Principles of
Quantum Computation and Quantum Information

Erwin Brüning
School of Mathematical Sciences

April 2013

Contents

Chapter 1

Introduction

Quantum computation and quantum information is a quite recent and very rapidly developing field of research. Effectively this field is based on fundamental ideas from the following fields:

1. quantum mechanics;
2. computer science;
3. information theory;
4. cryptography.

Accordingly quantum computation and quantum information is an interdisciplinary field with ground breaking developments in the experimental and theoretical side. It is based on profound experimental and theoretical progress in quantum physics (2012 Physics Nobel price winners: S. Haroche and D.

Wineland: made tremendous advances in our understanding of quantum entanglement, with beautiful experiments to show how atomic systems can be manipulated to exhibit the most extraordinary coherence properties).

Quantum computation requires some knowledge of the working of a classical computer. But this will not be discussed here.

1.1 History

Here we mention only the most important dates in its history. The following table presents some highlights. Some of these will be discussed in more detail later.

1920 - 1930	quantum mechanics emerged in its present form.
1936	Alan Turing - existence of a **Universal Turing Machine**.
1936 - 1939	**Church-Turing thesis**: A function is algorithmically computable if and only if it is computable by a Turing machine.
1947	transistor is developed by Bardeen, Brattain, and Shockley.

1948	C. Shannon: foundations of a mathematical theory of information and communication.
$\approx$1970	computational complexity theory: efficient algorithms run in polynomial time (with the size of the problem); strengthened version of the Church-Turing thesis: *Any algorithmic process can be simulated efficiently using a Turing machine*.
$\geq$1970	experimental control of single quantum systems (atoms, electrons), in particular designer arrays of atoms.
$\approx$1975	primality test of integers by a **randomized algorithm**.
$\approx$1976	R. Rivest, A. Shamir, L. Adleman: public key crypto system - RSA cryptosystem.
1982	Wootters and Zurek, Dieks: no-cloning theorem for quantum states.
1984	CH. Bennet, G. Brassard: BB84 protocol - quantum key exchange, security of communication based on quantum mechanics - quantum cryptography, 3.

1985	R. Feynman: suggestion of quantum mechanical computers, 11.
1985	D. Deutsch: **Universal Quantum Computer**, i.e., a computational device based on the principles of quantum mechanics which is capable of efficiently simulating an arbitrary physical system, 8, 9.
1993	G. Brassard C. Crépeau R. Jozsa A. Peres Bennett, C.H. and W.K. Wootters: First suggestion of teleportation with qubits, 7.
1994	L. Vaidman: continuous variable teleportation suggested, 14.
1994	P. Shor: algorithm for finding prime factors of an integer on a quantum computer (QC), 13.
1995	L. Grover: algorithm for search through unstructured search space on a QC, 12.

1996	R. Calderbank, P. Shor, A. Steane: quntum error-correcting codes - CSS codes, 6.
1997	D. Bouwmeester , JW Pan, K. Mattle, M. Eible, H. Weinfurter, A. Zeilinger: Experimental quantum teleportation with photon polarization, 4.
1998	S.L. Braunstein and H.J. Kimble: Teleportation of continuous quantum variables, 5.
2004	world's first bank transfer using quantum key distribution in Vienna and several other quantum key distribution networks were started (USA, Switzerland).

1.2 Overview

Chapter 2

Quantum Mechanics in Hilbert space

In this chapter we recall briefly the basics of quantum mechanics in a form and in a notation used later.

2.1 Hilbert spaces

In the following $\mathcal{H}$ denotes a complex separable Hilbert space. For most parts of this lecture $\mathcal{H}$ will actually be finite dimensional, i.e., isomorphic to the Hilbert space of n-tuples of complex numbers, for $n = 2,3,\ldots$:

$$\mathcal{H} \simeq \mathbb{C}^n.$$

A **Hilbert space** is a complete normed space whose norm is defined in terms of an inner product according to the relation

$$\|\psi\| = \sqrt{\langle\psi|\,\psi\rangle}, \quad \psi \in \mathcal{H}.$$

Here $\langle\cdot|\,\cdot\rangle$ denotes the **inner product** on $\mathcal{H}$ which in our convention is anti-linear in the first argument and linear in the second. Separability means that there is a **complete othonormal system** in $\mathcal{H}$, i.e., a sequence of vectors e_j, $j \in \mathbb{N}$, with $\langle e_j|\, e_i\rangle = \delta_{ji}$ such that for every $\psi \in \mathcal{H}$ one has

$$\psi = \sum_{j=0}^{\infty} \langle e_j|\,\psi\rangle\, e_j\,. \tag{2.1}$$

A unit vector $e \in \mathcal{H}$ defines a one dimensional subspace

$$\mathcal{H}_e = \mathbb{C}e\,,$$

and the **projection operator** P_e onto this subspace is defined by

$$P_e\psi = \langle e|\,\psi\rangle\, e, \quad \psi \in \mathcal{H}\,.$$

Following physicists tradition and using Dirac's bra and ket notation, this projection can be written as

$$P_e = |e\rangle\langle e|\,. \tag{2.2}$$

A **linear operator** A on $\mathcal{H}$ is a linear map $D(A) \to \mathcal{H}$ where $D(A)$ is a dense linear subspace of $\mathcal{H}$. Such a linear operator has a unique **adjoint operator** A^* defined by

$$\begin{aligned} D(A^*) &= \{\phi \in \mathcal{H} : \exists C < \infty, \, |\langle \phi | A\psi \rangle| \leq C \|\psi\| \ \forall \psi \in D(A)\}, \\ \langle A^*\phi | \psi \rangle &= \langle \phi | A\psi \rangle \, , \forall \psi \in D(A), \forall \phi \in D(A^*) \, . \end{aligned} \tag{2.3}$$

A linear operator A is **bounded** iff $D(A) = \mathcal{H}$ and $\|A\psi\| \leq \text{const}\|\psi\|$ for all $\psi \in \mathcal{H}$ and then the norm of A is defined by $\|A\| = \sup\{\|A\psi\| : \psi \in \mathcal{H}, \|\psi\| \leq 1\}$.

A bounded linear operator A is **isometric** iff $\|A\psi\| = \|\psi\|$ for all $\psi \in \mathcal{H}$. A bounded linear operator A is **unitary** iff A is isometric and onto $\mathcal{H}$. A unitary operator A on $\mathcal{H}$ is characterized by the identities

$$A^*A = AA^* = I$$

where I denotes the identity operator on $\mathcal{H}$.

Note that for a finite dimensional Hilbert space all linear operators are bounded.

2.2 States and Observables

The **observables** a of a quantum mechanical system Σ are realized as **self-adjoint operators** A in a complex separable Hilbert space $\mathcal{H}$. Recall that a

linear operator A is called self-adjoint iff it equals its adjoint: $A = A^*$ (note that this equality includes the equality of the domains of definition, i.e., $D(A) = D(A^*)$ as defined above).

Many observables have to be realized by unbounded self-adjoint operators, for instance the momentum operator P or the energy operator or Hamiltonian H.

Every self-adjoint operator A on a complex Hilbert space $\mathcal{H}$ has a unique **spectral representation**, i.e., there is a unique **spectral family** $E_\lambda, \lambda \in \mathbb{R}$, on $\mathcal{H}$ (the E_λ are orthogonal projections on $\mathcal{H}$) such that

$$D(A) = \left\{ x \in \mathcal{H} : \int_{\mathbb{R}} \lambda^2 d \left\| E_\lambda x \right\|^2 < \infty \right\}, \quad Ax = \int_{\mathbb{R}} \lambda dE_\lambda x,\ x \in D(A)\,. \tag{2.4}$$

The spectral representation of a self-adjoint operator A allows to calculate many functions $f(A)$ of A according to the formula

$$\begin{aligned} D(f(A)) &= \left\{ x \in \mathcal{H} : \int_{\mathbb{R}} |f(\lambda)|^2 d \left\| E_\lambda x \right\|^2 < \infty \right\}, \\ f(A)x &= \int_{\mathbb{R}} f(\lambda) dE_\lambda x,\ x \in D(f(A))\,. \end{aligned} \tag{2.5}$$

whenever the above integrals exist. This is certainly the case for all bounded continuous functions f on $\mathbb{R}$. If we do this for all the self-adjoint operators A

which correspondent to all the observables a of a system Σ we can form the C^*-algebra $\mathcal{O} = \mathcal{O}(\Sigma)$of all observables of Σ.

Recall that the **states of a quantum mechanical system** Σ are realized as normalized positive linear functionals on its C^*-algebra of observables $\mathcal{O}$, i.e., linear functions $\phi : \mathcal{O} \to \mathbb{C}$ satisfying

$$\phi(A^*A) \geq 0, \quad A \in \mathcal{O}, \quad \phi(I) = 1 \,.$$

Under certain technical assumptions such functionals are of the form

$$\phi(A) = \mathrm{Tr}\,(WA),\ A \in \mathcal{O} \tag{2.6}$$

where W is a **density matrix** on $\mathcal{H}$. A bounded linear operator W on $\mathcal{H}$ is a density matrix iff a) $W \geq 0$, i.e., $\langle x|\, Wx\rangle \geq 0$ for all $x \in \mathcal{H}$ and
b) W is of trace class and of trace 1, i.e., $\sum_{j=0}^{\infty} \left\langle e_j \right|\, We_j \rangle = 1$ for some (and then for any) orthonormal basis e_j of $\mathcal{H}$.

Every density matrix W in $\mathcal{H}$ has the following spectral representation: There is an orthonormal basis e_j, $j \in \mathbb{N}$, of $\mathcal{H}$ and a sequence of numbers $\sigma_j \geq 0$ with $\sum_{j=0}^{\infty} \sigma_j = 1$ such that

$$W = \sum_{j=0}^{\infty} \sigma_j P_{e_j} \tag{2.7}$$

where P_{e_j} denotes the orthogonal projector onto the subspace spanned by the vector e_j as given in (2.2). Note that some of the eigen-values σ_j of W can be 0. For such a W Formula (2.6) takes the form

$$\phi(A) = \sum_{j=0}^{\infty} \sigma_j \left\langle e_j \right| A e_j \rangle \tag{2.8}$$

2.3 Time evolution

As in classical mechanics the time evolution of a quantum system is generated by the Hamiltonian H according to the **Schrödinger equation**

$$\mathrm{i}\hbar \frac{d}{dt}\psi(t) = H\psi(t) \tag{2.9}$$

for a **wave function** $\psi : \mathbb{R} \to \mathcal{H}$. Since the Hamiltonian H is a self-adjoint operator it generates a unitary group

$$U(t) = e^{-\mathrm{i}\frac{t}{\hbar}H}, \quad t \in \mathbb{R} \tag{2.10}$$

and the solution of Schrödinger's equation for the initial condition is $\psi(0) = \psi$ is

$$\psi(t) = U(t)\psi.$$

In our context the Hamiltonian is always time independent and usually we work in units where $\hbar = \frac{\text{Planck's constant}}{2\pi} = 1$. Note that

$$U(t)^* = U(-t), \quad U(t_1)U(t_2) = U(t_1 + t_2), \quad U(0) = I$$

holds.

2.4 Measurements

2.4.1 General description of the measuring process

All the information which we have about a physical system is obtained from observations and measurements. Observations consist in bringing the system under examination in contact with some other system, the observer, or some measuring device M, and observing the reaction of the system on the observer.

Two important features of the measuring process:

1. Back-effect of the measuring device on the system: The measuring device M must interact somehow with the system. But an interaction always acts both ways, hence M also acts on the system, producing an effect on the system with no particularly desirable consequences. This back-effect on the system seems to be the cause of the difficulty in the interpretation of quantum mechanics.

2. Appearance of the “conscious observer": If a measurement is to be useful there must be a further observation on M, namely “reading the scale". Such further observations may be made at a later time by examining a permanent record of some sort, but in any case, these further observations must enter the consciousness of a scientific observer. (Schrödinger: Knowledge which nobody has, is no knowledge!)

2.4.2 Born’s rule

Suppose we measure an observable a of a quantum system in a state with the density matrix W. If this observable is represented by the self-adjoint operator A with the spectral representation (2.4) then the probability $p_W^A(a,b)$ that the measured value lies in the interval (a,b) is

$$p_W^A(a,b) = \mathrm{Tr}\,(AWE(a,b)), \quad E(a,b) = \int_a^b dE_\lambda\,. \tag{2.11}$$

If we take the form (2.7) for W into account this probability equals

$$p_W^A(a,b) = \sum_{j=0}^{\infty} \sigma_j \left\langle e_j \right| AE(a,b)e_j \left\rangle\,.\right.$$

In particular, if the observable A has a purely discrete spectrum, i.e., $A = \sum_i a_i P_i$ and the system is in the pure state given by the unit vector $\psi \in \mathcal{H}$, then this probability is

$$p_W^A(a,b) = \sum_{a_i \in (a,b)} a_i \langle \psi | P_i \psi \rangle . \tag{2.12}$$

2.5 Measurements

Born's rule gives the probability for a specific measurement outcome. But often in quantum mechanics one also needs to know the **post measurement state** of the system at which the measurement has been performed. We recall here the basic rules which we are going to use later. We restrict ourselves to the case of a finite dimensional state Hilbert space $\mathcal{H}$.

For an orthonormal basis $\{e_j\}$ of $\mathcal{H}$ denote by $[e_j] = P_{e_j}$ the orthogonal projector onto the one dimensional subspace spanned by the vector e_j. Typically these basis vectors are the eigen-vectors of some self-adjoint operator in $\mathcal{H}$, i.e., of an observable. According to (2.12) the probability for the outcome j of the measurement is $\| [e_j]\psi \|^2 = \langle \psi | [e_j]\psi \rangle = |\langle e_j | \psi \rangle|^2$ if our system is in the pure

state $\psi \in \mathcal{H}$ and the state of the system after the measurement is

$$\frac{[e_j]\psi}{\langle \psi |\, [e_j]\psi \rangle} = \frac{\langle e_j |\, \psi \rangle}{\langle \psi |\, [e_j]\psi \rangle} e_j \,. \tag{2.13}$$

If our system is in a general state with density matrix W, then the post measurement state is

$$\frac{[e_j]W[e_j]}{\mathrm{Tr}([e_j]W[e_j])} \,. \tag{2.14}$$

2.6 Heisenberg's Uncertainty Principle

This principle states that in a quantum system only one property of a pair of *conjugate properties* can be known with certainty. Recall that conjugate properties are represented by self-adjoint operators which do *not commute*. Heisenberg formulated this principle originally for the position and momentum of a particle. The operators of position Q and momentum P satisfy the commutation relation

$$[Q, P] = QP - PQ \subset \mathrm{i}I \tag{2.15}$$

where I denotes as usual the identity operator and where the symbol $\subset$ expresses the fact that the above identity holds on a dense subspace of the Hilbert

space $\mathcal{H}$, not on all of $\mathcal{H}$, since (2.15) involves unbounded operators.

The mean values or *expected value* of an observable A in a state W is

$$E(A,W) = \mathrm{Tr}\,(AW) = \langle A \rangle_W.$$

Here and in the following we assume that the operator products are of trace class.

The *uncertainty* of an observable A in a state W then is defined as

$$\Delta_W(A) = \sqrt{\mathrm{Tr}(A^2W) - \langle A \rangle_W^2}.$$

Now using the observation that

$$(A,B) \to \mathrm{Tr}(A^*BW)$$

defines a positive semi-definite sesquiliear form for which the Cauchy-Schwarz inequality holds one shows Heisenberg uncertainty principle in general form

$$\frac{1}{2}|\mathrm{Tr}([A,B]W)| \leq \Delta_W(A)\Delta_W(B), \tag{2.16}$$

hence in the case of position and momentum one has for a pure state ψ

$$\frac{1}{2} \leq \Delta_\psi(Q)\Delta_\psi(P).$$

2.7 Composite systems and entanglement

Suppose that a quantum system Σ is composed of two subsystems Σ_i with respective Hilbert spaces $\mathcal{H}_i$, , $i = 1,2$. Then the Hilbert $\mathcal{H}$ of Σ is the (Hilbert) **tensor product** of the spaces $\mathcal{H}_1$ and $\mathcal{H}_2$:

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$$

Note that in this formula the symbol $\otimes$ denotes the completion of the algebraic tensor product of the vector spaces $\mathcal{H}_i$. If e_j (f_k) is an orthonormal basis of $\mathcal{H}_1$ ($\mathcal{H}_2$) then

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 = \left\{ \sum_{j,k=1}^{\infty} c_{jk} e_j \otimes f_k \,:\, c_{jk} \in \mathbb{C}, \sum_{j,k=1}^{\infty} |c_{jk}|^2 < \infty \right\}, \tag{2.17}$$

i.e., $\mathcal{H}_1 \otimes \mathcal{H}_2$ is the Hilbert space with $e_j \otimes f_k$, $j,k \in \mathbb{N}$, as an orthonormal basis. Thus elements $\psi \in \mathcal{H}$ are given by double series according to (2.17). However, according to a much used result, each element $\psi \in \mathcal{H}$ has also a representation by a single series.

Schmidt decomposition: For every $\psi \in \mathcal{H}$ there are non-negative numbers p_n

and orthonormal bases e'_n of $\mathcal{H}_1$ respectively f'_n of $\mathcal{H}_2$ such that

$$\psi = \sum_{n=1}^{\infty} p_n e'_n \otimes f'_n, \quad \sum_{n=1}^{\infty} p_n^2 = \|\psi\|^2 . \tag{2.18}$$

States $\psi \in \mathcal{H}$ are called **separable** iff there are $\psi_1 \in \mathcal{H}_1$ and $\psi_2 \in \mathcal{H}_2$ such that

$$\psi = \psi_1 \otimes \psi_2.$$

States $\psi \in \mathcal{H}$ are called inseparable or **entangled** iff they are not separable.

Entanglement of two quantum systems occurs when these systems (for instance photons, electrons, molecules) interact physically and then become separated; the type of interaction is such that each resulting member of a pair is properly described by the same quantum mechanical description (state), which is indefinite in terms of important factors such as position, momentum, spin, polarization, etc.

The concept of entanglement was suggested by E. Schrödinger in a reply to the EPR paradox, a thought experiment by which Einstein, Podolsky and Rosen claimed to prove that the quantum-mechanical description of physical reality given by wave functions is not complete. Schrödinger stated:

> I would not call [entanglement] **one** but rather **the** characteristic trait

of quantum mechanics, the one that enforces its entire departure from classical lines of thought.

Quantum systems can become entangled through various types of interactions (see section on methods below). If entangled, one object cannot be fully described without considering the other(s). They remain in a quantum superposition and share a single quantum state until a measurement is made.

For example entanglement occurs when subatomic particles decay into other particles. These decay events obey the various conservation laws, and as a result, pairs of particles can be generated so that they are in some specific quantum states. Thus, entanglement is an experimentally verified and accepted property of nature. Non-locality and hidden variables are two proposed mechanisms that enable the effects of entanglement. And, as we will learn later, entanglement is a (physical) resource, for instance for quantum teleportation and to superdense coding.

2.7.1 Measurements on entangled states

We conclude this chapter with an explicit demonstration of the **amazing consequences of entanglement**. Suppose that a two qubit system is in the (gen-

eral) state

$$\psi = \sum_{i,j=0,1} \alpha_{ij} |ij\rangle_{12}, \quad |ij\rangle_{12} = |i\rangle_1 \otimes |j\rangle_2, \quad \|\psi\|^2 = \sum_{i,j=0,1} |\alpha_{ij}|^2 = 1 \tag{2.19}$$

where the subscripts 1,2 refer to qubit 1 and qubit 2. On such a two qubit system various *measurements* can be performed.

1. Before any measurement the state of this system is uncertain.
2. After the measurement the state of the system is certain, it is $|00\rangle_{12}$, $|01\rangle_{12}$, $|10\rangle_{12}$, $|11\rangle_{12}$, with probability $|\alpha_{00}|^2$, $|\alpha_{01}|^2$, $|\alpha_{10}|^2$, or $|\alpha_{11}|^2$.
3. What conclusions can be drawn when we observe only the first (or only the second) qubit? We expect the system to be left in an uncertain state, because we did not measure the second qubit that can still be in a continuum of states.
4. The first qubit can be in the state
   - $|0\rangle_1$ with probability $|\alpha_{00}|^2 + |\alpha_{01}|^2$, or
   - $|1\rangle_1$ with probability $|\alpha_{10}|^2 + |\alpha_{11}|^2$.

5. Denote by $\left|\psi_0^I\right\rangle$ ($\left|\psi_1^I\right\rangle$) the post-measurement state when we measure the first qubit and find it to be in state $|0\rangle_1$ ($|1\rangle_1$). According to (2.13) these states are

$$\left|\psi_0^I\right\rangle = \frac{\alpha_{00}|00\rangle_{12} + \alpha_{01}|01\rangle_{12}}{\sqrt{|\alpha_{00}|^2 + |\alpha_{01}|^2}}, \quad \left|\psi_1^I\right\rangle = \frac{\alpha_{10}|10\rangle_{12} + \alpha_{11}|11\rangle_{12}}{\sqrt{|\alpha_{10}|^2 + |\alpha_{11}|^2}} \tag{2.20}$$

6. Now consider the case of fully entangled states, for instance the Bell state (see (5.5)) $|\phi^+\rangle_{12}$ which is the special case of (2.19) with $\alpha_{00} = \alpha_{11} = 1/\sqrt{2}$ and $\alpha_{01} = \alpha_{10} = 0$, i.e.,

$$\psi = \frac{1}{\sqrt{2}}(|00_{12}\rangle + |11\rangle_{12}) = \left|\phi^+\right\rangle_{12}.$$

   When in this Bell state we measure the *first* qubit we get the post measurement states

$$\left|\psi_0^I\right\rangle = |00\rangle_{12}, \quad \left|\psi_1^I\right\rangle = |11\rangle_{12}.$$

7. When in this Bell state we measure the *second* qubit we get in a similar way the post measurement states

$$\left|\psi_0^I\right\rangle = |00\rangle_{12}, \quad \left|\psi_1^I\right\rangle = |11\rangle_{12}.$$

Important conclusions:

1. The two measurements mentioned above are correlated; once we measure the first qubit we get the same result as when we measure the second qubit.
2. This result is quite astonishing since the two qubits need not be physically constrained to be at the same location (they could be far apart) and yet, because of the strong coupling between them, measurement on the first qubit allow us to determine the state of the second.
3. This effect is known since the early days of quantum mechanics, and in particular A. Einstein was quite unhappy with this **spooky action at a distance**.

Chapter 3

Qubits and Quantum Circuits

In our lecture we assume that you have some background in classical computers. In the 1930s C. Shannon studied switching circuits and observed that one could apply the rules of **Boole's algebra** in this setting and he introduced the concept **switching algebra** as a way to analyze and design circuits by algebraic means in terms of **logic gates**.

Boolean algebra deals with the values 0 and 1 which can be thought of as two integers, or as the truth values *false* and *true* respectively. They are called **bits** or binary digits in contrast to the decimal digits $0, 1, \ldots, 9$.

A **logic gate** is a device implementing a Boolean function, i.e., it performs a logical operation on one or more logical inputs and produces a single logic

output. Such gates are primarily implemented using diodes or transistors as electronic switches. Logic gates can be put together to form compound logic gates or **logic circuits**, for instance in a present day computer.

In a classical computer the only reversible logic gate is the NOT gate. An n-bit datum is a string of bits $x_1, x_2, \ldots, x_n$ of length n. They are stored in an n-bit register. The set of n-bit data is the space $\{0,1\}^n$ which consists of 2^n strings of 0's and 1's. Thus we can formulate

> An n-bit reversible gate is a bijective mapping f from the set $\{0,1\}^n$ onto itself.

3.1 Bits and Qubits

As a bit is the basic unit of classical computation and classical information a **qubit** or **quantum bit** is the basic unit of quantum computing and quantum information, realized as a two-state quantum mechanical system. The two states in which a qubit may be measured are called basis states (or vectors) and traditionally are denoted as $|0\rangle$ and $|1\rangle$ (computational basis states).

The process of quantization in this context, i.e., the transition from bits to

qubits, means the following:

$$0 \longrightarrow |0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \in \mathbb{C}^2, \quad 1 \longrightarrow |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \in \mathbb{C}^2, \tag{3.1}$$

The decisive difference is that a bit must be either 0 or 1 a qubit can be either $|0\rangle$ or $|1\rangle$ or a combination

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \quad \alpha,\beta \in \mathbb{C},\ |\alpha|^2 + |\beta|^2 = 1 \tag{3.2}$$

of both, as a consequence of the **superposition principle** of quantum mechanics. Accordingly a qubit is a unit vector in a Hilbert space which is isomorphic to $\mathbb{C}^2$.

Here we use Dirac's **bra** and **ket** notation; thus we also write

$$|\psi\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \quad \langle\psi| = (\alpha^*,\beta^*) = \alpha^*\langle 0| + \beta^*\langle 1|$$

and for $|\psi_j\rangle = \alpha_j|0\rangle + \beta_j|1\rangle$

$$\langle\psi_1|\,\psi_2\rangle = \alpha_1^*\alpha_2 + \beta_1^*\beta_2, \quad |\psi_1\rangle\langle\psi_2| = \begin{pmatrix} \alpha_1\alpha_2^* & \alpha_1\beta_2^* \\ \alpha_2\alpha_2^* & \alpha_2\beta_2^* \end{pmatrix}, \quad \mathrm{Tr}(|\psi_1\rangle\langle\psi_2|) = \langle\psi_2|\,\psi_1\rangle\ .$$

Note that any two-level quantum system can be used as a qubit. Here is a list of some of the physical implementations of qubits:

Physical system	**system property**	$\|0\rangle$	$\|1\rangle$
photon	polarization of light	horizontal	vertical
coherent state of light	squeezed light	amplitude-squeezed	phase-squeezed
electrons	electronic spin	spin up	spin down
nucleus	nuclear spin addressed through NMR	spin up	spin down
optical lattices	atomic spin	spin up	spin down

Josephson junction	superconducting charge	uncharged superconducting island, $Q = 0$	charged superconducting island, $Q = 2e$, one extra Cooper pair
singly charged quantum dot pair	electron localization	electron on left dot	electron on right dot

As we know the Hilbert space of a composite system is the tensor product of the Hilbert spaces of its component. Accordingly the Hilbert space of a two qubit system is a space isomorphic to $\mathbb{C}^2 \otimes \mathbb{C}^2$. The basis vectors of this space are

$$|0\rangle \otimes |0\rangle, |0\rangle \otimes |1\rangle, |1\rangle \otimes |0\rangle, |1\rangle \otimes |1\rangle.$$

Usually one writes for $i, j \in \{0,1\}$

$$|i,j\rangle = |i\rangle \otimes |j\rangle . \tag{3.3}$$

Recall that composite quantum systems can be **entangled**. In particular two qubits can be entangled (in contrast to bits). An example of a state for two entangled qubits is

$$\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle).$$

Such a state is an equal superposition since the probabilities for measuring either $|00\rangle$ or $|11\rangle$ are equal, namely $1/2$.

Suppose now that the two entangled qubits are (spatially) separated, one to a location A, the other to a distant location B. If now a measurement of the one qubit is made in A and the result $|0\rangle$ is found (or with equal probabilty $|1\rangle$) a subsequent measurement of the other qubit at B will give the result $|0\rangle$ since $|00\rangle$ is the only state where the qubit in A is $|0\rangle$ (or the result $|1\rangle$ since $|11\rangle$ is the only state where the qubit in B is $|1\rangle$).

This observation about entangled qubits is the core of the quantum teleportation protocol (see later). Entanglement is also the basis of quantum computation and quantum information in general.

3.2 Quantum gates

There are two basic operations on pure qubit states.

- A **quantum logic gate** operates on a qubit $|\psi\rangle$ and produces another qubit $|\psi\rangle'$. Mathematically this is realized by a unitary transformation of the state space.
- Another operation on qubits is a standard basis measurement. The result of such a measurement on (3.2) will be either $|0\rangle$ with probability $|\alpha|^2$ or $|1\rangle$ with probability $|\beta|^2$.

Quantum logic gates are reversible, unlike many classical logic gates. However classical computing can be done using only reversible gates, since it is known that the reversible **Toffoli gate** (or **CCNOT gate**) can implement all Boolean functions. This controlled-controlled-not gate has a direct quantum equivalent, hence **quantum circuits** which are built out of quantum gates can perform all operations performed by classical logic circuits.

As indicated above quantum gates are represented by unitary matrices on the corresponding Hilbert space. The Hilbert space for one qubit is $\mathbb{C}^2$, as the quantized version of the one bit space $\{0,1\}$. More generally the **quantized**

version of the classical n-bit space $\{0,1\}^n$ is the space

$$\mathcal{H}_{nq} = \mathbb{C}^{\{0,1\}^n} = \mathbb{C}^{2^n} \tag{3.4}$$

of all functions on $\{0,1\}^n$ with values in the complex numbers $\mathbb{C}$. Elements of this space are called n-qubits and are written as $|x_1, x_2, \cdots, x_n\rangle$ when $x_1, x_2, \cdots, x_n$ is a classical n-bit string. Thus $|x_1, x_2, \cdots, x_n\rangle$ is the function which maps the classical bit $x_1, x_2, \cdots, x_n$ to 1 and all other n-bits to 0. These are 2^n special n-qubits, called **computational basis states**.

Accordingly an n-qubit (reversible) quantum gate is a unitary mapping on

$$\mathcal{H}_{nq} = \mathbb{C}^{2^n}.$$

There are a number basic quantum gates which are commonly used and which we describe now:

3.2.1 Hadamard gate H

This gate acts on a single qubit and maps the basis state $|0\rangle$ to the state $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and the basis state $|1\rangle$ $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$; hence the Hadamard gate is represented by the **Hadamard matrix**

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad (3.5)$$

circuit representation

3.2.2 Pauli-X gate

This gate also acts on a single qubit and is the quantum equivalent of the NOT gate. Its action is $|0\rangle \longrightarrow |1\rangle$ and $|1\rangle \longrightarrow |0\rangle$; hence its matrix representation is the Pauli X matrix:

$$X = \sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad (3.6)$$

3.2.3 Pauli-Y gate

Its action is $|0\rangle \longrightarrow \mathrm{i}|1\rangle$ and $|1\rangle \longrightarrow -\mathrm{i}|0\rangle$; hence its matrix representation is the Pauli Y matrix:

$$Y = \sigma_y = \begin{bmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{bmatrix} \quad (3.7)$$

3.2.4 Pauli-Z gate

Its action is $|0\rangle \longrightarrow |0\rangle$ and $|1\rangle \longrightarrow -|1\rangle$; hence its matrix representation is the Pauli Z matrix:

$$Z = \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{3.8}$$

Note that the matrices I_2, X, Y, Z are a basis in the space $M_2(\mathbb{C})$ of 2×2 matrices with complex entries. I_2 denotes the 2×2 unit matrix.

3.2.5 Phase shift gates

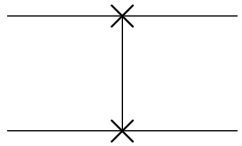
These gates leave the basis state $|0\rangle$ fixed while the basis state $|1\rangle$ is mapped to $e^{i\theta}|1\rangle$. Hence the probability of measuring $|0\rangle$ or $|1\rangle$ is unchanged after an application of this gate. Its matrix representation is

$$\begin{bmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{bmatrix} \quad \theta \text{ phase shift} \tag{3.9}$$

3.2.6 Swap gate

This gate swaps two qubits; hence its matrix representation is

$$SWAP = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.10)$$

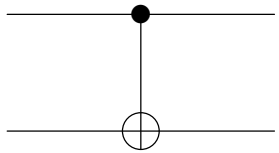


circuit representation

3.2.7 Controlled gates

These are gates which operate on two or more qubits and where one (or more) qubits act as a control. The **controlled NOT** gate or **CNOT** operates on two qubits and it performs the NOT operation on the second qubit only when the first qubit is $|1\rangle$, otherwise the second qubit is left unchanged. The matrix and circuit representations are

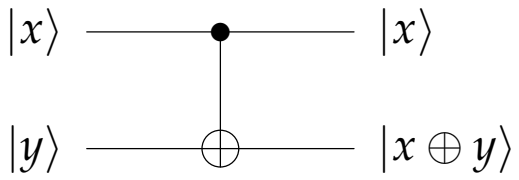
$$CNOT = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (3.11)$$



CNOT circuit

As a 2 qubit gate the CNOT gate acts on two incoming states and produces a

two state output:



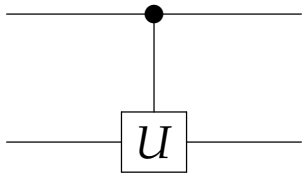
where $\oplus$ denotes addition of bits, i.e., addition modulo 2.

A straightforward generalization of the CNOT gate is the **controlled-U gate**. Here the first qubit serves as a control in the following way (recall (3.3)):

$$
\begin{aligned}
|00\rangle &\longrightarrow |00\rangle \\
|01\rangle &\longrightarrow |01\rangle \\
|10\rangle &\longrightarrow |1\rangle U|0\rangle = |1\rangle(x_{00}|0\rangle + x_{10}|1\rangle) \\
|11\rangle &\longrightarrow |1\rangle U|1\rangle = |1\rangle(x_{01}|0\rangle + x_{11}|1\rangle) \\
&\quad .
\end{aligned}
$$

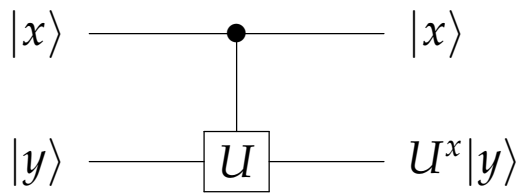
Accordingly its matrix and circuit representations are

$$C(U) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & x_{00} & x_{01} \\ 0 & 0 & x_{10} & x_{11} \end{bmatrix} \quad (3.12)$$



controlled -U gate

The circuit representation on two incoming states $|x\rangle, |y\rangle$ thus is



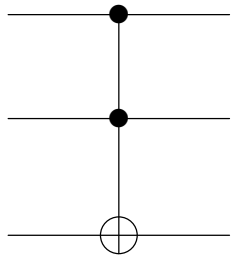
The CNOT gate is the special case $U^x = X$.

3.2.8 Toffoli gate

Recall that the Toffoli or **CCNOT** gate is universal for classical computation. It is a 3-bit gate. The quantum Toffoli gate is the same but is defined for 3 qubits. It is the gate which maps $|x_1, x_2, x_3\rangle$ to $|x_1, x_2, x_3 + x_1 x_2\rangle$. Thus, if the first two qubits are in the state $|1\rangle$ it applies the Pauli-X gate to the third qubit. In all

other conditions it does nothing. Accordingly its matrix and circuit representations are

$$CCNOT = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad (3.13)$$

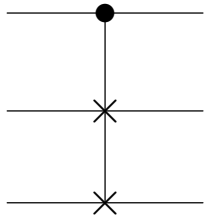


CCNOT gate

3.2.9 Fredkin gate

As the Toffoli gate the Fredkin or **CSWAP** gate is universal for classical computation. Its quantum analogue acts on 3 qubits. Its matrix and circuit representations are as follows:

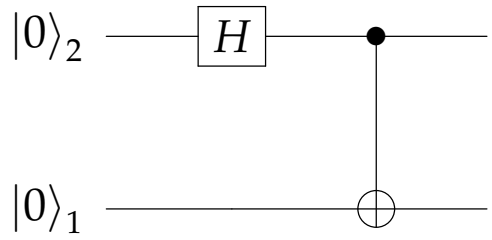
$$CSWAP = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.14)$$



CSWAP gate

3.2.10 Example of an entangling circuit

By means of a simple example we illustrate how these quantum circuits help to visualize certain operations on qubits. Which operations does the quantum circuit below indicate?



Recall the action of the Hadamard gate H

$$H : |0\rangle \longrightarrow (|0\rangle + |1\rangle)/\sqrt{2}, \quad H : |1\rangle \longrightarrow (|0\rangle - |1\rangle)/\sqrt{2}$$

and that of the $CNOT$ gate ($j = 0, 1$):

$$CNOT(|0\rangle_2 \otimes |j\rangle_1) = |0\rangle_2 \otimes |j\rangle_1, \quad CNOT(|1\rangle_2 \otimes |j\rangle_1) = |0\rangle_2 \otimes |j'\rangle_1,$$

where $j' = 1$ for $j = 0$ and $j' = 0$ for $j = 1$. Thus this circuits represents the following calculations:

$$\begin{aligned} CNOT(H \otimes I_2(|0\rangle_2 \otimes |0\rangle_1)) &= CNOT(|0\rangle_2 \otimes |0\rangle_1 + |1\rangle_2 \otimes |0\rangle_1)/\sqrt{2} \\ &= \frac{1}{\sqrt{2}}(|0\rangle_2 \otimes |0\rangle_1 + |1\rangle_2 \otimes |1\rangle_1) \end{aligned} \tag{3.15}$$

Thus the above circuit transforms the separable state $|0\rangle_2 \otimes |0\rangle_1$ into the fully entangled state (3.15).

Chapter 4

Quantum Information Theory

4.1 Classical Information Theory

Classical information theory as developed by C. Shannon and his successors addresses the following type of questions and suggests solutions.

Suppose a certain message is to be transmitted from a location A to a location B.

1. What resources are need for this transmission?

2. If we have a transmission channel with capacity of c bits per second, how long will it take?

3. If the transmission introduces errors (e.g. noise) what can be done about this?

Recall that a string of n bits can represent $N = 2^n$ messages. Thus, for the transmission of one of N distinct messages it is sensible to define the amount of information carried by a single message to be

$$\log_2 N \quad \text{bits.}$$

But typically one has to transmit a message from a given collection of N messages *repeatedly*, and if the messages can be assigned a non-uniform *probability distribution*, then, on average it is possible to use fewer than $\log N$ bits per second in order to transmit or store them.

Data compression (for instance 'gzip') is a known method to store messages efficiently. The key observation is to encode more common messages by using short strings of bits while less common message are encoded by longer strings.

The **Shannon entropy** is a logarithmic information measure, typically denoted $H(X)$ if X is a collection of labels x for a set of N messages. Suppose that $p(x)$ is the probability for message x, with $\sum_x p(x) = 1$. Or X is a random variable, i.e., a numerical function on some sample space; each x may correspond to several points in the sample space and $p(x)$ is the sum of the

probabilities associated with these points. Then one defines

$$H(X) = H(p) = -\sum_x p(x) \log_2 p(x). \tag{4.1}$$

Basic properties:

- $H(X) \geq 0$ and $H(X) = 0$ iff there is some x_0 such that $p(x_0) = 1$ and $p(x) = 0$ for $x \neq x_0$.
- If x can take only k values then $H(X) \leq \log k$ and $H(X) = \log k$ iff $p(x) = 1/k$ for all x.

An intuitive interpretation of $H(X)$ is the amount of information *on average* conveyed by an observation of x. If x is a message, then $H(X)$ is the *average missing information* about the message before it is received, and thus the average information conveyed by the message, since after a message is received (and read), the missing information about this particular message is 0.

One can also think of $H(X)$ as the difference, on average, of the information possessed by someone who knows what the actual message is, over against someone who knows the probability distribution but does not know the message.

Next, send a large number M of messages, all from the same collection X, one after the other. One expects that the total information conveyed by all messages is $MH(X)$.

$MH(X)$ can also be interpreted as the *minimum number of bits required to transmit M messages* when M is large.

4.1.1 The case of two random variables

In this case some new aspects emerge which we discuss briefly in the simplest setting. Suppose that we are given two random variables X and Y, each with a finite number of discrete values. Suppose furthermore that X is sent from a location A to a location B through a noisy channel without memory so that the output is Y. Then the output is related to the input by *conditional probabilities*: Given an input x, the probability that y emerges is $p(y|x)$. These conditional probabilities $p(y|x)$ are characteristics of the given channel.

Basic properties of these conditional probabilities are:

$$p(y|x) \geq 0, \quad \sum_y p(y|x) = 1 \,. \tag{4.2}$$

Furthermore, the probability $p(x)$ that a message x is sent into the channel is

determined by the ensemble X. Once $p(x)$ is given, the *joint probability* $p(x,y)$ that x enters the channel and y emerges, and the (marginal) probability $p(y)$ that y emerges are given by

$$p(x,y) = p(y|x)p(x), \quad p(y) = \sum_x p(x,y)\,. \tag{4.3}$$

Given these probabilities we can define the various information entropies according to (4.1):

$$H(X) = H(p(x)), \quad H(Y) = H(p(y)), \quad H(X,Y) = H(p(x,y)), \tag{4.4}$$

thus in particular

$$H(X,Y) = -\sum_{x,y} p(x,y)\log(p(x,y))\,. \tag{4.5}$$

4.1.2 Conditional Entropies and mutual Information

The *conditional entropies* $H(Y|x)$ and $H(Y|X)$ are defined by

$$H(Y|x) = -\sum_y p(y|x)\log p(y|x), \quad H(Y|X) = \sum_x p(x)H(Y|x) \tag{4.6}$$

and similarly for $H(X|y)$ and $H(X|Y)$

$$H(X|y) = -\sum_y p(x|y)\log p(x|y), \quad H(X|Y) = \sum_y p(y)H(X|y)$$

where $p(x|y) = p(x,y)/p(y)$. Alternatively one can write

$$H(Y|X) = H(X,Y) - H(X), \quad H(X|Y) = H(X,Y) - H(Y)\,. \tag{4.7}$$

In the context of these formulae it is assumed that both at location A (Alice) and B (Bob) the joint probability distribution $p(x,y)$ and hence all marginals and conditionals are known. What they do not know until they see it is what actually occurs in a particular case.

Interpretation of (4.6): When Alice puts a message x into the channel, she cannot be sure what y will emerge, since the channel is noisy. Then on average her ignorance about y is given by $H(Y|x)$. Avering this over all possible input messages x gives the overall average $H(Y|X)$ of information which Alice lacks about outputs when she knows the inputs.

By interchanging the roles of Bob and Alice a similar interpretation results for $H(X|y)$ and $H(X|Y)$.

Interpretation of (4.7): The above H quantities can be thought of as missing

information. Before Alice knows what x will go though the channel, her (average) ignorance about both x and y is measured by $H(X,Y)$. When message x actually appears and she sees it, her ignorance is reduced on average by $H(X)$, thus $H(X,Y) - H(X)$ is the information about the pair (x,y) that she is still lacking, and which, since she knows x, is missing information about y.

The *mutual information*

$$I(X:Y) = H(Y) - H(Y|X) = H(X) - H(X|Y) = H(X) + H(Y) - H(X,Y) \tag{4.8}$$

is the average amount of information which Alice, knowing x, has about the output y resulting from this x. It is Alice's (average) ignorance about y before knowing x, minus her ignorance about y when she knows x, and therefore the amount of information that she learns about y on average from observing x.

The second and third version in (4.8) follow from (4.7). The third version actually shows that the mutual information is symmetrical:

$$I(X:Y) = I(Y:X)\,.$$

Hence, the average amount which Bob learns about x by observing y is the same as the average amount which Alice knows about y when sending x. This important symmetry is intuitively not obvious.

By tracing the various definitions $I(X:Y)$ can also be expressed directly in terms of probabilities as

$$I(X:Y) = -\sum_{x,y} p(x,y) \log \frac{p(x)p(y)}{p(x,y)} \,. \tag{4.9}$$

This formula shows that $I(X:Y)$ is a measure of correlation in the sense of statistical independence, namely one can show:

a) $I(X:Y) \geq 0$;

b) $I(X:Y) = 0$ iff X and Y are statistically independent, i.e., iff $p(x,y) = p(x)p(y)$.

Remark 4.1.1 *The definitions and motivations for the various entropies and the mutual information have been given with reference to the transmission of information through a channel. It is important to realize that the same definitions can be made for two random variables X and Y which have a joint probability distribution $p(x,y)$, since then one can deduce the marginals $p(x)$ and $p(y)$ and the above formulae can be used.*

4.1.3 Channel capacity

The above interpretations also show that $I(X:Y)$ can be identified with the *average rate* at which information is being transmitted through the given channel; hence, if the channel is used M times (with M large) the information passing through it is about $MI(X:Y)$ bits.

In our discussion we had mentioned that the conditional probability $p(y|x)$ is a characteristic of the channel. According to (4.9) the mutual information $I(X:Y)$ depends on $p(y|x)$ and $p(x)$. Denote by $Q_{p(y|x)}$ all probabilities $p(x)$ which are compatible with $p(y|x)$ and define the *channel capacity* C by

$$C = \sup_{Q_{p(y|x)}} I(X:Y)\,. \tag{4.10}$$

The capacity C of a channel is the maximum possible rate at which information can be reliably (using appropriate error correction) transmitted through a noisy channel , measured in bits of information per uses of the channel. Here it is assumed that a "memoryless channel" is used, i.e., $p(y|x)$ is the same every time the channel is used, independently of what was previously sent through the channel.

4.2 Quantum Information Theory

Naturally, quantum information theory is to be the quantum analogue of classical information theory. Accordingly the quantum counter parts of classical information theory have to be defined, i.e., quantum information, quantum channels, measures of quantum information.

4.2.1 Quantum samples

The basis of classical probability theory and thus of classical information theory is the *sample space*. Accordingly we begin by defining the *quantum sample space* as a *decomposition of the identity* on the Hilbert space $\mathcal{H}$ of our quantum system:

$$\sum_j P_j = I = \mathrm{id}_{\mathcal{H}}, \quad P_j P_k = \delta_{jk} P_j \tag{4.11}$$

where the P_j are orthogonal projectors on $\mathcal{H}$ (i.e., $P_j^* = P_j = P_j^2$), different from 0. Such a decomposition represents a collection of mutually-exclusive properties or “events", one and only one of which is “true" or “occurs". The corresponding *event algebra* contains all projectors which can be written as a sum of these, and in addition the zero projector and the identity I.

To such a quantum sample $\{P_j\}$ one assigns *probabilities* $\{p_j\}$. Typically these probabilities are generated through the use of the Born rule, see (2.11) and (2.12). Recall that the probability that a quantum observable has a particular value is equal to the probability assigned to the projector onto the eigenspace of this eigen-value. Thus this projector must be part of some decomposition of the identity to which one has managed to assign probabilities. This is often referred to as the probability that this observable will have this particular value “if measured".

4.2.2 Compatible and incompatible quantum samples

Two quantum samples $\{P_j\}$ and $\{Q_k\}$ are *compatible* iff for every j,k

$$P_j Q_k = Q_k P_j \,,$$

otherwise they are *incompatible*.

The commutativity of the projection operators from one quantum sample or two compatible quantum samples implies that all results of classical (Shannon) information theory carry over to the quantum domain, as expected.

Incompatible quantum samples must not be combined.

4.2.3 Mutually-unbiased quantum samples

Two distinct quantum samples for a single qubit are always incompatible (think of the eigen-projections onto the eigen-spaces of the spin matrices σ_z and σ_x). But the degree of incompatibility may differ. The following definition expresses this in quantitative terms.

Observe that given an orthonormal basis $\{|e_j\rangle\}$ of a Hilbert space $\mathcal{H}$ of dimension m the family $\{[e_j]\}$ of orthogonal projectors onto the subspaces spanned by the e_j form a decomposition of the identity and thus a quantum sample.

Two orthonormal bases $\{|e_j\rangle\}$ and $\{|\bar{e}_j\rangle\}$ of $\mathcal{H}$ are called *mutually unbiased* iff for every j and every k

$$|\langle e_j|\,\bar{e}_k\rangle| = \frac{1}{\sqrt{m}}, \quad \text{or} \quad \mathrm{Tr}([e_j][\bar{e}_k]) = \frac{1}{m}\,. \tag{4.12}$$

4.2.4 Quantum channels

Recall that a classical channel (see figure below) is characterized by the conditional probability $p(y|x)$ that a y emerges when a message x has been put into the channel.

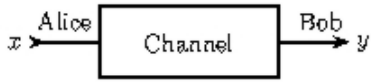


Figure: A channel transmitting x to y

An *ideal* or *perfect* or *noise-free* classical channel is one for which the output is an exact reproduction of the input, i.e., $p(y|x) = \delta_{xy}$.

A *perfect quantum channel* is a channel which does not change the internal state of a particle going through it. Thus for instance when a spin half particle enters such a channel with $S_z = +1/2$ it will emerge with $S_z = +1/2$, if it enters in the state $S_y = -1/2$ it exits with $S_y = -1/2$.

In the context of information theory we are only interested in the internal state of a particle, as information this particles carries.

A channel in which the internal state of the particle that enters and the particle that emerges can be described using a two-dimensional Hilbert space is a *one qubit channel*. Such a channel could be perfect or noisy.

Suppose that a particle enters a channel in the internal quantum state $|\psi\rangle$ and emerges in a state $U|\psi\rangle$, where U is a unitary operator independent of $|\psi\rangle$. Such a channel is called an *ideal* quantum channel. Consider the example of a qubit channel for which the unitary operator U equals the Hadamard gate H (see(3.5)) . If a particle enters such a channel in the state $S_z = +1/2$, it will

emerge in the state $S_z = +1/2$; if it enters in the state $S_y = -1/2$, it will emerge in the state $S_y = +1/2$.

A quantum channel which transmits one type of quantum information, i.e., one quantum sample corresponding to a particular orthonormal basis of the state space, whereas all mutually unbiased quantum samples are turned into pure noise (that is no information is transmitted), is a *perfect classical* channel or *perfectly decohering* channel. If a quantum channel transmits *all* types of quantum information perfectly, it is a perfect quantum channel.

Remark 4.2.1 *One can show: If a quantum channel perfectly transmits two mutually unbiased quantum samples then it perfectly transmits all other quantum samples as well.*

4.2.5 von Neumann entropy

The counterpart of Shannon's entropy (4.1) has to provide a measure which quantifies "missing information" for any specific quantum sample. This can be done as follows: Given a quantum sample (4.11) assign a probability distribution $p = (p_1, p_2, \ldots)$ to it and calculate the Shannon entropy (4.1) for this distribution.

A very useful entropy measure which is independent of the quantum information type is the *von Neumann entropy* which is defined for general density matrices ρ by

$$S(\rho) = -\mathrm{Tr}(\rho \log \rho) \tag{4.13}$$

using spectral calculus (2.5). Since every density matrix ρ has the spectral representation

$$\rho = \sum_j \rho_j [e_j]$$

with eigen-values $\rho_j > 0$ and eigen-vectors e_j one easily finds

$$S(\rho) = -\sum_j \rho_j \log \rho_j, \quad \rho = \sum_j \rho_j [e_j] \,. \tag{4.14}$$

Note the basic properties of the von Neumann entropy:

- $S(\rho) \geq 0$ for every density matrix ρ and $S(\rho) = 0$ iff ρ is pure, i.e., of the form $\rho = |\psi\rangle\langle\psi|$ for a unit vector $|\psi\rangle$.
- For any invertible matrix U and any density matrix ρ one has $S(U\rho U^{-1}) = S(\rho)$, i.e., the von Neumann entropy is invariant under similarity transformations.

Often a density matrix ρ is considered a “pre-probability" and one can show that the von Neumann entropy $S(\rho)$ represents the *minimum* missing information associated with a pre-probability ρ.

Chapter 5

Dense Coding, no Cloning and Teleportation

5.1 Fully entangled states and local unitaries

Recall that we mentioned earlier that entanglement is peculiar to quantum mechanics and in quantum information theory entanglement is used as an important resource. The first two cases we discuss are dense coding and teleportation. Thus we collect here some basic facts about a special class of entangled states.

Suppose that $\mathcal{H}_a$ and $\mathcal{H}_b$ are two Hilbert spaces of dimension d. Recall that

any state $|\psi\rangle \in \mathcal{H} = \mathcal{H}_a \otimes \mathcal{H}_b$ has a Schmidt representation

$$|\psi\rangle = \sum_{j=1}^{d} \lambda_j \big|a_j\big\rangle \otimes \big|b_j\big\rangle \tag{5.1}$$

with suitable orthonormal bases $\big|a_j\big\rangle$ respectively $\big|b_j\big\rangle$ of $\mathcal{H}_a$ respectively of $\mathcal{H}_b$. Such a state is called *fully entangled* or *maximally entangled* iff all coefficients are equal, i.e., $\lambda_j = 1/\sqrt{d}$ for all j in the case that $|\psi\rangle$ is of norm 1.

Suppose that $|\psi\rangle \in \mathcal{H}$ is normalized. The reduced density operators on $\mathcal{H}_a$ respectively $\mathcal{H}_b$ are defined

$$\rho_a = \mathrm{Tr}_b([\psi]), \quad \rho_b = \mathrm{Tr}_a([\psi]) \tag{5.2}$$

where Tr_b denotes the partial trace of the density operator $[\psi] = |\psi\rangle\langle\psi|$ on $\mathcal{H}_a \otimes \mathcal{H}_b$ with respect to the Hilbert space $\mathcal{H}_b$ and similarly for Tr_a. An easy calculation then shows that $|\psi\rangle$ is fully entangled iff

$$\rho_a = \frac{1}{d} I_a \quad \text{and} \quad \rho_b = \frac{1}{d} I_b \tag{5.3}$$

where I_a (I_b) denoted the identity operator on $\mathcal{H}_a$ ($\mathcal{H}_b$).

Since the transition from one orthonormal basis to another is effected by a unitary matrix another easy calculation shows: If $|\psi\rangle$ and $|\phi\rangle$ are both normal-

ized fully entangled states on $\mathcal{H}_a \otimes \mathcal{H}_b$ there a unitary operators U_a on $\mathcal{H}_a$ and U_b on $\mathcal{H}_b$ such that

$$|\phi\rangle = (U_a \otimes U_b)|\psi\rangle \,. \tag{5.4}$$

Since the subsystem A with the Hilbert space $\mathcal{H}_a$ and the subsystem B with the Hilbert space $\mathcal{H}_b$ are often located in two separate laboratories one refers to the unitary operators U_a and U_b in (5.4) as *local unitaries*. And we can say for instance that a local operation in A can change a fully entangled state $|\psi\rangle$ also in the distant location B!

For two qubits one has $d = 2$ and the states

$$\begin{aligned} \left|\phi^+\right\rangle = |B_0\rangle &= (|00\rangle + |11\rangle)/\sqrt{2} \\ \left|\psi^+\right\rangle = |B_1\rangle &= (|01\rangle + |10\rangle)/\sqrt{2} \\ \left|\phi^-\right\rangle = |B_2\rangle &= (|00\rangle - |11\rangle)/\sqrt{2} \\ \left|\psi^-\right\rangle = |B_3\rangle &= (|01\rangle - |10\rangle)/\sqrt{2} \end{aligned} \tag{5.5}$$

are fully entangled and are easily seen to be an orthonormal basis of $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$. These states are called *Bell states* .

Recall the action of the Pauli gates X and Z introduced in (3.6) and (3.7):

$$X: \ |0\rangle \longrightarrow |1\rangle \,, |1\rangle \longrightarrow |0\rangle; \quad Z: \ |0\rangle \longrightarrow |0\rangle \,, |1\rangle \longrightarrow -|1\rangle \,.$$

Now a straight forward calculation shows that the Bell states B_1, B_2, B_3 can be obtained from the Bell state B_0 by applying local unitaries:

$$\begin{aligned} |B_1\rangle &= (X \otimes I_b)|B_0\rangle \\ |B_2\rangle &= (Z \otimes I_b)|B_0\rangle \\ |B_3\rangle &= (ZX \otimes I_b)|B_0\rangle \end{aligned} \tag{5.6}$$

5.2 Dense Coding

Dense coding (or often also called super dense coding) and teleportation are two processes which can be considered as the starting point of modern quantum information theory. Both demonstrated completely new features of quantum information as opposed to classical information and both are based on the use of (fully) entangled states. The original papers 1 and 7 provided explicit examples using qubits. Later many extensions of their schemes were found. An early systematic approach is proposed in 16. There it is shown in particular that each of the published teleportation schemes also works as a dense coding scheme, and conversely (for tight schemes): Sender (Alice) and receiver (Bob) merely have to swap the equipment they use.

We are going to explain the basic ideas for dense coding and teleportation only for the simplest cases of qubits.

The essence of dense coding is: Suppose A and B have a *quantum* channel over which A can send qubits to B. One way to send her message is to encode 0 as $|0\rangle$ and 1 as $|1\rangle$.

If A and B share a Bell state, then A can send two classical bits of information using only one qubit.

The details are as follows: Suppose A and B share the Bell state $|\phi^+\rangle$, see (5.5). Depending on the message Alice wants to send, she applies a suitable gate to her qubit and then sends it to Bob. If Alice wants to send

$$\begin{bmatrix} 00 \\ 01 \\ 10 \\ 11 \end{bmatrix} \text{ she applies } \begin{bmatrix} I_2 \otimes I_2 \\ Z \otimes I_2 \\ X \otimes I_2 \\ XZ \otimes I_2 \end{bmatrix} \text{ to } |\phi^+\rangle \text{ and gets } \begin{bmatrix} |\phi^+\rangle \\ |\phi^-\rangle \\ |\psi^+\rangle \\ |\psi^-\rangle \end{bmatrix}$$

where we used (5.6). After receiving the message from Alice, Bob has one of the four mutually orthogonal Bell states. When he applies a measurement he can distinguish between them with certainty and thus can determine Alice's message.

Note that Alice used two qubits in total so send two classical bits, since Alice and Bob started with a shared Bell state. However, the first qubit, i.e., Bob's half of the Bell state, could have been sent well before Alice decided what message she wanted to send. Only after she had decided on her message, she sent the second qubit.

And one can show that one cannot do better. Two qubits are necessary to send two classical bits. Dense coding allows half the qubits to be sent before the message has been chosen.

5.3 Teleportation

In the process of **quantum teleportation** or **entanglement assisted teleportation** the state of a qubit is replaced by that of another. The state is "transmitted" by setting up an entangled state-space of three qubits and then removing two qubits from the entanglement (via measurement). Since the information of the source qubit is preserved by these measurements that "information" (i.e. state) ends up in the final third, destination qubit. This occurs without the source and destination qubit ever directly interacting. The interaction occurs via entanglement.

It is unrelated to the popular science fiction term of teleportation.

The idea of teleporting qubits from one location A to a distant location B was first suggested in 1993 in the seminal paper (7). Since then various extensions and generalizations (using photons or atoms) have been suggested, also experimentally.

Distances of more than 100 km have been used in quantum teleportation experiments.

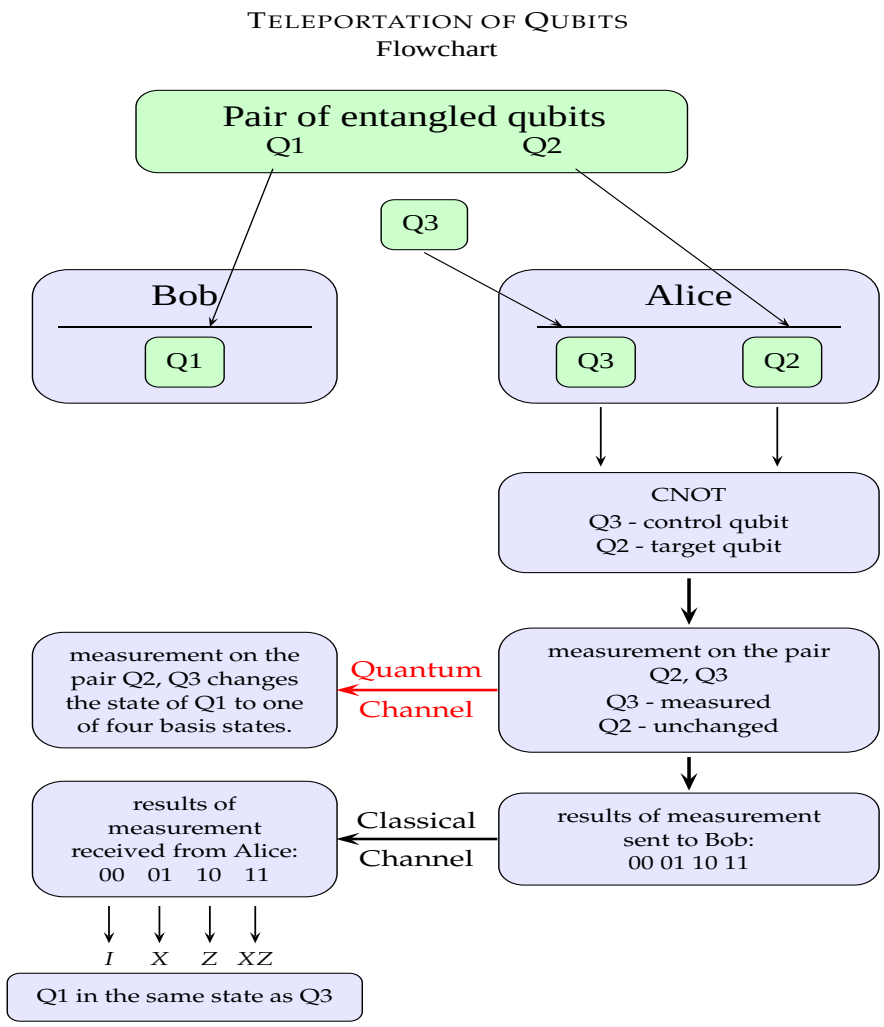
The established **protocol** for teleportation of qubits between two distant locations A and B reads:

1. An EPR pair (i.e., two entangled qubits in one of the Bell states, usually $|\phi^+\rangle$) is generated, and one qubit is send to location A, the other to location B.

2. At A there are two qubits, one to be teleported, the other the qubit from the EPR pair. A Bell measurement of these two qubits is performed yielding two classical bits and destroying these qubits in the process.

3. The two bits are sent from location A to B, using a classical channel, for instance a telephone line.

4. The qubit at B from the EPR pair is sent through a suitable quantum gate (determined by the two bits received from A) to produce a qubit which is identical to the one to be teleported.


TELEPORTATION OF QUBITS
Flowchart
Pair of entangled qubits
Q1
Q2
Q3
Bob
Q1
Alice
Q3
Q2
CNOT
Q3 - control qubit
Q2 - target qubit
measurement on the pair Q2, Q3 changes the state of Q1 to one of four basis states.
Quantum Channel
measurement on the pair
Q2, Q3
Q3 - measured
Q2 - unchanged
results of measurement received from Alice:
00 01 10 11
Classical Channel
results of measurement sent to Bob:
00 01 10 11
I X Z XZ
Q1 in the same state as Q3

Detailed computation.

Generating an EPR pair: The quantum circuit from Example 3.2.10 generates the Bell state $|\phi^+\rangle$ for two qubits in the states $|0\rangle_j$, $j = 1,2$.

$$|\phi^+\rangle_{21} = \frac{1}{\sqrt{2}}(|0\rangle_2 \otimes |0\rangle_1 + |1\rangle_2 \otimes |1\rangle_1) = CNOT(H \otimes I_2(|0\rangle_2 \otimes |0\rangle_1)). \quad (5.7)$$

Qubit 2 is sent to A and qubit 1 to B.

Bell measurement at A: Suppose that qubit 3 at A which is to be teleported is in the state

$$|\psi\rangle_3 = a|0\rangle_3 + b|1\rangle_3. \quad (5.8)$$

The state of our 3 qubit system then is

$$|\psi\rangle_3 \otimes |\phi^+\rangle_{21} = \frac{1}{\sqrt{2}}(a|0\rangle_3 \otimes (|00\rangle_{21} + |11\rangle_{21}) + b|1\rangle_3 \otimes (|00\rangle_{21} + |11\rangle_{21}))$$

which we write as

$$
\begin{aligned}
&= \frac{1}{\sqrt{2}}\big(a|00\rangle_{32}\otimes|0\rangle_1 + a|01\rangle_{32}|1\rangle_1 + b|10\rangle_{32}\otimes|0\rangle_1 + b|11\rangle_{32}\otimes|1\rangle_1\big)\\
&\qquad = \frac{a}{2}\big(|\phi^+\rangle_{32} + |\phi^-\rangle_{32}\big)\otimes|0\rangle_1 + \frac{a}{2}\big(|\psi^+\rangle_{32} + |\psi^-\rangle_{32}\big)\otimes|1\rangle_1\\
&\qquad + \frac{b}{2}\big(|\psi^+\rangle_{32} - |\psi^-\rangle_{32}\big)\otimes|0\rangle_1 + \frac{b}{2}\big(|\phi^+\rangle_{32} - |\phi^-\rangle_{32}\big)\otimes|1\rangle_1\\
&\qquad\quad = \frac{1}{2}|\phi^+\rangle_{32}(a|0\rangle_1 + b|1\rangle_1) + \frac{1}{2}|\phi^-\rangle_{32}(a|0\rangle_1 - b|1\rangle_1)\\
&\qquad\quad + \frac{1}{2}|\psi^+\rangle_{32}(a|1\rangle_1 + b|0\rangle_1) + \frac{1}{2}|\psi^-\rangle_{32}(a|1\rangle_1 - b|0\rangle_1)\\
&\qquad = \frac{1}{2}|\phi^+\rangle_{32}|\psi\rangle_1 + \frac{1}{2}|\psi^+\rangle_{32}X|\psi\rangle_1 + \frac{1}{2}|\phi^-\rangle_{32}Z|\psi\rangle_1 + \frac{1}{2}|\psi^-\rangle_{32}(-\mathrm{i})Y|\psi\rangle_1
\end{aligned}
$$

Since $(-\mathrm{i})\sigma_2 = \sigma_x\sigma_z$ we conclude

$$
\begin{aligned}
&|\psi\rangle_3\otimes|\phi^+\rangle_{21} =\\
&\frac{1}{2}\Big[|\phi^+\rangle_{32}\otimes I|\psi\rangle_1 + |\psi^+\rangle_{32}\otimes X|\psi\rangle_1 + |\phi^-\rangle_{32}\otimes Z|\psi\rangle_1 + |\psi^-\rangle_{32}\otimes XZ|\psi\rangle_1\Big]
\end{aligned}
\tag{5.9}
$$

Hence, when Alice performs a Bell measurement on her two-qubits system

$Q2, Q3$, i.e., when she measures the two commuting observables

$$\sigma_x^{(3)} \otimes \sigma_x^{(2)}, \quad \sigma_z^{(3)} \otimes \sigma_z^{(2)} \tag{5.10}$$

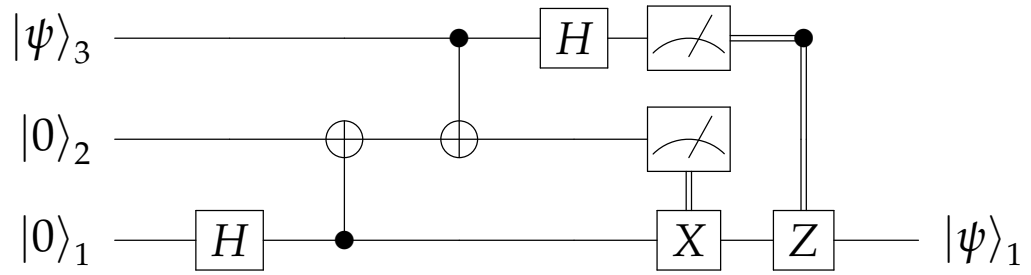
she finds with probability $\frac{1}{4}$ that her system is in one of the four Bell states which appear in (5.9). As expressed in (5.9) Alice Bell measurement establishes a correlation between $Q1$ and $Q3$. She sends her measurement result through a classical channel to Bob who applies a suitable gate according to the information he received.

$$\text{If Alice measures} \quad \begin{matrix} |\phi^+\rangle_{32} \\ |\psi^+\rangle_{32} \\ |\psi^-\rangle_{32} \\ |\phi^-\rangle_{32} \end{matrix} \quad \text{Bob applies} \quad \begin{matrix} I \\ X \\ XZ \\ Z \end{matrix}$$

and gets according to (5.9) the state $|\psi\rangle_1$ for his qubit $Q1$, since $X^2 = Y^2 = Z^2 = I$.

With the help of the results form subsection 2.7.1 we can easily read off from (5.9) the results of a Bell measurement of the subsystem formed by qubit 2 and 3 at A. The following quantum circuit is drawn for the case that Alice found

$|\psi^-\rangle_{32}$ and Bob applies the gates XZ.



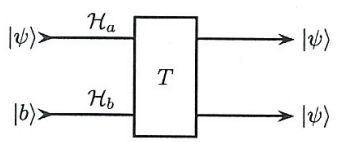
5.4 No Cloning

As we are going to explain quantum information cannot be perfectly copied or “cloned". This fact is stated in various no-cloning theorems. The basic version states:

> There is no quantum copying machine that can make two perfect copies (or one perfect copy and a remaining perfect original) of *two (or more) nonorthogonal states*.

Note that this says nothing about making as many perfect copies as one wants of *mutually orthogonal states* using a quantum copy machine. And within the

standard model of a quantum coping machine or hypothetical cloning machine this is actually a very elementary result. Here is the usual model.



$\mathcal{H}_a$ and $\mathcal{H}_b$ Hilbert space of the same dimension, T unitary operator on $\mathcal{H}_a \otimes \mathcal{H}_b$

Hypothetical cloning machine

In practice one works under the assumption that $\mathcal{H}_a = \mathcal{H}_b$ and that T acts according to the equation

$$T(|\psi\rangle \otimes |b\rangle) = |\psi\rangle \otimes |\psi\rangle \tag{5.11}$$

where $|b\rangle$ is some reference unit vector in $\mathcal{H}_b$ and $|\psi\rangle$ varies in $\mathcal{H}_a$. It is easy to see that for a particular $|\psi\rangle \in \mathcal{H}_a$ one can always construct a unitary operator T such that (5.11) holds. But we will show that there is no fixed T which accomplishes this for all possible inputs.

Suppose that $|\psi_j\rangle \in \mathcal{H}_a$ are normalized and satisfy (5.11), i.e.,

$$\begin{aligned} T(|\psi_1\rangle \otimes |b\rangle) &= e^{i\phi_1}|\psi_1\rangle \otimes |\psi_1\rangle \\ T(|\psi_2\rangle \otimes |b\rangle) &= e^{i\phi_2}|\psi_2\rangle \otimes |\psi_2\rangle \end{aligned} \tag{5.12}$$

where ϕ_j are some phases. Taking the inner product of these two identities gives, since T is unitary,

$$\langle \psi_1 | \, \psi_2 \rangle \, \langle b | \, b \rangle = e^{\mathrm{i}(\phi_2 - \phi_1)} \, \langle \psi_1 | \, \psi_2 \rangle^2$$

and thus

$$| \, \langle \psi_1 | \, \psi_2 \rangle \, | = | \, \langle \psi_1 | \, \psi_2 \rangle \, |^2 \, .$$

The only solutions to the last equation are $| \, \langle \psi_1 | \, \psi_2 \rangle \, | = 0$ or $| \, \langle \psi_1 | \, \psi_2 \rangle \, | = 1$. In the first case the two vectors are orthogonal and in the second case the two vectors are identical apart from a phase factor. This proves our claim.

Some further versions of no-cloning theorems are known; the above version gives just the essential core.

Chapter 6

Quantum Cryptography

We recall here only very briefly a few basic facts from (classical) cryptography. Cryptography is about sending messages between two parties in such a way that its contents cannot be understood by someone other than the indended recipient. The original message or *plaintext* is encrypted using an *encryption rule* typically based on an encryption *key* to produce an unintelligible *cybertext*. The recipient then applies a *decryption rule*, using the same key to the cybertext in order to recover the original plaintext message.

The *one-time pad* (OTP) is a type of (classical) encryption which has been proven to be impossible to crack if used correctly. Each bit or character from the plaintext is encrypted by modular addition with a bit or character from a

secret random key of the same length as the plaintext, producing the cybertext. If the key is truly random, as large or greater than the plaintext, never used in whole or part, and kept secret, the cybertext will be impossible to decrypt or break without knowing the key. But obviously there are several practical problems and thus the OTP's are not widely used.

Quantum cryptography is concerned with the distribution of encryption keys for cryptography where the distribution of this key is protected by basic principles of quantum mechanics. Nowadays there are several prominent *quantum key distribution protocols* which we will discuss:

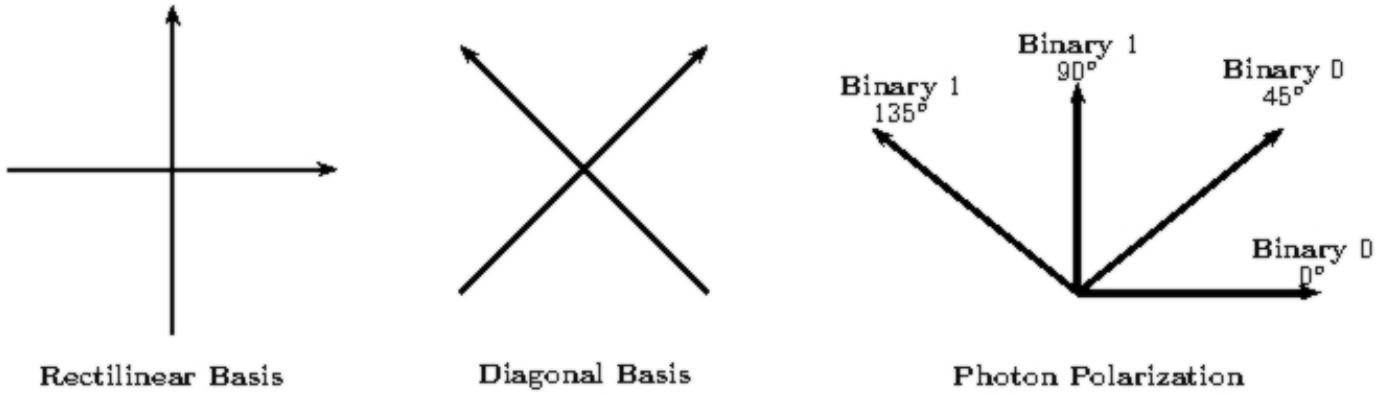
Protocols utilizing Heisenberg's uncertainty principle:	BB84 protocol B92 protocol
Protocols utilizing quantum entanglement:	Ekert's protocol Entangled BB84 variants.

We discuss in some detail only the BB84 and Ekert's protocols.

6.1 The BB84 Scheme

C. Bennet and G. Brassard published in 1984 3 the first QKD (Quantum Key Distribution) protocol. Today it is still one of the most prominent protocols. The basic idea of all HUP based protocols is as follows: Alice can transmit a random secret key to Bob by sending a string of photons where the secret key's bits are encoded in the polarization of photons. Heisenberg's Uncertainty Principle is used to guarantee that an eavesdropper cannot measure these photons and transmit them to Bob without disturbing the state of the photons in a detectable way and thus revealing her presence.

In the BB84 protocol the bits are encoded in the polarization states of a photon. Usually one defines a binary 0 as a polarization of 0 degrees in the rectilinear basis or 45 degrees in the diagonal basis. And similarly a binary 1 corresponds to 90 degrees in the rectilinear basis and 135 degrees in the diagonal basis. Thus a bit can be represented by polarizing the photon with respect to one of these two bases.



The rectilinear basis R is given by the vectors

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

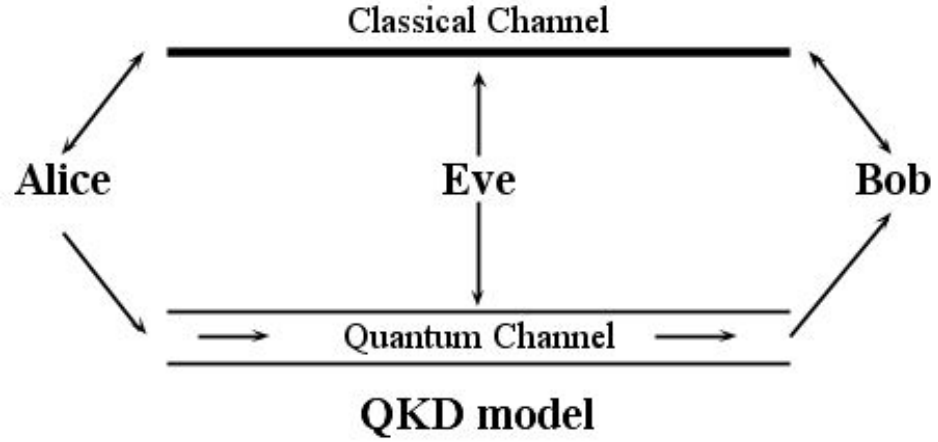
Then the vectors of the diagonal basis D are

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

and thus

$$|0\rangle = \frac{1}{\sqrt{2}}(|+\rangle + |-\rangle), \quad |1\rangle = \frac{1}{\sqrt{2}}(|+\rangle - |-\rangle).$$

If we perform measurements with respect to the bases R and D our formula (2.13) for the post-measurement state shows the following: A qubit can be either in the state $|0\rangle$ or $|1\rangle$ in the R basis or in the state $|+\rangle$ or $|-\rangle$ in the D basis. When a qubit is in one or the other state of the R basis then nothing can be said about its state in the D basis. When a qubit is in one or the other state of the D basis then nothing can be said about its state in the R basis.
More precisely we can say for instance: If a qubit is in the state $|+\rangle$ and we measure in the D basis then this state is reproduced, but if we measure in the R basis we find that the state $|+\rangle$ is destroyed and with probability $1/2$ the qubit is either in the state $|0\rangle$ or in the state $|1\rangle$, according to (2.13). The figure below shows the basic scheme for the BB84 protocol.



QKD model

The various steps of the **BB84 Quantum Key Distribution** protocol are:

1. Alice and Bob decide (publicly) on an acceptable key length N, taking a sensible error margin into account.

2. Secretly Alice chooses a random string of length $4N$ of data bits $d_1, d_2, \ldots, d_{4N}$ and a random string of length $4N$ of letters $a_1, a_2, \ldots, a_{4N}$, $a_j \in \{R, D\}$.

3. Bob too chooses secretly a random string of length $4N$ of letters $b_1, b_2, \ldots, b_{4N}$, $b_j \in \{R, D\}$.

4. Alice now does the following: For $j \in \{1, 2, \ldots, 4N\}$ she prepares the j^{th} qubit in the state $\left|d_j(a_j)\right\rangle$, i.e., the data value of the state is given by d_j and the basis is specified by a_j. After this preparation Alice sends the j^{th} qubit to Bob, through the quantum channel.

5. Bob receives the j^{th} qubit and measures it in the basis b_j and gets a classical bit e_j, for $j \in \{1, 2, \ldots, 4N\}$.

6. Alice and Bob exchange in public their basis label strings $a_1, a_2, \ldots, a_{4N}$ and $b_1, b_2, \ldots, b_{4N}$. Now both know the indices at which $a_1, a_2, \ldots, a_{4N}$ and $b_1, b_2, \ldots, b_{4N}$ agree, respectively disagree. They both discard those ele-

ments where there is disagreement. This leaves a common string $c_1, c_2, \ldots, c_{\ 2N}$, $c_j \in \{R, D\}$, which is typically of length about $2N$.

7. Alice discards all elements of the string $d_1, d_2, \ldots, d_{4N}$ that do not correspond with elements of the string $c_1, c_2, \ldots, c_{\ 2N}$. This gives a string of bits $D_1, D_2, \ldots, D_{\ 2N}$ which is typically of length about $2N$.

8. Bob discards the elements of the string $e_1, e_2, \ldots, e_{4N}$ that do not correspond to elements of the string $c_1, c_2, \ldots, c_{\ 2N}$. This leave him with a string of bits $E_1, E_2, \ldots, E_{\ 2N}$, again typically of length about $2N$.

9. Note that c_j, for each $j \in \{1, 2, \ldots, 2N\}$, is a basis name randomly chosen the same for the j^{th} qubit by Alice for preparation and by Bob for measurement. Thus the value E_j measured by Bob equals the value D_j prepared and sent by Alice. Therefore the two binary strings are equal: $D_1, D_2, \ldots, D_{\ 2N} = E_1, E_2, \ldots, E_{\ 2N}$. This common string can thus serve as a candidate secret key for communication between Alice and Bob.

10. Alice and Bob choose publicly a randomly selected subsequence of $c_1, c_2, \ldots, c_{\ 2N}$, typically of length about N and exchange publicly the subsequences of $D_1, D_2, \ldots, D_{\ 2N}$ and $E_1, E_2, \ldots, E_{\ 2N}$ that correspond to these values. They

should agree perfectly if there is no noise and/or eavesdropping.

11. If Eve has been eavesdropping, then about 25% of these values will disagree*. In this case Alice and Bob have to start again.

12. If there was no eavesdropping the remaining subsequences of $D_1, D_2, \ldots, D_{2N}$ and $E_1, E_2, \ldots, E_{2N}$, each of typical length about N, constitute a common sequence of bits $K_1, K_2, \ldots, K_N$ which is secretly shared by Alice and Bob and can serve as a secret key.

* Since Eve does not know the basis which has been assigned to each qubit she is likely to guess the basis incorrectly 50% of the time, and thus when she measures, any time she guesses wrong she will destroy the original state of the qubit and the classical information she gets, i.e., the bits, will be wrong 50% of the time.

If one assumes a noiseless quantum channel and that there are no measurement errors, a disagreement in any of the bits which are compared would indicate the presence of an eavesdropper on the quantum channel. If Eve the eavesdropper would attempt to determine the key, she would have no choice but to measure the photons sent by Alice before sending them to Bob. This is the case since the *no-cloning theorem* assures that she cannot replicate a particle

of unkown state. Since Eve will not know which bases Alice used to encode the bit until after Alice and Bob discuss their measurements, Eve has to guess. If she measures on the incorrect basis, *Heisenberg's Uncertainty Principle* ensures that the information encoded in the other basis is now lost. Thus when the photon reaches Bob, his measurement will now be random and he will read a bit incorrectly 50% of the time. Given that Eve will choose the measurement basis incorrectly on average 50% of the time, 25% of Bob's measured bits will differ from Alice's. If Eve has eavesdropped on all the bits, then after n bits comparisons by Alice and Bob, they will reduce the probability that Eve will not be detected to $(\frac{3}{4})^n$. Hence the chance that an eavesdropper can be successful will become negligible, if a sufficiently long sequence of bits are compared.

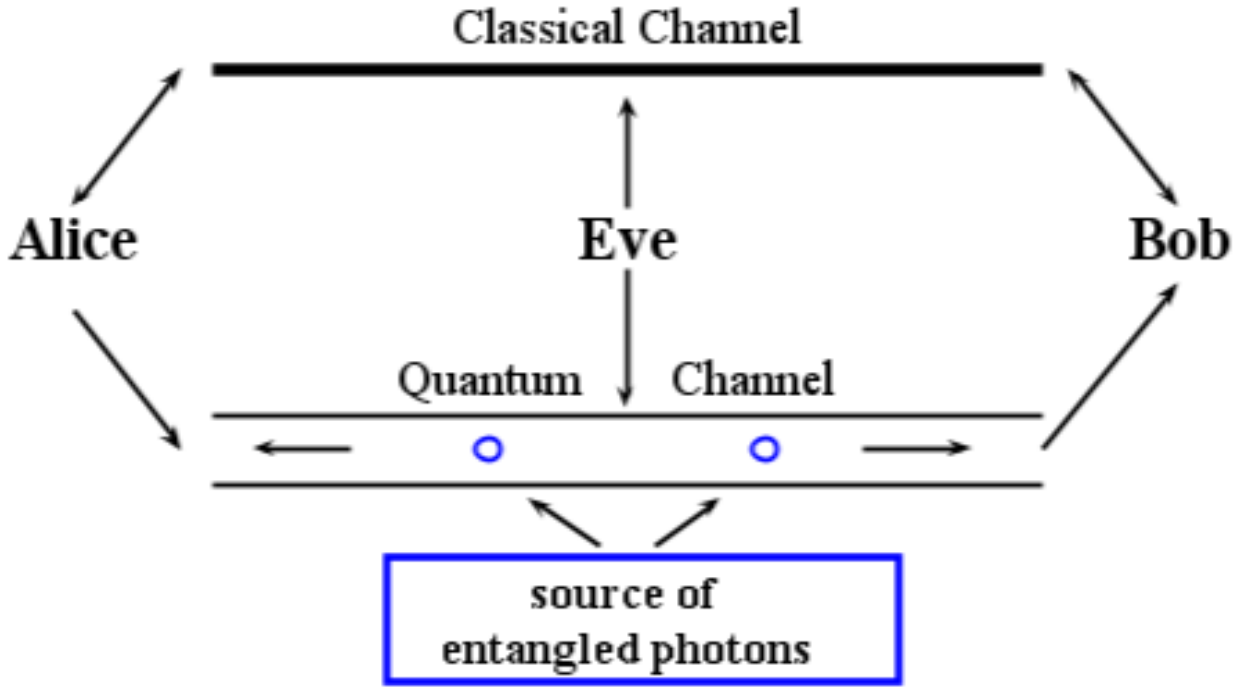
Here is an **example for the BB84 protocol**:

$1,2,\ldots,4N$	1	2	3	4	5	6	7	8	9	10	11	12
Bob's b_j's	D	R	D	D	R	D	R	R	D	D	D	R
Alice's a_j's	R	R	D	R	R	R	D	R	D	D	D	R
Alice's d_j's	0	1	1	0	1	1	1	0	1	0	1	0
Alice's sends	$\|0\rangle$	$\|1\rangle$	$\|-\rangle$	$\|0\rangle$	$\|1\rangle$	$\|1\rangle$	$\|-\rangle$	$\|0\rangle$	$\|-\rangle$	$\|+\rangle$	$\|-\rangle$	$\|0\rangle$
$a_j = b_j$?		y	y		y			y	y	y	y	y
Bob measures		$\|1\rangle$	$\|-\rangle$		$\|1\rangle$			$\|0\rangle$	$\|-\rangle$	$\|+\rangle$	$\|-\rangle$	$\|0\rangle$
Security test			1					0		0	1	
Secret key		1			1				1			0

Remark 6.1.1 *The B92 protocol was proposed by C. Bennet in 1992 2. It is essentially a simplified version of the BB84 protocol which uses only two non-orthogonal states instead of the 4 polarization states of the BB84 protocol. This* 2*-state encoding of the B92 protocol can be done as follows:* 0 *is encoded as* 0^o *in the rectilinear basis and* 1 *is encoded as* 45^o *in the diagonal basis.*

6.2 Ekert's protocol E91

This protocol was suggested in 1991 by Artur Ekert (Oxford and Singapore). It is based on quantum entanglement. The basic scheme for this protocol is:



Entanglement based QKD model

The source emits entangled particles, typically polarized photons which are

spatially separated. In detail the **E91 quantum key distribution protocol** reads:

0. Alice and Bob share $4N$ maximally entangled qubit pairs, i.e. Alice has one qubit of each entangled pair and Bob has the other.

1. Alice and Bob publicly decide an an acceptable key length N taking a sensible error margin into account.

2. Alice secretly chooses a random string of length $4N$ of letters $a_1, a_2, \ldots, a_{4N}$, $a_j \in \{R, D\}$.

3. Bob secretly chooses a random string of length $4N$ of letters $b_1, b_2, \ldots, b_{4N}$, $b_j \in \{R, D\}$.

4. For each j, $1 \leq j \leq 4N$, Alice measures her qubit of the j^{th} pair in the basis a_j and gets a classical bit d_j.

5. For each j, $1 \leq j \leq 4N$, Bob measures his qubit of the j^{th} pair in the basis b_j and gets a classical bit e_j.

6. Alice and Bob exchange publicly their basic label strings $a_1, a_2, \ldots, a_{4N}$ and $b_1, b_2, \ldots, b_{4N}$. They both know now the indices at which $a_1, a_2, \ldots, a_{4N}$ and

$b_1, b_2, \ldots, b_{4N}$ agree and the indices at which $a_1, a_2, \ldots, a_{4N}$ and $b_1, b_2, \ldots, b_{4N}$ disagree. Alice and Bob discard the elements that disagree and they are left with a common string (typically of length about $2N$) $c_1, c_2, \ldots, c_{\ 2N}$.

7. Alice discards the elements of $d_1, d_2, \ldots, d_{4N}$ that do not correspond to $c_1, c_2, \ldots, c_{\ 2N}$ and gets a string of bits (typically of length about $2N$) $D_1, D_2, \ldots, D_{\ 2N}$.

8. Bob discards the elements of $e_1, e_2, \ldots, e_{4N}$ that do not correspond to $c_1, c_2, \ldots, c_{\ 2N}$ and gets a string of bits (typically of length about $2N$) $E_1, E_2, \ldots, E_{\ 2N}$.

9. For each $j \in \{1, 2, \ldots, 2N\}$, c_j is the name of a basis chosen the same by Alice and Bob for the j^{th} pair of maximally entangled qubits , the value E_j measured by Bob, equals the value D_j measured by Alice. Hence the two binary strings are equal: $D_1, D_2, \ldots, D_{\ 2N} = E_1, E_2, \ldots, E_{\ 2N}$. Thus they can serve as a candidate secret key for communication between Alice and Bob.

10. Alice and Bob choose publicly a randomly selected subsequence of $c_1, c_2, \ldots, c_{\ 2N}$ (typically of length about N), and exchange in public the subsequences of $D_1, D_2, \ldots, D_{\ 2N}$ and $E_1, E_2, \ldots, E_{\ 2N}$ that correspond to these values. Ideally they should agree perfectly.

11. If Eve has been eavesdropping, or the environment has degraded the max-

imal entanglement of the qubit pairs, a significant proportion of these values will disagree. In this case, Alice and Bob must start again.

12. If not, the remaining subsequences of $D_1, D_2, \ldots, D_{2N}$ and $E_1, E_2, \ldots, E_{2N}$ (each typically of length about N) constitute a common sequence of bits $K_1, K_2, \ldots, K_N$, which is secretly shared by Alice and Bob, and thus can serve as a secret key.

Here is a simple example for this protocol.

$1,2,\ldots,4N$	1	2	3	4	5	6	7	8	9	10	11	12
Alice's a_j's	R	R	D	R	R	R	D	R	D	D	D	R
Bob's b_j's	D	R	D	D	R	D	R	R	D	D	D	R
$a_j = b_j$?		y	y		y			y	y	y	y	y
Alice's d_j's	0	1	1	0	1	1	1	0	1	0	1	0
Bob's e_j's	1	1	1	0	1	0	0	0	1	0	1	0
Alice's D_j's		1	1		1			0	1	0	1	0
Bob's E_j's		1	1		1			0	1	0	1	0
Security test			1					0		0	1	
Secret key		1			1				1			0

6.3 Commercial implementations

Quantum key distribution is nowadays beyond the experimental phase. Beside active research programs in various international companies (IBM, HP, Mitsubishi, NEC, NTT) there are three companies offering commercial quantum key distribution systems:

1. id Quantique (Geneva),
2. MagiQ Technologies (New York),
3. QuintessenceLabs (Australia).

Furthermore there are several quantum networks in operation.

1. *DARPA* Quantum Network: 10 nodes, running since 2004 in Massachusetts (USA); BBN Technologies, Harvard University, Boston University and QinetiQ.
2. *SECOQC* (Secure Communication based on Quantum Cryptography); the world's first computer network protected by quantum key distribution; implemented in 2008 in Vienna, using more than 200 km of standard fibre optic cables.

Special event: In 2004, the world's first bank transfer using quantum key distribution was carried out in Vienna, Austria.

3. *SwissQuantum*: started in 2007 by Id Quantique in Geneva, first for transmission of ballot results; from 2009 running reliably and stable for about 2 years, confirming the viability of QKD as a commercial encryption technology.

4. *Tokyo QKD Network* : The Tokyo QKD Network was inaugurated on the first day of the UQCC2010 conference. The network involves an international collaboration between 7 partners; NEC, Mitsubishi Electric, NTT and NICT from Japan, and participation from Europe by Toshiba Research Europe Ltd. (UK), Id Quantique (Switzerland) and All Vienna (Austria). "All Vienna" is represented by researchers from the Austrian Institute of Technology (AIT), the Institute for Quantum Optics and Quantum Information (IQOQI) and the University of Vienna.

5. *Durban Quantum Network*: since 2009 running on the municipal standard fibre optic cables network;

   special event in 2010: the connection of the football world cup soccer stadium to the municipal network was secured by QKD;

implemented by part of our group in Durban;
see http://quantum.ukzn.ac.za/

Chapter 7

Shor Factorization

On a classical computer the multiplication of large prime numbers can be implemented quite efficiently while the inverse function, i.e., finding the prime factors of a large number has not yet been solved by an efficient algorithm. Here an algorithm is called *efficient* if its execution time (the number of elementary operations) is asymptotically polynomial in the length of its input measured in bit. The classical *quadratic sieve* algorithm needs

$$O\left(e^{(\frac{64}{9})^{1/3}N^{1/3}(\ln N)^{2/3}}\right)$$

operations for factoring a binary number of N bits, i.e., this algorithm scales exponentially with the input size. It seems to be the best algorithm for this problem.

Exercise: Using only pen and paper find the (prime) factors of

$$29083$$

This means that the multiplication of large prime numbers is essentially a *one-way* function which is the basis of the cryptographic algorithm developed in 1978 by Rivest, Shamir, and Adelman. This method became the most popular public key system (*RSA encryption*).

It is believed that efficient prime factorization is impossible on a classical computer. On a quantum computer however an efficient algorithm has been proposed by P. Shor, inspired by work of D. Simon in 1994. This algorithm finds the factors of a large composite number N, i.e., a number which can be written in the form

$$N = pq \tag{7.1}$$

where the numbers p,q are assumed to be *relatively prime* which means that their greatest common divisor (gcd) is 1.

Here we will briefly explain the basic ideas for this factorization which takes time $O((\log N)^3)$ and hence the integer factorization can be efficiently solved on a quantum computer and is in the complexity class **BQP** (bounded error quantum polynomial time).

The realizability of Shor's algorithm has first been shown at IBM in 2001 where the number 15 was factored into 5×3 using an NMR implementation of a quantum computer with 7 qubits. In 2012 a group in Bristol (UK) achieved the factorization of 21. At present it is not yet possible to factor really large numbers since only relatively small quantum computers can be build.

The factorization method of Shor has two main parts, a classical part based on results from number theory and classical computation, and a quantum part. Under suitable restrictions the factorization problem is reduced in the classical part to that of a 'period finding problem'. For a solution of the period finding problem a suitable quantum computer is constructed which realizes the quantum Fourier transform through which the discrete spectrum of the period function in question can be determined approximately. Some post-processing then allows to calculate the period.

We begin by recalling some mathematical background, then proceed to explain the reduction of the factorization problem to the period finding problem and in the main part present the core of Shor's quantum algorithm.

7.1 Mathematical background

7.1.1 Some number theory

The Chinese remainder theorem

Suppose that $m_1, m_2, \ldots, m_k$ are positive integers which are coprime, i.e.,

$$\gcd(m_i, m_j) = 1 \quad \text{for all} \quad i \neq j\,.$$

Then, for any given sequence of integers $a_1, a_2, \ldots, a_k$, there exists a unique integer x, $1 \leq x \leq m_1 m_2 \cdots m_k = M$ solving the following system of simultaneous congruences:

$$x \equiv a_j \bmod \mathrm{m_j}, \quad \mathrm{j} = 1, 2, \ldots, \mathrm{k} \tag{7.2}$$

The solution is explicitly given as follows: Define $M_j = M/m_j$, then $\gcd(m_j, M_j) = 1$ by our assumption and M_j has an inverse $N_j \equiv M_j^{-1} \bmod m_j$, i.e.,

$$N_j M_j \equiv 1 \bmod m_j\,.$$

The unique solution of (7.2 is

$$x \equiv (a_1 N_1 M_1 + a_2 N_2 M_2 \cdots a_k N_k M_k) \bmod M\,. \tag{7.3}$$

Euclid's algorithm

This is an efficient method for calculating the *greatest common divisor* $\gcd(n_1, n_2)$ of two given integers n_1, n_2. We assume $n_1 \geq n_2$. Write n_1 as a multiple k_0 of n_2 and a remainder r_1:

$$n_1 = k_0 n_2 + r_1, \quad r_1 < n_2 .$$

Do the same with n_2 and r_1,

$$n_2 = k_1 r_1 + r_2, \quad r_2 < r_1 ,$$

and then repeat with the two r's,

$$r_1 = k_2 r_2 + r_3, \quad r_3 < r_2 ,$$

$$n_2 = k_3 r_3 + r_4, \quad r_4 < r_3 ,$$

until the remainder is zero (which occurs since the remainders are stricly decreasing): So for some l,

$$r_{l-1} = k_l r_l + r_{l+1}, \quad r_{l+1} < r_l ,$$

$$r_l = k_{l+1} r_{l+1} + 0 .$$

The greatest common divisor $\gcd(n_1, n_2)$ is then given by the last nonzero remainder

$$\gcd(n_1, n_2) = r_{l+1} .$$

Euler's φ function

For any integer $N > 1$ the set of integers $n \in \{1, \ldots, N-1\}$ which are relatively prime to N form a group Z_N^* under multiplication mod N. Its order is given by the value $\varphi(N)$ of Euler's φ function. For example take $N = 15$. Then Z_{15}^* is the set $\{1,2,4,7,8,11,13,14\}$ and thus its order is $\varphi(15) = 8$.

Since Z_N^* is a multiplicative group, for any $a \in Z_N^*$ all powers a^m belong to Z_N^* and thus constitute the subgroup $\langle a \rangle$ of Z_N^* generated by a. The order of this subgroup is the smallest integer $r \geq 1$ such that

$$a^r = 1 \mod \mathrm{N}\,. \tag{7.4}$$

This order r always divides $\varphi(N)$ since the order of a subgroup always divides the order of the group. For $a = 2$ the order of the subgroup $\langle 2 \rangle$ in our example Z_{15}^* is $r = 4$ since $2^4 = 1 \bmod 15$ and clearly $r = 4$ divides $\varphi(15) = 8$.

Continued fractions

Given positive integers $a_0, a_1, \ldots, a_N$ an expression of the form

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\cdots + \cfrac{1}{a_N}}}}} \tag{7.5}$$

is called a finite *continued fraction*, usually abbreviated as $[a_0, a_1, \ldots, a_N]$. For $n \leq N$ its n th convergent is $[a_0, a_1, \ldots, a_n]$ which can be written as p_n/q_n and one has the recurrence

$$\begin{aligned} p_0 &= a_0, \quad p_1 = a_1 a_0 + 1, \quad p_n = a_n p_{n-1} + p_{n-2}, \\ q_0 &= 1, \quad q_1 = a_1, \qquad\quad q_n = a_n q_{n-1} + q_{n-2}. \end{aligned} \tag{7.6}$$

If the convergents p_n/q_n are calculated by this iteration, they are always in their lowest term, i.e., $\gcd(p_n, q_n) = 1$.

Next we recall the efficient algorithm which calculates the continued fraction representation for any positive rational number x. Denote by $\llcorner x \lrcorner$ the

greatest integer less than or equal to x. Then, with $a_0 = \llcorner x \lrcorner$, write $x = a_0 + \xi_0$ for some $0 \leq \xi_0 < 1$. If $\xi_0 > 0$, define $a_1 = \llcorner 1/\xi_0 \lrcorner$ and thus $1/\xi_0 = a_1 + \xi_1$ for some $0 \leq \xi_1 < 1$. If $\xi_1 > 0$ define $a_2 = \llcorner 1/\xi_1 \lrcorner$, etc. Since x is assumed to be rational this process terminates and we get $x = [a_0, a_1, \ldots, a_n]$.

Eq. (7.5) shows that

$$[a_0, a_1, \ldots, a_{N-1}, a_N] = [a_0, a_1, \ldots, a_{N-1}, a_N - 1, 1]$$

holds. Thus if we impose the condition $a_N > 1$ the representation of x as a continued fraction is unique.

We are going to use the following result.

Theorem 7.1.1 *Suppose that p/q is a rational number satisfying*

$$\left|\frac{p}{q} - x\right| < \frac{1}{2q^2}.$$

Then p/q is a convergent of the continued fraction of x.

7.1.2 Quantum Fourier transform

The quantum Fourier transform is the classical discrete Fourier transform acting on the vector of amplitudes of a quantum state. Recall that the classical

discrete Fourier transform sends a vector $(x_0, x_1, \ldots, x_{N-1}) \in \mathbb{C}^N$ to the vector $(y_0, y_1, \ldots, y_{N-1}) \in \mathbb{C}^N$ with

$$y_k = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} \omega^{jk} x_j,$$

where ω denotes the canonical Nth primitive root of unity in the complex plane, i.e., $\omega = e^{\frac{2\pi \mathrm{i}}{N}}$. Accordingly the quantum Fourier transform sends the quantum state $\sum_{j=0}^{N-1} x_j |j\rangle$ to the quantum state $\sum_{j=0}^{N-1} y_j |j\rangle$ where the y_j's are given by the same formula.

Naturally the quantum Fourier transform can be described as a unitary matrix acting on qubit states. The gate QFT_n on n qubits is defined as the $N \times N$ unitary matrix, $N = 2^n$,

$$QFT_n = \left(\frac{\omega^{jk}}{\sqrt{N}}\right)_{0 \le j,k \le N-1} = \frac{1}{\sqrt{N}} \cdot \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega^1 & \omega^2 & \cdots & \omega^{(N-1)\cdot 1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{(N-1)\cdot 2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \omega^{2\cdot(N-1)} & \ldots & \omega^{(N-1)^2} \end{pmatrix} \tag{7.7}$$

It is a simple calculation to show that this matrix is indeed unitary.

For applications it is important that the quantum Fourier transform can be implemented by a relatively simple quantum circuit which we describe now.

Consider an n qubit system. Its states can be represented in binary form

$$|x\rangle = |x_1, x_2, \ldots, x_n\rangle = |x_1\rangle \otimes |x_2\rangle \otimes \cdots \otimes |x_n\rangle, \quad x_j \in \{0,1\}$$

where the x_j are the binary digits in the representation $x = \sum_{j=1}^{n} x_j 2^{n-j}$. Introduce the abbreviation $[0.x_1 \cdots x_n]$ for the binary fraction $\sum_{j=1}^{n} x_j 2^{-j}$ and abbreviate

$$\lambda_j(x) = e^{2\pi \mathrm{i} \cdot [0.\mathrm{x_j} \cdots \mathrm{x_n}]}.$$

This allows to realize the quantum Fourier transform in an simple way:

Proposition 7.1.2 *The quantum Fourier transform on n qubits is given by the map*

$$|x\rangle \longrightarrow \frac{1}{\sqrt{N}} \bigotimes_{j=0}^{n-1} \left(|0\rangle + \lambda_{n-j}(x)|1\rangle\right) \tag{7.8}$$

The proof is a straight forward calculation.

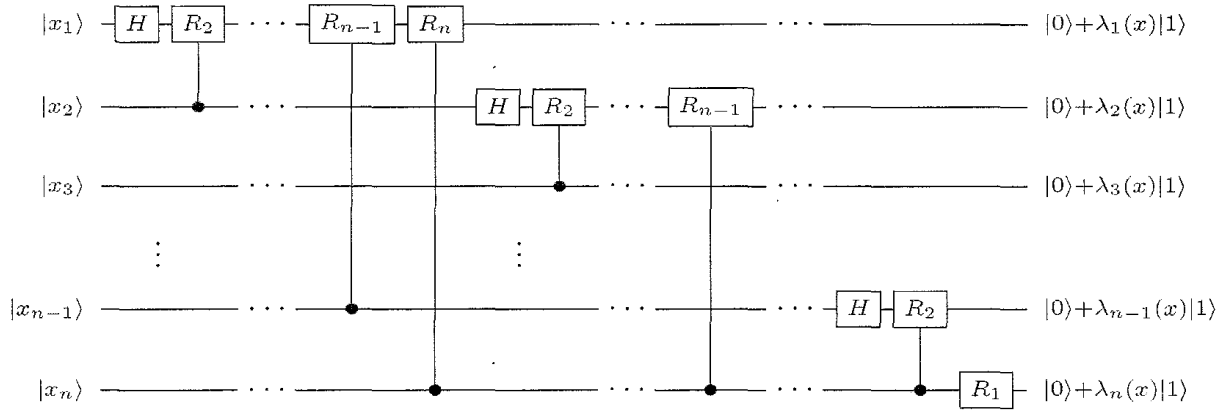
$$\begin{aligned}
\sqrt{N}\cdot|x_1,x_2,\ldots,x_n\rangle &\longrightarrow \sum_{k=0}^{N-1} e^{\frac{2\pi i x k}{N}}|k\rangle \\
&= \sum_{k_1\in\{0,1\}}\cdots\sum_{k_n\in\{0,1\}} e^{2\pi i x\frac{\sum_{j=1}^{n} k_j 2^{n-j}}{2^n}}|k_1,\ldots,k_n\rangle \\
&= \sum_{k_1\in\{0,1\}}\cdots\sum_{k_n\in\{0,1\}} e^{2\pi i x\sum_{j=1}^{n} k_j 2^{-j}}|k_1,\ldots,k_n\rangle \\
&= \sum_{k_1\in\{0,1\}}\cdots\sum_{k_n\in\{0,1\}} \bigotimes_{j=1}^{n} e^{2\pi i x k_j 2^{-j}}\left|k_j\right\rangle \\
&= \bigotimes_{j=1}^{n}\Big(\sum_{k_j\in\{0,1\}} e^{2\pi i x k_j 2^{-j}}\left|k_j\right\rangle\Big) \\
&= \bigotimes_{j=1}^{n}(|0\rangle + e^{2\pi i x 2^{-j}}|1\rangle) = \bigotimes_{j=0}^{n-1}(|0\rangle + e^{\pi i x 2^{-j}}|1\rangle) \\
&= \bigotimes_{j=0}^{n-1}(|0\rangle + e^{\pi i[0.x_{n-j}\cdots x_n]}|1\rangle)
\end{aligned}$$

This result shows that the quantum Fourier transform can be calculated through a combination of simple gates on individual qubits.

Recall the controlled U-gate (3.12) and introduce k-rotation gate R_k by

$$R_k = \begin{pmatrix} 1 & 0 \\ 0 & e^{\frac{2\pi \mathrm{i}}{2^k}} \end{pmatrix}.$$

Theorem 7.1.3 *Up to a reordering of the qubits, the quantum circuit depicted in the following figure computes the quantum Fourier transform:*



7.2 Shor's reduction of the factorization problem to order finding

Let $N \in \mathbb{N}$ be given; consider the quadratic equation

$$x^2 \equiv 1 \bmod \mathrm{N}. \tag{7.9}$$

This equation always has the trivial solutions $x \equiv 1 \bmod \mathrm{N}$. If N is an odd prime number p, then these are the only solutions, since multiplication modulo p has inverses and $x^2 - 1 \equiv (x+1)(x-1) \equiv 0 \bmod p$ implies $x - 1 \equiv 0$ or $x + 1 \equiv 0 \bmod p$. For a composite number N there are however also non-trivial pairs of solutions $x = \pm a \bmod N$ as we show now: Assume that N is of the form $N = n_1 n_2$ with $\gcd(n_1, n_2) = 1$ and consider the following set of equivalences:

$$\begin{aligned}
(a)\ &\begin{cases} x_1 \equiv 1 & \bmod \mathrm{n}_1 \\ x_1 \equiv 1 & \bmod \mathrm{n}_2 \end{cases} \qquad (a)\ \begin{cases} x_2 \equiv -1 & \bmod \mathrm{n}_1 \\ x_2 \equiv -1 & \bmod \mathrm{n}_2 \end{cases} \\
(c)\ &\begin{cases} x_3 \equiv 1 & \bmod \mathrm{n}_1 \\ x_3 \equiv -1 & \bmod \mathrm{n}_2 \end{cases} \qquad (d)\ \begin{cases} x_4 \equiv -1 & \bmod \mathrm{n}_1 \\ x_4 \equiv 1 & \bmod \mathrm{n}_2 \end{cases}
\end{aligned} \tag{7.10}$$

In all four sets of equivalences $x_j \equiv 1 \bmod n_1$ and $\bmod n_2$; each x_j satisfies Equation (7.9). By the Chinese remainder theorem each set has a unique solution mod N. From (a) and (b) we get the trivial solutions of Equation (7.9)

$x_1 \equiv 1$ and $x_2 \equiv -1 \bmod N$, while from (c) and (d) we get a non-trivial pair $x_3 \equiv a$ and $x_4 \equiv -a \bmod N$ of Equation (7.9). It follows that $(a+1)(a-1) \equiv 0 \bmod N$ and $a \pm 1$ are nonzero. Hence N divides $(a+1)(a-1)$, but does not divide $a \pm 1$, since $a \pm 1 \leq N+1$. Hence the greatest common divisor of N and $a \pm 1$ for $a \neq \pm 1$ is a nontrivial factor of N. And Euclid's algorithm allows to determine the greatest commen divisor of two given numbers efficiently.

How can we find a nontrivial solution x of Equation (7.9)? Choose a random $y < N$. If y and N are coprime , let r be the order of $y \bmod N$. And this is the period of the function for y and N, i.e., for

$$F_N(a) = y^a \bmod \mathrm{N}. \tag{7.11}$$

Therefore

$$y^r = 1 \bmod \mathrm{N}. \tag{7.12}$$

If r is an even number and if we set

$$x = y^{r/2}, \tag{7.13}$$

then we have $x^2 \equiv 1 \bmod N$ and thus x is a candidate for a nontrivial solution of Equation (7.9).

Certainly this procedure may fail if we have chosen a y which has an odd order r, or if r is even but $y^{r/2}$ turns out to be a trivial solution of Equation (7.9).

Fortunately the following result shows that the probability for this to happen is suitably small if y is chosen at random.

Theorem 7.2.1 *Let N be an odd number with prime factorization*

$$N = p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k}. \tag{7.14}$$

Choose a number y, $1 \leq y \leq N$ at random, satisfying $gcd(y,N) = 1$. If r is the order of y mod N, then

$$\mathrm{Prob}(\mathrm{r}\ \textit{is even and}\ \mathrm{y}^{\mathrm{r}/2} \not\equiv \pm 1 \bmod \mathrm{N}) \geq 1 - \frac{1}{2^{\mathrm{k}-1}}. \tag{7.15}$$

Note that in the case that N is even the factor 2 is easily detected and removed.

Remark 7.2.2 *For the proof of the above theorem see 10. It can be extended to show that*

$$\mathrm{Prob}(\mathrm{r}\ \textit{is even and}\ \mathrm{y}^{\mathrm{r}/2} \not\equiv \pm 1 \bmod \mathrm{N}) \geq \frac{1}{2} \tag{7.16}$$

holds for all N which are not of the form p^m or $2p^m$. In these cases the above probability is zero. But pure prime powers p^m are known to be efficiently recognizable by a classical probabilistic algorithm.

One can also show that for a random selection of y, $1 \leq y \leq N$, the probability of $gcd(y,N) = 1$ is greater than $1/\log N$. Thus, if we apply the above process to

a randomly chosen y for which we can compute the order r, we obtain a nontrivial factor of N with probability greater than $1/2\log N$.

Let us consider a simple *example* of the above factoring method. Take $N = 15$. We know $Z_{15}^* = \{1,2,4,7,8,11,13,14\}$. Pick for instance $y \in Z_{15}^*$, $y = 11$. The values of 11^a mod 15 for $a = 1,2,3,\ldots$ are $11,1,11,1,11,\ldots$. Thus the order of 11 modulo 15 is $r = 2$. This give $x = y^{r/2} = 11$. The largest common factors $\gcd(x \pm 1, N)$ are in this case $\gcd(10,15) = 5$ and $\gcd(12,15) = 3$, indeed the two prime factors of 15.

Summary:

1. Pick randomly an integer $y < N$.
2. Using the Euclidean algorithm, compute $\gcd(y,N)$.
3. If $\gcd(y,N) \neq 1$ there is a nontrivial factor of N, and we are done.
4. Otherwise we use the period-finding subroutine to find the period r of the function (7.11).
5. If r is odd, go back to step 1.

6. If $y^{r/2} \equiv -1 \bmod N$, go back to step 1.

7. Otherwise $\gcd(y^{r/2} \pm 1, N)$ is a nontrivial factor of N.

This part thus shows how to obtain factors from periods of a suitably chosen function.

7.3 Shor's quantum algorithm

This probabilistic algorithm runs in polynomial time, i.e., it requires polynomial$(\log n)$ steps. It computes the order r of a randomly chosen y with $\gcd(y, N) = 1$ with any prescribed probability of success $1 - \epsilon, \epsilon > 0$.

Given N take $q = 2^L$ such that $N^2 \leq q < 2N^2$ (sometimes a q such that $N^5 \leq q < 2N^5$ is used). Initialize two quantum registers of $[\log q]$ respectively $[\log N]$ qubits, each in the state $|0\rangle$, i.e., initialize the quantum computer in the state

$$|\psi\rangle = |0,0\rangle$$

and start the algorithm:

1. Apply the Hadamard gate H to each qubit in the first register to get the

state

$$\frac{1}{\sqrt{q}}\sum_{a=0}^{q-1}|a\rangle|0\rangle.$$

2. Compute $F_N(a) = y^a \bmod \mathrm{N}$ in the second register, which gives the state

$$\frac{1}{\sqrt{q}}\sum_{a=0}^{q-1}|a\rangle|y^a \bmod \mathrm{N}\rangle.$$

3. Measure the second register. It can be in a base state $|k\rangle$ where k is some power of x mod N. Denote by A the set of all $a < q$ such that x^a mod N equals k and denote by M the number of elements in A. Then the post-measurement state is

$$|\psi\rangle = \frac{1}{\sqrt{M}}\sum_{a\in A}|a,k\rangle.$$

Note that A has the representation

$$A = \{a_0, a_0 + r, a_0 + 2r, \ldots, a_0 + (M-1)r\}$$

with $M \approx \frac{q}{r} \gg 1$. Thus the post-measurement state can be written as

$$|\psi\rangle = \frac{1}{\sqrt{M}}\sum_{d=0}^{M-1}|a_0 + dr, k\rangle.$$

4. Apply the quantum Fourier transform QFT_q to the first register of this post-measurement state to get the state

$$\frac{1}{\sqrt{qM}}\sum_{c=0}^{q-1}\sum_{d=0}^{M-1} e^{2\pi \mathrm{ic}(\mathrm{a}_0+\mathrm{dr})/\mathrm{q}}|c,k\rangle = \sum_{c=0}^{q-1}\frac{e^{2\pi \mathrm{ica}_0/\mathrm{q}}}{\sqrt{qM}}\sum_{d=0}^{M-1}\zeta^d|c,k\rangle$$

where $\zeta = e^{2\pi \mathrm{icr}/\mathrm{q}}$.

5. Measure register 1. We observe register 1 to be in a particular state $|c\rangle$ with probability

$$Pr(c) = \frac{1}{qM}\Big|\sum_{d=0}^{M-1}\zeta^d\Big|^2 = \frac{1}{qM}\Big|\frac{1-\zeta^M}{1-\zeta}\Big|^2 = \frac{1}{qM}\frac{\sin^2(\pi Mcr/q)}{\sin^2(\pi cr/q)}. \qquad (7.17)$$

If $\frac{cr}{q}$ is not very 'close' to an integer, then powers of ζ very nearly cancel out ('destructive interference') and such states $|c\rangle$ are extremely unlikely to be observed. In this case this probability is small.

However, if for some integer d one has

$$\frac{cr}{q} \approx d$$

then $\zeta \approx 1$ and therefore, using periodicity of $\sin^2(\theta + k\pi) = \sin^2(\theta)$ for $k \in \mathbb{Z}$

$$Pr(c) \approx \frac{1}{qM} M^2 = \frac{M}{q}$$

is much larger. Hence the observed probability distribution is concentrated around values of c for which

$$\frac{c}{q} \approx \frac{d}{r} \quad \text{for some integer } d\,. \tag{7.18}$$

6. For the observed value of c one uses a classical computer to find fractions d/r very close to c/q, expecting that this will give the true order r of x mod N. For this one uses the method of continued fractions to compute the convergents d_j/r_j to c/q, see Theorem 7.1.1.

7. Here some more details of Step 6. According to (7.17) there are exactly r values of c mod q which satisfy

$$-r/2 \leq rc \bmod \mathrm{q} \leq \mathrm{r}/2 \tag{7.19}$$

and we wish to extract the value of r , given a value of c satisfying (7.19).

Note that (7.19) is equivalent to

$$\left|\frac{c}{q} - \frac{c'}{r}\right| \leq \frac{1}{q} \tag{7.20}$$

where c and q are known and $r \leq N$, $q \geq N^2$.

Because of $q \leq N^2$, there is exactly one fraction c'/r with denominator at most N in the range determined by (7.20). And this fraction may be found by using the continued fraction expansion of c/q as one of its convergents c'/q. Hence, if $\gcd(c', r) = 1$, we get the value of r.

Remark 7.3.1 *The essential point of this period finding algorithm is the ability of a quantum computer to be in many states simultaneously (superposition of states). This allows to compute the period of a function F by evaluating its values simultaneously at all points.*

As we know in quantum physics we get access to this information only through measurement. But a measurement will give only one of all possible values and destroys all others. The no cloning theorem forbids to make suitable copies before the measurement takes place. Therefore the superposition has been carefully transformed to another state that will return the correct answer with high probability. For this transformation the quantum Fourier transformation has been used.

Remark 7.3.2 *In order to implement the function (7.11) as a quantum transform one uses repeated squaring for the modular exponentiation transformation. This step is actually more difficult to implement than the quantum Fourier transform and it requires ancillary qubits and substantially more gates to implement. This results in a considerable slow down of the algorithm in concrete realizations.*

Chapter 8

Other important Topics

8.1 Deutsch: Universal Quantum Computer

8.2 Grover's search algorithm for unsorted database

8.3 Quantum Error Correction

8.4 Quantum Complexity Theory

8.5 Continuous variable QKD

8.6 Physical Implementations

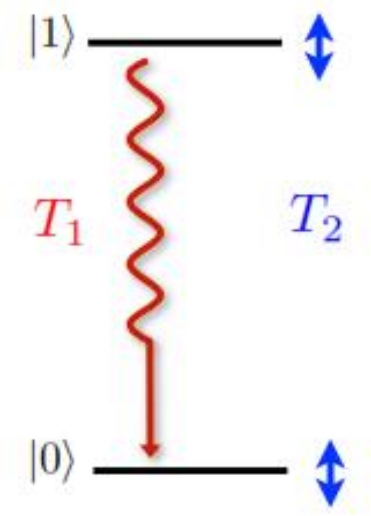
8.6.1 DiVincenzo Criteria

The DiVincenzo Criteria formulate requirements for the physical implementation of quantum computing 15.

1. A scalable physical system with well characterized qubits;
2. The ability to initialize the state of the qubits to a simple fiducial state, such as $|000\cdots\rangle$;
3. Long relevant decoherence times, much longer than the gate operation time;
4. A universal set of quantum gates such as single qubit rotations, C-Not C-Phase, see earlier section;
5. A qubit-specific measurement capability;
6. The ability to interconvert stationary and flying qubits;
7. The ability faithfully to transmit flying qubits between specified locations.

Comments about 'long relevant decoherence times':

Coherence times for qubits are characterized by the timescales: (1) for a change in the probability of occupation of either qubit state; and
(2) for a randomization of the phase in superposition states.



state: $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$

density matrix:

$$\rho = |\psi\rangle\langle\psi| = \begin{pmatrix} |\alpha|^2 & \alpha\beta^* \\ \alpha^*\beta & |\beta|^2 \end{pmatrix}$$

timescale T_1 characterizes changes in $|\alpha|^2 + |\beta|^2$;
timescale T_2 characterizes changes in $\alpha\beta^*$ and $\alpha^*\beta$ (loss of purity);
usually $T_2 < T_1$.

For ensemble measurements (e.g. repeated measurements with fluctuating parameters or multiple qubits in inhomogeneous environments), the system may appear to decohere, due to averaging on a timescale $T_2^* < T_2$.

8.6.2 Possible qubits

- Neutral atoms (1 electron outside closed shell, as in Alkali atoms e.g., Rb, Li, K, Cs,....; qubits are encoded on long-lived hyperfine states);
- Trapped ions
- Optical lattices (dipole traps for single atoms can be used to trap individual atoms for quantum computing purposes, Optical lattices allow preparation of a whole register at once, in contrast, e.g., to trapped ions);
- Colour centres (e.g., NV-centers in diamond);
- Quantum dots
- Superconducting qubits (charge, phase, flux)
- NMR (Nuclear Magnetic Resonance)
- Optical qubits
- Topological qubits

Bibliography

[1] C.H. Bennet and S.J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Phys. Rev. Lett.*, 69:2881 – 2884, 1992.

[2] C. H. Bennett. Quantum cryptography using any two nonorthogonal states. *Phys. Rev. Lett.*, 68:3121 – 3124, 1992.

[3] C. H. Bennett and G. Brassard. Proceedings of the IEEEInternational Conference on Computers, systems and signal processing, bangalore, india. In *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India*, pages 175 – 179, New York, 1984. IEEE.

[4] J.-W. Pan K. Mattle M. Eibl H. Weinfurter Bouwmeester, D. and A. Zeilinger. Experimental quantum teleportation. *Nature (London)*, 390:575, 1997.

[5] S.L. Braunstein and H.J. Kimble. Teleportation of Continuous Quantum Variables. *Phys. Rev. Lett.*, 80:869–872, 1998.

[6] A.R. Calderbank and B.P. W. Shor. Good quantum error-correcting codes exist. *Phys. Rev. A*, 54:1086, 1996.

[7] C.H. Bennett, C. Crépeau, R. Jozsa, A. Peres, G. Brassard and W.K. Wootters. Teleporting unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.*, 70:1895, 1993.

[8] D. Deutsch. Quantum theory, the Church-Turing Principle and universal quantum computer. *Proc. R. Soc. London A*, 400:11–20, 1985.

[9] D. Deutsch. Conditional quantum dynamics and logic gates. *Phys. Rev. Letters*, 74:4083–4036, 1995.

[10] A. Ekert and R. Jozsa. Quantum computation and Shor's factoring algorithm. *Reviews of Modern Physics*, 68:733 – 753, 1996.

[11] R. P. Feynman. Quantum mechanical computers. *Optics News,*, 11(2):11–20, 1985.

[12] Grover L.K. A fast quantum mechanical algorithm for database search. In *Proceedings, 28th Annual ACM Symposium on the Theory of Computing*, page 212, 1996.

[13] P. W. Shor. Algorithm for quantum computation: discrete logarithms and factoring. In *Proc. 35th Annual Symposium on Foundation of Computer Science*,, Los Alamitos CA, 1994. IEEE Press.

[14] L. Vaidman. Teleportation of quantum states. *Phys. Rev. A*, 49:1473 – 1476, 1994.

[15] P.D. DiVincenzo. The Physical Implementation of Quantum Computation. *Fortschritte der Physik*, 48:771, 2000.

[16] R.F. Werner. All teleportation and dense coding schemes. *J. Phys. A: Math. Gen.*, 34:7081, 2001.