varphi\rangle\langle\varphi|U_A^T|i\rangle\right|^2\\ &= \frac{1}{2}\left|\langle j|U_B|\varphi\rangle\right|^2\cdot\left|\langle\varphi|U_A^T|i\rangle\right|^2\end{aligned}$$

If Alice and Bob know $|\varphi\rangle$, then Alice can locally sample from $\Pr[\text{output } i] = \left|\langle\varphi|U_A^T|i\rangle\right|^2$ and Bob from $\Pr[\text{output } j] = \left|\langle j|U_B|\varphi\rangle\right|^2$. Every choice of $|\varphi\rangle$ yields a local hidden variable model with perfect detector efficiency, since both parties produce an outcome. Such a model cannot violate Ineq. (5.1). However, Gisin and Gisin [62] showed that if Alice only produces an output with probability $|\langle\varphi|A|\varphi\rangle|^2$ where A is the observable she measures, Bob always produces an output, and $|\varphi\rangle$ is chosen uniformly at random, then Eq. (5.2) holds just like in the quantum case. Alice's overall detector efficiency is $1/2$ in this case. In fact, in our restricted setting it is sufficient to have $|\varphi\rangle = |0\rangle$ or $|\varphi\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ with probability $1/2$ each. This means a single shared random bit is sufficient to violate the CHSH inequality. With little additional overhead, the marginal detector efficiency of Alice and Bob can be made the same, barring obvious ways to identify this local hidden variable model and thus rendering it more plausible.

Communication complexity The amount of communication needed to solve computational tasks is a well-studied problem in computer science. *Communication complexity* was introduced by Abelson and Yao [1, 119]. Alice has an n-bit string x and Bob has an n-bit string y and their goal is to compute some function $f : \{0,1\}^n \times \{0,1\}^n \mapsto \{0,1\}$, minimizing the number of bits they communicate to each other. The area of communication complexity is well studied; see for example the books by Kushilevitz and Nisan [80] and Hromkovič [76]. Cleve and Buhrman [43] and Buhrman, Cleve, and Wigderson [29] initiated the study of quantum communication complexity where Alice and Bob can exchange qubits or share entangled parts of a quantum state.

Ideas from quantum communication complexity have been used by Brassard et al. [25], Massar [86], and us [34] to propose new nonlocality experiments and to bound the maximum detector efficiency, minimum noise, and hidden communication using which the results can be explained by means of a classical local model. The goal is to construct an experiment that demonstrates the nonlocal character of quantum mechanics even when the experiments are faulty and make errors.

Deutsch-Jozsa correlations To demonstrate how ideas from combinatorics can be used to propose new nonlocality experiments, we take another look at the Deutsch-Jozsa problem (see Subsection 1.3.2, p. 19 ff.). The first gap for two-party qubit communication complexity was demonstrated by Buhrman, Cleve, and Wigderson [29]; their quantum protocol is inspired by the Deutsch-Jozsa algorithm.

The problem is as follows: Alice has $x \in \{0,1\}^n$, Bob has $y \in \{0,1\}^n$ and they are promised that either $x = y$ or x and y differ in exactly $n/2$ positions. Their task is to find out which of the two is the case. This amounts to figuring out whether $x_1 \oplus y_1 \dots x_n \oplus y_n$ is constant or balanced, since in the constant 0 case $x = y$ and in the balanced $x \neq y$. So if we set $X_i = x_i \oplus y_i$ we are back at the Deutsch-Jozsa problem.

If Alice could somehow obtain the final state from Eq. (1.22) on page 22, with 2^n now n and n now $\ell := \log n$,

$$\frac{1}{\sqrt{n}} \sum_{i \in \{0,1\}^\ell} \frac{1}{\sqrt{n}} \sum_{j \in \{0,1\}^\ell} (-1)^{X_i + i \cdot j} |j\rangle|1\rangle \ ,$$

she would do a final measurement and know the answer. To this end Bob prepares the following state:

$$\frac{1}{\sqrt{n}} \sum_{i \in \{0,1\}^\ell} |i\rangle \frac{1}{\sqrt{2}} (|0 \oplus y_i\rangle - |1 \oplus y_i\rangle)$$

and sends these $\log(n) + 1$ qubits to Alice. Alice then performs the unitary transformation that changes state $|i\rangle|b\rangle$ to $|i\rangle|b \oplus x_i\rangle$ resulting in state:

$$\frac{1}{\sqrt{n}} \sum_{i \in \{0,1\}^\ell} |i\rangle \frac{1}{\sqrt{2}} (|0 \oplus y_i \oplus x_i\rangle - |1 \oplus y_i \oplus x_i\rangle) \ ,$$

which can be rewritten precisely to the state from Eq. (1.20):

$$\frac{1}{\sqrt{n}} \sum_{i \in \{0,1\}^\ell} (-1)^{X_i} |i\rangle \frac{1}{\sqrt{2}} (|0\rangle - |1\rangle)$$

Next Alice proceeds as in the Deutsch-Josza algorithm and applies $H^{\otimes \log(n)+1}$ and measures the final state.

Following proposals by Brassard et al. [25] and Massar [86], the Deutsch-Jozsa communication problem can be turned into a nonlocality experiment.

This time, Alice and Bob cannot communicate, but they start out sharing a quantum state, receive classical bit strings $x, y \in \{0,1\}^n$, respectively; both Alice and Bob produce outputs, $a, b \in \{0,1\}^\ell$, respectively, and we are interested in the correlations between these outputs, namely the probability distributions $\Pr[a, b \mid x, y]$ of Alice outputting a and Bob outputting b given that Alice got input x and Bob input y. Recall that the "trick" in turning the Deutsch-Jozsa algorithm into a communication protocol was to let Bob perform the first steps of the algorithm and then send the quantum state to Alice who completed the steps with her input. Now, since Alice and Bob cannot communicate, we replace the quantum channel by EPR pairs. Alice and Bob start out with the following state comprised of $\ell = \log(n)$ EPR pairs and two auxiliary qubits:

$$\frac{1}{2\sqrt{n}} \sum_{i \in \{0,1\}^\ell} |i\rangle \left(|0\rangle - |1\rangle\right) |i\rangle \left(|0\rangle - |1\rangle\right)$$

Here, Alice has the first $\ell+1$ qubits and Bob the remaining $\ell+1$ qubits. Now both Alice and Bob pretend that they are in a local execution of the Deutsch-Jozsa algorithm before the oracle query, as given in Eq. (1.19). Accordingly, they perform the operation $|i\rangle|b\rangle \mapsto |i\rangle|b \oplus y_i\rangle$ on their part of the state, resulting in the following global state:

$$\frac{1}{2\sqrt{n}} \sum_{i \in \{0,1\}^\ell} (-1)^{x_i + y_i} |i\rangle \left(|0\rangle - |1\rangle\right) |i\rangle \left(|0\rangle - |1\rangle\right)$$

Then they apply the Hadamard operation on their $\ell + 1$ qubits, yielding the state

$$\begin{aligned} \frac{1}{n\sqrt{n}} \sum_{i \in \{0,1\}^\ell} (-1)^{x_i + y_i} &\left(\sum_{a \in \{0,1\}^\ell} (-1)^{(i,a)} |a\rangle \right) |1\rangle \left(\sum_{b \in \{0,1\}^\ell} (-1)^{(i,b)} |b\rangle \right) |1\rangle \\ &= \frac{1}{n\sqrt{n}} \sum_{a,b \in \{0,1\}^\ell} \left(\sum_{i \in \{0,1\}^\ell} (-1)^{x_i + y_i + (i, a \oplus b)} \right) |a\rangle|1\rangle|b\rangle|1\rangle \end{aligned}$$

Now they both measure and output their measurement result. By the laws of quantum mechanics, the probability for Alice to observe $|a\rangle|1\rangle$ and Bob $|b\rangle|1\rangle$ is

$$\Pr[a, b \mid x, y] = \frac{1}{n^3} \left(\sum_{i \in \{0,1\}^\ell} (-1)^{x_i + y_i + (i, a \oplus b)} \right)^2$$

If $x = y$, then

$$\Pr[a, b \mid x, y] = \begin{cases} \frac{1}{n} & \text{if } a = b \\ 0 & \text{if } a \neq b \end{cases}$$

whereas for $\Delta(x, y) = n/2$ and $a = b$ we have $\Pr[a, b \mid x, y] = 0$. Hence, the outputs are correlated in that whenever $x = y$, we always see $a = b$ and whenever $\Delta(x, y) = n/2$, we never see $a = b$.

Can these correlations be realized by a classical protocol with shared randomness and no communication? No, since then Bob could send his output to Alice, solving the communication problem with $\mathrm{O}(\log n)$ bits, which is ruled out by the *Holevo bound* [73], which implies that for transmitting n bits over a quantum channel $\Omega(n)$ qubits are needed. Then, how closely can they be realized approximately, i.e., how precise does an experiment need to be? For the detection loophole, it is assumed that every measurement succeeds with probability at least η and if it fails, there will be no output. Then η^2 is the probability that both Alice's and Bob's measurements succeed. If the world is classical, we have an adversary who is trying to reproduce the correlations without communication using the possibility not to produce an output on an η^2 fraction of the runs of the experiment. By the Yao principle [118] there will be for every distribution on the inputs a classical local *deterministic* strategy that produces a correct output for an η^2 fraction of the inputs. Consider the input distribution where $x \in \{0, 1\}^n$ is chosen uniformly at random and $y = x$; fix the best deterministic strategy. Let $Z_a = \{x : \text{Alice and Bob output } a\}$, then

$$\eta^2 2^n \leq \sum_{a \in \{0,1\}^\ell} |Z_a|$$

Moreover, for each $a \in \{0, 1\}^\ell$, $Z_a \subseteq \{0, 1\}^n$ must not contain x, y with $\Delta(x, y) = n/2$, therefore, by a combinatorial theorem by Frankl and Rödl [59], $|Z_a| \leq 2^{0.993n}$. This implies $\eta^2 2^n \leq n 2^{0.993n}$ or $\eta \leq \sqrt{n 2^{-0.007n}}$. Hence, with growing n, the detector efficiency at which there still exists a classical local model decreases exponentially. So if the quality of the measurement equipment does not decrease too fast with growing n, the detection loophole can be "closed" with an experiment for the Deutsch-Jozsa correlations.

There are several shortcomings in this approach. In a nonlocality experiment, the input distribution should be a product distribution so that it can be implemented locally in the lab. Furthermore, there are very efficient classical *bounded-error* protocols for equality, implying that the quantum correlations above can be very well simulated classically if the experiment is subject to noise. And finally, an asymptotic analysis is often too coarse since the region where the bounds kick in may be out of reach experimentally.

In the remainder of this chapter, we describe our results that address some of these questions. In Section 5.2 we give formal definitions for the investigation of multiparty nonlocality experiments; building on these, we prove in Section 5.3 that a multiparty nonlocality experiment is asymptotically robust both with inefficient detectors and noise. Section 5.4 provides a new upper bound on the amount of communication needed to reproduce quantum correlations in a classical world. We discuss our results and interesting open questions in Section 5.5.

5.2 Definitions

Consider the following situation. There are n spatially separated parties; party i receives an input $x_i \in \{1, \ldots, k\}$ and produces an output $a_i \in \{1, \ldots, \ell\}$. With $x = (x_1, \ldots, x_n)$ and $a = (a_1, \ldots, a_n)$, let $P(a|x)$ denote the probability of output a given input x. The inputs are distributed according to the probability distribution $\mu(x)$. We formalize this situation as follows.

5.2.1. Definition. An (n, k, ℓ) *correlation problem with input distribution* μ is a family of probability distributions $P(\cdot|x)$ on the "outputs" $\{1, \ldots, \ell\}^n$, for each "input" $x \in \{1, \ldots, k\}^n$ with $\mu(x) > 0$. We denote the support of μ by $D := \{x : \mu(x) > 0\}$.

Note that we usually only consider product distributions for μ—otherwise, a nonlocality experiment would have trouble selecting x according to μ when the detectors are timelike separated.

We are interested in correlation problems obtained from measurements on multipartite entangled quantum states. We define these as follows.

5.2.2. Definition. An (n, k, ℓ) *measurement scenario* is a correlation problem in which the parties share an entangled state $|\psi\rangle$; each input x_i determines a positive operator valued measure (POVM) $\hat{x}_i = \{\hat{x}_i^1, \ldots, \hat{x}_i^\ell\}$ with $\hat{x}_i^j \geq 0$, $\sum_{j=1}^{\ell} \hat{x}_i^j = \mathbb{1}_i$. If the measurement of party i produces outcome $\hat{x}_i^j$, then it outputs $a_i = j$. The probability $P_{\mathrm{QM}}(a|x)$ to obtain outcome a given input x is

$$P_{\mathrm{QM}}(a|x) = \langle\psi|\hat{x}_1^{a_1} \otimes \cdots \otimes \hat{x}_n^{a_n}|\psi\rangle \ .$$

Our aim is to study what classical resources are required to reproduce such measurement scenarios. Let us first consider classical models in which the parties cannot communicate after they have received the inputs. Such models are called *local*. The best the parties can do in this case is to randomly select in advance a deterministic strategy. This motivates the following definition.

5.2.3. Definition. A *deterministic local hidden variable (lhv) model* is a family of functions $\lambda = (\lambda_1, \ldots, \lambda_n)$ from the inputs to the outputs: $\lambda_i : \{1, \ldots, k\} \to \{1, \ldots, \ell\}$. Each party outputs $a_i = \lambda_i(x_i)$.

A *probabilistic lhv model* (or just *lhv model*) is a probability distribution $\nu(\lambda)$ over all deterministic lhv models for given (n, k, ℓ).

Thus in probabilistic lhv models the parties first randomly choose a deterministic lhv model λ using the probability distribution ν. Each party then outputs $a_i = \lambda_i(x_i)$.

We also consider classical models with communication. In such models, the parties may communicate over a possibly superluminal classical broadcast channel in order to reproduce the quantum correlations P_{QM}. Different communication models exist depending on whether the parties do not have access to randomness, possess local randomness only, or share randomness. These notions are adapted from the corresponding definitions in communication complexity.

5.2.4. Definition. Consider n parties who each receive an input $x_i \in \{1, \ldots, k\}$, communicate over a classical broadcast channel, and each produce an output $a_i \in \{1, \ldots, \ell\}$.

A *deterministic classical model with communication* is a rooted "communication protocol" tree $\mathcal{P}$; each internal node u is labeled with the party $i_u \in \{1, \ldots, n\}$ whose turn it is to broadcast a message; each edge e from u to a descendant is labeled with a set $\S_e \subseteq \{1, \ldots, k\}$ so that the $\S_e$ form a partition of $\{1, \ldots, k\}$; each leaf v is labeled with a lhv model λ_v. An execution of the protocol on input x starts at the root of tree; until a leaf is reached, the execution proceeds from node u to the descendant of u that is reached via the edge e with $x_{i_u} \in \S_e$. It is understood that the choice of the edge is broadcast to all parties so that all parties know at each moment at which node the execution is. When the execution has reached the leaf v, each party i outputs $\lambda_{v,i}(x_i)$ and the execution terminates. If there are m leaves and if the number of children of the nodes on the path from the root to the final leaf is $t_1, \ldots, t_m$, the number of bits broadcast is $c = \lceil \log t_1 \rceil + \cdots + \lceil \log t_m \rceil$.

A *classical model with shared randomness* is an arbitrary probability distribution $\nu(\mathcal{P})$ over deterministic classical models. An execution of such a model first probabilistically selects a deterministic model and then evaluates the deterministic model.

In a *classical model with local randomness*, the distribution $\nu(\mathcal{P})$ is constrained to be a product distribution of the individual strategies of the parties.

Of course, a classical model that always uses 0 bits of communication is just a lhv model.

5.2.5. Definition. For a correlation problem P with input distribution μ, we denote by $D(P)$, $R(P)$, and $R^{\mathrm{pub}}(P)$, respectively, the minimum number of bits that must be broadcast in order to perfectly reproduce the correlations P when the parties are deterministic, have local randomness only, or have shared randomness.

Where the choice of the correlation problem P is clear from the context, we drop it and write D, R, and R^{pub}.

Clearly, $D(P) \geq R(P) \geq R^{\text{pub}}(P)$. Since the results of quantum measurements are inherently random, it is in impossible to reproduce the quantum correlations using deterministic lhv models or using deterministic models with communication. Thus $D(P)$ is meaningless when trying to simulate quantum measurement scenarios. However, deterministic models are a very useful tool for studying the probabilistic models because properties of *all* deterministic models necessarily also hold for *all* probabilistic models, because the probabilistic models are just probabilistic mixtures of deterministic models. Note also that Massar et al. [87] showed that $R(P)$ can be infinite when P arises from a quantum measurement scenario.

In general, classical models cannot reproduce the quantum correlations P_{QM} unless communication is possible, the detector efficiency η is sufficiently small, or they are allowed to make errors. Let us consider now the situation where the detectors are inefficient.

In the case of inefficient detectors we enlarge the space of outputs to $a_i \in \{1, \ldots, \ell\} \cup \{\perp\}$, where $a_i =\perp$ is the event that the ith detector does not produce an output ("click"). We suppose that each measurement $\hat{x}_i$ has probability η of giving a result and a probability $1 - \eta$ of not giving a result. Whether a detector clicks or does not click is independent of the other detectors. This affects the probabilities in a more structured way than simply decreasing the probability that all detectors click simultaneously. This issue has been discussed by Massar and Pironio [88]; for simplicity we will consider here only the two extreme cases, namely that all detectors click (which occurs with probability η^n) or that at least one detector does not click. We define detector efficiency accordingly.

5.2.6. Definition. Let $P(\cdot|x)$ be a fixed (n, k, ℓ) correlation problem with input distribution μ. Let

$$C := \{a : \forall i \; a_i \neq \perp\}$$

denote the output vectors where all detectors click. With slight abuse of notation, we also use C as the indicator random variable of the event $a \in C$. We define the *detection efficiency* η of the correlations to be the expectation

$$\eta := \left(\mathrm{E}_\mu \left[\sum_a P(a|x) C \right] \right)^{1/n} .$$

Note that here the atomic events are tuples (x, a) of an input and an output vector with a joint distribution of the form Pr[input x and output a] =

$\mu(x)P(a|x)$. The expectation above is over the marginal distribution μ of the inputs.

We are also interested in the possibility that the lhv model makes errors.

5.2.7. Definition. Suppose that some classical model produces a probability distribution $P(a|x)$, which should approximate the probability distribution produced by a measurement scenario $P_{QM}(a|x)$. The *total-variation distance* is a measure for how much these two distributions differ:

$$\varepsilon_{\text{var}} := \mathrm{E}_{\mu}\left[\sum_{a} |P_{QM}(a|x) - P(a|x)| \, \frac{C}{\eta^{n}}\right]$$

The inclusion of the factor C/η^n takes care of the possible finite efficiency of the detectors, assumed to be the same for $P_{QM}(a|x)$ and for $P(a|x)$.

We will be particularly interested in quantum correlations that exhibit "pseudo telepathy", i.e., such that $P_{QM}(a|x) = 0$ for some a and x. For such correlations it is convenient to define the error probability as follows.

5.2.8. Definition. Let

$$F := \{(a,x) : P_{QM}(a|x) = 0\}$$

and again we also denote by F the indicator random variable of the event $P_{QM}(a|x) = 0$. The *error probability* is

$$\varepsilon := \mathrm{E}_{\mu}\left[\sum_{a} P(a|x) F \frac{C}{\eta^{n}}\right] \quad .$$

Thus ε is the probability to observe in one run an event that cannot occur in the quantum mechanical model. It is immediate to check that

$$\varepsilon_{\text{var}} \geq \varepsilon \quad .$$

For an (n, k, ℓ) correlation problem $P(\cdot|x)$ with input distribution μ, we denote by η^* the maximum detector efficiency of any lhv model that reproduces the quantum correlations, and by η^*_ε the maximum detector efficiency that reproduces the quantum correlations up to error ε. Similarly, we can define D_ε, R_ε, $R^{\text{pub}}_\varepsilon$ the amounts of communication required to reproduce the correlation problem P in the presence of error. We are interested in η^*_ε and by $R^{\text{pub}}_\varepsilon$. Below, we will generally drop the subscript and just write ε as it is clear that throughout the following discussion we allow the possibility of error.

We can map every communication model with c bits of communication with shared randomness into a model with inefficient detectors with efficiency

$\eta^n = 2^{-c}$: the shared randomness determines the conversation between the parties. Thus they all agree on the conversation. Each party i checks whether its input x_i is compatible with the conversation and, if yes, produces output a_i according to the communication model and otherwise produces no output, i.e., $\perp$. The total probability that all detectors click is equal to the probability that x belongs to the conversation. Since each input belongs to one and only one conversation, the probability that all detectors click is equal to one over the number of conversations. Note that in this model the probability that a specific detector, say detector i, clicks may depend on the input x_i. However, the probability that all detectors click remains independent of the input.

5.2.9. Theorem. *Consider lhv models where the probability that all detectors click is independent of the input, but where the probability that each detector clicks, say detector i, may depend on its input x_i. Then there exists a lhv model if the probability η^n that all detectors click is at most $2^{-R^{pub}}$. This implies that in these models,*

$$(\eta^*)^n \geq 2^{-R^{pub}} \quad . \tag{5.3}$$

This result was given in [35] in the absence of error, but it also holds when errors are present.

5.3 Bounds on Multiparty Nonlocality

5.3.1 Combinatorial bounds

We now introduce some definitions and notation, which allow us to state and then prove our result concerning a general relation between c, η and ε. We are concerned with pseudo-telepathy type correlations for which there are some $P(a|x)$ that vanish.

5.3.1. Definition. Let $P(\cdot|x)$ be a fixed (n, k, ℓ) correlation problem with input distribution μ. We define the sets of inputs that admit output a as

$$\mathrm{adm}(a) := \{x : P(a|x) > 0\}$$

for all $a \in C$. Moreover, for a set $S \subseteq \{1, \ldots, k\}^n$ of inputs and a specific output $a \in \{1, \ldots, \ell\}^n$, the *$a$-advantage of S* is

$$\mathrm{adv}_a(S) := \frac{\mu(S \cap \mathrm{adm}(a))}{\mu(S)}$$

for all $a \in C$.

For sets $A_1, \ldots, A_n$, a subset R of the Cartesian product $A_1 \times \cdots \times A_n$ is called a *rectangle* if there are $R_1 \subseteq A_1, \ldots, R_n \subseteq A_n$ such that $R = R_1 \times \cdots \times R_n$, i.e., R is a Cartesian product itself. The importance of rectangles is that for a deterministic lhv model $\lambda = (\lambda_1, \ldots, \lambda_n)$, the set $R_\lambda(a) := \{x : \lambda(x) = a\}$ of all inputs x leading to output a is a rectangle: $R_\lambda(a) = \lambda_1^{-1}(a_1) \times \cdots \times \lambda_n^{-1}(a_n)$.

5.3.2. Theorem. *Let P be a fixed (n, k, ℓ) correlation problem with input distribution μ. If for some δ $(0 \leq \delta \leq 1)$, all rectangles R with $\mathrm{adv}_a(R) \geq \delta$ have $\mu(R) \leq r$ for every $a \in C$, then for every classical model $\nu(\mathcal{P})$ with c bits of communication holds*

$$\frac{1}{2^c}\eta^n\left(1 - \varepsilon\frac{1}{1-\delta}\right) \leq \ell^n r.$$

This shows the strong relation between the detection efficiency and the amount of classical communication required to reproduce the correlations. Indeed one quantity can be traded for the other.

Proof of Theorem 5.3.2. Let $R_{\mathcal{P},v,a}$ denote the set of inputs x for which the deterministic protocol $\mathcal{P}$ terminates in leaf v and outputs a. Every $R_{\mathcal{P},v,a}$ is a rectangle. Let $L := \{(\mathcal{P}, v, a) : \mathrm{adv}_a(R_{\mathcal{P},v,a}) \geq \delta\}$. Then

$$\begin{aligned}
\eta^n(1-\varepsilon) &= \sum_{\mathcal{P},x} \nu(\mathcal{P})\mu(x)C(1-F) \\
&= \sum_{\mathcal{P},v,a} \nu(\mathcal{P})\mu(R_{\mathcal{P},v,a} \cap \mathrm{adm}(a)) \\
&= \sum_{\mathcal{P},v,a} \nu(\mathcal{P})\mu(R_{\mathcal{P},v,a})\,\mathrm{adv}_a(R_{\mathcal{P},v,a}) \\
&\leq \sum_{(\mathcal{P},v,a)\in L} \nu(\mathcal{P})r + \sum_{(\mathcal{P},v,a)\notin L} \nu(\mathcal{P})\mu(R_{\mathcal{P},v,a})\delta \\
&\leq 2^c d^n r + \delta \sum_{(\mathcal{P},v,a)\notin L} \nu(\mathcal{P})\mu(R_{\mathcal{P},v,a})
\end{aligned}$$

where the v range over the leafs of $\mathcal{P}$ and the a over $\{1, \ldots, \ell\}^n$. Similarly,

$$\begin{aligned}
\eta^n\varepsilon &= \sum_{\mathcal{P},v,a} \nu(\mathcal{P})\mu(x)CF \\
&= \sum_{\mathcal{P},v,a} \nu(\mathcal{P})\mu\Big(R_{\mathcal{P},v,a} \cap \big(\{1, \ldots, k\}^n \setminus \mathrm{adm}(a)\big)\Big) \\
&= \sum_{\mathcal{P},v,a} \nu(\mathcal{P})\mu(R_{\mathcal{P},v,a})\big(1 - \mathrm{adv}_a(R_{\mathcal{P},v,a})\big)
\end{aligned}$$

$$
\begin{aligned}
&\geq 0 + \sum_{(\mathcal{P},v,a)\notin L} \nu(\mathcal{P})\mu(R_{\mathcal{P},v,a})(1-\delta) \\
&= (1-\delta) \sum_{(\mathcal{P},v,a)\notin L} \nu(\mathcal{P})\mu(R_{\mathcal{P},v,a})
\end{aligned}
$$

Hence,

$$\eta^n(1-\varepsilon) \leq 2^c \ell^n r + \frac{\delta}{1-\delta}\eta^n \varepsilon \ ,$$

which implies Theorem 5.3.2. □

5.3.2 Application to the GHZ correlations

In this measurement scenario each of the n parties has a two-dimensional quantum system. The overall state of the n qubits is

$$|\psi\rangle = \frac{|0^n\rangle + |1^n\rangle}{\sqrt{2}} \tag{5.4}$$

where $|i^n\rangle = |i\rangle \otimes \ldots \otimes |i\rangle$ with n terms in the product. Each party receives as input $x_i \in \{0, \ldots, k-1\}$. Each party then measures his qubit in the basis

$$|\varphi_\pm\rangle = \frac{|0\rangle \pm e^{\pi \,\mathrm{i}\, x_i/k}|1\rangle}{\sqrt{2}} \tag{5.5}$$

If the qubit is projected onto state $|\varphi_+\rangle$, then party i outputs $a_i = 0$; and if the qubit is projected onto state $|\varphi_-\rangle$, party i outputs $a_i = 1$. As we explain below, the outputs are correlated to the inputs as follows:

$$
\begin{aligned}
\text{if} \qquad & \sum_{i=1}^n x_i \mod k = 0 \\
\text{then} \qquad & \sum_{i=1}^n a_i \mod 2 = \frac{1}{k}\left(\sum_{i=1}^n x_i \mod 2k\right) \ .
\end{aligned} \tag{5.6}
$$

For $n = 3$ and $k = 2$ this constitutes the GHZ paradox as formulated by Mermin [93]. The case $k = 2$, arbitrary n was studied by Mermin [92]. In Buhrman et al. [31] and our earlier research [35] the case where the number of settings k is a power of two was considered. In [31] it was shown that the amount c of classical communication which the parties must broadcast in order to reproduce exactly the correlations Eq. (5.6) is $c = \mathrm{O}(n \log n)$ when $k = \mathrm{O}(n)$. And in [35] it was shown that the maximum detector efficiency η^* for which a local classical model can reproduce the correlations Eq. (5.6) decreases as $1/n$. Furthermore the arguments of [31, 35] show that for the correlations Eq. (5.6) these results are essentially optimal.

We now apply Theorem 5.3.2 to this measurement scenario with the assumption that $k = n^{1/6}$ is a power of two. Recall that in this example there are n parties, and each party i obtains input data $x_i \in \mathbb{Z}_k$. We call an input $x = (x_1, \dots, x_n)$ *valid* if it satisfies

$$\left(\sum_{i=1}^{n} x_i\right) \bmod k = 0 \tag{5.7}$$

and we let $D \subset \mathbb{Z}_k^n$ denote the set of all valid inputs. Let $F : \mathbb{Z}_k^n \to \{0,1\}$ denote the Boolean function on the valid inputs defined by

$$F(x) = \frac{1}{k}\left[\left(\sum_{i=1}^{n} x_i\right) \bmod 2k\right] .$$

The function F can be viewed as computing the $(1+\log k)$-th least significant bit of the sum of the x_i.

In a quantum setting, the parties can compute F easily. Assume that each party has a two-dimensional quantum system that is part of the entangled state

$$|\psi\rangle = (|0^n\rangle + |1^n\rangle)/\sqrt{2} \ .$$

Each party i carries out the following measurement on its subsystem: it performs the unitary transformation $|0\rangle \mapsto |0\rangle$, $|1\rangle \mapsto e^{2\pi\,\mathrm{i}\,x_i/2^k}|1\rangle$ and then measures an operator whose eigenstates are $(|0\rangle+|1\rangle)/\sqrt{2}$ and $(|0\rangle-|1\rangle)/\sqrt{2}$. The first outcome is assigned the value $a_i = 0$, the second the value $a_i = 1$. If Eq. (5.7) holds, then

$$\left(\sum_{i=1}^{n} a_i\right) \bmod 2 = \frac{1}{k}\left[\left(\sum_{i=1}^{n} x_i\right) \bmod 2k\right] = F(x) \ . \tag{5.8}$$

Hence, if each party broadcasts its measurement outcome then each party can locally compute $F(x)$.

5.3.3. Lemma. *In the model with prior entanglement and classical broadcast communication, the communication complexity of computing $F(x)$ is $\mathrm{O}(n)$.*

Moreover, the above measurement scenario will exactly reproduce the following $(n, k, 2)$ correlation problem (see Definition 5.2.1): let $\mu(x)$ be a distribution on the inputs that gives zero weight to the invalid inputs x, which do not satisfy Eq. (5.7), and let

$$P(a|x) := \begin{cases} \frac{1}{2^{n-1}} & \text{if } F(x) = a_1 + \cdots + a_n \mod 2 \\ 0 & \text{otherwise.} \end{cases}$$

for all $a \in \{0,1\}^n$ and $x \in D$.

A simple classical strategy for reproducing these correlations is for every party to broadcast its input. Hence, with $k = n^{1/6}$, the communication problem and the correlation problem can be solved exactly with $\mathrm{O}(n \log n)$ bits of communication. We show that this is essentially optimal, even allowing constant error probability.

5.3.4. Theorem. *Let μ be the uniform distribution on valid inputs. Then the number c of bits broadcast, the efficiency η and the error ε of every lhv model ν are constrained by*

$$\frac{1}{2^{c/n}}\eta\left(1-\varepsilon\left[2+\mathrm{O}\left(\frac{1}{n^{1/6}}\right)\right]\right)^{1/n}=\mathrm{O}\left(\frac{1}{n^{1/6}}\right).$$

5.3.5. Corollary. *Every bounded-error randomized public coin protocol for $F : \mathbb{Z}_k^n \to \{0,1\}$ with $k \geq n^{1/6}$ requires $\Omega(n \log n)$ bits of communication.*

We now turn to the proof of Theorem 5.3.4. We say a rectangle $R = A_1 \times \cdots \times A_n \subseteq k^n$ *involves m parties* if at least m of the n subsets A_i have size at least 2. Every rectangle involving at most m parties can have size at most k^m.

5.3.6. Lemma (Small rectangles are insignificant). *Every rectangle R involving at most $n^{5/6}$ parties satisfies $\log|R| \leq n^{5/6}\log k$.*

We say a rectangle R *has bias at most δ* if

$$|F^{-1}(1) \cap D \cap R| \leq (1+\delta)|F^{-1}(0) \cap D \cap R|$$

and

$$|F^{-1}(0) \cap D \cap R| \leq (1+\delta)|F^{-1}(1) \cap D \cap R|.$$

Note that for every a we have $\mathrm{adm}(a) \cap D = F^{-1}(a_1 + \cdots + a_n \mod 2) \cap D$. Therefore, if μ is a distribution that is uniform on D, then R has bias at most δ if and only if it has a-advantage at most $(1+\delta)/(2+\delta)$ for every a. The next lemma expresses that every "large" rectangle is almost unbiased.

5.3.7. Lemma (Large rectangles are almost unbiased). *Every rectangle involving at least $n^{5/6}$ parties has bias at most $\mathrm{O}(1/n^{1/6})$.*

The proof of Lemma 5.3.7 is based on addition theorems for cyclic groups and is given in the next subsection.

Proof of Theorem 5.3.4. Lemma 5.3.7 implies that each rectangle involving at least $n^{5/6}$ parties can have a-advantage at most $1/2 + \mathrm{O}(1/n^{1/6})$ for any a. Hence, rectangles with a-advantage greater than $1/2 + \mathrm{O}(1/n^{1/6})$ must involve less than $n^{5/6}$ parties. By Lemma 5.3.6, such a rectangle R has size less than $k^{n^{5/6}}$ and thus

$$\mu(R) = |R|/k^{n-1} \leq k^{n^{5/6}-n+1} = n^{-\frac{1}{6}(n-n^{5/6}-1)}.$$

Plugging these values into Theorem 5.3.2, we obtain

$$\frac{1}{2^c}\eta^n \left(1 - \varepsilon \left[2 + \mathrm{O}\left(\frac{1}{n^{1/6}}\right)\right]\right) \leq 2^{-\frac{1}{6}n \log n + \mathrm{O}(n)}$$

□

5.3.3 An addition theorem

Let $\mathbb{Z}_T$ denote the additive cyclic group of order T. Let $\mu_A(x)$ denote the multiplicity of an element x in the multiset A. For multisets A and B of $\mathbb{Z}_T$, let $A + B$ denote the multiset $\{a + b \,|\, a \in A, b \in B\}$.

5.3.8. Definition. We say a multiset A of $\mathbb{Z}_T$ has bias at most ε with respect to a subgroup $H \leqslant \mathbb{Z}_T$ if $\mu_A(a) \leq (1+\varepsilon)\mu_A(a+h)$ for all $a \in A$ and all $h \in H$.

5.3.9. Theorem (Addition Theorem). *Let $A_1, \ldots, A_r$ be subsets of $\mathbb{Z}_T$, each of size at least 2, with $r \geq T^3$ and $T = 2^t$ a power of 2. Then the multiset $A_1 + A_2 + \cdots + A_r$ has bias at most $\mathrm{O}(T^{3/2}/r^{1/2})$ with respect to the subgroup $\{0, 2^{t-1}\}$.*

Essentially, this theorem is derived by a sequence of simple reductions to the following observation: We may generate an almost uniformly distributed random number between 0 and $K - 1$ by flipping a fair coin K^2 times, and counting the number of heads modulo K.

5.3.10. Lemma. *For multisets A and B over $\mathbb{Z}_T$, if A has bias at most ε with respect to some subgroup H, then so does $A + B$. In particular, the multiset $A + \{d\}$ has the same bias as A.*

5.3.11. Lemma. *Let $f : \{0,1\}^s \to \mathbb{Z}_K$ be defined by*

$$f(a_1, \ldots, a_s) = \left(\sum_{i=1}^{s} a_i\right) \bmod K \ .$$

If $s \geq K^2$, then $|f^{-1}(x)| \leq \left(1 + 4\frac{K}{\sqrt{s}}\right) |f^{-1}(y)|$ for all $x, y \in \mathbb{Z}_K$.

Proof. First suppose $x \leq y$. Then

$$\begin{aligned}
|f^{-1}(x)| &= \sum_i \binom{s}{x+iK} \\
&= \sum_{i:y+iK<s/2} \binom{s}{x+iK} + \sum_{i:y+iK\geq s/2} \binom{s}{x+iK} \\
&\leq \sum_{i:y+iK<s/2} \binom{s}{y+iK} + \sum_{i:y+iK\geq s/2} \binom{s}{x+iK+K} + \binom{s}{s/2} \\
&\leq \sum_{i:y+iK<s/2} \binom{s}{y+iK} + \sum_{i:y+iK\geq s/2} \binom{s}{y+iK} + \binom{s}{s/2} \\
&= |f^{-1}(y)| + \binom{s}{s/2}.
\end{aligned}$$

Similarly, if $x > y$, then still $|f^{-1}(x)| \leq |f^{-1}(y)| + \binom{s}{s/2}$. Thus, for all $y \in \mathbb{Z}_K$, we have that $|f^{-1}(y)|$ is within $\binom{s}{s/2}$ of the average value of $\frac{2^s}{K}$. Hence,

$$\binom{s}{\frac{s}{2}} \leq \frac{4}{5}\frac{2^s}{K}\frac{K}{\sqrt{s}} \leq \frac{4}{5}\left(|f^{-1}(y)| + \binom{s}{\frac{s}{2}}\right)\frac{K}{\sqrt{s}} \ ,$$

from which follows

$$\binom{s}{\frac{s}{2}} \leq \frac{4}{5\frac{\sqrt{s}}{K} - 4}|f^{-1}(y)| \ .$$

□

5.3.12. Lemma. *Let $B_1 = \cdots = B_s = \{0, b\}$ be s identical size-2 subsets of $\mathbb{Z}_T$, with $s \geq T^2$. Then the multiset $B_1 + B_2 + \cdots + B_s$ has bias at most $4|H|/s^{1/2}$ with respect to the subgroup $H = \langle b \rangle$.*

Proof. Set $K = |H|$ and define function $f : \{0,1\}^s \to \mathbb{Z}_K$ by $f(a_1, \ldots, a_s) = \left(\sum_{i=1}^s a_i\right) \bmod K$. Then we may generate the multiset $B_1 + B_2 + \cdots + B_s$ as $b \cdot f(\{0,1\}^s)$. Applying Lemma 5.3.11 gives that f is almost unbiased on $\mathbb{Z}_K$ and hence $b \cdot f$ is almost unbiased with respect to H. □

5.3.13. Lemma. *Let $B_1, \ldots, B_r$ be size-2 subsets of $\mathbb{Z}_T$, with $r \geq T^3$. There exists a nontrivial subgroup $H \leqslant \mathbb{Z}_T$ such that $B_1 + B_2 + \cdots + B_r$ has bias at most $4T^{3/2}/r^{1/2}$ with respect to H.*

Proof. First suppose $0 \in B_i$ for all i. There exists some nontrivial element $b \in \mathbb{Z}_T$ such that $B_i = \{0, b\}$ for s of the subsets, with $s \geq r/T \geq T^2$. Applying Lemma 5.3.12 on these s subsets yields a multiset of bias at most

$4|\langle b\rangle|/s^{1/2} \leq 4T^{3/2}/r^{1/2}$ with respect to $\langle b\rangle$. By Lemma 5.3.10, adding the remaining $r-s$ subsets to this multiset does not increase the bias.

In general, we do not have that $0 \in B_i$ for all i. In this case, observe that by Lemma 5.3.10, adding any offset to a multiset does not change its bias, and thus we may reduce to the former case by adding an appropriate offset d_i to subset B_i such that $0 \in B_i + \{d_i\}$, for each i. □

Proof of Theorem 5.3.9. Let $B_i \subseteq_R A_i$ be a random size-2 subset of A_i, for each i. By Lemma 5.3.13, the sub-rectangle $R' = B_1 \times \cdots \times B_r$ is almost unbiased with respect to some nontrivial subgroup H'. Since H' is nontrivial, it contains $H = \{0, 2^{t-1}\}$, and hence R' is also almost unbiased with respect to H. By this selection process, every $(a_1, \ldots, a_r) \in A_1 \times \cdots \times A_r$ has the same probability of being selected and, hence, R itself is almost unbiased with respect to H. □

Proof of Lemma 5.3.7. Set $t = \frac{1}{6}\log n$ and $T = 2^t$. Consider any rectangle $R = A_1 \times \cdots \times A_n$ involving at least $r \geq n^{5/6} = T^5$ parties. By the Addition Theorem, the multiset $A_1 + \cdots + A_n$ has bias at most $\mathrm{O}(T^{3/2}/r^{1/2}) \subseteq \mathrm{O}(1/n^{1/6})$ with respect to $\{0, 2^{t-1}\}$. Hence, rectangle R has bias at most $\mathrm{O}(1/n^{1/6})$, too. □

5.4 Reproducing Quantum Correlations

In this section we investigate whether one can put general bounds on the amount of communication or on the threshold detection efficiency η^* required to reproduce quantum correlations, independently of the details of the measurement scenario. We will focus on the amount of classical communication required to reproduce the correlations, since Theorem 5.2.9 immediately provides a corresponding bound for η^*.

A first step is the observation that in an (n, k, ℓ) correlation problem, $c = n\log k$ bits of communication are always sufficient to reproduce the correlations classically: each party broadcasts its input. We proceed to prove a bound independent of the number of inputs k and of the number of outputs ℓ. Our bound depends solely on the number of parties and on the dimensionality of the quantum systems.

5.4.1. Theorem. *Consider a quantum measurement scenario involving n parties. The quantum system held by each party is of dimension d. Then*

$$2^{R^{\mathrm{pub}}_{\varepsilon_{\mathrm{var}}}} \leq \left(\frac{2nd^{n+3/2}}{\varepsilon_{\mathrm{var}}}\right)^{4dn} \left(1 + \mathrm{O}\left(\frac{1}{d^{n/3-1}}\right)\right) .$$

Thus $R^{\text{pub}}_{\varepsilon_{\text{var}}} = \mathrm{O}(n^2)$ for d fixed, $n \to \infty$, and $R^{\text{pub}}_{\varepsilon_{\text{var}}} = \mathrm{O}(d \log d)$ for n fixed, $d \to \infty$. The corresponding results for η^* follow immediately from Theorem 5.2.9. These results hold independently of the quantum state shared by the parties, of whether the measurements are von Neumann measurements or POVMs, of the number of inputs and of the number of outputs.

Note that the bound of Theorem 5.4.1 is independent both of the input distribution μ and of k, the number of possible inputs per party. The bound does not hold for arbitrary non-quantum correlations as we can model every multiparty communication problem by it and there are problems with $\Omega(\log k)$ required bits of communication (see [80]).

Proof of Theorem 5.4.1. We consider the situation where N parties each have a d-dimensional system. The overall state of the N systems is in an entangled state Ψ. Each party receives an input x_i. To each input x_i, party i associates a measurement with outcomes a_i. This measurement is a POVM described by its elements $x_i^{a_i}$, which are positive and sum to identity

$$x_i^{a_i} \geq 0 \quad , \quad \sum_{a_i} x_i^{a_i} = \mathbb{1}_i \ .$$

The probability of obtaining outcomes a_1 through a_N is

$$P(a_1, \ldots, a_N | x_1, \ldots, x_N) = \langle \Psi | x_1^{a_1} \otimes \ldots \otimes x_N^{a_N} | \Psi \rangle \ .$$

We first describe a classical protocol for exact simulation of this measurement scenario. The exact simulation may require infinite communication. In a second step, we approximate the exact simulation with finite precision and good bounds on the amount of communication.

Protocol for perfect simulation of quantum correlations

1. Each party has a classical description of the quantum state Ψ at its disposal, e.g., in form of the components of the state in some basis.

2. Without loss of generality we assume that the POVM elements have rank one and that the outcomes are positive numbers. Then each party i can write its POVM elements as

   $$x_i^{a_i} = |x_i^{a_i}| |x_i^{a_i}\rangle\langle x_i^{a_i}|$$

   where $|x_i^{a_i}\rangle$ are normalized states.

3. Denote $\psi^{(1)} = \Psi$.

4. For $k = 1$ to N,

5. Party k computes the probabilities

$$P(a_k) = \langle \psi^{(k)} | x_k^{a_k} \otimes \mathbb{1}_{k+1} \otimes \ldots \otimes \mathbb{1}_N | \psi^{(k)} \rangle$$

6. Party k randomly chooses outcome a_k using this probability distribution.

7. Party k broadcasts a classical description of the state $|x_k^{a_k}\rangle$ onto which his system has been projected

8. The parties compute the state $\psi^{(k+1)}$ as

$$\psi^{(k+1)} = \frac{\langle x_k^{a_k} | \psi^{(k)} \rangle}{\sqrt{\langle \psi^{(k)} || x_k^{a_k} \rangle \langle x_k^{a_k} | \otimes \mathbb{1}_{k+1} \otimes \ldots \otimes \mathbb{1}_N | \psi^{(k)} \rangle}}$$

   where $\langle x_k^{a_k} | \psi^{(k)} \rangle$ denotes the partial inner product: $|\psi^{(k)}\rangle = |x_k^{a_k}\rangle \otimes |\psi_1^{(k)}\rangle + |y\rangle \otimes |\psi_2^{(k)}\rangle$ for some not normalized $|\psi_1^{(k)}\rangle$, $|\psi_2^{(k)}\rangle$, and $|y\rangle$ with $\langle x|y\rangle = 0$; then $\langle x_k^{a_k} | \psi^{(k)} \rangle = |\psi_1^{(k)}\rangle$.

   The state $\psi^{(k+1)}$ is a normalized state belonging to the space of parties $k+1, \ldots, N$. It is the state that is obtained when parties $1, \ldots, k$ have carried out their measurement.

9. Next k

It is easy to check that the above protocol exactly reproduces the quantum correlations. Note that in the above protocol the only information that each party must broadcast is a classical description of the state on which his system has been projected. Note that the last party does not have to broadcast this information. Note also that it is not crucial for the states $|x_k^{a_k}\rangle$ to be normalized.

If the parties only give a finite-precision description of this state, then the amount of communication will be bounded, but the probabilities will not coincide exactly with the quantum probabilities. In order to analyze this in detail we set up a slightly modified measurement $\tilde{x}_i^{a_i}$. For this modified measurement the above simulation protocol requires only a finite amount of communication. We will then compare the amount of communication required to simulate the modified measurement to the amount by which the probabilities are modified. This will yield the upper bound on the amount of communication required in the presence of error.

Finite-precision approximation of the measurement As before we write the POVM elements as $x^a = |x^a||x^a\rangle\langle x^a|$ where $|x^a\rangle$ are normalized states. Suppose the states $|x^a\rangle$ are written in some fixed basis; then each component in this basis has a real and imaginary part, which we can approximate by a binary fraction. We write

$$|x^a\rangle = |\tilde{x}^a\rangle + |\bar{x}^a\rangle$$

where $|\tilde{x}^a\rangle$ is obtained by truncating the real and imaginary part of the components of $|x^a\rangle$ at the r-th bit. The number of bits required to describe the states $|\tilde{x}^a\rangle$ is $c = 2d(r+1)$ where we have taken into account that there are d real and d imaginary components, and that each component must be specified with its sign. The error on the state is bounded by $\langle \bar{x}^a|\bar{x}^a\rangle \leq 2d2^{-r}$.

We define new operators

$$\tilde{y}^a := |x^a||\tilde{x}^a\rangle\langle\tilde{x}^a| = |x^a|\left(|x^a\rangle\langle x^a| + z^a\right)$$

with

$$z^a := -|x^a\rangle\langle\bar{x}^a| - |\bar{x}^a\rangle\langle x^a| + |\bar{x}^a\rangle\langle\bar{x}^a| \ .$$

The operators $\tilde{y}^a$ are positive, but do not sum to identity. Let $|\varphi\rangle$ be an arbitrary normalized state. The largest eigenvalue of $\sum_a \tilde{y}^a$ is bounded by

$$\langle\varphi|\sum_a \tilde{y}^a|\varphi\rangle \leq 1 + \sum_a |x^a| \left(2\big|\langle\varphi|x^a\rangle\langle\bar{x}^a|\varphi\rangle\big| + \big|\langle\bar{x}^a|\varphi\rangle\big|^2\right) \leq 1 + d\Delta$$

where we applied the Cauchy-Schwarz inequality and let

$$\Delta := 2\sqrt{\langle\bar{x}^a|\bar{x}^a\rangle} + \langle\bar{x}^a|\bar{x}^a\rangle \leq 2\sqrt{2d}2^{-r/2} + \mathrm{O}(d2^{-r}) \ . \tag{5.9}$$

We now define the truncated POVM by the elements

$$\begin{aligned} \tilde{x}_a &:= \frac{\tilde{y}^a}{1+d\Delta} = \frac{|x^a|}{1+d\Delta}\left(|x^a\rangle\langle x^a| + z^a\right) \ , \\ R &:= \mathbb{1} - \sum_a \tilde{x}_a \end{aligned}$$

where R is an additional POVM element that is added to ensure that the POVM sums to the identity. Outcome R is interpreted as error, e.g., we can assume that output $\perp$ is produced. The probability of obtaining outcome R is bounded by

$$\langle\varphi|R|\varphi\rangle \leq \frac{2d\Delta}{1+d\Delta} \leq 2d\Delta \ .$$

Approximate measurements by n parties Let us now consider that there are n parties, each of which modifies his measurement as described above. Thus the measurements $x_i^{a_i}$ are modified into $\tilde{x}_i^{a_i}, R_i$. To estimate how much these modified measurements differ from the original measurement, simple arithmetic gives use the following bounds:

$$\begin{aligned}|P_{\text{exact}}(a|x) - P_{\text{approx}}(a|x)| &= |\langle\varphi|x_1^{a_1}\dots x_N^{a_N}|\varphi\rangle - \langle\varphi|\tilde{x}_1^{a_1}\dots\tilde{x}_N^{a_N}|\varphi\rangle| \\ &\leq \frac{|x^{a_1}|\dots|x^{a_n}|}{(1+d\Delta)^n}\big((1+d\Delta)^n + (1+\Delta)^n - 2\big) \\ &= |x^{a_1}|\dots|x^{a_n}|n(d+1)\Delta(1+\mathrm{O}(nd\Delta))\end{aligned}$$

for every a that is a vector of valid outputs. Thus,

$$\begin{aligned}\sum_a |P_{\text{exact}}(a|x) - P_{\text{approx}}(a|x)| &\leq \sum_a |x^{a_1}|\dots|x^{a_n}|n(d+1)\Delta(1+\mathrm{O}(nd\Delta)) \\ &= nd^n(d+1)\Delta(1+\mathrm{O}(nd\Delta)) \qquad (5.10)\end{aligned}$$

where we have used the fact that $\sum_a |x^{a_i}| = d$. Furthermore, the probability that at least one of the R_i results occur is

$$\begin{aligned}\Pr\big[\text{at least one } R_i \text{ result}\big] &\leq 1 - \left(1 - \frac{2d\Delta}{1+d\Delta}\right)^n \\ &\leq 2nd\Delta(1+\mathrm{O}(nd\Delta)) \ . \qquad (5.11)\end{aligned}$$

The total-variation distance ε_{var} is the sum of all, i.e., the sum of Eqs. (5.10) and (5.11),

$$\varepsilon_{\text{var}} \leq n\Delta(d^{n+1} + d^n + 2d)(1+\mathrm{O}(nd\Delta))$$

Thus the total-variation distance is small if $n\Delta d^{n+1}$ is small. Using Eq. (5.9) to replace Δ by its value in terms of the amount of communication, we obtain

$$\begin{aligned}\varepsilon_{\text{var}} &\leq n\Delta d^{n+1}(1+\mathrm{O}(nd\Delta)) \\ &\leq 2\sqrt{2}\frac{nd^{n+3/2}}{2^{r/2}}\left(1+\mathrm{O}\left(\frac{nd^{n+3/2}}{2^{r/2}}\right)\right) \ .\end{aligned}$$

Solving for $2^{r/2}$, we get

$$2^{r/2} \leq \frac{1}{\varepsilon_{\text{var}}}\sqrt{2}nd^{n+3/2}\left(1+\mathrm{O}\left(\frac{1}{\sqrt{nd^{n+3/2}}}\right)\right)$$

and now we can square both sides, multiply them by two, and take them to the power $2dn$:

$$2^{2dn(r+1)} \leq \left(\frac{2nd^{n+3/2}}{\varepsilon_{\text{var}}}\right)^{4dn}\left(1+\mathrm{O}\left(\frac{1}{d^{n/3-1}}\right)\right)$$

Since the $2dn(r+1)$ is the communication of our protocol, Theorem 5.4.1 follows. □

5.5 Conclusions

The work presented in this chapter aims at devising experiments for validating quantum nonlocality in the presence of noise and with imperfect detectors. Specifically we concentrated on the generalization of the GHZ paradox to n parties previously considered as a quantum communication complexity problem by Buhrman et al. [31].

The only prior asymptotic results in quantum communication complexity that hold in the presence of noise concern multi-round quantum communication protocols, such as the appointment-scheduling problem of Buhrman et al. [29] or the example due to Raz [100]. It appears that these results cannot be mapped to results concerning quantum nonlocality, whereas communication complexity problems with a single round of communication and nonlocal quantum correlations can generally be mapped one onto the other.

The multiparty problem considered by Buhrman et al. [31] was only proved in the absence of noise. We extended the classical lower bound to the bounded-error case and likewise made the corresponding correlation problem robust to noise. We considered the situation where there is a finite probability ε for an error to occur. We tied together the number of parties n, the number c of bits communicated via a superluminal channel, and the maximum detector efficiency η^* for which a local classical model exists:

$$\eta^* 2^{-c/n} = \mathrm{O}\left(n^{-1/6}\right) \ . \tag{5.12}$$

This implies that with bounded error and $\eta = 1$ we have $c = \Omega(n \log n)$; with bounded error and $c = 0$ holds $\eta^* = \mathrm{O}(n^{-1/6})$. Hence, the amount of communication and the detection efficiency can be traded one for the other. This result constitutes the first example in which the degree to which the quantum correlations are nonlocal increases with the size of the entangled system in the presence of noise.

There are several directions in which one may wish to improve the result Eq. (5.12). The first concerns the evaluation of the right-hand side of this relation. A detailed investigation of the proof shows that the right-hand side becomes nontrivial only for values of n that exceed a few hundred. Therefore our result will not be useful for the moderate values of n, say, $n \leq 10$, which may be attainable by real-world experiments in the next few years.

Another question concerns our notion of error, which is not entirely appropriate to a multiparty setting: one expects that each party may induce an error independently of the other parties. Thus it would be more natural to consider that the probability of an error goes as $\varepsilon = 1 - \delta^n$. We do not know whether a constraint of the form Eq. (5.12) holds in this case also.

To see whether the correlations Eq. (5.12) are amongst the strongest multiparty nonlocal quantum correlations, or whether there are other multiparty measurement scenario that exhibit much stronger nonlocality, we considered arbitrary measurement scenarios involving n parties, each holding a d-dimensional system. We derived bounds on the minimum amount c of classical communication that each party must broadcast in order to reproduce the quantum correlations or, alternatively, the maximum detector efficiency η^* for which a local classical model exists in the bounded error model:

$$c = \mathrm{O}\left(n^2\right) \tag{5.13}$$

$$\eta^* = \Omega\left(2^{-2dn\log d}\right) \quad . \tag{5.14}$$

These constraints are independent of the quantum state shared by the parties and of the number of inputs each party receives. A large gap remains open between our nonlocality experiment and the upper bounds of Eqs. (5.13) and (5.14): if we take as relevant quantity the average amount of communication broadcast by each party, c/n, then an exponential gap exists between our results Eq. (5.12) and Eq. (5.13). A corresponding exponential gap also exists for η^*. Closing this gap would either require to significantly improve Eqs. (5.13) and (5.14), or to find a completely different and much stronger example of multipartite quantum nonlocality.

Chapter 6

Quantum Coin Flipping

This chapter is based on joint research conducted with Ambainis, Buhrman, and Dodis [12].

6.1 Introduction

Research into quantum cryptography is motivated by two observations about quantum mechanics:

1. Nonorthogonal quantum states cannot be distinguished perfectly and parts of certain orthogonal quantum states cannot be distinguished if the remaining parts are inaccessible;

2. Measurement disturbs the quantum state. This is the so-called “collapse of the wave function.”

The second observation hints at the possibility of detecting eavesdroppers or other types of cheaters, whereas the first property appears to allow hiding data. Both rely on assumptions about the physical world, but are unhampered by unproven computational assumptions. Indeed, for the task of cooperatively establishing a random bit string between two parties in the presence of eavesdroppers, quantum key distribution [21, 89, 84] achieves security against the most general attack by an adversary that has unbounded computational power but has to obey the laws of quantum mechanics.

Initially, it was thought that these properties would admit protocols for the cryptographic primitive *bit commitment*. In bit commitment, there are two parties Alice and Bob; in the initial phase of the protocol, Alice has a bit b and communicates with Bob to “commit” to the value of b without revealing it. At a later time, Alice “unveils” her bit, allowing Bob to perform checks against the information obtained in the initial phase to test whether the revealed bit equals the committed bit. The properties sought of bit-commitment protocols are that they are *concealing* and *binding*: Bob does not learn anything about b in the initial phase and Bob will catch Alice trying to unveil $1-b$ instead of b.

Unfortunately, Mayers [90] and Lo and Chau [83] proved that perfect quantum bit commitment is impossible. Their impossibility result extends to *strong coin tossing* [91, 83], a weaker cryptographic primitive where the two parties want to agree on a random bit whose value cannot be influenced by either of them. Moreover, the impossibility extends even to the case of *weak coin tossing* [10], where outcome $b = 0$ is favorable for Alice and outcome $b = 1$ favorable for Bob, thus ruling out perfect quantum protocols for leader election. However, what turned out to be possible are coin-tossing protocols where there are guarantees on how much a cheater can bias the outcome.

Consider k parties out of which at least $g \geq 1$ are honest and at most $(k-g)$ are dishonest; which players are dishonest is fixed in advance but unknown to the honest players. The players can communicate over broadcast channels. Initially they do not share randomness, but they can privately flip coins; the probabilities below are with respect to the private random coins. A coin-flipping protocol establishes among the honest players a bit b such that

- if all players are honest, $\Pr[b=0] = \Pr[b=1] = 1/2$
- if at least g players are dishonest, then $\Pr[b=0], \Pr[b=1] \leq 1/2+\varepsilon$

ε is called the *bias*; a small bias implies that colluding dishonest players cannot strongly influence the outcome of the protocol. Players may abort the protocol. This allows the bad players to block outcomes they do not desire; therefore the quality of a coin-flipping protocol is measured in terms of the overall probability of forcing a fixed outcome. Frequent aborts reduce this figure of merit.

Classically, if a weak majority of the players is bad then no bias $< 1/2$ can be achieved and hence no meaningful protocols exist [104]. For example, if we only have two players and one of them is dishonest, then no protocols with bias $< 1/2$ exist. For a minority of bad players, quite non-trivial protocols exist. For example, Feige [52] elegantly showed that $(\frac{1}{2}+\delta)$-fraction of good players can achieve bias $\frac{1}{2} - \Omega(\delta^{1.65})$, while achieving bias better than $\frac{1}{2} - \delta$ is impossible.

Allowing qubits to be sent instead of classical bits changes the situation dramatically. Surprisingly, already in the two-party case coin flipping with bias $< 1/2$ is possible, as was first shown in [4]. The best known bias is $1/4$ and this is optimal for a special class of three-round protocols [10]; for a bias of ε at least $\Omega(\log\log(1/\varepsilon))$ rounds of communication are necessary

[10]. Kitaev (unpublished, see [79]) showed that in the two-party case no bias smaller than $1/\sqrt{2} - 1/2$ is possible.

A weak version of the coin-flipping problem is one in which we know in advance that outcome 0 benefits Alice and outcome 1 benefits Bob. In this case, we only need to bound the probabilities of a dishonest Alice convincing Bob that the outcome is 0 and a dishonest Bob convincing Alice that the outcome is 1. In the classical setting, a standard argument shows that even weak coin flipping with a bias $< 1/2$ is impossible when a majority of the players is dishonest. In the quantum setting, this scenario was first studied under the name *quantum gambling* [63]. Subsequently, Spekkens and Rudolph [111] gave a quantum protocol for weak coin flipping with bias $1/\sqrt{2} - 1/2$, i.e., no party can achieve the *desired outcome* with probability greater than $1/\sqrt{2}$. Notice that this is a better bias than in the best strong coin flipping protocol of [10].

We also remark that Kitaev's lower bound proof for strong coin flipping does not apply to weak coin flipping. Thus, weak protocols with arbitrarily small $\varepsilon > 0$ may be possible. The only known lower bounds for weak coin flipping are that the protocol of [111] is optimal for a restricted class of protocols [11] and that a protocol must use at least $\Omega(\log\log(1/\varepsilon))$ rounds of communication to achieve bias ε. This was shown in [10] for strong coin flipping but the proof also applies to weak coin flipping.

In this chapter, we focus on quantum coin flipping for more than two players. However, for our multiparty quantum protocols we will will first need a new two-party quantum protocol for *coin flipping with penalty for cheating.* In this problem, players can be heavily penalized for cheating, which will allow us to achieve lower cheating probability as a function of the penalty. This primitive and the quantum protocol for it are presented in Section 6.2; they may be of independent interest.

One way to classically model communication between more than two parties is by a primitive called *broadcast*. When a player sends a bit to the other players he broadcasts it to all the players at once [18]. However, when we deal with qubits such a broadcast channel is not possible since it requires to clone or copy the qubit to be broadcast and cloning a qubit is not possible [117]. In Section 6.3 we develop a proper quantum version of the broadcast primitive, which generalizes the classical broadcast. Somewhat surprisingly, we show that our quantum broadcast channel is essentially as powerful as a combination of pairwise quantum channels and a classical broadcast channel. This could also be of independent interest.

Using this broadcast primitive we obtain our main result:

6.1.1. THEOREM. *For k parties out of which g are honest, the optimal achievable bias is* $(\frac{1}{2} - \Theta(\frac{g}{k}))$.

We prove Theorem 6.1.1 by giving an efficient protocol with bias $(\frac{1}{2} - \Omega(\frac{g}{k}))$ in Section 6.4 and showing a lower bound of $(\frac{1}{2} - \mathrm{O}(\frac{g}{k}))$ in Section 6.5. Our protocol builds upon our two-party coin-flipping with penalties which we develop in Section 6.2, and the classical protocol of Feige [52] which allows to reduce the number of participants in the protocol without significantly changing the fraction of good players present. Our lower bound extends the lower bound of Kitaev [79].

6.2 Two-Party Coin Flipping with Penalty for Cheating

We consider the following model for coin flipping. We have two parties: Alice and Bob, among at least one is assumed to be honest. If no party is caught cheating, the winner gets 1 coin, the loser gets 0 coins. If honest Alice catches dishonest Bob, Bob loses v coins but Alice wins 0 coins. Similarly, if honest Bob catches dishonest Alice, she loses v coins but Bob wins 0 coins.

6.2.1. Theorem. *If Alice (Bob) is honest, the expected win by dishonest Bob (Alice) is at most $\frac{1}{2} + \frac{1}{\sqrt{v}}$, for $v \geq 4$.*

Proof. The protocol is as follows. Let $\delta = \frac{2}{\sqrt{v}}$. Define $|\psi_a\rangle = \sqrt{\delta}|a\rangle|a\rangle + \sqrt{1-\delta}|2\rangle|2\rangle$.

1. Alice picks $a \in \{0,1\}$ uniformly at random, generates the state $|\psi_a\rangle$ and sends the second register to Bob.
2. Bob stores this state in a quantum memory, picks $b \in \{0,1\}$ uniformly at random and sends b to Alice.
3. Alice then sends a and the first register to Bob and Bob verifies if the joint state of the two registers is $|\psi_a\rangle$ by measuring it in a basis consisting of $|\psi_a\rangle$ and everything orthogonal to it. If the test is passed, the result of coin flip is $a \oplus b$, otherwise Bob catches Alice cheating.

Theorem 6.2.1 follows from the following two claims.

6.2.2. Claim. *Bob cannot win with probability more than $\frac{1}{2} + \frac{1}{\sqrt{v}}$, thus his expected win is at most $\frac{1}{2} + \frac{1}{\sqrt{v}}$.*

Proof. Let ρ_a be the density matrix of the second register of $|\psi_a\rangle$. Then, for the trace distance between ρ_0 and ρ_1 we have $\|\rho_0 - \rho_1\|_t = 2\delta$.

Aharonov et al. [3] showed that the trace distance is a measure for the distinguishability of quantum states analogously to the total variation distance of probability distributions; in particular, the probability of Bob winning is at most $\frac{1}{2} + \frac{\|\rho_0 - \rho_1\|_t}{4} = \frac{1}{2} + \frac{\delta}{2} = \frac{1}{2} + \frac{1}{\sqrt{v}}$. □

6.2.3. Claim. *Dishonest Alice's expected win is at most $\frac{1}{2} + \frac{1}{\sqrt{v}}$.*

Proof. Without loss of generality, we can assume that Alice is trying to achieve $a \oplus b = 0$, which is equivalent to $a = b$. Since initially she has no information about the b that Bob is going to send, the state she sends in the first round is independent of b. So she prepares some pure quantum state $|\psi\rangle$, of which a part is sent to Bob. We can assume that this state is of the form

$$|\psi\rangle = \alpha_0|0\rangle|0\rangle + \alpha_1|1\rangle|1\rangle + \alpha_2|2\rangle|2\rangle$$

for some α_0, α_1, $\alpha_2 \geq 0$, because all that matters is the purification of the density matrix that Bob receives. Moreover, by symmetry we can assume that the amplitudes α_1 and α_2 have the same magnitude so

$$|\psi\rangle = \sqrt{\varepsilon}|0\rangle|0\rangle + \sqrt{\varepsilon}|1\rangle|1\rangle + \sqrt{1-2\varepsilon}|2\rangle|2\rangle$$

for some $\varepsilon \geq 0$. Since the state is symmetric with respect to switching $|0\rangle$ and $|1\rangle$, the maximum expected win that Alice can achieve is the same is she receives $b = 0$ from Bob and if she receives $b = 1$.

It suffices to consider the case when she receives $b = 0$. After receiving $b = 0$, Alice performs a measurement on her register. By $|\psi'_i\rangle$ we denote the projection of $|\psi\rangle$ to the subspace in which Alice answers $a = i$. Hence, $|\psi\rangle = |\psi'_0\rangle + |\psi'_1\rangle$. By symmetry, we can assume that

$$|\psi'_0\rangle = \sqrt{\varepsilon_0}|0\rangle|0\rangle + \sqrt{\varepsilon_1}|1\rangle|1\rangle + \sqrt{x-\varepsilon}|2\rangle|2\rangle \ ,$$

$$|\psi'_1\rangle = \sqrt{\varepsilon-\varepsilon_0}|0\rangle|0\rangle + \sqrt{\varepsilon-\varepsilon_1}|1\rangle|1\rangle + \sqrt{1-x-\varepsilon}|2\rangle|2\rangle$$

for some $\varepsilon_0, \varepsilon_1, x \geq 0$. The best strategy for Alice is just to send the first register to Bob unchanged. The probability with which Alice succeeds is $|\langle\psi'_0|\psi_0\rangle|^2$ for $a = 0$ and $|\langle\psi'_1|\psi_1\rangle|^2$ for $a = 1$. If $\varepsilon_1 > 0$, then changing ε_1 to 0 does not change $|\langle\psi'_0|\psi_0\rangle|^2$ and increases $|\langle\psi'_1|\psi_1\rangle|^2$. Similarly, changing ε_0 to ε does not change $|\langle\psi'_1|\psi_1\rangle|^2$ and increases $|\langle\psi'_0|\psi_0\rangle|^2$. Therefore, we can assume that $\varepsilon_0 = \varepsilon$, $\varepsilon_1 = 0$ and the states are

$$|\psi'_0\rangle = \sqrt{\varepsilon}|0\rangle|0\rangle + \sqrt{x-\varepsilon}|2\rangle|2\rangle,$$

$$|\psi'_1\rangle = \sqrt{\varepsilon}|1\rangle|1\rangle + \sqrt{1-x-\varepsilon}|2\rangle|2\rangle.$$

Let $|\psi_i''\rangle = \sqrt{1-\delta}|i\rangle|i\rangle - \sqrt{\delta}|2\rangle|2\rangle$ for $i \in \{0,1\}$. Then $|\psi_i''\rangle$ is orthogonal to $|\psi_i\rangle$, so we can assume that Bob's verification measurement has $|\psi_i''\rangle$ as one of the outcomes that indicate that Alice is cheating. Therefore, the probability of Alice caught cheating is at least $|\langle\psi_0''|\psi_0'\rangle|^2 + |\langle\psi_1''|\psi_1'\rangle|^2$.

Let $d = \max\{x, 1-x\} - \frac{1}{2}$. Then the probability of Alice claiming $a = 0$ (and hence forcing outcome $a \oplus b = 0$ as desired) is $\langle\psi_0'|\psi_0'\rangle = x \leq \frac{1}{2} + d$. However, she may be caught cheating. We claim

6.2.4. Claim. *The probability of Alice being caught by Bob is at least $\frac{d^2\delta}{2}$.*

Proof. Consider the two inner products

$$\langle\psi_0''|\psi_0'\rangle = \sqrt{\varepsilon}\sqrt{1-\delta} - \sqrt{x-\varepsilon}\sqrt{\delta},$$

$$\langle\psi_1''|\psi_1'\rangle = \sqrt{\varepsilon}\sqrt{1-\delta} - \sqrt{1-x-\varepsilon}\sqrt{\delta}$$

To compare their difference, note that

$$\sqrt{x-\varepsilon} - \sqrt{1-x-\varepsilon} \geq \sqrt{x} - \sqrt{1-x} = \frac{x-(1-x)}{\sqrt{x}+\sqrt{1-x}} \geq \frac{x-(1-x)}{\sqrt{2}}$$

where the first inequality follows from convexity of square root function and the second inequality follows from Cauchy-Schwartz. Therefore, $\langle\psi_0''|\psi_0'\rangle$ and $\langle\psi_1''|\psi_1'\rangle$ differ in absolute value by at least $\frac{|x-(1-x)|\sqrt{\delta}}{\sqrt{2}} = d\sqrt{2\delta}$. This implies that one of $|\langle\psi_0''|\psi_0'\rangle|^2$ and $|\langle\psi_1''|\psi_1'\rangle|^2$ is at least $\frac{d^2\delta}{2}$ and Alice gets caught with probability at least $\frac{d^2\delta}{2}$. □

Therefore, Alice's expected win is at most

$$\frac{1}{2} + d - \frac{d^2\delta v}{2} = \frac{1}{2} + d\left(1 - \frac{d\delta v}{2}\right).$$

Consider two cases. If $d\delta v \geq 2$, then $1 - \frac{d\delta v}{2} \leq 0$ and the expected win is at most $\frac{1}{2}$. If $d\delta v \leq 2$, then $d \leq \frac{2}{\delta v} = \frac{1}{\sqrt{v}}$ and Alice's expected win is at most $\frac{1}{2} + \frac{1}{\sqrt{v}}$. □

□

6.3 The Multiparty Model

6.3.1 Adversaries

We assume computationally unbounded adversaries. However, they have to obey quantum mechanics and cannot read the private memory of the honest players, but they can communicate secretly with each other. Moreover, we assume that they can only access the message space in between rounds or when according to the protocol it is their turn to send a message.

6.3.2 The broadcast channel

A classical broadcast channel allows one party to send a classical bit to all the other players. In the quantum setting this would mean that a qubit would be sent to all the other players. However, when there are more than two players in total we would have to *clone* or *copy* the qubit in order to send it to the other players. Even if the sender knows a classical preparation of the state he wants to send, we cannot allow him to prepare copies because he may be a cheater and send different states to different parties. It is well known that it is impossible to clone a qubit [117], because cloning is not a unitary operation. This means that we will have to take a slightly different approach. Quantum broadcast channels have been studied in an information-theoretic context before [14, 116] but not in the presence of faulty or malicious parties.

Our quantum broadcast channel works as follows. Suppose there are k players in total and that one player wants to broadcast a qubit that is in the state $\alpha|0\rangle + \beta|1\rangle$. What will happen is that the channel will create the k-qubit state $\alpha|0^k\rangle + \beta|1^k\rangle$ and send one of the k qubits to each of the other players. The state $\alpha|0^k\rangle + \beta|1^k\rangle$ can be easily created from $\alpha|0\rangle + \beta|1\rangle$ by taking $k-1$ fresh qubits in the state $|0^{k-1}\rangle$. This joint state can be written as $\alpha|0^k\rangle + \beta|10^{k-1}\rangle$. Next we flip the last $k-1$ bits conditional on the first bit being a 1, thus obtaining the desired state $\alpha|0^k\rangle + \beta|1^k\rangle$. This last operation can be implemented with a series of controlled-not operations. Note that this state is not producing k copies of the original state, which would be the k-fold product state $(\alpha|0\rangle + \beta|1\rangle) \otimes \ldots \otimes (\alpha|0\rangle + \beta|1\rangle)$.

6.3.1. Theorem. *In the following sense, a quantum broadcast channel between k parties is comparable to models where the parties have a classical broadcast channel and/or pairwise quantum channels:*

- *If all parties are honest:*

  1. *One use of the quantum broadcast channel can be simulated with* $2(k-1)$ *uses of pairwise quantum channels.*
  2. *One use of a classical broadcast channel can be simulated with one use of the quantum broadcast channel.*
  3. *One use of a pairwise quantum channel can be simulated by* $k+1$ *uses of the quantum broadcast channel.*

- *If all but one of the parties are dishonest, using one of the simulations above in place of the original communication primitive does not confer extra cheating power.*

Proof. We first give the simulations and argue that they work in case all players are honest.

1. The sender takes $k-1$ fresh qubits in state $|0^k\rangle$. He applies $k-1$ times CNOT where the subsystem to be broadcast is the control of the CNOT and the fresh qubits are the destination. He then sends each of the $k-1$ qubits via the pairwise quantum channels to the $k-1$ other parties. Each recipient j flips a private classical random bit r_j and if $r_j = 1$ performs a σ_z phase flip on the received qubit. Here $\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ is the Pauli matrix that multiplies the relative phase between the $|0\rangle$ and the $|1\rangle$ state by -1. He then sends r_j back to the sender. The sender computes the parity of the r_j and if it is odd, he performs a σ_z phase flip on his part of the broadcast state, thus restoring the correct relative phase. This randomization is a countermeasure; its utility is explained below.

2. When the sender wants to broadcast bit $b \in \{0,1\}$, he uses the quantum broadcast channel on qubit $|b\rangle$. The recipients immediately measure their qubit in the computational basis to obtain the classical bit.

3. The quantum broadcast channel can be used to create an EPR pair $(|00\rangle + |11\rangle)/\sqrt{2}$ between two players P_i and P_j with the assistance of the other $(k-2)$ players. i and j are determined by the protocol.

   First one player broadcasts the state $(|0\rangle + |1\rangle)/\sqrt{2}$, resulting in the k qubit state $|\varphi\rangle = (|0^k\rangle + |1^k\rangle)/\sqrt{2}$. Now one after the other, the $k-2$ remaining players perform a Hadamard transformation on their qubit, measure it in the computational basis, and broadcast the classical result. Next, if P_i receives a 1 he applies a phase flip σ_z to his part of $|\varphi\rangle$ (P_j does nothing). After this operation, $|\varphi\rangle$ will be an EPR state between P_i and P_j unentangled with the other $k-2$ parties. Using a shared EPR pair, a protocol called *teleportation* [19] can be used to simulate a private quantum channel between P_i and P_j. Teleportation requires the transmission of two bits of classical information.

For the case of all but one party being dishonest:

1. If the sender is honest, the recipients obtain exactly the same subsystems as for the quantum broadcast channel.

If one of the recipients is honest, he may receive an arbitrary quantum subsystem up to the randomized relative phase. However, exactly the same can be achieved with a quantum broadcast channel with $k-1$ cheating parties, who each perform a Hadamard transformation on their subsystem followed by a measurement in the computational basis.

2. If the sender is honest, all recipients obtain the same computational-basis state.

   If one of the recipients is honest, he obtains a classical bit that is possibly randomized in case the dishonest sender does not broadcast a basis state. Since the sender can flip a coin himself, this does not give more cheating power.

3. If the sender is honest, we can assume without loss of generality that all cheating action is done after the EPR pair has been established, because the merged cheaters can easily recreate the original broadcast state and also compensate phase flips of the honest sender. However, after the EPR pair has been established, the sender unilaterally performs his part of the teleportation circuit and measurements and sends the two bits of classical information. So the most general cheating action is to apply a quantum operation after the reception of the two classical bits. Furthermore, we can even assume that the cheating action is done *after* the correction circuit of teleportation. This is similar to the teleportation of quantum gates [67], and, hence, amounts to cheating on a pairwise quantum channel.

   If one of the recipients is honest, the best the cheaters can aim for is to give an arbitrary quantum state to the honest recipient. This they can also achieve over a pairwise quantum channel.

□

6.4 Multiparty Quantum Coin-Flipping Protocols

We will first consider the case of only one good player, i.e., $g = 1$, and later extend our results to general g.

One honest player Recall, we need to construct a protocol with bias $1/2-\Omega(1/k)$. Before proceeding to our actual protocol, let us consider a simple protocol which trivially extends the previous work in the two-party setting, but does not give us the desired result. The protocols is as follows: player 1 flips a random coin with player 2, player 3 flips a random coin with player 4 and so forth. In each pair, the player with the higher id wins if the coin is 1 and the one with the lower id if the coin is 0. The winners repeat the procedure. With each repetition of the tournament, half of the remaining players are eliminated. If there is an odd number of players at any moment, the one with the highest id advances to the next round. When there are only two players left, the coin they flip becomes the output of the protocol. Above we assume we have private point-to-point quantum channels and a classical broadcast channel, which is justified by Theorem 6.3.1.

Now, the elimination rounds can be implemented using the weak two-party coin-tossing protocol by Spekkens and Rudolph [111] and the last round by the the strong two-party coin-tossing protocol by Ambainis [10]. If there is only one good player, the probability that he makes it to the last round is $(1-1/\sqrt{2})^{\lceil -1+\log k\rceil}$; in this case, the probability that the bad players can determine the output coin is 3/4. In case the good player gets eliminated, the bad players can completely determine the coin. Hence, the overall probability that the bad players can determine the coin is $1-\frac{1}{4}(1-\frac{1}{\sqrt{2}})^{\lceil -1+\log k\rceil} \leq 1-\frac{1}{4k^{1.78}}$, which corresponds to bias $\frac{1}{2}-\Omega(1/k^{1.78})$.

To improve the above naive bound to the desired value $\frac{1}{2}-\Omega(1/k)$, we will use our coin-flipping protocol with penalty from Section 6.2. The idea is that in current quantum coin-flipping protocols for two parties, there are three outcomes for a given player: "win," "lose," and "abort." Now, looking at the elimination tournament above, if an honest player loses a given coin flipping round, he does not "complain" and bad player win the game. However, if the honest player detect cheating, he can and will abort the entire process, which corresponds to the failure of the dishonest players to fix the coin. Of course, if the are few elimination rounds left, bad players might be willing to risk the abort if they gain significant benefits in winning the round. However, if the round number is low, abort becomes prohibitively expensive: a dishonest player might not be willing to risk it given there are plenty more opportunities for the honest player to "normally lose". Thus, instead of regular two-party coin-tossing protocols, which do not differentiate between losing and abortion, we can employ our protocol for coin flipping with penalty, where the penalties are very high at the original rounds, and eventually get lower towards the end of the protocol. Specific penalties are chosen in a way which optimizes the final bias we get, and allows us to achieve the desired bias $1/2-\Omega(1/k)$.

6.4.1. THEOREM. *There is a strong quantum coin-tossing protocol for k parties with bias at most $1/2-c/k$ for some constant c, even with $(k-1)$ bad parties.*

Proof. We assume that $k=2^n$ for some $n>0$, as it changes c by at most a constant factor. Let Q_v be the maximum expected win in a two-party protocol with penalty v. Consider the following protocol with n rounds.

In the i^{th} round, we have 2^{n+1-i} parties remaining. We divide them into pairs. Each pair performs the two-party coin-flipping protocol with penalty $(2^{n-i}-1)$, with Alice winning if the outcome is 1 and Bob winning if the outcome is 0. The winners proceed to $(i+1)^{\text{st}}$ round.

In the $(n-2)^{\text{nd}}$ round, there are just 8 parties remaining. At this stage, they can perform three rounds of regular coin flipping with no penalty of [10, 77] in which no cheater can bias the coin to probability more than 3/4, which will result in maximum probability of 63/64 of fixing the outcome. The result of this last two-round protocol is the result of our 2^n-party protocol.

Assume that the honest player has won the first $(n-j)$ coin flips and advanced to $(j+1)^{\text{st}}$ round. Assume that the all other players in the $(j+1)^{\text{st}}$ round are dishonest. Let P_j be the maximum probability with which (2^j-1) dishonest players can fix the outcome to 0 (or 1).

6.4.2. Claim.

$$1-P_j \geq (1-P_{j-1})(1-Q_{2^{j-1}-1}) \tag{6.1}$$

Proof. Let p_w, p_l, p_c be the probabilities of the honest player winning, losing and catching the other party cheating in the $(j+1)^{\text{st}}$ round of the protocol. Notice that $p_w+p_l+p_c=1$. Then, the probability P_j of 2^j-1 dishonest parties fixing the coin is at most $p_l+p_wP_{j-1}$. If the honest player loses, they win immediately. If he wins, they can still bias the coin in $j-1$ remaining rounds to probability at most P_{j-1}. If he catches his opponent cheating, he exits the protocol and the dishonest players have no more chances to cheat him. Using $p_w=1-p_l-p_c$, we have

$$\begin{aligned} P_j \leq p_l+p_wP_{j-1} &= P_{j-1}+(1-P_{j-1})p_l-P_{j-1}p_c \\ &= P_{j-1}+(1-P_{j-1})\left(p_l-\frac{P_{j-1}}{1-P_{j-1}}p_c\right) \end{aligned} \tag{6.2}$$

Next, notice that $P_{j-1}\geq 1-\frac{1}{2^{j-1}}$. This is because $2^{j-1}-1$ bad players could just play honestly when they face the good player and fix the coin flip if two bad players meet in the last round. Then, the probability of the good player winning all $j-1$ rounds is $\frac{1}{2^{j-1}}$. Therefore, $\frac{P_{j-1}}{1-P_{j-1}}\geq 2^{j-1}-1$ and (6.2) becomes

$$P_j \leq P_{j-1}+(1-P_{j-1})(p_l-(2^{j-1}-1)p_c) \tag{6.3}$$

Finally, the term in brackets is at most $Q_{2^{j-1}-1}$, which gives

$$P_j \leq P_{j-1}+(1-P_{j-1})Q_{2^{j-1}-1} \tag{6.4}$$

which in turn is equivalent to the desired (6.1). □

By applying the claim inductively, we get

$$1 - P_n \geq \frac{1}{64} \prod_{j=4}^{n} (1 - Q_{2^{j-1}-1})$$

where the $\frac{1}{64}$ term comes from the naive protocol we use in the last three rounds. Now, using the bound in Theorem 6.2.1 we have

$$\begin{aligned} 1 - P_n &\geq \frac{1}{64} \prod_{j=3}^{n-1} (1 - Q_{2^j-1}) \geq \frac{1}{64} \prod_{j=3}^{n-1} \left(\frac{1}{2} - \frac{1}{\sqrt{2^j - 1}} \right) \\ &\geq \frac{1}{8 \cdot 2^n} \prod_{j=3}^{n-1} \left(1 - \frac{2}{\sqrt{2^j - 1}} \right) \ . \end{aligned}$$

The last term in the brackets is at least $\prod_{j=3}^{\infty}(1 - \frac{2}{\sqrt{2^j-1}})$ which is a positive constant. Therefore, for some constant $c > 0$ we have $1 - P_n \geq \frac{c}{2^n} = \frac{c}{k}$, which means that the bias is at most $\frac{1}{2} - \Omega(\frac{1}{k})$. □

Extending to many honest players We can extend Theorem 6.4.1 to every number $g \geq 1$ of good players by using the classical lightest-bin protocol of Feige [52]. This protocol allows us to reduce the total number of players until a single good player is left without significantly changing the fraction of good players, after which we can run the quantum protocol of Theorem 6.4.1 to get the desired result. Specifically, Lemma 8 from [52] implies that starting from $g = \delta k$ good players out of k players, the players can classically select a sub-committee of $\mathrm{O}(1/\delta) = \mathrm{O}(k/g)$ players containing at least one good player with probability at least 1/2. Now, this sub-committee can use the quantum protocol of Theorem 6.4.1 to flip a coin with bias $1/2 - \Omega(g/k)$, provided it indeed contains at least one honest player. But since the latter happens with probability at least 1/2, the final bias is at most $1/2 - (1/2) \cdot \Omega(g/k) = 1/2 - \Omega(g/k)$, as desired.

6.5 Lower Bound

6.5.1 The two-party bound

For completeness and to facilitate the presentation of our generalization, we reproduce here Kitaev's unpublished proof [79] that every two-party strong quantum coin-flipping protocol must have bias at least $1/\sqrt{2}$. The model here is that the two parties communicate over a quantum channel.

6.5.1. Definition. Let $\mathcal{H} := \mathcal{A} \otimes \mathcal{M} \otimes \mathcal{B}$ denote the Hilbert space of the coin-flipping protocol composed of Alice's private space, the message space, and Bob's private space. A $2N$-round two-party coin-flipping protocol is a tuple

$$(U_{A,1}, \ldots, U_{A,N}, U_{B,1}, \ldots, U_{B,N}, \Pi_{A,0}, \Pi_{A,1}, \Pi_{B,0}, \Pi_{B,1})$$

where

- $U_{A,j}$ is a unitary operator on $\mathcal{A} \otimes \mathcal{M}$ for $j = 1, \ldots, N$,
- $U_{B,j}$ is a unitary operator on $\mathcal{M} \otimes \mathcal{B}$ for $j = 1, \ldots, N$,
- $\Pi_{A,0}$ and $\Pi_{A,1}$ are projections from $\mathcal{A}$ onto orthogonal subspaces of $\mathcal{A}$, representing Alice's final measurements for outcome 0 and 1, respectively,
- $\Pi_{B,0}$ and $\Pi_{B,1}$ are projections from $\mathcal{B}$ onto orthogonal subspaces of $\mathcal{B}$, representing Bob's final measurements for outcome 0 and 1, respectively,

so that for

$$|\psi_N\rangle := (1_{\mathcal{A}} \otimes U_{B,N})(U_{A,N} \otimes 1_{\mathcal{B}})(1_{\mathcal{A}} \otimes U_{B,N-1})(U_{A,N-1} \otimes 1_{\mathcal{B}}) \cdots \\ \cdots (1_{\mathcal{A}} \otimes U_{B,1})(U_{A,1} \otimes 1_{\mathcal{B}})|0\rangle$$

holds

$$(\Pi_{A,0} \otimes 1_{\mathcal{M}} \otimes 1_{\mathcal{B}})|\psi_N\rangle = (1_{\mathcal{A}} \otimes 1_{\mathcal{M}} \otimes \Pi_{B,0})|\psi_N\rangle \tag{6.5}$$
$$(\Pi_{A,1} \otimes 1_{\mathcal{M}} \otimes 1_{\mathcal{B}})|\psi_N\rangle = (1_{\mathcal{A}} \otimes 1_{\mathcal{M}} \otimes \Pi_{B,1})|\psi_N\rangle \tag{6.6}$$
$$\|(\Pi_{A,0} \otimes 1_{\mathcal{M}} \otimes 1_{\mathcal{B}})|\psi_N\rangle\| = \|(\Pi_{A,1} \otimes 1_{\mathcal{M}} \otimes 1_{\mathcal{B}})|\psi_N\rangle\| \tag{6.7}$$

The first two conditions ensure that when Alice and Bob are honest, they both get the same value for the coin and the third condition guarantees that when Alice and Bob are honest, their coin is not biased. A player aborts if her or his final measurement does not produce outcome 0 or 1; of course, it is no restriction to delay this action to the end of the protocol.

6.5.2. Lemma. *Fix an arbitrary two-party quantum coin-flipping protocol. Let p_{1*} and p_{*1} denote the probability that Alice or Bob, respectively, can force the outcome of the protocol to be* 1 *if the other party follows the protocol. Denote by p_1 the probability for outcome* 1 *when there are no cheaters. Then $p_{1*}p_{*1} \geq p_1$.*

Hence, if $p_1 = 1/2$, then $\max\{p_{1*}, p_{*1}\} \geq 1/\sqrt{2}$. To prove Lemma 6.5.2, we construct the view of a run of the protocol from an honest Alice's point of view, with Bob wanting to bias the protocol towards 1. The problem of optimizing Bob's strategy is a semidefinite program (SDP).

Semidefinite programming is a generalization of linear programming. In addition to the usual linear constraints, it is allowed to require that a square matrix of variables is positive semidefinite, i.e., all its eigenvalues are nonnegative. The proof below makes use of the well-developed duality theory for SDPs. Let A, B, and C denote square matrices of the same dimension. If A is positive semidefinite, we write $A \geq 0$. We define $A \geq B :\Leftrightarrow A - B \geq 0$. The following properties are straightforward to verify:

$$\begin{gathered} A \geq B \Leftrightarrow \forall |\psi\rangle : \langle\psi|A|\psi\rangle \geq \langle\psi|B|\psi\rangle \\ A \geq B \Rightarrow \operatorname{tr}_{\mathcal{V}} A \geq \operatorname{tr}_{\mathcal{V}} B \text{ for every subspace } \mathcal{V} \\ A = B + C \text{ and } C \geq 0 \Rightarrow A \geq B \end{gathered}$$

6.5.3. Lemma. *The optimal strategy of Bob trying to force outcome* 1 *is the solution to the following SDP over the semidefinite matrices $\rho_{A,0}, \ldots, \rho_{A,N}$ operating on $\mathcal{A} \otimes \mathcal{M}$:*

$$\begin{aligned} &\text{maximize} \qquad \operatorname{tr}\left((\Pi_{A,1} \otimes 1_{\mathcal{M}})\rho_{A,N}\right) \text{ subject to} && (6.8) \\ &\operatorname{tr}_{\mathcal{M}} \rho_{A,0} = |0\rangle\langle 0|_{\mathcal{A}} && (6.9) \\ &\operatorname{tr}_{\mathcal{M}} \rho_{A,j} = \operatorname{tr}_{\mathcal{M}} U_{A,j}\rho_{A,j-1}U^*_{A,j} \quad (1 \leq j \leq N) && (6.10) \end{aligned}$$

Proof. Alice starts with her private memory in state $|0\rangle_{\mathcal{A}}$ and we permit Bob to determine the $\mathcal{M}$ part of the initial state. Therefore all Alice knows is that initially, the space accessible to her is in state $\rho_{A,0}$ with $\operatorname{tr}_{\mathcal{M}} \rho_{A,0} = |0\rangle\langle 0|_{\mathcal{A}}$. Alice sends the first message, transforming the state to $\rho'_{A,0} := U_{A,1}\rho_{A,0}U^*_{A,1}$. Now Bob can do an arbitrary unitary operation on $\mathcal{M} \otimes \mathcal{B}$ leading to $\rho_{A,1}$, so the only constraint is $\operatorname{tr}_{\mathcal{M}} \rho_{A,1} = \operatorname{tr}_{\mathcal{M}} \rho'_{A,0}$. In the next round, honest Alice applies $U_{A,2}$, then Bob can do some operation that preserves the partial trace, and so forth. The probability for Alice outputting 1 is $\operatorname{tr}((\Pi_{A,1} \otimes 1_{\mathcal{M}})\rho_{A,N})$ because the final state for Alice is $\rho_{A,N}$ and she performs an orthogonal measurement on $\mathcal{A}$ with projections $\Pi_{A,0}$, $\Pi_{A,1}$, and $1_{\mathcal{A}} - \Pi_{A,0} - \Pi_{A,1}$, which represents "abort." □

6.5.4. Lemma. *The dual SDP to the primal SDP in Lemma 6.5.3 is*

$$\begin{aligned} &\text{minimize} \qquad \langle 0|Z_{A,0}|0\rangle \text{ subject to} && (6.11) \\ &Z_{A,j} \otimes 1_{\mathcal{M}} \geq U^*_{A,j+1}(Z_{A,j+1} \otimes 1_{\mathcal{M}})U_{A,j+1} \quad (0 \leq j \leq N-1) && (6.12) \\ &Z_{A,N} = \Pi_{A,1} && (6.13) \end{aligned}$$

over the Hermitian matrices $Z_{A,0}, \ldots Z_{A,N}$ operating on $\mathcal{A}$.

Proof. In the Lagrange-multiplier approach, a “primal” optimization problem

$$\max_{x\geq 0} f(x) \text{ subject to } g(x) \leq a \text{ with } a > 0$$

reformulated as

$$\max_{x} \inf_{\lambda\geq 0} f(x) - \lambda \cdot (g(x) - a) \ ,$$

which is bounded from above by $\min_{\lambda\geq 0} \lambda \cdot a$ subject to $(f - \lambda \cdot g)(x) \leq 0$ for all $x \geq 0$. In linear programming, $(f - \lambda \cdot g)(x) \leq 0$ for all $x \geq 0$ if and only if $f - \lambda \cdot g \leq 0$, therefore the preceding optimization problem can be simplified to $\min_{\lambda\geq 0} \lambda \cdot a$ subject to $f - \lambda \cdot g \leq 0$. The same construction can be applied to SDPs; we form the dual of the SDP in Lemma 6.5.3 as follows: the dual is equivalent to maximizing over the $\rho_{A,j}$ the minimum of

$$\begin{aligned} &\operatorname{tr}((\Pi_{A,1} \otimes 1_{\mathcal{M}})\rho_{A,N}) - \operatorname{tr}(Z_{A,0}(\operatorname{tr}_{\mathcal{M}} \rho_{A,0} - |0\rangle\langle 0|_{\mathcal{A}})) \\ &\qquad - \sum_{j=1}^{N} \operatorname{tr}(Z_{A,j} \operatorname{tr}_{\mathcal{M}}(\rho_{A,j} - U_{A,j}\rho_{A,j-1}U_{A,j}^*)) - \sum_{j=0}^{N} \operatorname{tr}(Y_j \rho_{A,j}) \end{aligned} \tag{6.14}$$

subject to the operators $Z_{A,j}$ on $\mathcal{M}$ being Hermitian and the operators Y_j on $\mathcal{A} \otimes \mathcal{M}$ being positive semidefinite, for $0 \leq j \leq N$. In the above sum, the terms containing $\rho_{A,j}$ for $0 \leq j < N$ are

$$\begin{aligned} &- \operatorname{tr}(Z_{A,j}(\operatorname{tr}_{\mathcal{M}} \rho_{A,j})) + \operatorname{tr}(Z_{A,j+1} \operatorname{tr}_{\mathcal{M}}(U_{A,j+1}\rho_{A,j}U_{A,j+1}^*)) - \operatorname{tr}(Y_j \rho_{A,j}) = \\ &\qquad \operatorname{tr}\left(\left(-(Z_{A,j} \otimes 1_{\mathcal{M}}) + U_{A,j+1}^*(Z_{A,j+1} \otimes 1_{\mathcal{M}})U_{A,j+1} - Y_j\right) \rho_{A,j}\right) \end{aligned} \tag{6.15}$$

Since the primal constraints (6.9) and (6.10) are equality constraints, the dual constraint (6.15) must be equal to 0. However, since Y_j is positive semidefinite and does not appear anywhere else, we can drop it from (6.15) to arrive at the inequality (6.12).

For $j = N$, we obtain the dual equality constraint (6.13) and the dual objective function becomes the only summand of (6.14) that does not involve any $\rho_{A,j}$. □

Proof of Lemma 6.5.2. Let $Z_{A,j}$ and $Z_{B,j}$ $(0 \leq j \leq N)$ denote the optimal solutions for the dual SDPs for a cheating Bob and a cheating Alice, respectively. For each j, $0 \leq j \leq N$, let $|\psi_j\rangle := (1_{\mathcal{A}} \otimes U_{B,j})(U_{A,j} \otimes 1_{\mathcal{B}}) \cdots (1_{\mathcal{A}} \otimes U_{B,1})(U_{A,1} \otimes 1_{\mathcal{B}})|0\rangle$ denote the state of the protocol in round j when both parties are honest. Let $F_j := \langle\psi_j|(Z_{A,j} \otimes 1_{\mathcal{M}} \otimes Z_{B,j})|\psi_j\rangle$. We claim

$$p_{1*}p_{*1} = F_0 \tag{6.16}$$

$$F_j \geq F_{j+1} \qquad (0 \leq j < N) \tag{6.17}$$

$$F_N = p_1. \tag{6.18}$$

Combining (6.16)–(6.18), we obtain the desired $p_{1*}p_{*1} \geq p_1$. We now proceed to prove these claims.

Note that the primal SDP from Lemma 6.5.3 is strictly feasible: Bob playing honestly yields a feasible solution that is strictly positive. The strong-duality theorem of semidefinite programming states that in this case, the optimal value of the primal and the dual SDPs are the same, and therefore $p_{1*} = \langle 0|_{\mathcal{A}} Z_{A,0} |0\rangle_{\mathcal{A}}$ and $p_{*1} = \langle 0|_{\mathcal{B}} Z_{B,0}|0\rangle_{\mathcal{B}}$ and

$$\begin{aligned} p_{1*}p_{*1} &= \langle 0|_{\mathcal{A}} Z_{A,0}|0\rangle_{\mathcal{A}} \cdot \langle 0|_{\mathcal{M}} 1_{\mathcal{M}} |0\rangle_{\mathcal{M}} \cdot \langle 0|_{\mathcal{B}} Z_{B,0}|0\rangle_{\mathcal{B}} \\ &= \langle 0|(Z_{A,0} \otimes 1_{\mathcal{M}} \otimes Z_{B,0})|0\rangle = F_0. \end{aligned}$$

The inequalities (6.17) hold because of the constraints (6.12). Equality (6.18) holds because by constraint (6.13) we have

$$\langle \varphi|(Z_{A,N} \otimes 1_{\mathcal{M}} \otimes Z_{B,N})|\varphi\rangle = \|(\Pi_{A,1} \otimes 1_{\mathcal{M}} \otimes 1_{\mathcal{B}})(1_{\mathcal{A}} \otimes 1_{\mathcal{M}} \otimes \Pi_{B,1})|\varphi\rangle\|^2$$

for every $|\varphi\rangle$; $|\psi_N\rangle$ is the final state of the protocol when both players are honest, so by equation (6.6),

$$\|(\Pi_{A,1} \otimes 1_{\mathcal{M}} \otimes 1_{\mathcal{B}})(1_{\mathcal{A}} \otimes 1_{\mathcal{M}} \otimes \Pi_{B,1})|\psi_N\rangle\|^2 = \|(\Pi_{A,1} \otimes 1_{\mathcal{M}} \otimes 1_{\mathcal{B}})|\psi_N\rangle\|^2 = p_1.$$

□

6.5.2 More than two parties

We will now extend Kitaev's lower bound to k parties. As with the upper bounds, we first start with a single honest player ($g = 1$), and then extend the result further to every g.

6.5.5. Theorem. *Every strong quantum coin-tossing protocol for k parties has bias at least $1/2 - (\ln 2)/k - \mathrm{O}(1/k^2)$ if it has to deal with up to $(k-1)$ bad parties.*

We consider the model of private pairwise quantum channels between the parties; by Theorem 6.3.1 the results immediately carry over to the quantum broadcast channel. Before proving Theorem 6.5.5, we make the following detour.

6.5.6. Definition. Let $\mathcal{H} := \mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_k \otimes \mathcal{M}$ denote the Hilbert space composed of the private spaces of k parties and the message space. An N-round k-party coin-flipping protocol is a tuple

$$(i_1, \ldots, i_N, U_1, \ldots, U_N, \Pi_{1,0}, \Pi_{1,1}, \ldots, \Pi_{k,0}, \Pi_{k,1})$$

where

- i_j with $1 \le i_j \le k$, $1 \le j \le N$, indicates whose turn it is to access the message space in round j,
- U_j is a unitary operator on $\mathcal{A}_{i_j} \otimes \mathcal{M}$ for $j = 1, \ldots, N$,
- for $1 \le i \le k$, $\Pi_{i,0}$ and $\Pi_{i,1}$ are projections from $\mathcal{A}_i$ to orthogonal subspaces of $\mathcal{A}_i$, representing the measurement that party i performs to determine outcome 0 or 1, respectively,

so that for $|\psi_N\rangle := \tilde{U}_{i_N} \cdots \tilde{U}_{i_1}|0\rangle$ and each pair $1 \le i < i' \le k$ and every $b \in \{0,1\}$ holds

$$\tilde{\Pi}_{i,b}|\psi_N\rangle = \tilde{\Pi}_{i',b}|\psi_N\rangle \tag{6.19}$$

$$\|\tilde{\Pi}_{i,b}|\psi_N\rangle\| = \|\tilde{\Pi}_{i,1-b}|\psi_N\rangle\|. \tag{6.20}$$

Here $\tilde{U}_j$ denotes the extension of U_j to all of $\mathcal{H}$ that acts as identity on the tensor factors $\mathcal{A}_{i'}$ for $i' \neq i_j$; $\tilde{\Pi}_{i,b} := (1_{\mathcal{A}_1} \otimes \cdots \otimes 1_{\mathcal{A}_{i-1}} \otimes \Pi_{i,b} \otimes 1_{\mathcal{A}_{i+1}} \otimes \cdots \otimes 1_{\mathcal{A}_k})$ is the extension of $\Pi_{i,b}$ to $\mathcal{H}$.

6.5.7. Lemma. *Fix an arbitrary quantum coin flipping protocol. For $b \in \{0,1\}$, let p_b be the probability of outcome b in case all players are honest. Let $p_{i,b}$ denote the probability that party i can be convinced by the other parties that the outcome of the protocol is $b \in \{0,1\}$. Then*

$$p_{1,b} \cdot \ldots \cdot p_{k,b} \ge p_b$$

Proof of Lemma 6.5.7. The optimal strategy for $k-1$ bad players trying to force outcome 1 is the solution to the SDP from Lemma 6.5.3 where all the cheating players are merged into a single cheating player.

Let $(Z_{i,j})_{0 \le j \le N}$ denote the optimal solution for the dual SDP for good player i, $1 \le i \le k$. For each j, $0 \le j \le N$, let $|\psi_j\rangle := \tilde{U}_j \cdots \tilde{U}_1|0\rangle$ denote the state of the protocol in round j when all parties are honest. Let $F_j := \langle\psi_j|(Z_{1,j} \otimes \cdots \otimes Z_{k,j} \otimes 1_{\mathcal{M}})|\psi_j\rangle$. By a similar argument as in the proof of Lemma 6.5.2, we have

$$p_{1,1} \cdot \ldots \cdot p_{k,1} = F_0 \tag{6.21}$$

$$F_j \ge F_{j+1} \qquad (0 \le j < N) \tag{6.22}$$

$$F_N = p_1 \tag{6.23}$$

Hence, $p_{1,1} \cdot \ldots \cdot p_{k,1} \ge p_1$. Repeating the argument with the cheaters aiming for outcome 0 completes the proof. □

Now, Theorem 6.5.5 is an immediate consequence.

Proof of Theorem 6.5.5. Using the notation of Lemma 6.5.7, we have $p_0 = 1/2$. Let $q = \max_i p_{i,0}$ denote the maximum probability of any player forcing output 0. By Lemma 6.5.7, $q^k \geq p_{1,0} \cdot \ldots \cdot p_{k,0} \geq 1/2$, from which follows that $q \geq (1/2)^{1/k} \geq 1 - (\ln 2)/k - \mathrm{O}(1/k^2)$. By Theorem 6.3.1 this result applies both to private pairwise quantum channels and the quantum broadcast channel. □

Extending to many honest players Extension to any number of honest players follows almost immediately from Theorem 6.5.5. Indeed, take a protocol Π for k parties tolerating $(k - g)$ cheaters. Arbitrarily partition our players into $k' = k/g$ groups and view each each as one "combined player." We get an induced protocol Π' with k' "super-players" which achieves at least the same bias ε as Π, and can tolerate up to $(k' - 1)$ bad players. By Theorem 6.5.5, $\varepsilon \geq 1/2 - \mathrm{O}(1/k') = 1/2 - \mathrm{O}(g/k)$.

6.6 Summary

We showed that quantum coin flipping is significantly more powerful than classical coin flipping. Moreover, we give *tight* tradeoffs between the number of cheaters tolerated and the bias of the resulting coin achievable by quantum coin-flipping protocols. We also remark that the fact that we obtain tight bounds in the quantum setting is somewhat surprising. For comparison, such tight bounds are unknown for the classical setting.

Bibliography

[1] H. Abelson. Lower bounds on information transfer in distributed computations. *Journal of the ACM*, 27(2):384–392, 1980. Earlier version in FOCS'78. 95

[2] 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. 41

[3] D. Aharonov, A. Yu. Kitaev, and N. Nisan. Quantum circuits with mixed states. In *Proceedings of 30th ACM STOC*, pages 10–20, 1998, quant-ph/9806029. 121

[4] D. Aharonov, A. Ta-Shma, U. Vazirani, and A. Yao. Quantum bit escrow. In *Proceedings of 32nd ACM STOC*, pages 705–714, 2000, quant-ph/0004017. 118

[5] N. Alon. Testing subgraphs in large graphs. In *Proceedings of 42nd IEEE FOCS*, pages 434–441. IEEE, 2001. 57

[6] N. Alon, L. Babai, and A. Itai. A fast and simple randomized parallel algorithm for the maximal independent set problem. *Journal of Algorithms*, 7:567–583, 1986. 72

[7] N. Alon, E. Fischer, M. Krivelevich, and M. Szegedy. Efficient testing of large graphs. In *Proceedings of 40th IEEE FOCS*, pages 656–666. IEEE, 1999. 57

[8] N. Alon, I. Newman, M. Krivelevich, and M. Szegedy. Regular languages are testable with a constant number of queries. In *Proceedings of 40th IEEE FOCS*, pages 645–655, 1999. 57

[9] A. Ambainis. Quantum walk algorithmn for element distinctness. o, quant-ph/0311001. 35

[10] A. Ambainis. A new protocol and lower bounds for quantum coin flipping. In *Proceedings of 33rd ACM STOC*, pages 134–142, 2001, quant-ph/0204022. 118, 119, 126, 127

[11] A. Ambainis. Lower bound for a class of weak quantum coin flipping protocols. 2002, quant-ph/0204063. 119

[12] A. Ambainis, H. Buhrman, Y. Dodis, and H. Röhrig. Multiparty quantum coin flipping. Submitted, 2003, quant-ph/0304112. 117

[13] A. Aspect, J. Dalibard, and G. Roger. Experimental test of Bell's inequalities using time-varying analyzers. *Physical Review Letters*, 49(25):1804, 1982. 89

[14] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher. Noncommuting mixed states cannot be broadcast. *Physical Review Letters*, 76(15):2818–2821, 1996. 123

[15] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials. *Journal of the ACM*, 48(4):778–797, 2001. Earlier version in FOCS 98. 22, 41, 58, 71, 72, 73, 78, 82

[16] N. de Beaudrap, R. Cleve, and J. Watrous. Sharp quantum vs. classical query complexity separations. *Algorithmica*, 34(4):449–461, 2002, quant-ph/0011065. 22

[17] J. S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195–200, 1965. 16, 89

[18] M. Ben-Or and N. Linial. Collective coin-flipping. In *Randomness and Computation*, pages 91–115, 1990. 119

[19] C. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70:1895–1899, 1993. 124

[20] C. H. Bennett, E. Bernstein, G. Brassard, and U. Vazirani. Strengths and weaknesses of quantum computing. *SIAM Journal on Computing*, 26(5):1510–1523, 1997, quant-ph/9701001. 33, 41, 71

[21] C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing*, pages 175–179, 1984. 117

[22] E. Bernstein and U. Vazirani. Quantum complexity theory. *SIAM Journal on Computing*, 26(5):1411–1473, 1997. Earlier version in STOC'93. 1, 22, 57, 58, 59

[23] M. Blum, M. Luby, and R. Rubinfeld. Self-testing and self-correcting programs, with applications to numerical programs. *Journal of Computer and System Sciences*, 47:549–595, 1993. 57

[24] M. Boyer, G. Brassard, P. Høyer, and A. Tapp. Tight bounds on quantum searching. *Fortschritte der Physik*, 46(4–5):493–505, 1998, quant-ph/9605034. Earlier version in Physcomp'96. 31, 33

[25] G. Brassard, R. Cleve, and A. Tapp. The cost of exactly simulating quantum entanglement with classical communication. *Physical Review Letters*, 83(9):1874–1877, 1999, quant-ph/9901035. 90, 95, 96

[26] G. Brassard and P. Høyer. An exact quantum polynomial-time algorithm for Simon's problem. In *Proceedings of the 5th Israeli Symposium on Theory of Computing and Systems (ISTCS'97)*, pages 12–23, 1997, quant-ph/9704027. 22, 33, 62, 64

[27] G. Brassard, P. Høyer, M. Mosca, and A. Tapp. Quantum amplitude amplification and estimation. In Jr. S. J. Lomonaco and H. E. Brandt, editors, *Quantum Computation and Quantum Information: A Millennium Volume*, volume 305 of *Contemporary Mathematics*, pages 53–74. American Mathematical Society, 2002, quant-ph/0005055. 33

[28] G. Brassard, P. Høyer, and A. Tapp. Quantum counting. In *Proceedings of 25th ICALP*, volume 1443 of *Lecture Notes in Computer Science*, pages 820–831. Springer, 1998, quant-ph/9805082. 33, 52

[29] 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. 95, 96, 115

[30] H. Buhrman, R. Cleve, R. de Wolf, and Ch. Zalka. Bounds for small-error and zero-error quantum algorithms. In *Proceedings of 40th IEEE FOCS*, pages 358–368, 1999, cs.CC/9904019. 33

[31] H. Buhrman, W. van Dam, P. Høyer, and A. Tapp. Multiparty quantum communication complexity. *Physical Review A*, 60(4):2737–2741, 1999, quant-ph/9710054. 91, 105, 106, 115

[32] H. Buhrman, Ch. Dürr, M. Heiligman, P. Høyer, F. Magniez, M. Santha, and R. de Wolf. Quantum algorithms for element distinctness. In *Proceedings of 16th IEEE Conference on Computational Complexity*, pages 131–137, 2001, quant-ph/0007016. 35

[33] H. Buhrman, L. Fortnow, I. Newman, and H. Röhrig. Quantum property testing. In *Proceedings of 14th SODA*, pages 480–488, 2003, quant-ph/0201117. 57

[34] H. Buhrman, P. Høyer, S. Massar, and H. Röhrig. Combinatorics and quantum nonlocality. *Physical Review Letters*, 91(4):047903, 2003, quant-ph/0209052. 89, 95

[35] H. Buhrman, P. Høyer, S. Massar, and H. Röhrig. Combinatorics and quantum nonlocality with noise. Manuscript in preparation, 2003. 89, 103, 105, 106

[36] H. Buhrman, I. Newman, H. Röhrig, and R. de Wolf. Robust quantum algorithms and polynomials. Submitted, quant-ph/0309220. 75, 86

[37] H. Buhrman and H. Röhrig. Distributed quantum computing. In B. Rovan and P. Vojtas, editors, *Mathematical Foundations of Computer Science 2003*, volume 2747 of *Lecture Notes in Computer Science*, pages 1–20. Springer, 2003. 89

[38] H. Buhrman and R. de Wolf. Complexity measures and decision tree complexity: A survey. *Theoretical Computer Science*, 288(1):21–43, 2002. 77

[39] I. L. Chuang. Quantum computation with nuclear magnetic resonance. In H.-K. Lo, S. Popescu, and T. P. Spiller, editors, *Introduction to Quantum Computation and Information*, pages 311–339. World Scientific, 1998. 44

[40] B. S. Cirel'son. Quantum generalizations of Bell's inequality. *Letters in Mathematical Physics*, 4(2):93–100, 1980. 93

[41] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23:880–884, 1969. 90, 93

[42] R. Cleve. The query complexity of order-finding. In *Proceedings of 15th IEEE Conference on Computational Complexity*, pages 54–59, 2000, quant-ph/9911124. 23

[43] R. Cleve and H. Buhrman. Substituting quantum entanglement for communication. *Physical Review A*, 56(2):1201–1204, 1997, quant-ph/9704026. 95

[44] R. Cleve and J. Watrous. Fast parallel circuits for the quantum Fourier transform. In *Proceedings of 41st IEEE FOCS*, pages 526–536, 2000, quant-ph/0006004. 36

[45] J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19(90):297–301, 1965. 37

[46] D. Coppersmith. An approximate Fourier transform useful in quantum factoring. IBM Research Report No. RC19642, 1994, quant-ph/0201067. 36, 37

[47] 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. 1

[48] D. Deutsch and R. Jozsa. Rapid solution of problems by quantum computation. In *Proceedings of the Royal Society of London*, volume A439, pages 553–558, 1992. 1, 20

[49] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, 47:777–780, 1935. 12

[50] A. Ekert and R. Jozsa. Quantum algorithms: Entanglement enhanced information processing. In *Proceedings of the Royal Society of London*, volume A356, pages 1769–1782, 1998, quant-ph/9803072. 36, 37, 39

[51] F. Ergün, S. Kannan, S. Kumar, R. Rubinfeld, and M. Vishwanathan. Spot-checkers. *Journal of Computer and System Sciences*, 60(3):717–751, 2000. 57

[52] U. Feige. Noncryptographic selection protocols. In *Proceedings of 40th IEEE FOCS*, pages 142–152, 1999. 118, 120, 128

[53] U. Feige, P. Raghavan, D. Peleg, and E. Upfal. Computing with noisy information. *SIAM Journal on Computing*, 23(5):1001–1018, 1994. 75, 82

[54] R. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6/7):467–488, 1982. 1

[55] R. Feynman. Quantum mechanical computers. *Optics News*, 11:11–20, 1985. 1

[56] E. Fischer. The art of uninformed decisions: A primer to property testing. *The Bulletin of the European Association for Theoretical Computer Science*, 75:97–126, 2001. 57

[57] E. Fischer. Testing graphs for colorability properties,. In *Proceedings of 12th SODA*, pages 873–882, 2001. 57

[58] E. Fischer and I. Newman. Testing of matrix properties. In *Proceedings of 33rd ACM STOC*, pages 286–295, 2001. 57

[59] P. Frankl and V. Rödl. Forbidden intersections. *Transactions of the American Mathematical Society*, 300(1):259–286, 1987. 98

[60] E. Fredkin and T. Toffoli. Conservative logic. *International Journal of Theoretical Physics*, 21(3–4):219–253, 1982. 18

[61] K. Friedl, F. Magniez, M. Santha, and P. Sen. Quantum testers for hidden group properties. In *Proceedings of the 28th International Symposium on Mathematical Foundations of Computer Science*, 2003, quant-ph/0208184. 62, 73

[62] N. Gisin and B. Gisin. A local hidden variable model of quantum correlation exploiting the detection loophole. *Phys. Lett. A*, 260(5):323–327, 1999. 90, 95

[63] L. Goldenberg, L. Vaidman, and S. Wiesner. Quantum gambling. *Physical Review Letters*, 82:3356–3359, 1999. 119

[64] O. Goldreich. Combinatorial property testing (a survey), 1998. Manuscript. 58

[65] O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45(4):653–750, 1998. 57

[66] O. Goldreich and L. Trevisan. Three theorems regarding testing graph properties. In *Proceedings of 42nd IEEE FOCS*, pages 460–469. IEEE, Nevada, 2001. 57

[67] D. Gottesman and I. Chuang. Demonstrating the viability of universal quantum computation using teleportation and single qubit operations. *Nature*, 402(6760):390–393, 1999. 125

[68] D. M. Greenberger, M. A. Horne, and A. Zeilinger. Going beyond Bell's theorem. In M. Kafatos, editor, *Bell's Theorem, Quantum Theory, and Conceptions of the Universe*, volume 37 of *Fundamental Theories of Physics*, pages 69–72. Kluwer, 1989. 89

[69] L. K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of 28th ACM STOC*, pages 212–219, 1996, quant-ph/9605043. 23, 27, 41, 57, 76

[70] L. K. Grover. A framework for fast quantum mechanical algorithms. In *Proceedings of 30th ACM STOC*, pages 53–62, 1998, quant-ph/9711043. 33

[71] L. K. Grover. How fast can a quantum computer search? 1998, quant-ph/9809029. 41

[72] L. K. Grover. Rapid sampling through quantum computing. In *Proceedings of 32nd ACM STOC*, pages 618–626, 2000, quant-ph/9912001. 40

[73] A. S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problemy Peredachi Informatsii*, 9(3):3–11, 1973. English translation in *Problems of Information Transmission*, 9:177–183, 1973. 98

[74] P. Høyer. Fourier sampling. Private communication, 2001. 62, 73

[75] P. Høyer, M. Mosca, and R. de Wolf. Quantum search on bounded-error inputs. In *Proceedings of 30th ICALP*, volume 2719 of *Lecture Notes in Computer Science*, pages 291–299. Springer, 2003, quant-ph/0304052. 48, 76, 79

[76] J. Hromkovič. *Communication Complexity and Parallel Computing.* Springer, 1997. 95

[77] J. Kerenidis and A. Nayak. Weak coin flipping with small bias. 2002, quant-ph/0206121. 127

[78] A. Yu. Kitaev. Quantum error correction with imperfect gates. In O. Hirota, A. S. Holevo, and C. M. Caves, editors, *Quantum Communication, Computing, and Measurement*, Proceedings of the Third International Conference held in Shizuoka, Japan, September 25-30, 1996, pages 181–188. Kluwer Academic/Plenum Publishers, 1997. 41

[79] A. Yu. Kitaev. Quantum coin-flipping. Talk at QIP 2003 (slides and video at MSRI), December 2002. 119, 120, 128

[80] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 1997. 77, 95, 111

[81] A. K. Lenstra and H. W. Lenstra. *The Development of the Number Field Sieve*, volume 1554 of *Lecture Notes in Mathematics*. Springer, 1993. 23

[82] H. W. Lenstra and C. Pomerance. A rigorous time bound for factoring integers. *Journal of the American Mathematical Society*, 5:483–516, 1992. 23

[83] H. K. Lo and H. F. Chau. Why quantum bit commitment and ideal quantum coin tossing are impossible. *Physica D*, 120:177–187, 1998. 118

[84] H-K. Lo and H. F. Chau. Unconditional security of quantum key distribution over arbitrarily long distances. 3 Mar 1998, quant-ph/9803006. 117

[85] G. L. Long, Y. S. Li, W. L. Zhang, and C. C. Tu. Dominant gate imperfection in Grover's quantum search algorithm. *Physical Review A*, 61:042305, 2000, quant-ph/9910076. 45

[86] S. Massar. Nonlocality, closing the detection loophole, and communication complexity. *Physical Review A*, 65(032121), 2002. 90, 95, 96

[87] S. Massar, D. Bacon, N. Cerf, and R. Cleve. Classical simulation of quantum entanglement without local hidden variables. *Physical Review A*, 63(5):052305, 2001, quant-ph/0009088. 101

[88] S. Massar and S. Pironio. Violation of local realism vs detection efficiency. 2003, quant-ph/0210103. 101

[89] D. Mayers. Unconditional security in quantum cryptography. 10 Feb 1998, quant-ph/9802025. 117

[90] D. Mayers. Unconditionally secure quantum bit commitment is impossible. *Physical Review Letters*, 78:3414–3417, 1997, quant-ph/9605044. 118

[91] D. Mayers, L. Salvail, and Y. Chiba-Kohno. Unconditionally secure quantum coin tossing. 22 Apr 1999, quant-ph/9904078. 118

[92] N. D. Mermin. Extreme quantum entanglement in a superposition of macroscopically distinct states. *Physical Review Letters*, 65:1838–1840, 1990. 105

[93] N. D. Mermin. Simple unified form for the major no-hidden-variables theorems. *Physical Review Letters*, 65:3373–3376, 1990. 89, 105

[94] Y. I. Ozhigov and H. Röhrig. Fast quantum comparison of degeneracy degrees. Manuscript, 2000. 27

[95] B. Pablo-Norman and M. Ruiz-Altaba. Noise in Grover's quantum search algorithm. *Physical Review A*, 61:012301, 2000, quant-ph/9903070. 45

[96] C. H. Papadimitriou. *Computational Complexity.* Addison-Wesley, 1994. 17

[97] P. Pearle. Hidden-variable example based upon data rejection. *Phys. Rev. D*, 2(8):1418–1425, 1970. 90

[98] Roger Penrose. *Shadows of the mind: a search for the missing science of consciousness.* Oxford University Press, Oxford, 1994. 5

[99] J. Preskill. Lecture notes for a course on quantum computation. Manuscript., 1998–1999. 39

[100] R. Raz. Exponential separation of quantum and classical communication complexity. In *Proceedings of 40th IEEE FOCS*, pages 358–367, 1999. 115

[101] R. Rivest, A. Shamir, and L. Adleman. A method for obtaining digital signatures and public key cryptosystems. *Communications of the ACM*, 21:120–126, 1978. 23

[102] D. Ron. Property testing. In S. Rajasekaran, P. M. Pardalos, J. H. Reif, and J. D. P. Rolim, editors, *Handbook of Randomized Computing*, volume 9 of *Combinatorial Optimization.* Kluwer, 2001. 57

[103] R. Rubinfeld and M. Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, April 1996. 57

[104] M. Saks. A robust noncryptographic protocol for collective coin flipping. *SIAM J. Discrete Math.*, 2(2):240–244, 1989. 118

[105] N. Shenvi, K. R. Brown, and K. B. Whaley. Effects of noisy oracle on search algorithm complexity. submitted, 2003, quant-ph/0304138. 45

[106] P. W. Shor. Scheme for reducing decoherence in quantum memory. *Physical Review A*, 52:2493, 1995. 41

[107] P. W. Shor. Fault-tolerant quantum computation. In *Proceedings of 37th IEEE FOCS*, pages 56–65, 1996. 41

[108] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, 1997, quant-ph/9508027. Earlier version in FOCS'94. 1, 36, 37, 57

[109] D. Simon. On the power of quantum computation. *SIAM Journal on Computing*, 26(5):1474–1483, 1997. Earlier version in FOCS'94. 1, 22, 57, 58, 62

[110] R. Solovay and A. Yao, 1996. Manuscript. 71

[111] R. W. Spekkens and T. Rudolph. A quantum protocol for cheat-sensitive weak coin flipping. *Physical Review Letters*, 89:227901, 2002, quant-ph/0202118. 119, 126

[112] 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. 41

[113] M. Steiner. Towards quantifying non-local information transfer: finite-bit non-locality. *Physics Letters A*, 270(5):239–244, 2000, quant-ph/9902014. 90

[114] M. Szegedy. Private communication. 1999. 58, 73

[115] M. Szegedy and X. Chen. Computing Boolean functions from multiple faulty copies of input bits. In *Proceedings of 5th LATIN*, volume 2286 of *Lecture Notes in Computer Science*, pages 539–553, 2002. 76, 83

[116] R. Wilmink. *Quantum Broadcast Channels and Cryptographic Applications for Separable States.* PhD thesis, Universität Bielefeld, 2002. 123

[117] W. K. Wootters and W. H. Zurek. A single quantum cannot be copied. *Nature*, 299:802–803, 1982. 119, 123

[118] A. C-C. Yao. Probabilistic computations: Toward a unified measure of complexity. In *Proceedings of 18th IEEE FOCS*, pages 222–227, 1977. 60, 63, 98

[119] A. C-C. Yao. Some complexity questions related to distributive computing. In *Proceedings of 29th ACM STOC*, pages 209–213, 1979. 95

[120] Ch. Zalka. Grover's quantum searching algorithm is optimal. *Physical Review A*, 60:2746–2751, 1999, quant-ph/9711070. 33

Index

Samenvatting

Theoretische informatica bestudeert de sterkte en beperkingen van computers op basis van abstracte modellen van berekening. De keuze van modellen wordt geleid door drie overwegingen: (1) hoe ver is het model verwijderd van bestaande computers of computers die in principe kunnen worden geconstrueerd? (2) hoe geschikt is het model voor het bewijzen van interessante eigenschappen van computers? (3) hoe elegant het model wiskundig gezien?

"Quantum computation" doet een beroep op alle drie de criteria. De tegenwoordig gangbare mening is dat de natuurkundige theorie van de quantum mechanica de realiteit accuraat representeert op het niveau van zeer kleine schalen in lengte, tijd en energie. Klassieke probabilistische Turing machines kunnen worden begrepen als modellen voor computers die werken op basis van de klassieke natuurkunde; "quantum circuits" daarentegen zijn realistische computers die zich gedragen volgens de wetten van de quantum mechanica. "Query complexiteit," een variant van tijdscomplexiteit, heeft een nauwkeurig analogon voor quantum computers; net als in het klassieke geval zijn de ons bekende wiskundige werktuigen beter geschikt voor deze beperkte maat van complexiteit dan voor algemene tijdscomplexiteit. In sommige gevallen kan quantum query complexiteit zelfs nieuwe inzichten in de klassieke complexiteitstheorie leveren. Quantum mechanica is gebaseerd op de elegante wiskundig theorie van de functionaal-analyse; daardoor profiteert quantum computing van nieuwe toepassingen van lineaire algebra en matrix-analyse.

Dit proefschrift bekijkt de eigenschappen en toepassingen van quantum query complexiteit en het verwante begrip van communicatie complexiteit. Wij suggereren nieuwe cryptografische protocollen en nieuwe natuurkundige experimenten om de voorspellingen van de quantum mechanica te testen. Quantum toestanden zijn erg gevoelig; dit proefschrift onderzoekt hoe imperfecties en fouten kunnen worden overwonnen in verschillende situaties.

Quantum Query Complexiteit In het geval van query complexiteit meten we hoe vaak een algoritme input bits leest. Een beroemd resultaat in het vakgebied van quantum computing is Grover's algoritme. Dit kan een zoekruimte veel sneller doorzoeken dan elk klassiek algoritme. Wij bestudeerden het quantum zoekproces met verschillende soorten imperfecties.

"Property testing" heeft veel aandacht getrokken in de laatste jaren, zowel om theoretische redenen zoals de PCP stelling, als om praktische redenen voor toepassingen op grote data bestanden. Het uitgangspunt bij dit laatste, is dat de input zo lang is dat het niet mogelijk is om het geheel te lezen, we kunnen slechts een beperkt aantal input bits bekijken. Voor de meesten mogelijke eigenschappen van inputs, zijn zulke tests niet voldoende om een onderscheid te maken tussen het geval dat de input wel de eigenschap heeft, en het geval dat er ten minste één bit verschil is met iedere andere input die de eigenschap heeft. Maar een afgezwakte vorm van het testen van de eigenschap blijft denkbaar: we willen weten of de input hetzij de eigenschap heeft, hetzij in veel bit posities afwijkt van iedere andere input die de eigenschap heeft. "Property testing" bestudeert algoritmen die met weinig vragen aan de input kunnen onderscheiden tussen deze beiden gevallen. Onze bijdrage is het overzetten van property testing naar quantum computation: wij bewijzen dat voor het testen van bepaalde eigenschappen, quantum computers exponentieel veel efficiënter zijn dan klassieke computers, en wij laten zien dat er eigenschappen zijn die niet op een efficiënte manier kunnen worden getest met een quantum computer.

Daadwerkelijk een quantum computer bouwen is een moeilijke taak. Fouten in quantum geheugen en operaties zijn onvermijdelijk en moeten dus worden bestreden hetzij door software, hetzij door hardware. Een serie van fundamentele resultaten heeft laten zien dat de fragiele quantum toestand kan worden beschermd tegen bepaalden typen van fouten. Het is zelfs mogelijk om fout-tolerant quantum computation uit te voeren indien de fouten niet al te frequent zijn, en beperkt zijn tot bepaalde typen. Gecombineerd met recente experimentele vooruitgang, heeft dit de perspectieven voor quantum computers aanzienlijk verbetert. Deze fout-tolerante constructies zijn echter niet toegepast op het model van query complexiteit, waar die fouten kunnen optreden bij het lezen van input bits. Het is van belang zulke fouten te onderzoeken omdat zij in quantum geheugen en in de compositie van quantum algoritmen optreden. Wij formaliseren het begrip van "noisy" quantum geheugen met behulp van de definitie van "noisy queries." Wij laten zien dat met zulke queries bepaalde quantum algoritmen robuust kunnen worden gemaakt tegen fouten, soms zelfs efficiënter dan klassieke algoritmen robuust kunnen worden gemaakt. Tevens breiden wij het concept van benadering van Boolean functies door polynomen uit naar benadering door *robuust* polynomen.

Quantum Distributed Computing Niet-lokaliteit is en aspect van de quantum mechanica dat niet expliciet werd ingebouwd door de uitvinders. Einstein en collega's merkten op dat de axioma's van de quantum mechanica voorspellen dat twee objecten kunnen bestaan in een verstrengelde ("entangled") toestand waar elke manipulatie van het ene object en direct effect heeft op het andere object, zelfs als deze zich ver van het eerste object bevindt. In het begin werd dit effect beschouwd als een onrealistisch artefact en dus werd er gekeken naar alternatieve theorieën zonder niet-lokaliteit. Maar toen het technologisch mogelijk werd niet-lokaliteit experimenteel te testen, werden geen contradicties met quantum mechanica gevonden. Echter, als men wiskundig rekening houdt met de praktisch beperkingen in de precisie van hedendaagse experimenten, dan ziet man dat er wel vergezochte klassieke theorieën bestaan die met de experimentele data overeenstemmen. Dus wordt er nog steeds gezocht naar experimenten voor het aantonen van niet-lokaliteit die door geen enkele klassieke theorie verklaard kunnen worden. Wij gebruiken combinatorische technieken die oorspronkelijk ontwikkelt werden voor quantum communication complexiteit om nieuwe experimenten voor te stellen die resistent zijn tegen de meest voorkomende fout, namelijk inefficiënte detectoren. De experimenten die wij voorstellen zijn tegelijkertijd bestand tegen bepaalde algemene noise.

"Distributed computing" bestudeert computationele taken voor groepen van participanten. Bijvoorbeeld, verkiezingen of het uitzenden van één mededeling aan vele partijen, waarbij enkele participanten defect zijn of opzettelijk saboteren. Deze problemen en hun oplossingen hebben veel gemeen met cryptografie. Problemen zoals de onmogelijke quantum bit commitment en "coin flipping" kunnen worden afgezwakt tot versies die bij benadering werken. Wij ontwerpen het concept van quantum broadcast, introduceren een nieuw protocol voor twee partijen, en passen dit toe op verkiezingen in het geval waarbij er een grote meerderheid van "slechte" participanten is. We bewijzen dat dit nieuwe protocol optimaal is.

Abstract

In complexity theory, the strengths and limitations of computers are investigated on abstract models of computation. The choice of these models is governed by three considerations: (1) how close is the model to existing computers or computers that could be built in principle? (2) how well does it lend itself to proving interesting properties of computers? (3) how elegant is the model mathematically?

Quantum computation appeals to all three criteria. In functional analysis, quantum mechanics has a beautiful mathematical underpinning, which benefits quantum computing through new applications of linear algebra and matrix analysis. Nowadays it is a widely-held belief that the physical theory of “quantum mechanics” describes reality accurately at very small scales of length, time, and energy. Where classical probabilistic Turing machines may be seen as capturing the power of computers operating according to finite-precision classical physics, the computational model of “quantum circuits” aims at modeling what realistic computers in a quantum mechanical world can do. Query complexity, a variant of time complexity, has a close analogue for quantum computers; as in the classical case, our current mathematical tools are more amenable to this restricted complexity measure than to general time complexity. Sometimes, the implications of quantum query complexity shed new light even on classical complexity theory.

This thesis investigates the properties and applications of quantum query complexity and the related quantum communication complexity. It suggests new cryptographic protocols and new experiments for probing the predictions of quantum mechanics. Quantum states are very sensitive; this thesis examines ways to deal with imperfections and errors in a number of different situations.

Quantum Query Complexity In query complexity, we are concerned with the number of times an algorithm reads a bit of the input. A celebrated result of quantum computing is Grover's algorithm, which allows an entry to be found in an unordered database with significantly less queries than any classical computer. We studied quantum search and its generalizations, particularly in the presence of imperfections.

"Property testing" drew a lot of attention in recent years, both for theoretical applications in relation to the PCP theorem and for practical applications on large data sets. The premise is that the input is so large that it is not possible to consider it in its entirety, only sampling from it in a few places instead. For most properties, sampling is not sufficient to tell whether the input has that property or whether it differs from each input with the property in at least a single bit position. However, a relaxed notion of checking the property is still conceivable: we would like to know whether or not the input differs from all inputs with the property in *many* bit positions. "Property testing" is concerned with algorithms that distinguish between the two cases of being close or far from having a given property. Our contribution is to translate this concept to quantum computation: we prove that quantum computers can be exponentially more efficient than classical computers in testing certain properties and we also show that there are properties that are untestable even by quantum computers.

Building quantum computers will be a challenging task. Errors in the quantum memory and quantum operations are unavoidable and need to be dealt with either by hardware or software. Surprisingly, a chain of landmark results showed that the fragile quantum state can be protected against certain types of errors and it is even possible to perform fault-tolerant quantum computation, provided the noise is of a certain kind and the noise level not too high. Together with recent experimental progress, this improves the prospects of real-world quantum computers. However, the fault-tolerance constructions do not apply to errors in the query-complexity model caused by distorted access to the input. Errors of this type are of interest because they arise in the composition of quantum algorithms and because they model real-world errors in accesses to quantum memory. We formalize the notion of noisy access to the input by proposing models of "noisy queries." We show that for one such model (which corresponds to composing quantum algorithms) some quantum algorithms can actually be made robust at less cost than classical algorithms. We also extend the concept of approximating Boolean functions by polynomials to polynomials "robustly" approximating Boolean functions.

Quantum Distributed Computing Nonlocality is a feature of quantum mechanics that was not explicitly incorporated by its inventors. Instead, Ein-

stein and others remarked that the axioms of quantum mechanics predict that two distant objects can be in an "entangled" state where manipulations of one object have an immediate effect on the other object, no matter how far apart. At first, this effect was discounted as an unrealistic and hence undesirable property, which needed to be eliminated by a theory replacing or amending quantum mechanics. When it became technologically feasible to conduct experiments probing nonlocality, it turned out that the results do not contradict quantum mechanics. However, due to the difficulty of conducting such experiments, they are hampered by practical limitations. Taking noise into account, it is possible to explain the data from all experiments conducted up to now using contrived classical theories. Consequently, there is an ongoing effort to devise and conduct "loophole-free" nonlocality experiments. Using combinatorial techniques developed originally for the study of quantum communication complexity, we present abstract experiments that are resistant to the most common type of error, detector inefficiency, as well as some level of more general noise.

Distributed computing studies computational tasks to be accomplished by a group of people. Examples include voting and broadcasting the same message to many parties over point-to-point channels in presence of disabled or malevolent participants. These problems share many properties and techniques with cryptography. Problems such as the impossible quantum bit commitment can be relaxed to approximate coin tossing, which can be used for two-party leader election. We develop the notion of a quantum broadcast channel, introduce a new two-party protocol, and apply it to multiparty coin-flipping with an overwhelming majority of "bad" parties. We show that the new multiparty protocol is asymptotically optimal.

Titles in the ILLC Dissertation Series:

ILLC DS-1999-01: **Jelle Gerbrandy**
Bisimulations on Planet Kripke

ILLC DS-1999-02: **Khalil Sima'an**
Learning efficient disambiguation

ILLC DS-1999-03: **Jaap Maat**
Philosophical Languages in the Seventeenth Century: Dalgarno, Wilkins, Leibniz

ILLC DS-1999-04: **Barbara Terhal**
Quantum Algorithms and Quantum Entanglement

ILLC DS-2000-01: **Renata Wassermann**
Resource Bounded Belief Revision

ILLC DS-2000-02: **Jaap Kamps**
A Logical Approach to Computational Theory Building (with applications to sociology)

ILLC DS-2000-03: **Marco Vervoort**
Games, Walks and Grammars: Problems I've Worked On

ILLC DS-2000-04: **Paul van Ulsen**
E.W. Beth als logicus

ILLC DS-2000-05: **Carlos Areces**
Logic Engineering. The Case of Description and Hybrid Logics

ILLC DS-2000-06: **Hans van Ditmarsch**
Knowledge Games

ILLC DS-2000-07: **Egbert L.J. Fortuin**
Polysemy or monosemy: Interpretation of the imperative and the dative-infinitive construction in Russian

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