Quantum information

Lecture notes

Štěpán Holub

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

Contents

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

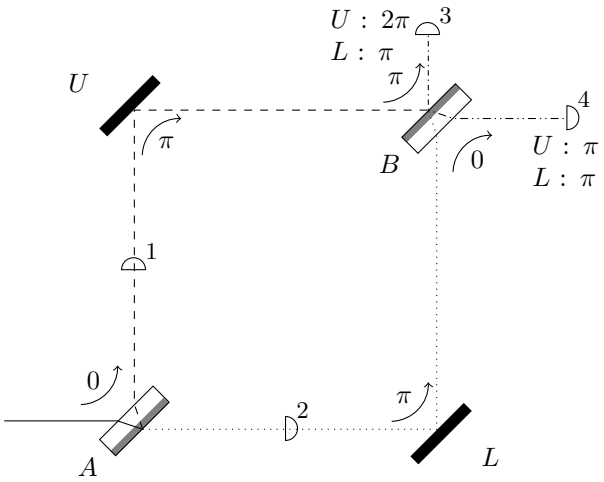


Figure 1: Mach-Zehnder interferometer.

1 Quantum mechanics

1.1 Example of quantum behavior

The following example illustrates the peculiar behavior of light, which not only contradicts the classical notion of its wave character, but raises much deeper doubts about the relationship between theoretical description, measurement, and physical reality. The device is shown in Figure 1.

A and B are semi-transparent mirrors through which half of the arriving beam passes and half is reflected. Such a mirror is simply a glass plate on which a layer of a substance, such as aluminum, of appropriate thickness is applied on one side. U and L are ordinary mirrors (in the experiment they only serve to direct the beam, the effect itself does not depend on them). Numbers 1 to 4 indicate the positions of detectors (obviously, if we want the light to reach the mirror B, we must remove detectors 1 and 2).

Light that is reflected by a surface changes its phase depending on whether it is reflected by an environment with a lower or higher optical density (corresponding to the speed of light in that environment). If light is reflected from an optically denser environment, it changes its phase by π (which corresponds to a shift of half a wavelength); when reflected from an optically thinner environment, the phase does not change. (In addition, the passage of light through the glass shifts the phase by a small value of φ, which is irrelevant to the experiment and can be neglected.)

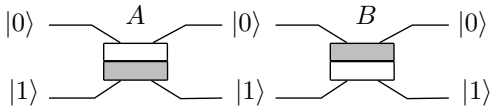
There are four options for the passage of light: A - U - B - 3, A - U - B - 4, A - L - B - 3 and A - L - B - 4. If we measure the intensity of light on detector 1 or 2, in both cases we measure half the input intensity, in accordance with

the assumed properties of the mirror. Measurements on detectors 3 and 4 (after removing detectors 1 and 2) show that there is no signal on detector 3, only on detector 4. This is due to light interference. Light traveling along the path A - U - B - 3 is canceled by destructive interference with light traveling along the path A - L - B - 3, because the first beam is shifted by half a phase relative to the second due to reflection on the mirror A. Conversely, there is constructive interference between the beams A - U - B - 4 and A - L - B - 4, because they were both reflected twice.

This is the classical wave description of the experiment. However, if the energy of light is reduced to a certain amount (i.e. a certain *quantum* denoted by the famous word *photon*), the light begins to behave as a particle in the sense that the signal is captured either on detector 1 or on the detector 2, the division into two half-intensity detection is excluded. The probability of both possibilities is equally one half. At first glance, this rehabilitates the old notion of light as a stream of particles. The classical measurement result can thus be interpreted as the fact that half of the photons pass and half are reflected. However, such an interpretation would require the photon, regardless of its trajectory, to behave in the same way on the B mirror. Each of the four paths would then have the same probability of one fourth, and detectors 3 and 4 should detect the signal both with a probability of one half (similarly to detectors 1 and 2). However, the result of the experiment turns out to be the same as in the classical case: no signal on the detector 3, signal always on the detector 4.

This mysterious phenomenon is an example of why we speak about the wave-corpuscular nature of light: it behaves partly as a wave and partly as a particle. But the problem is deeper. How can a photon, which is always captured one path only (if we decide to measure at 1 or 2), somehow interfere with itself if we delay the measurement?

We will show the answer of quantum mechanics. For clarity, we will forget about ordinary mirrors L and U and display the mirrors A and B schematically as follows:



The symbols $|0\rangle$ and $|1\rangle$ indicate the state of the photon at three moments in the experiment. If the diagram is related to the more detailed Figure 1, at the beginning of the experiment the photon is in the state $|0\rangle$. After interacting with the mirror A we say that the photon is in the state $|0\rangle$ if it travels through the mirror U, and after interacting with the mirror B it is in the state $|0\rangle$ if it is heading to detector 3. Similarly for states $|1\rangle$.

The basis of the quantum description of the experiment is the assumption that a photon can be in a *superposition* of states, which is mathematically expressed by their linear combination. Thus, although the photon did not divide, it is nevertheless in a state that indicates some kind of division. After passing

through the mirror A, the photon is in the state

$$\frac{|0\rangle + |1\rangle}{\sqrt{2}}.$$

(We divide the sum by $\sqrt{2}$ because we want to work with vectors of norm one.) If the photon coming into the mirror were in the state $|1\rangle$, i.e. from below, it would go into the state

$$\frac{|0\rangle - |1\rangle}{\sqrt{2}},$$

where $-1 = e^{\pi i}$ expresses the phase shift of π, caused by the reflection. Overall, therefore, the matrix of action of the mirror A can be expressed by a matrix

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

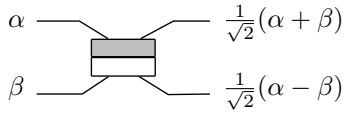
if we work in the basis

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

This matrix describes the effect of the mirror on photons in a superposition of basis states. Therefore, if a photon enters the mirror in the state $\alpha\,|0\rangle + \beta\,|1\rangle$, it leaves in the state

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

Schematically:



By the same considerations, we conclude that the action of the mirror B is expressed by the matrix

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

The overall effect of the system is then obtained by combining both mappings as

$$\frac{1}{\sqrt{2}}\begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \cdot \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

This explains the result of the experiment: a photon entering in the state $|0\rangle$ exits in the state $|1\rangle$. The missing negative sign, indicating the phase shift by

π, is caused by the fact that we neglected the ordinary mirrors U and L. The matrix of the action of these mirrors is obviously

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

which completes the description. It remains to explain the results of measurements on detectors 1 and 2. This is done using the quantum postulate of measurement, which is one of the strangest and most controversial aspects of the mainstream interpretation of quantum mechanics.

1.2 Postulates

Postulate 1. Associated to any isolated physical system is a unitary space known as the *state space* of the system. The system is completely described by its state vector, which is a unit vector in the system's state space.

Two vectors which differ just by a factor $e^{i\varphi}$, referred to as the *global phase*, are experimentally indistinguishable. In this sense, states are one dimensional spaces represented by a class of vectors of length one.

The natural basis of quantum informatics is a quantum system with two basis states, which are analogous to 0 and 1 used in classical information theory. Such a system is therefore called *qubit* and its basis states are denoted $|0\rangle$and $|1\rangle$. Taking into account the projective equivalence, the qubit is mathematically the *complex projective line* $\mathbb{P}^1(\mathbb{C})$. But we will more often denote it as $\mathbb{H}_2$ (thus ignoring the phase equivalence of states).

Postulate 2. The time evolution of the isolated quantum system $|u(t)\rangle$ is given by differential equation

$$i\hbar\frac{\partial}{\partial t}|u(t)\rangle = H|u(t)\rangle,$$

where $\hbar \in \mathbb{R}$ is the so-called reduced Planck constant and H is an Hermitian operator, called the *Hamiltonian* of the system.

This equation is called *Schrödinger's equation*. The physical significance of the Planck constant is the ratio between the energy and frequency of a photon. Since it is a real number, it is possible to omit it from the equation (and consider the Hamiltonian divided by this constant). Since Hamiltonian is Hermitian, Schrödinger's equation has a simple form for its eigenvectors (we omit the Planck constant)

$$\frac{\partial}{\partial t}|u(t)\rangle = -ir|u(t)\rangle,$$

where $r \in \mathbb{R}$ is the eigenvalue of the operator H. Assuming that the Hamiltonian does not change over time, it is easy to find a solution for the eigenvector $|u(t)\rangle$

$$|u(t)\rangle = e^{-irt}|u(t_0)\rangle.$$

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

Using our convention about functions of operators, we get a notation for the general vector $|v\rangle$

$$|v(t)\rangle = e^{-iHt}|v(t_0)\rangle.$$

It is easy to see that the operator e^{-iHt} has eigenvalues of size one (namely e^{-irt}), and is therefore unitary.

Because in quantum computers we want to perform precisely defined discrete operations on the input (on the input qubites), we can reformulate the second postulate in discrete form as follows:

Postulate 2'. The quantum state of an isolated quantum system $|\varphi\rangle$ changes during a time interval Δt to the state $U|\varphi\rangle$, where U is a unitary operator.

Postulate 3. A measurement is given by a Hermitian operator M, called *observable*. Let

$$M = \sum_i m_i P_i$$

be the spectral decomposition of M (i.e, m_i are the eigenvalues of M and P_i projections on the eigenspace corresponding to the eigenvalue m_i).

- The result of the measurement is one of the numbers m_i (which is real because the operator is Hermitian).
- The probability that the result of measuring the state $|\varphi\rangle$ will be $m_i$is equal to $\langle\varphi|P_i|\varphi\rangle$.
- If the result of the state measurement $|\varphi\rangle$ is equal to m_i, the system immediately after the measurement is in the state

$$\frac{P_i|\varphi\rangle}{\sqrt{\langle\varphi|P_i|\varphi\rangle}}$$

(we say the system *collapses* into this state).

This postulate describes the so-called *projective measurement* and does not describe the phenomenon of quantum measurement in general. For our purposes, however, this will be enough, moreover, it is true that each measurement can be converted to projective measurements with certain modifications. In the so-called *non-degenerate* case, the number of different eigenvalues is equal to the dimension of the system (there are no multiple eigenvalues) and all the mentioned subspaces are one-dimensional. Degenerate measurement is therefore characterized by the fact that the number of possible results is smaller than the dimension of the system, i.e. smaller than the measurement of other quantities. Note that we the dimension of the system is the maximum number of possible measurement results.

Note that the measurement is given by a set of projection operators P_i. Which one of them will be used is a random phenomenon determining the measurement result. The probability that the operator P_i will be used is given by the square of the size of the projection result, i.e. the square of the norm of the vector $P_i|\varphi\rangle$. This is equal to $\langle\varphi|P_i^\dagger P_i|\varphi\rangle$, which is equal to $\langle\varphi|P_i|\varphi\rangle$ since the projection is Hermitian and idempotent. Since $|\varphi\rangle = \sum_i P_i|\varphi\rangle$ holds, the sum of all probabilities is equal to one for a unit vector.

The result of the projection is standardized in the above formula by the square root of the probability of the result. Note that the normalization factor $|\varphi\rangle$ depends on the vector and causes the measurement to be a nonlinear mapping.

Each measurement captures some property of the system. The Hamiltonian, which occurs in the Schrödinger equation, for example, corresponds to the so-called total energy of the system (the time evolution of the system is therefore determined by this quantity).

Since the observable is Hermitian, it has an orthonormal basis of eigenvectors $|\mathbf{b}_i\rangle$. The projection on subspace P_i is then equal to

$$P_i = \sum_j |\mathbf{b}_j\rangle\langle\mathbf{b}_j|,$$

where we sum over all base vectors with eigenvalue m_i.

Writing the observable as one operator (i.e. not, for example, as a set of projections) enables, among other things, fast calculation of the mean value of the observable M on a specific state $|\varphi\rangle$ as

$$\mathbf{E}(M) = \sum_i m_i p(m_i) = \sum_i m_i \langle\varphi|\mathbf{b}_i\rangle\langle\mathbf{b}_i|\varphi\rangle = \langle\varphi|(\sum_i m_i|\mathbf{b}_i\rangle\langle\mathbf{b}_i|)|\varphi\rangle = \langle\varphi|M|\varphi\rangle.$$

Postulate 4. Let U and V be quantum systems. Then a system composed of U and V is described by the tensor product $U \otimes V$. If the system U is in the state $|u\rangle$ and the system V is in the state $|v\rangle$, then the state of the compound system is equal to $|u\rangle \otimes |v\rangle$.

2 Complex linear algebra

2.1 Complex unitary spaces

Complex unitary space of dimension n is a vector space $\mathbb{C}^n$ with scalar product. If $\alpha = a + bi$, $a, b \in \mathbb{R}$ we will denote the α^* number associated with α, i.e. $a - bi$.

Recall that the scalar product, which we will denote for a moment by the symbol $\odot$, is the mapping $\mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ satisfying the following relations:

- $u \odot (v + w) = u \odot v + u \odot w$;
- $u \odot (\alpha v) = \alpha(u \odot v)$;

- $v \odot u = (u \odot v)^*$;
- $u \odot u > 0$ pro $u \neq 0$.

The fourth condition silently assumes that $u \odot u$ is a real number, which is guaranteed by the third condition. The most important thing to realize is that the conditions yield $(\alpha u) \odot v = \alpha^*(u \odot v)$, so the scalar product **is not** linear in the first component. However, it is linear in the second component.

The *Hilbert space* of the dimension n, denoted by the symbol $\mathbb{H}_n$, is actually the n-dimensional complex unitary space. The difference between the terms unitary space and Hilbert space is given by the additional condition that Hilbert space must be complete with respect to the norm defined by the scalar product. However, this condition is always fulfilled for finite dimensional spaces, and therefore both concepts coincide on the finite dimension.

The fact that the variable u indicates an element of a vector space is sometimes referred to as $\vec{u}$. We will use the notation introduced by Dirac, common in quantum physics, which denotes the vector space element by the symbol $|u\rangle$.

As we have already said, the scalar product is linear in the second component, i.e. the mapping $\tilde{u} : \mathbb{C}^n \to \mathbb{C}$ given by the formula $\tilde{u}(v) = u \odot v$, is a linear form, or linear mapping from vector space to the field (or, equivalently, to one-dimensional vector space). Linear forms themselves form a vector space called *dual space*. Because it is, in matrix notation, a line vector space from $\mathbb{C}^n$, the dual space is isomorphic to $\mathbb{C}^n$, in which, by convention, we use column vectors. The dual vector $\tilde{u}$ to the vector u is written in Dirac notation as $\langle u|$. The origin of this notation is that the scalar product $u \odot v$ can now be written as $\langle u|v\rangle$ after omitting the $\odot$ sign, which is a notation commonly used for scalar product. The English word for the parentheses, *bracket*, gave rise to the designation *bra* -vector for elements $\langle u|$ of the dual space and *ket* -vector for elements $|v\rangle$ of the original space.

In finite-dimensional space, we are used to write vectors as n-tuples using their coordinates with respect to the chosen base. It is worth noting that in the case of an arithmetic vector space, such as $\mathbb{C}^n$, and with the choice of the canonical base $K = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$, a n-tuple understood as a vector is the same as a n-tuple understood as coordinates with respect to K. Formally,

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \left\{ \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \right\}_K .$$

The scalar product is easily expressed using coordinates in the orthonormal basis, i.e. in the basis $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ satisfying $\langle \mathbf{b}_i | \mathbf{b}_j \rangle = \delta_{ij}$. Then for

$$u = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

we have

$$\langle u| = (a_1^*, \ldots, a_n^*).$$

We can write this as $\langle u| = (|u\rangle^*)^{\mathrm{T}}$, which can be abbreviated as $\langle u| = |u\rangle^\dagger$.

We also have

$$\langle u|u\rangle = \sum_{i=1}^{n} |a_i|^2 .$$

Recall that the scalar product allows you to define the norm of the vector $\|u\|$ as $\sqrt{\langle u|u\rangle}$. Note that $|\alpha| = \|\alpha\|$, if we understand α once as a complex number and once as a one-dimensional vector.

2.2 Spectral properties of linear operators

Linear mappings (or homomorphisms) of vector spaces are also called (especially in physics) *linear operators*. Each operator $\varphi : \mathbb{C}^m \to \mathbb{C}^n$ can, as is well known, be represented as the multiplication by a matrix A of size $n \times m$. This matrix is given by the choice of bases M and N of spaces $\mathbb{C}^m$ and $\mathbb{C}^n$ and we have

$$A \cdot \{|u\rangle\}_M = \{\varphi|u\rangle\}_N.$$

It follows from the previous notation why we will understand vectors $\mathbb{C}^n$ as **columns, not rows**: it is more natural due to the convention that we multiply the vector by the matrix of the operator from the left. We are interested in matrices precisely because they are (along with multiplication) linear operators. So when we talk about a matrix, we mean the corresponding operator. Therefore, we will usually write $A|u\rangle$, instead of $A \cdot \{|u\rangle\}_M$.

For a operator φ we define the *adjoint* operator $\varphi^\dagger$ by the relation

$$\langle \varphi^\dagger(u)|v\rangle = \langle u|\varphi(v)\rangle,$$

where $\varphi^\dagger(u)$ is an abbreviation for $\varphi^\dagger|u\rangle$ for clarity, and $\varphi(v)$ for $\varphi|v\rangle$. This notation may be a bit confusing from a formal point of view (which physicians usually don't care so much about), but without Dirac's notation we can write it as

$$\varphi^\dagger(u) \odot v = u \odot \varphi(v).$$

It is not difficult to verify that in the matrix notation of the operator, the symbol † has the usual meaning of the Hermite-associated matrix (transposed and complex conjugated), which we already used above in the characterization of $\langle u|$. Especially in the context of quantum mechanics, the Hermite-associated matrix is simply called the *adjoint* matrix (although this term is often used in linear algebra for a matrix defined by subdeterminants).

The *eigenvalues* and *eigenvectors* are decisive for the properties of operators. The eigenvectors (which by definition are non-zero) determine one-dimensional subspaces that are mapped on themselves by the operator (they are therefore an invariant of the mapping). Therefore, if $|u\rangle$ is the eigenvector of an operator φ, then

$$\varphi|u\rangle = \lambda \cdot |u\rangle,$$

where λ is a complex number, called the eigenvalue corresponding to the (linear space spanned by the) vector $|u\rangle$. The set of eigenvalues is called the *spectrum* of the operator. The following list characterizes matrices of operators that have some nice spectral properties.

- The matrix A is *diagonalizable* if there exists a regular (i.e. invertible) matrix P such that $P^{-1}AP$ is diagonal. This occurs iff there is a basis of eigenvectors of the operator A. The matrix P is the matrix of the transition from the canonical basis to the basis of eigenvectors.

- The matrix A is called *normal* if the equality

  $$AA^\dagger = A^\dagger A,$$

  holds, that is, if the mapping commutes with its adjoint mapping. One of the most important theorems of the complex linear algebra is the **theorem on spectral decomposition of normal operators**, which says that an operator is normal if and only its eigenvectors form an orthonormal basis (since eigenvectors are given up to a scalar factor, it would be more accurate to say that it forms an orthogonal basis, which, however, can be converted into an orthonormal one by normalization). There are two important subclasses of normal matrices:

  – The matrix A is *Hermitian*, or *self-adjoint*, if

    $$A = A^\dagger.$$

    Hermitian matrices are obviously normal and have real eigenvalues.

  – The matrix U is called *unitary* if it preserves the scalar product. This is true when

    $$U^\dagger U = E,$$

    which is clearly shown by Dirac's notation:

    $$\langle u|v\rangle = \langle u|U^\dagger U|v\rangle.$$

    The equality $U^\dagger U = E$ also shows that the columns (rows) of the matrix U form an orthonormal basis. The unitary matrices are obviously normal.

  The theorem on spectral decomposition of normal operators can now also be formulated so that the operator is normal iff it is unitarily diagonalizable, i.e. when the corresponding transition matrix is unitary. This must be true because both the initial, i.e. canonical, and target bases of the eigenvectors are orthonormal. (The canonical base is orthonormal by definition; in other words, by convention, we always write operators in the base that is orthonormal in the given unitary space.)

Dirac notation provides an elegant notation for the projection operators P_v on the selected vector v. We have:

$$P_v = |v\rangle\langle v|.$$

The product of the arithmetic form (i.e. of the expression in coordinates) of the vectors $|v\rangle$ and $\langle v|$ in this order is a square matrix. That this is a projection operator can be seen from the formula

$$P_v|u\rangle = |v\rangle\langle v|u\rangle$$

and from the fact that the scalar product $\langle v|u\rangle$ determines the size of the projection of the vector u on the vector v.

It is also easy to see that each normal operator can be written as a linear combination of projections on its own vectors v_1, v_2, ..., v_n (forming an orthonormal basis). So

$$A = \sum_{i=1}^{n} a_i |v_i\rangle\langle v_i|,$$

where a_i is the eigenvalue of the corresponding vector v_i.

This also allows us to extend standard functions of complex numbers to operators. If $f : \mathbb{C} \to \mathbb{C}$ is a function, then $f(A)$ means the operator

$$f(A) = \sum_{i=1}^{n} f(a_i) |v_i\rangle\langle v_i|.$$

2.3 Tensor product and quantum registers

Tensor product of n-dimensional Hilbert space U with m-dimensional Hilbert space V is a bilinear mapping

$$\begin{aligned} U \times V &\to U \otimes V \\ (|u\rangle, |v\rangle) &\mapsto |u\rangle \otimes |v\rangle, \end{aligned}$$

where $U \otimes V$ is the Hilbert space generated by all images of this mapping, that is, by all elements $|u\rangle \otimes |v\rangle$. (The term "tensor product" is also commonly used to refer to the space $U \otimes V$ itself, and the element $|u\rangle \otimes |v\rangle$ is called the tensor product of vectors u and v). The scalar product is defined on $U \otimes V$ "component-wise", i.e. by extending the relation

$$\langle u_1 \otimes v_1 | u_2 \otimes v_2 \rangle := \langle u_1 | u_2 \rangle \langle v_1 | v_2 \rangle.$$

Bilinearity means that:

$$\begin{aligned} |w\rangle \otimes (|u\rangle + |v\rangle) &= |w\rangle \otimes |u\rangle + |w\rangle \otimes |v\rangle; \\ (|u\rangle + |v\rangle) \otimes |w\rangle &= |u\rangle \otimes |w\rangle + |v\rangle \otimes |w\rangle; \\ (\alpha|u\rangle) \otimes |v\rangle &= |u\rangle \otimes (\alpha|v\rangle) = \alpha(|u\rangle \otimes |v\rangle). \end{aligned}$$

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

We often shorten the tensor product of vectors $|u\rangle \otimes |v\rangle$ to $|u\rangle|v\rangle$ or even (especially for base vectors) to $|uv\rangle$.

If $|\mathbf{b}_i\rangle$, $i = 1, \ldots, n$, is a basis of U and $|\mathbf{c}_i\rangle$, $i = 1, \ldots, m$, a basis of V, then for the tensor product of vectors $|u\rangle \in U$ and $|v\rangle \in V$ we get

$$|u\rangle \otimes |v\rangle = \left(\sum_i \alpha_i |\mathbf{b}_i\rangle\right) \otimes \left(\sum_j \beta_j |\mathbf{c}_j\rangle\right) = \sum_{i,j} \alpha_i \beta_j |\mathbf{b}_i \mathbf{c}_j\rangle.$$

It can be seen that the space $U \otimes V$ is generated by vectors $|\mathbf{b}_i \mathbf{c}_j\rangle$. The definition of the tensor product is completed by the requirement that these vectors be linearly independent and thus form a basis of $U \otimes V$. From the point of view of universal algebra, the tensor product is the direct product of Hilbert spaces endowed with identities of bilinearity.

From the definition of the scalar product on the space $U \otimes V$ it follows that the base $|\mathbf{b}_i \mathbf{c}_j\rangle$ is orthonormal, and the scalar product on $U \otimes V$ satisfies

$$\left\langle \sum_{i,j} \alpha_{i,j} |\mathbf{b}_i \mathbf{c}_j\rangle \,\middle|\, \sum_{\ell,k} \beta_{\ell,k} |\mathbf{b}_\ell \mathbf{c}_k\rangle \right\rangle = \sum_{i,j} \alpha^*_{i,j} \beta_{i,j}.$$

We can make the tensor product of more than two spaces. Then we will require that the tensor product be associative, that is, that

$$(|u\rangle \otimes |v\rangle) \otimes |w\rangle = |u\rangle \otimes (|v\rangle \otimes |w\rangle),$$

which allows us to omit the parentheses and define tensor powers. In quantum informatics, mainly products of qubits, so-called *quantum registers*, are used. A quantum register $\mathbb{H}_2^{\otimes n}$ of n qubits has basis $|0\rangle \otimes \cdots \otimes |0\rangle$, $|0\rangle \otimes \cdots \otimes |1\rangle$, $\ldots$, $|1\rangle \otimes \cdots \otimes |1\rangle$, which according to the above convention can be shortened to $|0 \cdots 0\rangle$, $|0 \cdots 1\rangle$, $\ldots$, $|1 \cdots 1\rangle$. If we now understand zeros and ones as digits of binary notation, we get two different bases of size 2^n: one is the basis of the space $\mathbb{H}_2^{\otimes n}$, the other of the space $\mathbb{H}_{2^n}$. We thus obtain a natural tensor decomposition of the basis $|0\rangle$, $|1\rangle$, $\ldots$, $|2^n - 1\rangle$.

It is important to note that $U \otimes V$ also contains vectors that cannot be written as a tensor product of vectors from the original spaces. For example, the state $|00\rangle + |11\rangle$ is indecomposable; we have

$$(a|0\rangle + b|1\rangle) \otimes (c|0\rangle + d|1\rangle) = ac|00\rangle + ad|01\rangle + bc|10\rangle + bd|11\rangle$$

and it is easy to see that no a, b, c, d satisfy $c = bd = 1$ and $ad = bc = 0$. It is crucial for quantum phenomena that such *entangled* states of two or more systems are physically possible, the corresponding systems can even be spatially quite distant (e.g. by sending two entangled photons in different directions). The fact that spatially discontinuous particles can form a single system is called *nonlocal* character of quantum mechanics.

We can also make tensor products of operators. If $A : U_1 \to U_2$ and $B : V_1 \to V_2$ are two operators, their tensor product is a linear mapping $A \otimes B :$

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

$U_1 \otimes V_1 \to U_2 \otimes V_2$ defined by their values on the generating set of decomposable vectors as follows:

$$(A \otimes B)(|u\rangle \otimes |v\rangle) = (A|u\rangle) \otimes (B|v\rangle).$$

From the above properties of the scalar and tensor product, it is not difficult to verify that the tensor product of unitary operators is again unitary. The matrix of the operator $A \otimes B$ of the type $mp \times nq$ arises from the matrices A of the type $m \times n$ and B of the type $p \times q$ using the so-called Kronecker product, which is given as follows:

$$A \otimes B = \begin{pmatrix} a_{1,1}B & \dots & a_{1,n}B \\ \vdots & \ddots & \vdots \\ a_{m,1}B & \dots & a_{m,n}B \end{pmatrix}$$

For instance, for

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

we get

$$H^{\otimes 2} = \frac{1}{2} \left(\begin{array}{cc|cc} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ \hline 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{array} \right).$$

The operator H is called the *Hadamard* operator and we will later encounter its tensor powers. Let's see what the tensor power $H^{\otimes n}$ looks like. Its matrix is a square of size $2^n \times 2^n$ and if we factor out the coefficient $\left(\frac{1}{\sqrt{2}}\right)^n$, we get a matrix with entries 1 and -1. Let's index the rows and columns with the numbers $0, 1, \dots, 2^n - 1$ and look at the sign of the element $(H^{\otimes n})_{i,j}$. We can take advantage of the fact that the j-th column is a vector $H^{\otimes n}|j\rangle$. If we write j in binary, we get a tensor decomposition

$$\begin{aligned} H^{\otimes n}|j\rangle = H^{\otimes n}|j_1 j_2 \cdots j_n\rangle = H^{\otimes n}|j_1\rangle|j_2\rangle \cdots |j_n\rangle = \\ = \bigotimes_{k=1}^{n} H|j_k\rangle = \left(\frac{1}{\sqrt{2}}\right)^n \bigotimes_{k=1}^{n} (|0\rangle + (-1)^{j_k}|1\rangle). \end{aligned}$$

Multiplying out the last expression we find the required sign as a coefficient for the vector $|i\rangle = |i_1 i_2 \cdots i_n\rangle$. Minus signs are contributed to the product by the vectors $|i_k\rangle$ for which $i_k = j_k = 1$. Indeed, that's exactly when we take $|1\rangle$ from k-th expanded parentheses (because $i_k = 1$) and at the same time this $|1\rangle$ has a coefficient of -1 (because $j_k = 1$). Hence we have

$$(H^{\otimes n})_{i,j} = \left(\frac{1}{\sqrt{2}}\right)^n (-1)^{i_1 j_1 + i_2 j_2 + \cdots + i_n j_n} = \left(\frac{1}{\sqrt{2}}\right)^n (-1)^{i \cdot j},$$

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

where $i \cdot j$ denotes the dot product of the vectors of the binary expansion of digits i and j, i.e. the sum of $i_1 j_1 + i_2 j_2 + \cdots + i_n j_n$.

Note that if we understand the scalar product $\langle u|v\rangle$ as the application of the linear form $\langle u|$ on $|v\rangle$, the definition of the scalar product $\langle u_1 \otimes v_1 | u_2 \otimes v_2\rangle$ corresponds to the tensor product of the forms $\langle u_1|$ and $\langle u_2|$.

Note also that for endomorphisms A and B, the eigenvectors of the endomorphism $A \otimes B$ are vectors $|\mathbf{b}_i \otimes \mathbf{c}_j\rangle$ with eigenvalues $\lambda_i \cdot \kappa_j$ where $|\mathbf{b}_i\rangle$ is the eigenvector of A with eigenvalue λ_i and $|\mathbf{c}_j\rangle$ is the eigenvector of B with eigenvalue κ_j. Similarly, $|\mathbf{b}_i \otimes \mathbf{c}_j\rangle$ is an eigenvector of the endomorphism $A \otimes I + I \otimes B$ with eigenvalue $\lambda_i + \kappa_j$ (where I denotes identical operators of the appropriate size). These relations provide a handy proof of the commutative algebra fact that the integral elements of a ring form a ring.

2.4 Trace of a matrix and positive operators

The *trace* of a square matrix A is the sum of its diagonal elements. We denote it by $\mathrm{tr}(A)$. The trace satisfies the cyclic property:

$$\mathrm{tr}(AB) = \mathrm{tr}(BA).$$

Note that it does not follow, and in general does not hold, that $\mathrm{tr}(ABC) = \mathrm{tr}(BAC)$. However, it follows from here that $\mathrm{tr}(A) = \mathrm{tr}(QAQ^{-1})$ for any regular matrix Q. The trace is thus the same for similar matrices, which also suggests its importance: it is a property of the linear operator, not just of its particular matrix form. For example, if the operator A has a basis of eigenvectors, it is possible to sum the diagonal (in any basis) to obtain the sum of the eigenvalues (including multiplicity).

Another useful property of trace is the relation $\langle\psi|A|\psi\rangle = \mathrm{tr}\,(A|\psi\rangle\langle\psi|)$, which is valid for any unit vector $|\psi\rangle$, as can be easily seen from the equation

$$\mathrm{tr}\,(A|\psi\rangle\langle\psi|) = \sum_i \langle i|A|\psi\rangle\langle\psi|i\rangle = \langle\psi|A|\psi\rangle,$$

where the sum runs over some orthonormal basis containing $|\psi\rangle$.

We call an operator A *positive* if for every vector $|\psi\rangle$, $\langle\psi|A|\psi\rangle \geq 0$ holds. (Note that this is shorthand for an operator, or matrix, that is positively semidefinite.) For the positive operator A, specifically, $\langle\psi|A|\psi\rangle$ is a real number for every $|\psi\rangle$. It follows that the positive operators are Hermitian (see note below).

It is easy to verify that operators of the form $A^\dagger A$ are positive, and thus in particular also all projections $|\psi\rangle\langle\psi|$. Such operators have therefore diagonal form with non-negative eigenvalues. The diagonal form of $|\psi\rangle\langle\psi|$ is a matrix with a single non-zero coefficient 1 on the diagonal corresponding to the eigenvector $|\psi\rangle$.

Remark 2.1. For diagonalizable positive operators, it is obvious that they are Hermitian. For a general operator, the positivity implies $\langle x|A|x\rangle \in \mathbb{R}$ for any $|x\rangle$, and hence $\langle x|A|x\rangle = \langle x|A^\dagger|x\rangle$. Now we can use the polarization relation:

$$\begin{aligned} 4\langle x|A|y\rangle = &\langle x+y|A|x+y\rangle - \langle x-y|A|x-y\rangle \\ &- i\langle x+iy|A|x+iy\rangle + i\langle x-iy|A|x-iy\rangle\,. \end{aligned}$$

We abuse Dirac notation and write $|x+\lambda y\rangle$ instead of $|x\rangle + \lambda|y\rangle$; and $\langle x+\lambda y|$ instead of $(|x\rangle + \lambda|y\rangle)^\dagger$.

Alternatively, note that $A = C+iD$, where C and D are Hermitian operators

$$C = \frac{A+A^\dagger}{2},\ D = \frac{iA^\dagger - iA}{2}\,.$$

Then $\langle x|A|x\rangle = \langle x|C|x\rangle + i\langle x|D|x\rangle$, where $\langle x|C|x\rangle, \langle x|D|x\rangle \in \mathbb{R}$. Thus $\langle x|D|x\rangle = 0$ for each x. Since D is Hermitian and hence diagonalizable, we get $D = 0$.

3 Black-box algorithms

3.1 Deutsch algorithm

Deutsch's algorithm is the simplest example of quantum computers being capable of computations that go beyond the capabilities of classical computers. Suppose that the function $f : \{0,1\} \mapsto \{0,1\}$ is given by some oracle (that is, a "black box", which returns the value $f(x)$ at the input x without revealing anything about how to calculate this value). The task is to decide whether f is constant or not. In the classical case, it is obvious that we have to perform two queries, i.e. to find out both values of the function f. On the other hand, notice that the question is about a single bit of information: "constant yes or no"? However, there is no way to ask the oracle just this question. This is exactly the point at which the quantum computer has the upper hand.

The situation becomes somewhat complicated by the question of what a quantum oracle should look like. It follows from the postulates of quantum mechanics that it should be some unitary transformation. The problem, however, is that the function f need not be injective, i.e. not regular, let alone unitary. The standard solution to this problem is to introduce an auxiliary cubit that represents the input value. The function f will therefore correspond to the two-bit operator U_f, which is defined for $x, y \in \{0,1\}$ by the relation

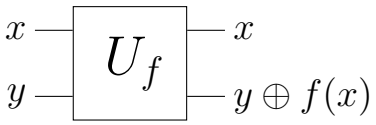
$$|x\rangle \otimes |y\rangle \overset{U_f}{\longmapsto} |x\rangle \otimes |y \oplus f(x)\rangle,$$

where the symbol $\oplus$ denotes a binary sum (sum in $\mathbb{Z}_2$). Note that the matrix U_f permutes the four basis states $|00\rangle, |01\rangle, |10\rangle, |11\rangle$, and is therefore obviously unitary (moreover, $U_f \circ U_f = \mathrm{Id}$).

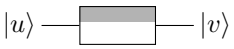
The construction of the quantum oracle above is the basis of the extended capabilities of the quantum algorithm. It is therefore easy to get the impression

that the quantum algorithm is more successful due to the more relaxed definition of the oracle. This impression is only partially justified. The quantum oracle has no advantage over the classical one in terms of the **basis states** $|0\rangle$ and $|1\rangle$, on which the function is defined. Extended capabilities are not so much given by the construction of the oracle as by the typically quantum fact that the oracle can also process **superpositions**. This involves some kind of "illegal" information about the inner workings of the oracle, namely that it behaves **linearly** with respect to state superpositions.

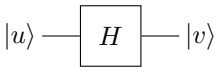
Following the example of algorithmic schemes, we can display U_f as a logical gate:



This notation should not be confused with the scheme we used for the beamsplitter in the description of the Mach-Zehnder interferometer. It was a single-cubic operator, which should be drawn as a gate as



where $|v\rangle = H|u\rangle$, or better yet

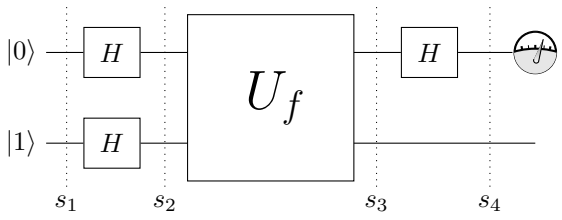


to make it clear that we don't care whether operator

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

is realized by a beamsplitter, or otherwise. As we have already pointed out, this H operator plays an important role in quantum computers and is called the *Hadamard gate*.

The quantum circuit implementing the Deutsch algorithm is relatively simple. It consists, in addition to the oracle U_f, of three Hadamard gates:



In the figure, the vertical lines indicate the four phases of the calculation. At the beginning, the two-bit register is in the state

$$s_1 = |01\rangle.$$

In the second phase we get

$$s_2 = \frac{|0\rangle + |1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}}.$$

The result of the oracle, of course, depends on the f function. The simplest case is $f(0) = f(1) = 0$, where U_f is the identity. Then we have

$$s_3 = \frac{|0\rangle + |1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}}.$$

If $f(0) = f(1) = 1$, the action U_f is given by the relation

$$|00\rangle \mapsto |01\rangle \qquad |01\rangle \mapsto |00\rangle \qquad |10\rangle \mapsto |11\rangle \qquad |11\rangle \mapsto |10\rangle.$$

So

$$s_3 = \frac{1}{2}U_f(|00\rangle - |01\rangle + |10\rangle - |11\rangle) = \frac{1}{2}(|01\rangle - |00\rangle + |11\rangle - |10\rangle) =$$
$$= -\frac{|0\rangle + |1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}}.$$

We can proceed similarly in other cases and get the overall expression

$$s_3 = \begin{cases} \pm\frac{|0\rangle+|1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle-|1\rangle}{\sqrt{2}}, & \text{if } f(0) = f(1), \\ \pm\frac{|0\rangle-|1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle-|1\rangle}{\sqrt{2}}, & \text{if } f(0) \neq f(1). \end{cases}$$

Finally

$$s_4 = \begin{cases} \pm|0\rangle \otimes \frac{|0\rangle-|1\rangle}{\sqrt{2}}, & \text{if } f(0) = f(1), \\ \pm|1\rangle \otimes \frac{|0\rangle-|1\rangle}{\sqrt{2}}, & \text{if } f(0) \neq f(1). \end{cases}$$

Now is the right time to **measure the first cubit**. The eigenvalue corresponding to $|0\rangle$ will mean that f is constant, the eigenvalue of $|1\rangle$ the opposite answer.

Deutch's algorithm, in its simplicity, shows the basic idea of all quantum algorithms: the superposition of states allows, in a sense, to compute many values simultaneously. Note that the Hadamard transform brings about the evaluation the balanced superposition of both values. This, on the other hand, does not mean that we have direct access to any functional value. For example, if we measure the first cubite in the phase s_3, we get a worthless random result, independent of the function f.

3.2 Deutsch-Jozsa algorithm

A more general form of the algorithm is called the *Deutsch-Jozsa algorithm*. There is a Boolean function $f : \{0,1\}^n \to \{0,1\}$, which is either constant or

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

balanced (i.e. exactly half of the arguments take the value 1 and the other half the value 0). The task is again to find out which of the options applies.

The circuit looks the same as in Deutsch's algorithm, only at the input there is a register $|0\rangle^{\otimes n}$ instead of $|0\rangle$ and also the corresponding Hadamard transformation of this register is the tensor product: $H^{\otimes n}$. We get a somewhat more complicated description of the individual phases. At the beginning we have the state

$$s_1 = |0^n 1\rangle.$$

and in the second phase

$$s_2 = \left(\frac{1}{\sqrt{2}}\right)^n \bigotimes_{i=1}^{n}(|0\rangle + |1\rangle)\left(\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right) = \left(\frac{1}{\sqrt{2}}\right)^n \left(\sum_{x\in\{0,1\}^n} |x\rangle\right)\left(\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right).$$

The case analysis from the Deutsch algorithm can be written succinctly. Note that

$$U_f\left(|x\rangle\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right) = |x\rangle\frac{|0 \oplus f(x)\rangle - |1 \oplus f(x)\rangle}{\sqrt{2}} = (-1)^{f(x)}|x\rangle\frac{|0\rangle - |1\rangle}{\sqrt{2}}.$$

So after the application of the oracle we get

$$s_3 = \left(\frac{1}{\sqrt{2}}\right)^n \left(\sum_{x\in\{0,1\}^n} (-1)^{f(x)}|x\rangle\right)\left(\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right).$$

Recall that

$$H^{\otimes n}|x\rangle = \left(\frac{1}{\sqrt{2}}\right)^n \sum_{z\in\{0,1\}^n} (-1)^{x\cdot z}|z\rangle,$$

where $x \cdot z = x_1x_2 \ldots x_n \cdot z_1z_2 \ldots z_n$ denotes the dot product of the vectors of the binary development digits, i.e.

$$\sum_{i=1}^{n} x_i z_i.$$

So for the final phase of the algorithm we get

$$s_4 = \frac{1}{2^n}\left(\sum_{x\in\{0,1\}^n}\sum_{z\in\{0,1\}^n} (-1)^{x\cdot z + f(x)}|z\rangle\right)\left(\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right).$$

What are the possible results of the measurement of the first register? Note that each basis state appears 2^n times in the sum, with different signs. However, the signs for the state $|0^n\rangle$ depend only on $f(x)$. Thus, if f is constant, the amplitude of the state $|0^n\rangle$ is equal to 1 or -1. Conversely, if f is balanced, the number of positive terms is the same as the number of negative ones and the amplitude of

the probability is 0. Therefore, the measurement result will correspond to the state $|0^n\rangle$ if and only if f is constant.

Note that a more general rule applies, which roughly states that the probability of measuring zero increases as the function f gets closer to a constant.

The Deutsch-Jozsa algorithm allows a correct answer after a single oracle query. This means an exponential speed up compared to the deterministic classical algorithm, which needs $2^{n-1}+1$ queries for a (certain) answer. However, there is a classical probabilistic algorithm with probability of error $\frac{1}{2^k}$, for which k queries are enough: after k random queries we answer "constant" just when all the results are the same. The possibility of error only exists for balanced functions and is obviously less than the required error bound. From this point of view, especially when we consider the susceptibility of quantum phenomena to errors, the acceleration of the quantum computer is only constant.

4 Quantum Fourier Transform

4.1 Multiplicative Characters and Discrete Fourier Transform

The most important quantum algorithm is the *Discrete Fourier transform* (DFT). There are two reasons for this:

- DFT is exponentially faster for quantum computers than for classical computers;
- DFT allows (among other things) to factorize of natural numbers.

The discrete Fourier deals with the mappings from a finite commutative group G to $\mathbb{C}$. Any such mapping f can be understood as a vector

$$(f(g_1), f(g_2), \ldots, f(g_n)),$$

where $n = |G|$ and g_i are elements of G. The set of all mappings thus forms the vector space $\mathbb{C}^n$, and the base to which the notation relates is the basis of the characteristic functions of individual elements, i.e. the functions $b_1, b_2, \ldots, b_n$ defined by the relation $b_i(g_j) = \delta_{ij}$ (where "the Kronecker delta" δ_{ij} is equal to 1 or 0 according to whether $i = j$ or not).

DFT is a transition from a " chronological " notation of a function in this base, to a notation in the base of the so-called group of *characters* G, which expresses the "frequency" decomposition of a function. To understand DFT, it is therefore necessary to first discuss characters of finite groups.

Let $(G, \cdot)$ be a commutative group. Each group homomorphism

$$\chi : (G, \cdot) \to (\mathbb{C}, \cdot)$$

is called a *character* of the group G. We shall further consider only finite groups G.

Characters also form a group, with multiplication defined by

$$(\chi_1 \cdot \chi_2)(g) = \chi_1(g) \cdot \chi_2(g).$$

The unit element of this group of characters is the identical one, which we denote by ε and we call it the *trivial* character.

Theorem 4.1. *Let X be the group of characters of a finite commutative group G. Then*

$$X \cong G.$$

Proof. Because G is finite, all its elements must be mapped to a unit circle. More precisely, the g element must be mapped to the r-th root of 1, where r is the order of g. So we have

$$\chi(g) = \exp\left[2\pi i \frac{k}{r}\right],$$

for some $k \in \mathbb{Z}_r$.

Let $\{h_1, \ldots, h_m\}$ be some minimal set of generators of the group G, where h_j has order r_j. Then

$$G \cong \mathbb{Z}_{r_1} \times \mathbb{Z}_{r_2} \times \cdots \times \mathbb{Z}_{r_m}$$

and the character χ is determined by the choice of

$$(k_1, \ldots, k_m) \in \mathbb{Z}_{r_1} \times \mathbb{Z}_{r_2} \times \cdots \times \mathbb{Z}_{r_m}$$

such that

$$\chi(h_j) = \exp\left[2\pi i \frac{k_j}{r_j}\right].$$

It is easy t see that the mapping $\chi \mapsto (k_1, \ldots, k_m)$ yields the required isomorphism. □

Since we move on a unit circle, we have

$$\chi^{-1}(g) = \chi(g)^{-1} = \chi(g)^*.$$

It is also useful to note that for $g, h \in \mathbb{Z}_{r_1} \times \mathbb{Z}_{r_2} \times \cdots \times \mathbb{Z}_{r_m}$ we have

$$\chi_h(g) = \chi_g(h).$$

Indeed, bth sides are equal to

$$\exp\left[2\pi i \sum_{j=1}^{m} \left(\frac{k_j \ell_j}{r_j}\right)\right], \qquad (\diamond)$$

where $g = (k_1, k_2, \ldots, k_m)$ a $h = (\ell_1, \ell_2, \ldots, \ell_m)$.

The following statement is crucial for computation with characters.

Lemma 4.1. *For any non-trivial character χ of the group G we have*

$$\sum_{g\in G}\chi(g)=0.$$

Proof. Let χ be nontrivial, and choose $h\in G$ such that $\chi(h)\neq 1$. Since $g\mapsto hg$ is a permutation of the group G, we have

$$\sum_{g\in G}\chi(g)=\sum_{g\in G}\chi(hg)=\chi(h)\sum_{g\in G}\chi(g),$$

hence

$$\sum_{g\in G}\chi(g)=0.$$

□

The following statement shows that characters form an orthogonal set with respect to the standard scalar product (so we work in the Hilbert space $\mathbb{H}_n$, not only in $\mathbb{C}_n$).

Lemma 4.2. *Let χ_1 a χ_2 be two distinct characters of the group G. Then*

$$\sum_{g\in G}\chi_1(g)^*\chi_2(g)=0.$$

Proof. Since $\chi_1\neq\chi_2$, the character $\chi_1^*\chi_2=\chi_1^{-1}\chi_2$ is nontrivial, and the claim follows from the previous lemma. □

The norm of each character χ is

$$\sqrt{\sum_{g\in G}\chi(g)^*\chi(g)}=\sqrt{n}.$$

We then see that the set

$$\left(\frac{1}{\sqrt{n}}\chi_1,\frac{1}{\sqrt{n}}\chi_2,\ldots,\frac{1}{\sqrt{n}}\chi_n\right)$$

is an orthonormal basis of $\mathbb{H}_n$, which we call the *basis of characters*.

As mentioned at the beginning, the Discrete Fourier Transform is the conversion of the representation $f:G\to\mathbb{C}$ from the notation in the basis of characteristic functions to the notation in the basis of characters. Because both bases are orthonormal, the operator is unitary. The DFT matrix is thus the inverse of the transition matrix from the canonical basis to the basis of characters. Since the inverse matrix of a unitary matrix is its adjoint matrix, we have

$$[\mathrm{DFT}]_{k,\ell}=\frac{1}{\sqrt{n}}\chi_{g_k}(g_\ell)^*,$$

where $g_1,g_2,\ldots,g_n$ is some numbering of elements of the group G. Due to the interchangeability of indices shown above, the DFT and DFT^{-1} are complex conjugate, and we usually considered the inverse transformation IFT to simplify the notation (this allows to omit minus signs in the exponent).

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

4.2 Quantum decomposition of the DFT

Quantum realization of the Discrete Fourier Transform consists in the construction of the circuit calculating the DFT operator, i.e. in the decomposition of DFT into small operators.

We have defined the DFT for a general group G. The most important and most common is the DFT for the cyclic group $(\mathbb{Z}_N, +)$, and unless explicitly stated otherwise, the term DFT means this case.

To illustrate the concept and to become familiar with it, however, we first perform the DFT on the group $(\mathbb{Z}_2^m, +)$. We have $M = 2^m$. The k-th - basis element of $\mathbb{H}_M$ is as usual denoted by $|k\rangle = |k_1 k_2\, dots k_m\rangle$, where $k_1 k_2 \ldots k_m$ is a binary expansion of k. We will also assume that the numbering of the group $\mathbb{Z}_2^m$ corresponds to this notation, so that the k-th element is just $(k_1, k_2, \ldots, k_m)$.

According to $(\diamond)$ we then have

$$[\mathrm{DFT}]_{k,\ell} = \frac{1}{\sqrt{2^m}} \exp\left[2\pi i \sum_{j=1}^{m} \frac{k_j \ell_j}{2}\right] = \frac{1}{\sqrt{2^m}} (-1)^{\sum_{j=1}^{m} k_j \ell_j} = \frac{1}{\sqrt{2^m}} (-1)^{k \cdot \ell}.$$

However, this is a matrix we already know from the Deutsch-Jozsa algorithm above; over $\mathbb{Z}_2^m$ we therefore get an easy decomposition

$$\mathrm{DFT} = H^{\otimes m}.$$

Let us now turn to the case $(\mathbb{Z}_M, +)$. We will use the remark at the end of the previous section and decompose IFT, where

$$[\mathrm{IFT}]_{k,\ell} = \frac{1}{\sqrt{M}} \exp\left[2\pi i \frac{k\ell}{M}\right].$$

he circuit is always defined on the basis elements. So we want to construct a circuit that maps the input $|k\rangle = |k_1\rangle|k_2\rangle \cdots |k_m\rangle$ to:

$$|k\rangle \mapsto \frac{1}{\sqrt{M}} \sum_{\ell=0}^{M-1} \exp\left[2\pi i \frac{k\ell}{M}\right] |\ell\rangle,$$

which after the decomposition

$$|\ell\rangle = \bigotimes_{j=1}^{m} |\ell_j\rangle, \qquad \frac{\ell}{M} = \sum_{j=1}^{m} \frac{\ell_j}{2^j}$$

yields

$$|k\rangle \mapsto \frac{1}{\sqrt{M}} \sum_{\ell_1=0}^{1} \sum_{\ell_2=0}^{1} \cdots \sum_{\ell_m=0}^{1} \bigotimes_{j=1}^{m} \exp\left[2\pi i \frac{k}{2^j} \ell_j\right] |\ell_j\rangle.$$

This can be decomposed as the product of m sums of two terms

$$|k\rangle \mapsto \frac{1}{\sqrt{M}} \bigotimes_{j=1}^{m} \sum_{\ell_i=0}^{1} \exp\left[2\pi i \frac{k}{2^j} \ell_j\right] |\ell_j\rangle = \bigotimes_{j=1}^{m} \frac{1}{\sqrt{2}} \left(|0\rangle + \exp\left[2\pi i \frac{k}{2^j}\right] |1\rangle\right).$$

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

The factor at $|1\rangle$ can be expanded as:

$$\exp\left[2\pi i\frac{k}{2^j}\right] = \exp\left[2\pi i\frac{\sum_{t=1}^{m} 2^{m-t}k_t}{2^j}\right] = \exp\left[2\pi i\sum_{t=1}^{m} 2^{(m-t-j)}k_t\right]$$

and from the periodicity of the exponential function we get

$$\exp\left[2\pi i\frac{k}{2^j}\right] = \exp\left[2\pi i\sum_{t=m-j+1}^{m} 2^{(m-t-j)}k_t\right]$$

To make the notation more readable, it is convenient to extend the binary expansion even beyond the decimal (or rather "binary") dot, and write

$$0, a_1a_2\cdots = \sum_j \frac{a_j}{2^j}.$$

The IFT can then be expressed as:

$$|k\rangle \mapsto \frac{|0\rangle + \exp\left[2\pi i(0,k_m)\right]|1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle + \exp\left[2\pi i(0,k_{m-1}k_m)\right]|1\rangle}{\sqrt{2}} \otimes \cdots$$
$$\otimes \frac{|0\rangle + \exp\left[2\pi i(0,k_2\cdots k_{m-1}k_m)\right]|1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle + \exp\left[2\pi i(0,k_1\cdots k_{m-1}k_m)\right]|1\rangle}{\sqrt{2}}.$$

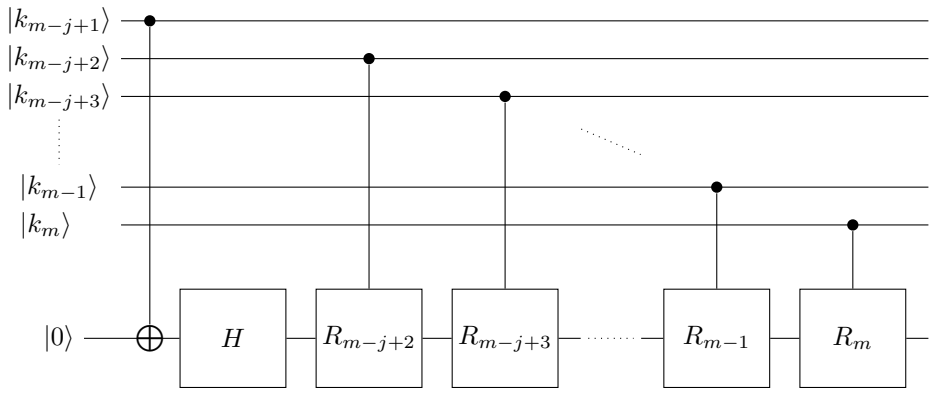
It is no more difficult to construct a circuit computing the IFT. First note that

$$\frac{|0\rangle + \exp\left[2\pi i(0,a)\right]|1\rangle}{\sqrt{2}} = H|a\rangle,$$

where H is the Hadamard operator. Moreover, we need the relative phase shift matrices

$$R_t = \begin{pmatrix} 1 & 0 \\ 0 & e^{2\pi i/2^t} \end{pmatrix},$$

which we apply controlled by the bit on the t-th position beyond the "binary" dot. The construction of the j-th output qubit now looks like this:

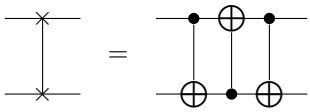


It is enough to use m auxiliary qubits, initially in the state $ket0$, as the output register of the transformation.

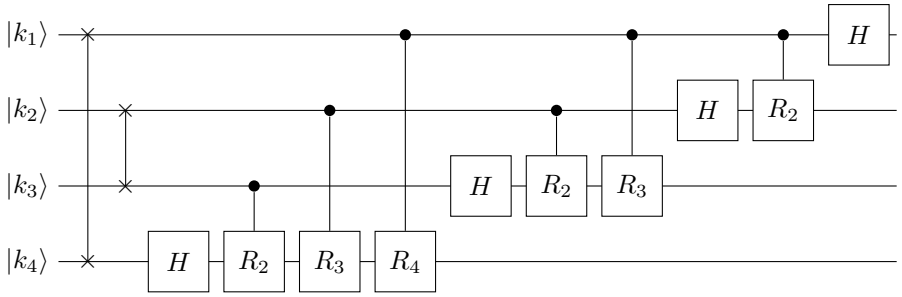
However, it is also possible to save auxiliary cubes if we notice that the first qubit of the input is needed only to calculate the n-th output qubit, the second qubit of the input only to calculate the last two output qubits, etc. Using this observation we can start with the first qubit of the output, and thus gradually construct in the j-th input qubit the j-th output counted from the back.

We then get the Fourier transform "upside down", which is certainly not a serious problem. If we also want to remove this inaccuracy, just reverse the order of the input qubits at the beginning using $\lfloor \frac{m}{2} \rfloor$ transpositions. The transposition of base qubits is, of course, unitary (like any permutation), it is denoted as

and it is easy to see that



The whole IFT circuit for $\mathbb{Z}_{2^4}$ is shown in the following picture.



It is obvious that the complexity of the algorithm (number of gates) is $\mathcal{O}(m^2) = mathcalO(\ log^2 M)$. The fastest classical algorithm, the so-called *fast Fourier transform*, has complexity of $\mathcal{O}(M \log M)$. In this case, therefore, quantum computers bring exponential speed up.

4.3 Examples of the Discrete Fourier Transform

Consider the mapping $k \mapsto 2^k \mod 15$ over the group $\mathbb{Z}_{16}$. This mapping in the canonical basis (the list of values) is

$$(1, 2, 4, 8, 1, 2, 4, 8, 1, 2, 4, 8, 1, 2, 4, 8)$$

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

It is a periodic vector whose period divides its length. So it's actually just a vector of length four, repeated several times. Its Fourier decomposition therefore contains only those Fourier base vectors which themselves have a period four. The expression in the Fourier base is:

$$(15, 0, 0, 0, -3-6i, 0, 0, 0, -5, 0, 0, 0, -3-6i, 0, 0, 0)\,.$$

The aperiodic vector

$$(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15)$$

has an ugly Fourier transform

$$\begin{array}{llll}
(30, & -2-10.05i, & -2-4.83i, & -2-2.99i, \\
-2-1.34i, & -2-0.83i, & -2-0.40i, & -2, -2+0.40i, \\
-2+1.34i, & -2+2i, & -2+2.99i, & -2+4.83i, \\
& & & -2+10.05i)\,.
\end{array}$$

A "partly periodic" vector is obtained in Shor's algorithm by factoring the number 21 from the mapping $k \mapsto 5^k \mod 21$ over the group $\mathbb{Z}_{32}$:

$$(1, 5, 4, 20, 16, 17, 1, 5, 4, 20, 16, 17, 1, 5, 4, 20,$$
$$16, 17, 1, 5, 4, 20, 16, 17, 1, 5, 4, 20, 16, 17, 1, 5)$$

It has a period six, which, however, does not divide the length of the vector. The Fourier coefficients are

$$\begin{array}{llll}
(56.75, & -2.75+0.11i, & -3.08+0.24i, & -3.87+0.44i, \\
-6.03+0.91i, & -21.74+3.97i, & 9.91-2.09i, & 3.65-0.85i, \\
2.12-0.53i, & 1.45-0.37i, & 1.09-0.28i, & 0.87-0.21i, \\
0.72-0.16i, & 0.63-0.11i, & 0.57-0.07i, & 0.54-0.03i, \\
-19.27, & 0.54+0.03i, & 0.57+0.07i, & 0.63+0.11i, \\
0.72+0.16i, & 0.87+0.21i, & 1.09+0.28i, & 1.45+0.37i, \\
2.12+0.53i, & 3.65+0.85i, & 9.91+2.09i, & -21.74-3.97i, \\
-6.03-0.91i, & -3.87-0.44i, & -3.08-0.24i, & -2.75-0.11i)\,.
\end{array}$$

5 Shor's factorization algorithm

The Shor factorization algorithm is the most important application of the quantum Fourier transform and one of the main reasons for the interest in quantum computers. The algorithm would allow probabilistic polynomial factorization of large numbers.

From the number theoretical point of view, this is nothing new: the basis of Shor's algorithm is Fermat's factorization algorithm, in which the factorization of . N is obtained from the knowledge of two numbers a, b, satisfying $a^2 \equiv b^2 \mod N$, thanks to the relation

$$(a+b)(a-b) \equiv 0 \mod N.$$

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

Fermat's procedure can be used, in particular, if we know some element a and its even order r in the multiplicative group $\mathbb{Z}_N$. Then we have

$$(a^{\frac{r}{2}} + 1)(a^{\frac{r}{2}} - 1) \equiv 0 \mod N,$$

which provides factorization of N if and only if $a^{\frac{r}{2}}$ is not equal to $-1 \mod N$. Thus, Shor's factorization algorithm for composite odd N looks like this:

- choose $a \in \mathbb{Z}_N^*$ at random (choosing a non-invertible element leads to a factorization immediately)
- find the order r of the element a in $\mathbb{Z}_N^*$
- if r is odd or if $a^{\frac{r}{2}} \equiv -1 \mod N$, then fail
- otherwise return a factor $\gcd(N, a^{\frac{r}{2}} - 1)$

We know from the number theory that the number of elements a that do not lead to failure is sufficient (at least one half). However, the impracticality of this algorithm stems from the fact that it is difficult to determine the order of the element in the group $\mathbb{Z}_N^*$. The quantum essence of Shor's algorithm is thus the search for the order of the element. For this task, the Fourier transform is suitable, and it is polynomial on a quantum computer.

5.1 Finding the order

The exponentiation of the element a modulo N, i.e. $k \mapsto a^k \mod N$, is the mapping $f : \mathbb{N} \to \mathbb{Z}_N^*$ with the period r. This gives a basic idea of why the Fourier transform can be useful for finding the order.

Quantum exponentiation must take place on finite binary registers. So let $n = \lceil \log N \rceil$ be the number of bits in the binary expansion of the number N, and choose some $M = 2^m$ large enough (the size of m will affect the probability of success of the algorithm).

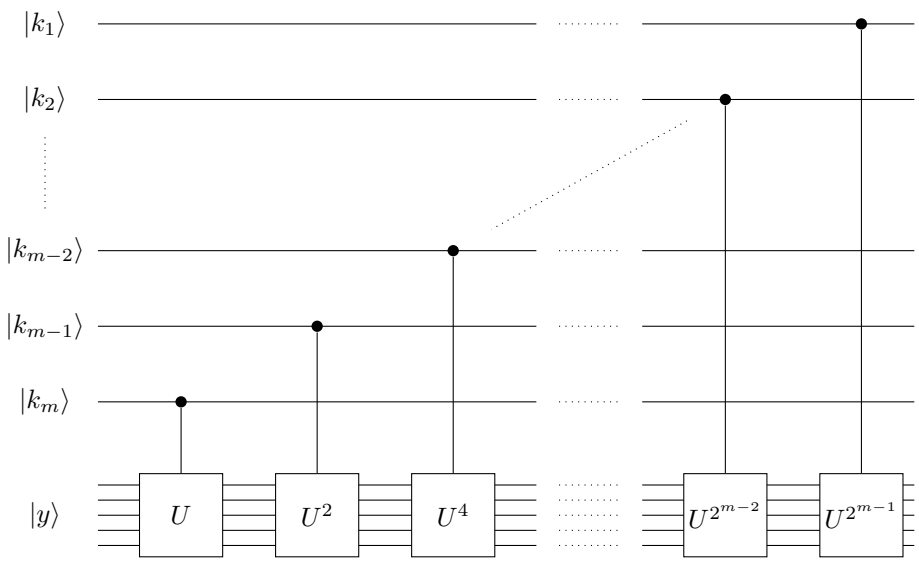
The exponentiation is simulated by the operator

$$\begin{aligned} W : \mathbb{H}_2^m \otimes \mathbb{H}_2^n &\to \mathbb{H}_2^m \otimes \mathbb{H}_2^n \\ |k\rangle|y\rangle &\mapsto |k\rangle|ya^k \mod N\rangle \end{aligned}$$

where for $N \leq y \leq 2^n - 1$, i.e. for elements for which the remainder would be repeated, we define $W|k\rangle|y\rangle := |k\rangle|y\rangle$. Because a relatively prime to N, the operator W permutes base elements and is therefore unitary. Implementation of the W operator is possible using modular exponentiation. If U is an operator for which we have controlled powers U^{2^j}, then the following circuit exponentiates U, that is, it realizes the mapping

$$|k\rangle|y\rangle \mapsto |k\rangle U^k|y\rangle,$$

in this way:

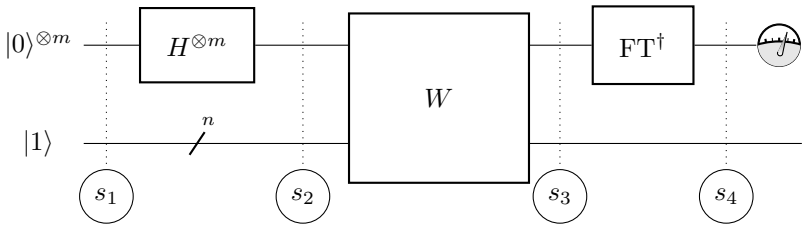


In the case of the operator W, U corresponds to the multiplication by the element a in the group $\mathbb{Z}_N$, i.e. the transformation

$$\begin{aligned} U : \mathbb{H}_2^n &\to \mathbb{H}_2^n \\ |y\rangle &\mapsto |ay \mod N\rangle, \end{aligned}$$

where again $U|y\rangle := |y\rangle$ pro $y \geq N$.

The basic idea of the order-revealing algorithm is the standard one: evaluate W on all values of $|k\rangle$ simultaneously. Because the exponentiation function is periodic, we apply the Fourier transform to it and we should get information about the period. The whole algorithm looks like this:



Note that the state $|1\rangle$ (or $|y\rangle$ for $y = 1$) is a base element of the n-qubit register with the number 1, ie $|0\rangle^{(n-1)}|1\rangle = |0\cdots01\rangle$. The first three phases give

$$s_1 : \ |0\rangle^{\otimes m}|0\cdots01\rangle \qquad s_2 : \ \frac{1}{\sqrt{M}}\sum_{k=0}^{M-1}|k\rangle|0\cdots01\rangle \qquad s_3 : \ \frac{1}{\sqrt{M}}\sum_{k=0}^{M-1}|k\rangle|a^k\rangle,$$

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

thereby preparing the desired uniform superposition of the values of the function $k \mapsto a^k$. By applying the Fourier transform to the first register we get

$$s_4 : \quad \frac{1}{M} \sum_{k}^{M-1} \sum_{z}^{M-1} \exp\left[2\pi i \frac{kz}{M}\right] |z\rangle |a^k\rangle$$

We will now measure the first register. The probability that the measurement result will correspond to some selected $|z\rangle$ is obtained according to the postulate of the measurement as a square of the size of the projection on the subspace of the result, i.e. on the vector of components containing $|z\rangle$. It is thus the sum of the squares of the magnitude of the probability amplitudes for all terms in which $|z\rangle$ occurs. There are r such terms, namely $|z\rangle|a^0\rangle$, $|z\rangle|a^1\rangle$, $\ldots$, $|z\rangle|a^{r-1}\rangle$, where the coefficient at $|z\rangle|a^t\rangle$ is the sum of the coefficients for all $|z\rangle|a^k\rangle$, where k is of the form $sr + t \mod N$. So all terms containing some fixed $|o\rangle f$ are

$$\frac{1}{M} \sum_{t=0}^{r-1} \left(\sum_{s=0}^{\ell_t} \exp\left[2\pi i \frac{(sr+t)z}{M}\right] \right) |z\rangle |a^t\rangle$$

and the corresponding probability is

$$\begin{aligned} P(z) &= \frac{1}{M^2} \sum_{t=0}^{r-1} \left| \sum_{s=0}^{\ell_t} \exp\left[2\pi i \frac{(sr+t)z}{M}\right] \right|^2 = \\ &= \frac{1}{M^2} \sum_{t=0}^{r-1} \left| \exp\left[2\pi i \frac{tz}{M}\right] \right|^2 \left| \sum_{s=0}^{\ell_t} \exp\left[2\pi i \frac{srz}{M}\right] \right|^2 = \\ &= \frac{1}{M^2} \sum_{t=0}^{r-1} \left| \sum_{s=0}^{\ell_t} \exp\left[2\pi i \frac{rz}{M} s\right] \right|^2 . \end{aligned}$$

The number ℓ_t is the largest such that $\ell_t r + t$ is less than M, i.e.

$$\ell_t = \left\lfloor \frac{M-1-t}{r} \right\rfloor .$$

The values of ℓ_t may differ by one for different t's. The complication is that r does not generally divide M; if it divided it, ℓ would simply be equal to $M/r - 1$. This irregularity is of deeper importance. Note that we are performing a Fourier transform on the group $\mathbb{Z}_M$, not $\mathbb{Z}_N$! The result will have some inaccuracy, because the mapping $k \mapsto a^k \mod N$ is not completely periodic on $\mathbb{Z}_M$: around zero, the periodicity is broken (if r does not divide M). For large M, however, this inaccuracy will be negligible.

These general considerations are specified in the calculation of the value of $P(z)$. We will show that the following is true

$$\left| \sum_{s=0}^{\ell_t} \exp\left[2\pi i \frac{rz}{M} s\right] \right| \approx \begin{cases} \frac{M}{r} & \text{if } rz \approx pM \text{ for some integer } p, \\ 0 & \text{otherwise.} \end{cases} \qquad (*)$$

In the above-mentioned ideal case, where r divides M, the sum above ranges over values of some character of the group $\mathbb{Z}_M$, and the relation $(*)$ therefore holds with equality in the place of $\approx$. So we will measure z, which is of the form $p \cdot \frac{M}{r}$, where $p \in \{0, 1, 2, \ldots, r-1\}$. For each such z, the probability $P(z)$ is equal to $\frac{1}{r}$, as is easily calculated. From z we obtain the fraction

$$\frac{z}{M} = \frac{p}{r},$$

whose denominator is r if p and r are disjoint. This occurs for $r > 19$ with a probability of at least $\frac{1}{4}\frac{1}{\log\log r}$. If p s r has a common factor, we get at least some factor of r. By repeating the procedure several times, we will most likely eventually obtain r.

In the general case, that is, if r does not divide M, the measured z is most likely close to some multiple of $\frac{M}{r}$, so that

$$\frac{z}{M} \approx \frac{p}{r}.$$

An interesting question arises as to how to find all fractions with a limited numerator that are close to a given value of α. The answer is the continued fraction expansion. It holds that if the distance between α and the fraction $\frac{p}{r}$ is less than $\frac{1}{2r^2}$, then this fraction is present in a continued fraction convergent of the number α (see the lecture in Czech on continued fractions within Number Theory and RSA, especially the application to Shor's algorithm on page 8, translated at the end of this chapter). If we assume that z is the rounded value of $p\frac{M}{r}$, that is, that

$$\left| z - p\frac{M}{r} \right| \leq \frac{1}{2},$$

then

$$\left| \frac{z}{M} - \frac{p}{r} \right| \leq \frac{1}{2M},$$

which leads to the choice of M to be approximately N^2 ensuring the detection of the corresponding $\frac{p}{r}$ using continued fractions.

It remains to show with what precision the estimate $(*)$ holds in these circumstances. Denote

$$\varphi = \frac{rz}{M} - p$$

the aproximation "error", which, according to our assumption, satisfies

$$|\varphi| \leq \frac{r}{2M}.$$

We approximate the sum of the geometric series:

$$\left| \sum_{s=0}^{\ell} \exp\left[2\pi i \frac{rz}{M} s\right] \right|^2 = \left| \sum_{s=0}^{\ell} \exp\left[2\pi i \varphi s\right] \right|^2 = \frac{\left|\exp\left[2\pi i \varphi(\ell+1)\right] - 1\right|^2}{\left|\exp\left[2\pi i \varphi\right] - 1\right|^2} = \frac{\sin^2 \pi\varphi(\ell+1)}{\sin^2 \pi\varphi},$$

where the last equality follows from the relation

$$\left|e^{ix}-1\right|^2 = (e^{ix}-1)(e^{-ix}-1) = 2(1-\cos x) = 4\sin^2\frac{x}{2}.$$

It is not difficult to verify that the value decreases with increasing φ, which is consistent with φ being a measure of inaccuracy: the maximum M/r is reached in our ideal case that corresponds to $\varphi = 0$. In addition, since $\sin^2$ is an even function, we get

$$\frac{\sin^2 \pi\varphi(\ell+1)}{\sin^2 \pi\varphi} \geq \frac{\sin^2 \frac{\pi}{2}\frac{r(\ell+1)}{M}}{\sin^2 \frac{\pi}{2}\frac{r}{M}}.$$

It follows from the definition of ℓ that $M - r < r(\ell+1) < M + r$. The numerator of the fraction is therefore very close to one (for $r/M < 1/100$ differs from one by less than a thousandth) a the denominator, which, on the other hand, is very small, can be estimated quite accurately from above relation $\sin x < x$. In total we get

$$\left|\sum_{s=0}^{\ell} \exp\left[2\pi i \frac{rz}{M} s\right]\right|^2 > 0.999 \cdot \frac{4}{\pi^2}\frac{M^2}{r^2} > \frac{2}{5}\frac{M^2}{r^2}$$

and

$$P(z) > \frac{2}{5}\frac{1}{r}.$$

We can conclude that with a probability of at least $\frac{2}{5}$ we measure z, for which is $\frac{p}{r}$ present in the continued fraction expansion of $\frac{z}{M}$.

The overall success rate of the algorithm is summarized in the following table:

success condition	probability
choosing a suitable a	$\frac{1}{2}$
z is close to $p\frac{M}{r}$	$\frac{2}{5}$
p is coprime with r	$\frac{1}{4}\frac{1}{\log\log n}$

So the total success rate is at least $\frac{1}{20}\frac{1}{\log\log n}$. E.g. for the RSA module of length 4096, the success rate of one round of the algorithm is at least 0.6%, so four hundred rounds gives more than 90% probability of success. This estimate is unnecessarily pessimistic especially in the requirement that r and $p$are coprime; even if r and p have common factors, we get some of them in each round and after several attempts it is likely to reconstruct r as the least common multiple of the factors found.

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

5.2 Example: RSA

Continued fractions are an effective tool for the rational approximation of irrational numbers. However, they are also important for the approximation of rational numbers. Suppose we have an inaccurate value of a fraction, caused by, for example, rounding or measurement inaccuracy. An example of such a situation is Shor's quantum factorization algorithm. To reveal the original fraction, we use the continued fraction expansion of an inaccurate value.

Example: We have the value $h = 0.15328$, which we know is the rounding (to the nearest hundredth of a thousand) of a proportion of at most eight-bit numbers. The continued fraction expansion of h is $[0, 6, 1, 1, 9, 1, 10]$ with convergents:

$$\left(0, \frac{1}{6}, \frac{1}{7}, \frac{2}{13}, \frac{19}{124}, \frac{21}{137}, \frac{229}{1494}, \frac{479}{3125}\right).$$

Of the fractions with a denominator and a numerator of at most eight bits, only $^{21}/_{137}$ is equal to h when rounded to the nearest hundredth of thousand.

zlomek	zaokrouhlení
$\frac{1}{6}$	0.16667
$\frac{1}{7}$	0.14286
$\frac{2}{13}$	0.15385
$\frac{19}{124}$	0.15323
$\frac{21}{137}$	0.15328

Of course, the question arises as to whether we have not missed a fraction with the same rounding in the continued fraction. The following statement is relevant to this question.

Theorem: If

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{2q^2},$$

then the fraction $^p/_q$ is a convergent of α.

In the above example, the denominator is less than 256 and the rounding error is at most $5 \cdot 10^{-6}$. Because

$$5 \cdot 10^{-6} < \frac{1}{2 \cdot 256^2},$$

we see that the fraction sought is indeed one of the convergents.

6 Complex projective line

6.1 Representaions

According to the postulates of quantum mechanics,a qubit is an element of $\mathbb{C}^2$ of size one, and it is possible to ignore the global phase. Thus, a qubit can be

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

understood as an element of a one-dimensional complex projective space $\mathbb{C}\mathbf{P}^1$. By definition, the elements $\mathbb{C}\mathbf{P}^1$ are a pair of complex numbers (*alpha*, *beta*) with equivalence

$$(\alpha, \beta) \sim \lambda(\alpha, \beta), \qquad \lambda \in \mathbb{C}.$$

We usually represented a qubit by any vector of magnitude one with preserved ambiguity regarding the global phase, i.e. regarding multiplication by a complex unit.

If we want to represent the element of a complex line unambiguously, several possibilities are available:

1. The pair (α, β), where we expand the requirement of the unit norm $\alpha\alpha^* + \beta\beta^* = 1$ by the assumption that α is real and non-negative. With this assumption, we actually choose the global phase, except for the situation where $\alpha = 0$, for which we choose the pair $(0, 1)$. Note that for any unit vector (α, β) one can find $\psi \in [0, \pi/2]$ such that $|\alpha| = \cos\psi$ a $|\beta| = \sin\psi$. Our choice of representative can then be written as $(\cos\psi, e^{-i\varphi}\sin\psi)$, where $\varphi \in [0, 2\pi)$ is given uniquely except for the case $(1, 0)$, where we put $\varphi = 0$ (similarly, the above convention selects $\varphi = 0$ for $(0, 1)$).

Remark 6.1. In the literature, φ is often chosen so that the representative is $(\cos\psi, e^{i\varphi}\sin\psi)$. We violate this convention to bring it into line with the usual concept of stereographic projection below.

2. Qubit can therefore also be represented by the pair (ψ, φ), which can be understood as polar coordinates of one half of a unit sphere. To extend such a representation to the whole sphere, let's put $\vartheta = 2\psi$. Then we have

$$(\alpha, \beta) = \left(\cos\frac{\vartheta}{2},\, e^{-i\varphi}\sin\frac{\vartheta}{2}\right)$$

and qubits uniquely correspond to the set of pairs (ϑ, φ), $\vartheta \in (0, \pi)$ and $\varphi \in [0, 2\pi)$ extended by $(0, 0)$, $(\pi, 0)$, according to the above conventions.

It turns out that the complex projective line can be represented by a real unit sphere $\mathbb{S}^2$ (in mathematics we speak of the Riemann sphere, in quantum physics we speak of the Bloch sphere).

3. Kubit lze tedy také reprezentovat trojicí reálných čísel (x, y, z), splňujících $x^2 + y^2 + z^2 = 1$, které představují standardní kartézské souřadnice Blochovy sféry. Vztah k úhlovým souřadnicím je za předpokladu, že úhel φ měříme v rovině os x a y počínaje osou x a úhel ϑ měříme od osy z, dán jako

3. Qubit can therefore also be represented by three real numbers (x, y, z), satisfying $x^2 + y^2 + z^2 = 1$, which represent the standard Cartesian coordinates of the Bloch sphere. The relation to polar coordinates is given as

$$(x, y, z) = (\cos\varphi\sin\vartheta, \sin\varphi\sin\vartheta, \cos\vartheta)\,,$$

under the assumption that we measure the angle φ in the plane of the axes x and y starting from the axis x and measuring the angle ϑ from the axis z.

4. The pair (α, β) can finally be replaced by a number

$$\frac{\alpha}{\beta},$$

which is a representative of the projective line

$$\left(\frac{\alpha}{\beta}, 1\right),$$

where $(1,0)$ is the improper point of the projective line, naturally called ∞. The qubits are thus represented by the *extended complex plane* $\mathbb{C} \cup \{\infty\}$.

In total, we have the following representations for elements of $\mathbb{C}\mathbf{P}^1$:

$$\left(\cos\frac{\vartheta}{2},\, e^{-i\varphi}\sin\frac{\vartheta}{2}\right) \in \mathbb{C}^2,$$
$$(\cos\varphi\sin\vartheta, \sin\varphi\sin\vartheta, \cos\vartheta) \in \mathbb{S}^2,$$
$$e^{i\varphi}\cot\frac{\vartheta}{2} \in \mathbb{C}\cup\{\infty\}\,.$$

From the point (x,y,z) on the Bloch sphere, the corresponding element of the extended complex plane can be obtained by the *stereographic projection*, let's denote it $\mathcal{S}$, where each point on the sphere corresponds to its image projected from the North Pole to the plane given by the axes x and y understood as a complex plane with the imaginary axis y (we assign the point ∞ to the north pole itself). From the similarity we simply see that

$$\mathcal{S}(x,y,z) = \frac{x+iy}{1-z}\,.$$

For $(\alpha,\beta) = \left(\cos\frac{\vartheta}{2},\, e^{-i\varphi}\sin\frac{\vartheta}{2}\right)$ a $(x,y,z) = (\cos\varphi\sin\vartheta, \sin\varphi\sin\vartheta, \cos\vartheta)$ we therefore have

$$\mathcal{S}(x,y,z) = \frac{x+iy}{1-z} = \frac{e^{i\varphi}\sin\vartheta}{1-\cos\vartheta} = e^{i\varphi}\frac{2\sin\frac{\vartheta}{2}\cos\frac{\vartheta}{2}}{2\sin^2\frac{\vartheta}{2}} = e^{i\varphi}\cot\frac{\vartheta}{2} = \frac{\alpha}{\beta}\,.$$

Then the following diagram of qubit representations commute:

$$\begin{array}{ccc} (\alpha,\beta) & \longrightarrow & \left(\cos\frac{\vartheta}{2},\, e^{-i\varphi}\sin\frac{\vartheta}{2}\right) \\ \downarrow & & \downarrow \\ \frac{\alpha}{\beta} & \xleftarrow{\;\mathcal{S}\;} & (\cos\varphi\sin\vartheta, \sin\varphi\sin\vartheta, \cos\vartheta) \end{array}$$

The picture 2 shows the inverse stereographic projection $P = \mathcal{S}^{-1}$. It is given by:

$$\mathcal{S}^{-1} : a+bi \mapsto \left(\frac{2a}{a^2+b^2+1}, \frac{2b}{a^2+b^2+1}, \frac{a^2+b^2-1}{a^2+b^2+1}\right).$$

Finally, for a unit (α,β), the corresponding point on the Bloch sphere can be obtained directly as

$$(\mathrm{Re}(2\alpha^*\beta), \mathrm{Im}(2\alpha^*\beta), |\alpha|^2 - |\beta|^2).$$

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

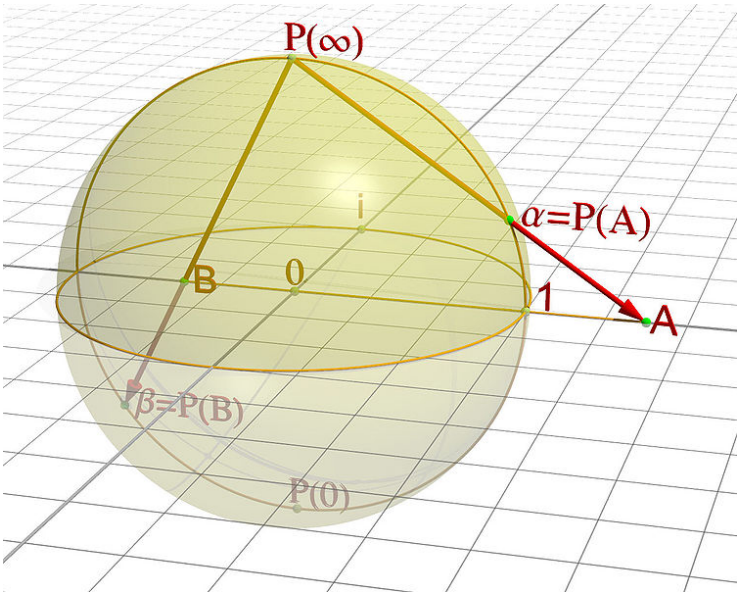


Figure 2: Inevrse stereographic projection (from Wikipedia)

5. The last important representation of the qubit ψ is using the projection operator $|\psi\rangle\langle\psi|$. It is also called *density operator* in quantum mechanics and plays an important role when working with so-called mixed systems. Note first that the density operator actually represents a unique representative, because two states differing by the global phase have the same operator. At the same time, it has decomposition using Pauli matrices, which is related to the above geometric representations. Indeed, for

$$|\psi\rangle = \begin{pmatrix} \cos\frac{\vartheta}{2} \\ e^{i\varphi}\sin\frac{\vartheta}{2} \end{pmatrix}$$

we have

$$|\psi\rangle\langle\psi| = \frac{1}{2}\begin{pmatrix} 1+\cos\vartheta & e^{-i\varphi}\sin\vartheta \\ e^{i\varphi}\sin\vartheta & 1-\cos\vartheta \end{pmatrix} = \frac{1}{2}(E + xX + yY + zZ),$$

where

$$\begin{cases} x = \sin\vartheta\cos\varphi, \\ y = \sin\vartheta\sin\varphi, \\ z = \cos\vartheta\,. \end{cases}$$

Therefore, if we put $\sigma = (X, Y, Z)$, we can write

$$|\psi\rangle\langle\psi| = \frac{1}{2}(E + r_\psi \cdot \sigma),$$

where r_ψ is the representative $|\psi\rangle$ on the Bloch sphere.

6.2 The Hopf fibration

If

$$\alpha = a + bi, \quad \beta = c + di,$$

then the unit vector (α, β) corresponds to

$$\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} \in \mathbb{R}^4,$$

which is an element of $\mathbb{S}^3$, (three-dimensional) sphere in $\mathbb{R}^4$. This representation is not unique, we identify vectors that differ by a global phase. One can easily verify that $(e^{i\varphi}\alpha, e^{i\varphi}\beta)$ correspond to points

$$\begin{pmatrix} \cos\varphi & -\sin\varphi & 0 & 0 \\ \sin\varphi & \cos\varphi & 0 & 0 \\ 0 & 0 & \cos\varphi & -\sin\varphi \\ 0 & 0 & \sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix},$$

which form in $\mathbb{R}^4$ a circle $\mathbb{S}^1$ around the origin. This circle corresponds to a single point on the Bloch sphere. We thus get the mapping

$$\mathbb{S}^1 \hookrightarrow \mathbb{S}^3 \twoheadrightarrow \mathbb{S}^2,$$

called the *Hopf fibration*. In words, the three-dimensional sphere decomposes into (disjoint) circles, which correspond to points on two-dimensional sphere.

7 Projective unitary operators

7.1 Quaternions

Recall that quaternions are a four-dimensional algebra (that is a vector space with a distributive vector multiplication) $\mathbb{K}$ over real numbers generated by the elements $\{1, \ell, j, k\}$, which satisfy

$$\ell^2 = j^2 = k^2 = \ell jk = -1.$$

First of the imaginary generators is usually denoted as i, but to avoid the confusion caused by identifying with a complex unit, we will use ℓ. Multiplying the equality $\ell jk = -1$ from both sides by k we get $k\ell j = -1$. Similarly, $jk\ell = -1$. Imaginary generators are therefore cyclically interchangeable. Multiplying by only one k we also get $\ell j = k$, and symmetrically $jk = \ell$ and $k\ell = j$. Further, multiplying $\ell j = k$ by ℓ from the left, we get $j = -\ell k$ and similarly $k = -j\ell$ a $kj = -\ell$.The generators are therefore anti-commutative. However, each quaternion obviously commutes with a real number (which is itself a quaternion).

For $q = a + b\ell + cj + dk$ we define the *adjoint element* $q^* = a - b\ell - cj - dk$.

Norm of q is defined as $N(q) := qq^* = a^2 + b^2 + c^2 + d^2 = |q|^2$, where $|q|$ is the Euclidean norm in $\mathbb{R}^4$. The sphere $\mathbb{S}^3$ is therefore naturally identified with unit quaternions $\mathbb{K}_1$ (that is, quaternions of norm one).

We have $(pq)^* = q^*p^*$. This implies $N(pq) = N(p)N(q)$, and unit quaternions form a multiplicative group. Thus, the inverse element of the quaternion q has the form $q^{-1} = q^*/N(q)$, or $q^{-1} = q^*$ for unit quaternions.

Quaternions of the form $b\ell + cj + dk$ are called *imaginary*. Unit imaginary quaternions can be identified with the sphere $\mathbb{S}^2$ and they satisfy $p^2 = -1$ (similarly as the generators), because $p^{-1} = -p$.

We will now show the most important property of quaternions. Conjugation of an imaginary quaternion by any quaternion corresponds to the rotation of three-dimensional space.

Theorem 7.1. *For $0 \neq q = (r + x\ell + yj + zk) \in \mathbb{K}$, the mapping*

$$\begin{aligned} \rho_q : \mathbb{R}^3 &\to \mathbb{R}^3 \\ (b, c, d) &\mapsto (b', c', d') \end{aligned}$$

defined by

$$b'\ell + c'j + d'k = q(b\ell + cj + dk)q^{-1}$$

is the rotation around the axis passing through the point (x, y, z) by the angle

$$\omega = 2 \arccos \frac{r}{\sqrt{N(q)}}\,.$$

Proof. Since $qpq^{-1} = (tq)p(tq)^{-1}$ arbitrary real t, we can w.l.o.g. assume, that q is a unit quaternion and $qpq^{-1} = qpq^*$.

Conjugation is an automorphism of $\mathbb{K}$. In addition, it is an identity on real numbers, because a real number commutes with any quaternion. Moreover,

$$N(qpq^{-1}) = N(q)N(p)N(q^{-1}) = N(p).$$

Thus, conjugation can be understood as an orthonormal mapping $\mathbb{R}^4$, preserving the first coordinate. Therefore it is also orthonormal on the orthogonal complement of the first component. Let $q = r + v$, that is, v is the imaginary part of q. Then

$$qvq^* = (r + v)v(r - v) = (r + v)(rv - vv) = (r + v)(r - v)v = N(q)v = v\,.$$

We can see that ρ_q is an isometry with a fixpoint (x, y, z).

Let us write q as

$$q = \cos\frac{\omega}{2} + \sin\frac{\omega}{2}(\ell \sin\vartheta\cos\varphi + j\sin\vartheta\sin\varphi + k\cos\vartheta)\,,$$

where

$$v' = \ell \sin\vartheta\cos\varphi + j\sin\vartheta\sin\varphi + k\cos\vartheta$$

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

is a unitary imaginary quaternion expressing the axis of rotation using its polar coordinates. Denote

$$\kappa = \cos\frac{\omega}{2} + \sin\frac{\omega}{2}k,$$

which is the case with $v' = k$. Direct calculation of images $\rho_\kappa(\ell)$, $\rho_\kappa(j)$ and $\rho_\kappa(k)$ yields

$$[\rho_\kappa]_{\ell,j,k} = \begin{pmatrix} \cos\omega & -\sin\omega & 0 \\ \sin\omega & \cos\omega & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and the theorem holds for this particular case.

Similarly (or from symmetry) we get validity for the cases $v' = \ell$ and $v' = j$, ie for rotations around the second and third axes $\mathbb{R}^3$.

Consider now quaternions

$$q_\varphi = \cos\frac{\varphi}{2} + k\sin\frac{\varphi}{2}\,,$$
$$q_\vartheta = \cos\frac{\vartheta}{2} + j\sin\frac{\vartheta}{2}\,.$$

Their action corresponds to the respective rotations, so

$$q_\varphi q_\vartheta k q_\vartheta^* q_\varphi^* = v'\,,$$

and thus

$$q_\varphi q_\vartheta \kappa q_\vartheta^* q_\varphi^* = q\,.$$

From here we deduce

$$qpq^* = q_\varphi(q_\vartheta(\kappa(q_\vartheta^*(q_\varphi^* p q_\varphi)q_\vartheta)\kappa^*)q_\vartheta^*)q_\varphi^*$$

that is

$$\rho_q = \rho_\varphi \circ \rho_\vartheta \circ \rho_\kappa \circ \rho_\vartheta^{-1} \circ \rho_\varphi^{-1}\,,$$

and ρ_q is the mapping similar to ρ_κ, in other words, it is a rotation by the angle ω with respect to different othonormal basis. In particular

$$[\rho_q]_{\rho_\varphi^* \circ \rho_\vartheta^*(\ell,j,k)} = [\rho_\kappa]_{\ell,j,k}\,.$$

Since we already know the fixpoint of ρ_q the proof is complete. □

Remark 7.1. A direct calculation of images $\varphi_q(\ell)$, $\varphi_q(j)$ and $\varphi_q(k)$ yields (for unit q) the matrix

$$[\varphi_q]_{\ell,j,k} = \begin{pmatrix} 1-2(y^2+z^2) & 2(xy-rz) & 2(ry+xz) \\ 2(xy+rz) & 1-2(x^2+z^2) & 2(yz-rx) \\ 2(xz-ry) & 2(rx+yz) & 1-2(x^2+y^2) \end{pmatrix}.$$

We can verify that it is orthogonal with determinant 1.

7.2 Geometry of the action

If we identify unitary operators that have the same action on classes given by the projective equivalence, we get the *Projective Unitary Group*, which we denote by PU(2) (two denotes the dimension). We thereby identify the operator U with the operator $e^{i\varphi}U$. (Recall that $e^{i\varphi}$ here represents the so-called *scalar matrix*, i.e. a diagonal matrix with all indices on the diagonal equal to $e^{i\varphi}$, thus having the determinant $e^{i2\varphi}$.)

First, let us explore what the general unitary operator U looks like. Its first column is some unit vector $\begin{pmatrix} a \\ b \end{pmatrix}$. The second column is then perpendicular to it, so it is the vector $\begin{pmatrix} -b^* \\ a^* \end{pmatrix}$ up to multiplication by a complex unit. The general form of a unitary matrix is therefore

$$U = \begin{pmatrix} a & -e^{i\psi}b^* \\ b & e^{i\psi}a^* \end{pmatrix},$$

with determinant $e^{i\psi}$. In the basis of eigenvectors, the U is of the form

$$\begin{pmatrix} e^{i\varphi_1} & 0 \\ 0 & e^{i\varphi_2} \end{pmatrix},$$

where $\psi = \varphi_1 + \varphi_2$. The matrix U is projectively equivalent to the matrix

$$e^{-i\frac{\psi}{2}}U = \begin{pmatrix} e^{-i\psi/2}a & -e^{i\psi/2}b^* \\ e^{-i\psi/2}b & e^{i\psi/2}a^* \end{pmatrix} = \begin{pmatrix} c & -d^* \\ d & c^* \end{pmatrix},$$

where $c = e^{-i\psi/2}a$ a $d = e^{-i\psi/2}b$, with determinant one and the diagonal form

$$\begin{pmatrix} e^{-i\omega/2} & 0 \\ 0 & e^{i\omega/2} \end{pmatrix},$$

where $\omega = \varphi_2 - \varphi_1$. It is therefore natural to choose this simple representative of unitary operators projectively equivalent with U. It is an element of the *Special Unitary group* denoted SU(2). However, there are two such representatives! Namely $\pm e^{-i\psi/2}U$.

Remark 7.2. Another natural choice is the matrix $e^{-i\varphi_1}$, with the diagonal form

$$\begin{pmatrix} 1 & 0 \\ 0 & e^{i\omega} \end{pmatrix}.$$

Note an interesting difference. While the mapping

$$R(\omega) = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\omega} \end{pmatrix}$$

has the period 2π, the mapping

$$T(\omega) = \begin{pmatrix} e^{-i\frac{\omega}{2}} & 0 \\ 0 & e^{i\frac{\omega}{2}} \end{pmatrix}$$

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

has the period 4π, and the matrices $T(\omega)$ and $T(\omega+2\pi)$ differ by the sign, being two representatives of SU(2) in PU(2).

Writing $c = p - ti$ and $d = s - ri$, where $p, r, s, t \in \mathbb{R}$, $(p, r, s, t) \in \mathbb{S}^3$, we have

$$\begin{pmatrix} p - ti & -s - ri \\ s - ri & p + ti \end{pmatrix}.$$

The sign in SU(2) can now be chosen in order to make p non-negative. (If $p = 0$ we will decide according to t or even s.) The advantage of this expression is the equality

$$\begin{pmatrix} p - ti & -s - ri \\ s - ri & p + ti \end{pmatrix} = pE + r(-iX) + s(-iY) + t(-iZ),$$

which provides a decomposition into matrices E, $-iX$, iY, $-iZ$ which satisfy the defining relations of quaternion units $1, \ell, j, k$. We can therefore identify $\ell = -iX$, $j = -iY$, $k = -iZ$ and we obtain a one-to-one correspondence between unit quaternions with non-negative real part and PU(2). Every element $U \in \mathrm{PU}(2)$ can be uniquely expressed as

$$U = \cos\frac{\omega}{2}E + (x\ell + yj + zk)\sin\frac{\omega}{2},$$

where $(x, y, z) \in \mathbb{S}^2$ and $\omega \in [0, \pi)$. In quantum mechanics, this is often written using so called *extended Euler's formula*

$$U = e^{-i\frac{\omega}{2}\,\xi\cdot\sigma} = E\cos\frac{\omega}{2} - i\,\xi\cdot\sigma\sin\frac{\omega}{2} = \cos\frac{\omega}{2}E + (x\ell + yj + zk)\sin\frac{\omega}{2},$$

where $\xi = (x, y, z)$ a $\sigma = (X, Y, Z)$. Each pair ξ, ω defines the rotation $R(\xi, \omega)$ of $\mathbb{R}^3$ around the axis ξ by the angle ω. These rotations make the *Special Orthonormal group* SO(3), that is, the group of matrices whose columns (and rows) form an orthonormal basis, and their determinant is one. Each non-identity rotation is thereby defined by two pairs due to the equality $R(\xi, \omega) = R(-\xi, -\omega)$.

Remark 7.3. Identity matrix E brings about some technical difficulties, sins its "axis" can be chosen arbitrarily (and $\omega = 0$). It is natural to ineoduce the convention for E that $x = y = z = 0$, that is, $\xi = \vec{0}$.

Therefore we have a bijection between $\mathbb{S}^2 \times (0, 2\pi)$ and $\mathrm{SU}(2) \setminus \{E\}$, where always two elements correspond to a single rotation in SO(3), or in PU(2). The above considerations can be summarized as follows:

$$\mathrm{PU}(2) \cong \mathrm{SU}(2)/\mathbb{Z}_2 \cong \mathbb{S}^3/\mathbb{Z}_2 \cong \mathbb{K}_1/\mathbb{Z}_2 \cong \mathbb{S}^2 \times (0, 2\pi)/\mathbb{Z}_2 \cup (\vec{0}, 0) \cong \mathrm{SO}(3)\,.$$

By $\cong$ we loosely mean the above described identifications.

The first and the last elements of the are nevertheless related in a much more precise way, which is given by the relation between rotations and quaternion conjugations formulated in the following theorem.

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

Theorem 7.2. *The mapping*

$$\begin{aligned}\Phi : \& \operatorname{SO}(3) &\to \operatorname{PU}(2)\\ R(\xi,\omega) &\mapsto e^{-i\frac{\omega}{2}\,\xi\cdot\sigma}\end{aligned}$$

is a group isomorphism. Moreover, for each rotation $\rho \in \operatorname{SO}(3)$ we have

$$\rho = \mathcal{S}^{-1} \circ \Phi(\rho) \circ \mathcal{S},$$

where $\mathcal{S} : \mathbb{S}^2 \to \mathbb{C}\mathbf{P}^1$ is the stereographic projection.

Proof. For $U = e^{-i\frac{\omega}{2}\,\xi\cdot\sigma}$ we have

$$R(\xi,\omega) = \rho_U \overset{\Phi}{\longmapsto} U.$$

The mapping is therefore injective and surjective, and the composition of rotations corresponds to the matrix multiplication. It remains to show that $R(\xi,\omega)$ acts on $\mathcal{S}^{-1}(|\psi\rangle)$ in the same way as U on $|\psi\rangle$. Here we exploit the density operator. Operator of the image $U|\psi\rangle$ is of the form

$$U|\psi\rangle\langle\psi|U^\dagger = \frac{1}{2}E + \frac{i}{2}U(b\ell + cj + dk)U^\dagger\,.$$

From the theorem about the action of quaternion conjugation we deduce that $\mathcal{S}^{-1}\left(U|\psi\rangle\right)$ is indeed equal to $\rho_U(b,c,d)$. Hence the following diagram commutes.

$$\begin{array}{ccc} \mathbb{C}\mathbf{P}^1 & \xleftarrow{\ \mathcal{S}\ } & \mathbb{S}^2 \\ {\scriptstyle e^{-i\frac{\omega}{2}\,\xi\cdot\sigma}}\downarrow & & \downarrow{\scriptstyle R(\xi,\omega)} \\ \mathbb{C}\mathbf{P}^1 & \xleftarrow{\ \mathcal{S}\ } & \mathbb{S}^2 \end{array}$$

□

7.3 Self-adjoint unitary operators and the AXBXC decomposition

In the previous chapter, we introduced without explanation the extended Euler formula

$$e^{-i\frac{\omega}{2}\,\xi\cdot\sigma} = \cos\frac{\omega}{2}E + (x\ell + yj + zk)\sin\frac{\omega}{2}\,.$$

Let us now show that this formula corresponds to the definition of an operator function as defined on normal operators, i.e. as a function applied to eigenvalues.

The operators in the domain of our function are of the form

$$\xi\cdot\sigma = xX + yY + zZ,$$

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

where $\xi = (x, y, z)$ is a unit real vector. The following lemma shows that these matrices, together with identity, form an intersection of two popular classes of normal matrices, namely unitary and Hermitian (that is,self-adjoint) matrices. These are matrices for which $A = A^\dagger$ (Hermitian) and $A^{-1} = A^\dagger$, or $A = A^{-1} = A^\dagger$. Such is, for example, the ubiquitous Hadamard matrix. In particular, $A^2 = E$ also holds, which, obviously, is true for a diagonalizable matrix just when it has a diagonal form

$$\begin{pmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{pmatrix}.$$

Note, however, that $A^2 = E$ also holds for some non-diagonalizable matrices, such as the matrix $\begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}$.

Theorem 7.3. *The matrix A is a Hermitian unitary, just when it is equal to $\pm E$, or is of the form $xX + yY + zZ$, where $(x, y, z) \in \mathbb{S}^2$.*

Proof. Since all three Pauli matrices are Hermitian, $(xX + yY + zZ)^\dagger = xX + yY + zZ$ holds, so the matrix $xX + yY + zZ$ is Hermitian. The relation $(xX + yY + zZ)^2 = E$ results by a direct calculation from the fact that Pauli matrices are involutive and anticommutative, i.e. that they satisfy

$$X^2 = Y^2 = Z^2 = E, \qquad XY = -YX, \qquad YZ = -ZY, \qquad ZX = -XZ.$$

Conversely, if A is a Hermitian unitary, it has a diagonal form

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

that is, $\pm E$ or $\pm Z$. If the diagonal form is $\pm E$, then $A = \pm E$ (the identity has the same form for all bases). If the diagonal form is $\pm Z$, then the determinant is -1. From the characterization of unitary matrices (see chapter *Geometry of projective unitary operators*) it is now easy to see that a unitary matrix with determinant -1, which is also Hermitian, is of the form

$$\begin{pmatrix} z & x - iy \\ x + iy & -z \end{pmatrix} = xX + yY + zZ,$$

where $x^2 + y^2 + z^2 = 1$. □

For the operator A with real eigenvalues r_1 nda r_2 the expression $e^{-i\frac{\omega}{2}A}$ denotes the operator that has a diagonal form

$$\begin{pmatrix} e^{-i\frac{\omega}{2}r_1} & 0 \\ 0 & e^{-i\frac{\omega}{2}r_2} \end{pmatrix} = \begin{pmatrix} \cos\frac{\omega}{2}r_1 - i\sin\frac{\omega}{2}r_1 & 0 \\ 0 & \cos\frac{\omega}{2}r_2 - i\sin\frac{\omega}{2}r_2 \end{pmatrix}.$$

In addition, if r_1 i r_2 are equal to ± 1, we indeed get (in the base of eigenvectors) due to cosine evenness and sinus oddity

$$e^{-i\frac{\omega}{2}A} = \cos\frac{\omega}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - i\sin\frac{\omega}{2}\begin{pmatrix} r_1 & 0 \\ 0 & r_2 \end{pmatrix} = \cos\frac{\omega}{2}E - iA\sin\frac{\omega}{2}.$$

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

The formula is therefore correct for Hermitian unitary matrices.

For rotations around the main axes we get, according to the previous formula,

$$R_X(\omega) := R_{(1,0,0)}(\omega) = E\cos\frac{\omega}{2} - iX\sin\frac{\omega}{2} = \begin{pmatrix} \cos\frac{\omega}{2} & -i\sin\frac{\omega}{2} \\ -i\sin\frac{\omega}{2} & \cos\frac{\omega}{2} \end{pmatrix},$$

$$R_Y(\omega) := R_{(0,1,0)}(\omega) = E\cos\frac{\omega}{2} - iY\sin\frac{\omega}{2} = \begin{pmatrix} \cos\frac{\omega}{2} & -\sin\frac{\omega}{2} \\ \sin\frac{\omega}{2} & \cos\frac{\omega}{2} \end{pmatrix},$$

$$R_Z(\omega) := R_{(0,0,1)}(\omega) = E\cos\frac{\omega}{2} - iZ\sin\frac{\omega}{2} = \begin{pmatrix} \exp(-i\frac{\omega}{2}) & 0 \\ 0 & \exp(i\frac{\omega}{2}) \end{pmatrix}.$$

The explicit form of the representative of the projective class of unitary matrices corresponding to the rotation by the angle ω around the axis $\xi = (x, y, z)$ is

$$R_\xi(\omega) = \begin{pmatrix} \cos\frac{\omega}{2} - iz\sin\frac{\omega}{2} & -i(x-iy)\sin\frac{\omega}{2} \\ -i(x+iy)\sin\frac{\omega}{2} & \cos\frac{\omega}{2} + iz\sin\frac{\omega}{2} \end{pmatrix}.$$

Since $XYX = -Y$ and $XZX = -Z$, we get a useful relationship

$$XR_Y(\omega)X = XEX\cos\frac{\omega}{2} - iXYX\sin\frac{\omega}{2} = R_Y(-\omega)$$

and similarly

$$XR_Z(\omega)X = R_Z(-\omega).$$

Now we can prove the theorem we need to construct a controlled operator for the general U.

Theorem 7.4. *Each unitary operator U is projectively equivalent to the operator $AXBXC$, where $ABC = E$.*

Proof. We know that the U operator is projectively equivalent to a form operator

$$\begin{pmatrix} \cos\frac{\vartheta}{2} & -e^{i(\psi-\varphi)}\sin\frac{\vartheta}{2} \\ e^{i\varphi}\sin\frac{\vartheta}{2} & e^{i\psi}\cos\frac{\vartheta}{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\varphi} \end{pmatrix} \begin{pmatrix} \cos\frac{\vartheta}{2} & -\sin\frac{\vartheta}{2} \\ \sin\frac{\vartheta}{2} & \cos\frac{\vartheta}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & e^{i(\psi-\varphi)} \end{pmatrix},$$

which is projectively equivalent to an operator

$$R_Z(\varphi)R_Y(\vartheta)R_Z(\psi)R_Z(-\varphi)\,.$$

Using the above derived properties of conjugation by the operator X we get

$$\begin{aligned} &R_Z(\varphi)R_Y(\vartheta)R_Z(\psi)R_Z(-\varphi) = \\ &= R_Z(\varphi)R_Y\left(\frac{\vartheta}{2}\right)R_Y\left(\frac{\vartheta}{2}\right)R_Z\left(\frac{\psi}{2}\right)R_Z\left(\frac{\psi}{2}\right)R_Z(-\varphi) = \\ &= R_Z(\varphi)R_Y\left(\frac{\vartheta}{2}\right)\left(XR_Y\left(-\frac{\vartheta}{2}\right)X\right)\left(XR_Z\left(-\frac{\psi}{2}\right)X\right)R_Z\left(\frac{\psi}{2}\right)R_Z(-\varphi) = \\ &= \left(R_Z(\varphi)R_Y\left(\frac{\vartheta}{2}\right)\right)X\left(R_Y\left(-\frac{\vartheta}{2}\right)R_Z\left(-\frac{\psi}{2}\right)\right)X\left(R_Z\left(\frac{\psi}{2}\right)R_Z(-\varphi)\right) \end{aligned}$$

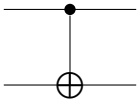
and now it is enough to put

$$A = R_Z(\varphi)R_Y\left(\frac{\vartheta}{2}\right), \quad B = R_Y\left(-\frac{\vartheta}{2}\right)R_Z\left(-\frac{\psi}{2}\right), \quad C = R_Z\left(\frac{\psi}{2}\right)R_Z(-\varphi)\,.$$

□

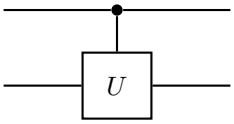
8 Universal set of gates

In this chapter we will show the basic principle of the construction of the quantum computer, namely the fact that any unitary operator can be constructed with the help of one-cubit operators and a single two-cubit CNOT operator, that is, the controlled negation, which we denote



8.1 Controlled single-qubit operators.

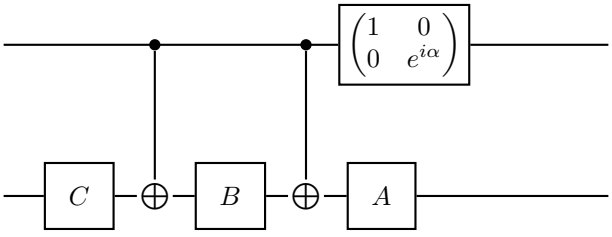
The first step is the construction of arbitrary controlled single-qubit operators. These correspond to the conditional construction "if the first qubit is one, perform the operation U on the second qubit", schematically:



The key to the construction is to decompose any operator using X and some operators A, B and C such that

$$U = e^{i\alpha}AXBXC, \qquad\qquad ABC = E.$$

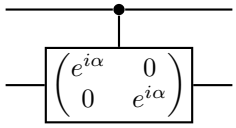
Thanks to this decomposition, we get the controlled operator U using the circuit



It is straightforward to verify that $|0\rangle \otimes |\varphi\rangle$ maps to $|0\rangle \otimes |\varphi\rangle$ and $|1\rangle \otimes |\varphi\rangle$ maps to $|1\rangle \otimes U|\varphi\rangle$. Note that the matrix

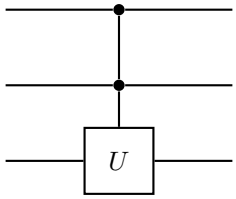
$$\begin{pmatrix} 1 & 0 \\ 0 & e^{i\alpha} \end{pmatrix}$$

applied to the first qubit is equivalent to the controlled multiplication of the scalar matrix $e^{i\alpha}$:

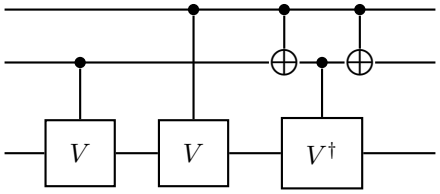


8.2 Two-controlled single-qubit operators.

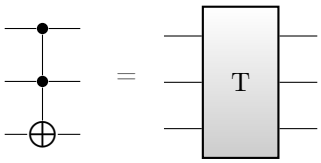
Next important step is the construction of two-checked operators, that is, operators that are executed just when both controlling values are one. Schematically:



This requires the square root of the operator U, that is, an operator $V = \sqrt{U}$, such that $V^2 = U$ (see below). Two-controlled operator U is then implemented by the circuit

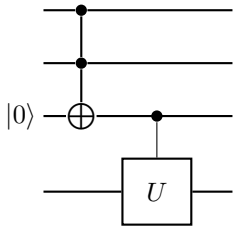


Two-controlled operator is actually an operator controlled by the conjunction of two values. Therefore, we would not need to emphasize its construction if we could implement an AND circuit, which, as we know, is possible in a reversible way using the Toffoli gate. But this is actually itself a double-checked negation (and therefore sometimes also referred to as CCNOT): :

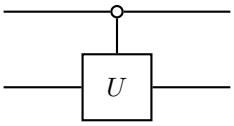


The Toffoli gate is therefore a special case of this construction and thanks to it we have all Boolean functions available, because the Toffoli gate is universal. Thus, the two-checked operator U could also be expressed by a more complex

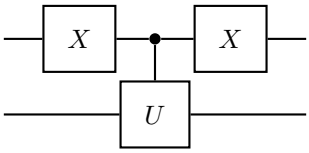
circuit with one auxiliary cubit as:



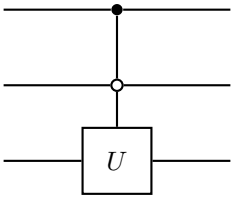
Similarly, operators controlled by any Boolean function can be constructed. If we want the operator to be applied if the value of the controlling cubite is zero, not one, then we will write schematically



which, in fact, is a shortcut for



The two options can also be combined, for instance as



On the example of the Toffoli gate, we shall show the construction of the operator V, that is, a "square root" of negation. Finding such an operator is a special case of application of a function to a normal operator. For any function $f : \mathbb{R} \to \mathbb{C}$ and a normal operator A we defined $f(A)$ as an operator satisfying

$$f(A)|u_\lambda\rangle = f(\lambda)|u_\lambda\rangle$$

for each eigenvector u_λ of the operator A, where λ is the corresponding eigenvalue. The negation is given by the Pauli operator

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

which can be written by projections on eigenvectors as

$$X = |+\rangle\langle +| - |-\rangle\langle -|,$$

where

$$|+\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}, \qquad |-\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}.$$

From here we have

$$V = \sqrt{1}|+\rangle\langle +| + \sqrt{-1}|-\rangle\langle -|.$$

We have four options for choosing the pair of square roots. For $\sqrt{1} = 1$ a $\sqrt{-1} = i$ we get

$$V = \frac{1}{2}\begin{pmatrix}1 & 1\\1 & 1\end{pmatrix} + \frac{i}{2}\begin{pmatrix}1 & -1\\-1 & 1\end{pmatrix} = \frac{1-i}{2}\begin{pmatrix}i & 1\\1 & i\end{pmatrix}.$$

8.3 Conversion of two-level operators to single-qubit controlled operators.

Consider the unitary operator U on a four-dimensional space given by the matrix

$$U = \begin{pmatrix}a & b & 0 & 0\\c & d & 0 & 0\\0 & 0 & 1 & 0\\0 & 0 & 0 & 1\end{pmatrix}$$

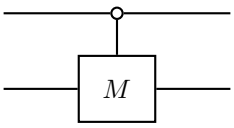
The operator acts non-identically only on the basis vectors $|00\rangle$ and $|01\rangle$, as follows:

$$|0\rangle \otimes |0\rangle \mapsto |0\rangle \otimes (a|0\rangle + c|1\rangle), \qquad |0\rangle \otimes |1\rangle \mapsto |0\rangle \otimes (b|0\rangle + d|1\rangle).$$

If we denote

$$M = \begin{pmatrix}a & b\\c & d\end{pmatrix}$$

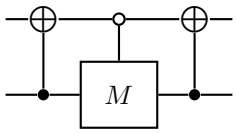
it is therefore possible to construct U as



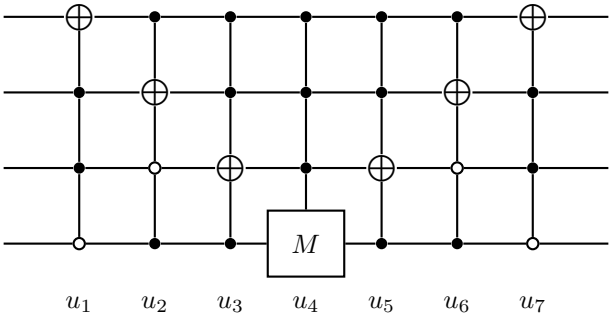
Operators acting non-identically on only two base vectors are called two-level. However, not every two-level operator has such a simple circuit as the U operator above. E.g., the operator

$$U' = \begin{pmatrix}a & 0 & 0 & b\\0 & 1 & 0 & 0\\0 & 0 & 1 & 0\\c & 0 & 0 & d\end{pmatrix}$$

acts non-identically on the basis vectors $|00\rangle$ a $|11\rangle$, which differ in more than one place, and therefore cannot be simply written as a controlled matrix M. It is necessary to first change the basis so that the non-identically mapped vectors differ only in one place. We therefore perform a permutation that swaps $|11\rangle$ and $|01\rangle$, which is the CNOT on the first qubit controlled by the second one. Then we can proceed as in the case of U and then swap back $|11\rangle$ and $|01\rangle$. The whole circuit looks like this



n the case of a general two-level matrix acting non-identically on basis vectors $\mathbf{b} = |k_{n-1}k_{n-2}\dots k_0\rangle$ and $\mathbf{b}' = |\ell_{n-1}\ell_{n-2}\dots \ell_0\rangle$, we have to map these vectors to basis elements that differ only in one cubit. We then perform the controlled operation on it and convert the base back to its original form. In total, this means a series of operations controlled by all but one qubit that varies. Suppose, for example, that the matrix M acts non-identically on qubits $\mathbf{b} = |0110\rangle$ and $\mathbf{b}' = |1001\rangle$. We can choose base vectors, differing in only one cubit, on which we will perform the controlled operation M; for example, choose $|1110\rangle$ and $|1111\rangle$. So we have to change the first three cubits: the first in the base vector $\mathbf{b}$, the second and the third in the base vector $\mathbf{b}'$. The circuit will look like this



- u_1 and u_7: transposition $|0110\rangle \leftrightarrow |1110\rangle$
- u_2 and u_6: transposition $|1001\rangle \leftrightarrow |1101\rangle$
- u_3 and u_5: transposition $|1101\rangle \leftrightarrow |1111\rangle$
- u_4: transformation $|1110\rangle \mapsto a|1110\rangle + b|1111\rangle$; $|1111\rangle \mapsto c|1110\rangle + d|1111\rangle$

8.4 Decomposition into two-level operators.

It remains to show that any unitary operator can be decomposed into unitary two-level operators. The process of such decomposition is similar to the

Gaussian elimination, and the two-level matrices sought are matrices of the corresponding elementary transformations. These are always two-level: they only manipulate two lines. Unlike the classical Gaussian elimination, however, we still have to ensure that they are unitary. This is certainly true if we only swap lines (to get a non-zero element on the diagonal). Let's study the case when we want to subtract an element outside the diagonal. Let the matrix to be modified be of the form

$$U = \begin{pmatrix} a & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ b & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix},$$

where $U_{1,1} = a \neq 0$ and $U_{j,1} = b \neq 0$ and other elements are arbitrary. We may have ensured that a is non-zero by a permutation of rows, if needed. We now want to get rid of the element b, i.e. to set the position $(j, 1)$ to zero. We can do this by adding an appropriate multiple of the first line to the j-th line. In the case of classical Gaussian elimination we would use the matrix of elementary transformation

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ -b/a & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Note that if we want to avoid division (e.g. when manipulating an integer matrix), we can also use the matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ b & 0 & 0 & -a & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Although neither of these matrices is unitary, it is not difficult to complete it to a unitary one by normalizing j-th row and changing the first line to a orthogonal unit vector:

$$U_1 = \begin{pmatrix} a^*/c & 0 & 0 & b^*/c & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ b/c & 0 & 0 & -a/c & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

where $c = \|(a, b)\| = \sqrt{aa^* + bb^*}$. Multiplying we obtain

$$U_1 \cdot U = \begin{pmatrix} c & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} .$$

In this way we gradually convert the matrix U to

$$U' = \begin{pmatrix} a' & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} .$$

Since the resulting matrix is still unitary (we multiplied it by unitary matrices), we have $|a'| = 1$. In addition, it is clear from the last step of the elimination that $a' = 1$. Because also the rows of a unitary matrix have the norm of one, U' is actually of the form

$$U' = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} .$$

Repeating the procedure for the smaller matrix, we finally get the identity matrix. We then have

$$U_k \cdots U_2 U_1 \cdot U = I \,,$$

where U_i are two-level unitary operators (some of them may be permutation matrices swapping rows). Thus we have the desired decomposition of U into two-level operators

$$U = U_1^\dagger U_1^\dagger \cdots U_k^\dagger \,.$$

9 Quantum entropy

9.1 Mixed states

By the *mixed* state of a system we mean the probabilistic combination of the states we have discussed so far, called *pure*. (Note that formally a pure state is a special case of a mixed state). For computational reasons, mixed states are

usually represented by a density matrix rather than a vector. Thus, the density matrix of a mixed state is of the form

$$\rho = \sum p_i \rho_i,$$

where p_i are non-negative numbers with sum one (i.e., discrete probability distributions) and ρ_i are density matrices. The set $\{(\rho_i, p_i)\}$ where $\sum p_i = 1$ is called an *ensemble of states*. A mixed state thus represents a situation where, in addition to the underlying quantum-mechanical uncertainty about the measurement result, there is also "ordinary" probabilistic uncertainty about what pure state the system is in. Note that we get a mixed state even if the states ρ_i are themselves mixed. This shows a fundamental advantage of this approach to mixed states: we can treat the density matrix as a single state, regardless of whether it is mixed or pure. It is easy to see that this is also true of the evolution of a system: if ρ is the density matrix of a system, then $U\rho U^\dagger$ is the density matrix of that system after applying the unitary operation U (including the appropriate probabilistic interpretation).

The previous observation can be reinforced: we can operate on a mixed state **without knowing what pure states it consists of**. Indeed, the same density matrix can arise from different sets of states. However, we noted that not knowing the "correct" decomposition of the density matrix into pure states does not prevent us from computing the evolution of the system. We now show that the same is true for measurements: the density matrix uniquely determines the results of the measurements (i.e., their probability distribution). To this end, we define a more general notion of measurement than that of projective measurement, to which we have restricted ourselves in formulating the relevant postulate.

Postulate 3' The measurement of a quantum system is given by a system of operators M_i satisfying the condition

$$\sum_i M_i^\dagger M_i = I.$$

After measuring the state $|\psi\rangle$, the system with probability $\langle\psi|M_i^\dagger M_i|\psi\rangle$ is in the state

$$\frac{M_i|\psi\rangle}{\sqrt{\langle\psi|M_i^\dagger M_i|\psi\rangle}}.$$

If the use of the operator M_i is associated with the measured value m_i, then the expected value of the measurement is

$$E(m) = \sum_i p_i m_i = m_i \langle\psi|M_i^\dagger M_i|\psi\rangle = \langle\psi|M|\psi\rangle\,,$$

where

$$M = \sum m_i M_i^\dagger M_i$$

is called *observable*.

Using the density matrix $\rho = |\psi\rangle\langle\psi|$ and the relation $\langle\psi|A|\psi\rangle = \mathrm{tr}(A|\psi\rangle\langle\psi|)$, we obtain the equivalent condition that the i-th state measurement ρ will occur with probability $\mathrm{tr}(M_i^\dagger M_i \rho)$ and the system will be in the state after such a measurement

$$\frac{M_i \rho M_i^\dagger}{\mathrm{tr}(M_i^\dagger M_i \rho)} .$$

The mean value of the measurement is $\mathrm{tr}(M\rho)$.

Thus, the state after the measurement can be seen as a mixed state $\sum_i M_i \rho M_i^\dagger$. From linearity we get the desired property that the above holds even if the original measured state ρ was mixed, independently of the particular set of states. Let us illustrate this fact on the expected value. If $\rho = \sum p_i \rho_i$, where ρ_i are the pure states, then the expected outcome of the measurement of the pure state ρ_i is $\mathrm{tr}(M\rho)$. Since the state ρ_i is measured with probability p_i, the mean of the mixed state measurement is

$$\sum p_i \,\mathrm{tr}(M\rho_i) = \mathrm{tr}(M \sum p_i \rho_i) = \mathrm{tr}(M\rho) .$$

Note further that while we need to know the measurement operators to compute the state of the system after the measurement, for the statistics of the results it is sufficient to know the set $\{M_i^\dagger M_i\}$, which is the decomposition of the identity into positive operators. For a one-time measurement where we are not interested in the state of the system after the measurement (e.g., this is naturally true for destructive measurements such as photon detection), it is sufficient to specify such a set E_i. A measurement defined in this way is called POVM (positive operator value measurement) in the literature.

The diagonal form of the pure state density matrix contains exactly one 1 on the diagonal. It follows that the trace of the pure, and hence of any mixed state matrix, is equal to one. Since density matrices are positive operators, their diagonal form represents a discrete probability distribution. In other words, each mixed matrix can be viewed as a probabilistic combination of projections onto some orthonormal basis.

Recall that the density matrix of a general pure qubit is of the form $\frac{1}{2}(E + xX + yY + zZ)$, where (x, y, z) is a unit vector. The mixed state of the qubit is therefore of the form

$$\frac{1}{2} \sum p_i \left(E + x_i X + y_i Y + z_i Z\right) .$$

Vector

$$\sum p_i \left(x_i, y_i, z_i\right)$$

is the weighted average (center of gravity) of the points (x_i, y_i, z_i). The convexity of the sphere implies that it is a vector less than one. Conversely, each point

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

in the unit sphere represents a mixed state. Thus the Bloch ball represents all mixed states, with the pure states lying on the surface.

Let ρ^{AB} be the density matrix of the composite system. We define the *reduced density matrix* on system A as

$$\rho^A := \mathrm{tr}_B\left(\rho^{AB}\right),$$

where tr_B is the so-called *partial trace* defined by a linear extension of the relation

$$\mathrm{tr}_B(\rho_a \otimes \rho_b) = \mathrm{tr}(\rho_b)\rho_a$$

or, written in basis vectors,

$$\mathrm{tr}_B\left(|a_i\rangle\langle a_j| \otimes |b_k\rangle\langle b_\ell|\right) = \delta_{k\ell}|a_i\rangle\langle a_j|\,.$$

The reduced density matrix ρ^A has a clear and important physical meaning. It captures the properties of the system A understood in isolation. More precisely, the measurement results of the system A alone are the same for the state ρ^A as for the state ρ^{AB}. Even more precisely, the measurement of the state ρ^A given by the operators (M_j) has the same properties as the measurement results of the state ρ^{AB} given by the operators $(M_j \otimes E)$. Another physical view of the same statement is that the matrix ρ^A describes the state of system A after (any!) measurement of system B for which we do not know the outcome (this uncertainty translates into a mixed state ρ^A). Indeed, any such measurement releases system A from entanglement with system B. These two views are equivalent: if we are restricted to measurements of system A, we cannot know whether system B has been measured or not.

9.2 Definition

Entropy is a measure of the information content of a random variable. Classical entropy, measured in bits and called *Shannon entropy* after its inventor, is for a discrete random variable X with probabilities $\Pr[X = i] = p_i$, defined by the formula

$$H(X) = -\sum_i p(i) \log p_i.$$

This value can be interpreted informally as the average number of bits of the address of the random event that occurred. The basic property of Shannon entropy that shows that it actually expresses information content is the source coding theorem, also called the compression theorem or the noiseless channel capacity theorem. It shows that a sequence n of independent copies of a random variable X can be encoded by a sequence of bits of length $nH(X)$ with error probability asymptotically going to zero for large n. (The theorem is proved by noting that the vast majority of sequences have the expected distribution of the number of letters, and showing that there are almost exactly $2^{nH(X)}$ of such typical sequences of length n.)

The total information of a pair of random variables $H(X,Y)$ is at most the sum of $H(X)+H(Y)$, but can be smaller. For example, if Y is a function of X, then (X,Y) is completely determined by X and $H(X,Y)=H(X)$. The remaining information content of Y given knowledge of X is the conditional entropy $H(Y\,|\,X)$. (We should correctly say "average entropy", since it is the average over the different values of X. The entropy $H(Y)$ is itself an average over different values of Y.) So the value $H(Y)-H(Y\,|\,X)$ expresses how much information we learn about Y if we know X. Similarly, $H(X)-H(X\,|\,Y)$ is a measure of the information Y reveals about X. The expected relation $H(X,Y)=H(X)+H(Y\,|\,X)=H(Y)+H(X\,|\,Y)$ holds. Value

$$I(X:Y) := H(X)+H(Y)-H(X,Y)=H(X)-H(X\,|\,Y)=H(Y)-H(Y\,|\,X)$$

is thus a measure of the dependence of the two quantities and is called *mutual information*. Note that this value is symmetric in X and Y.

The information content of a quantum system is given by the uncertainty about the measurement results. In other words, the result of a given measurement of a given system is a random variable with some entropy. However, the entropy of a quantum state must take into account all possible measurements. If the system is in a pure state, there is a measurement whose result is given uniquely (it is any measurement in the basis containing the measured state). Thus, the entropy of a quantum system is non-zero only in the case of mixed states and comes from the uncertainty about the prepared state. However, it is also affected by the nature of the ensemble. Let us illustrate this with the states that occur in the BB84 protocol. If we know what base Alice encodes in, but we don't know what bit she encodes, the system is in a state

$$\frac{1}{2}|0\rangle\langle 0|+\frac{1}{2}|1\rangle\langle 1|=\frac{1}{2}E\,.$$

Measuring in the canonical basis is equivalent to accepting a random bit. The fact that the value was encoded using quantum states rather than classically plays no role here. The entropy of such a state should therefore be equal to one. If, on the other hand, we know that Alice encoded zero, but we do not know in what basis, we get a matrix

$$\rho=\frac{1}{2}|0\rangle\langle 0|+\frac{1}{2}|+\rangle\langle +|=\begin{pmatrix}\frac{1}{2}&0\\0&0\end{pmatrix}+\begin{pmatrix}\frac{1}{4}&\frac{1}{4}\\\frac{1}{4}&\frac{1}{4}\end{pmatrix}=\begin{pmatrix}\frac{3}{4}&\frac{1}{4}\\\frac{1}{4}&\frac{1}{4}\end{pmatrix}\sim\begin{pmatrix}\frac{2+\sqrt{2}}{4}&0\\0&\frac{2-\sqrt{2}}{4}\end{pmatrix}.$$

The diagonal form is with respect to the (normalized) eigenvectors

$$|v_1\rangle=\frac{1}{\sqrt{4-2\sqrt{2}}}\begin{pmatrix}-1\\\sqrt{2}+1\end{pmatrix},\qquad |v_2\rangle=\frac{1}{\sqrt{4+2\sqrt{2}}}\begin{pmatrix}1\\\sqrt{2}-1\end{pmatrix}.$$

Thus, we get the same density matrix if we choose v_1 or v_2 with probabilities $\frac{2+\sqrt{2}}{4}$ and $\frac{2-\sqrt{2}}{4}$. The classical entropy of such a random variable is

$$-\frac{2+\sqrt{2}}{4}\log\frac{2+\sqrt{2}}{4}-\frac{2-\sqrt{2}}{4}\log\frac{2-\sqrt{2}}{4}\doteq 0.6\,.$$

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

The entropy is not equal to one because the selected bit was encoded into two states that are not completely distinguishable, thus some information was lost.

We can also use this example to illustrate the independence of the measurement result from the way the density matrix was created. Let us examine the probability of obtaining a result corresponding to $|0\rangle$, $|1\rangle$ when measuring in the $|0\rangle$ basis. By Postulate 3', this is $\mathrm{tr}(|0\rangle\langle 0|\rho) = 3/4$. This corresponds to simple reasoning: with probability one-half we have a state $|0\rangle$, and then the measurement result corresponds to $|0\rangle$ with certainty; with probability one-half we have a state $|+\rangle$, where the result corresponds to $|0\rangle$ with probability one-half. Similarly, we could verify that if $|v_1\rangle = \alpha_1|0\rangle + \beta_1|1\rangle$ and $|v_2\rangle = \alpha_2|0\rangle + \beta_2|1\rangle$, then

$$\frac{2+\sqrt{2}}{4}|\alpha_1|^2 + \frac{2-\sqrt{2}}{4}|\alpha_2|^2 = \frac{3}{4}\,.$$

These examples lead to the definition of the *von Neumann entropy* density matrix ρ. It is the entropy of the random variable corresponding to the choice of the eigenvectors ρ, which can be written concisely as

$$S(\rho) = -\,\mathrm{tr}(\rho \log \rho)\,.$$

Let us list some properties of quantum entropy. For the entropy of the mixed state $\rho = \sum_i p_i \rho_i$, the following holds

$$S(\rho) \leq H(X) + \sum_i p_i S(\rho_i),$$

where X is the random variable of the state selection ρ_i, i.e. a discrete random variable with probability $\Pr[X = i] = p_i$. Equality holds if and only if the states ρ_i are distinguishable, i.e. if they are defined on mutually orthogonal spaces. The entropy of a state ρ is at most the entropy of the corresponding choice of ρ_i plus the average entropy contained in ρ_i itself. If the states ρ_i are pure, we get $S(\rho) \leq H(X)$. If they are pure and distinguishable, we have $S(\rho) = H(X)$. Then it is a classical random variable, it does not matter that we encode its values by quantum states.

For complex systems, *strong subadditivity* holds:

$$S(\rho^{ABC}) + S(\rho^B) \leq S(\rho^{AB}) + S(\rho^{BC})\,.$$

Let us define the mutual information of two quantum states as

$$S(\rho^A : \rho^B) := S(\rho^A) + S(\rho^B) - S(\rho^{AB})\,.$$

9.3 Holevo bound

The measurement postulate implies that the von Neumann entropy corresponds exactly to the entropy of the random variable that is the result of the measurement of a given mixed state in the basis of its eigenvectors. However, as we have already said, a "properly" defined entropy should take into account all possible

measurements. The random variable about which we are trying to obtain information by measurement is the random variable with the distribution defining an ensemble of states. The exact relation of von Neumann entropy to the information obtainable by arbitrary measurements is unknown. The most important result in this respect is the so-called Holevo bound.

Theorem 9.1 (Holevo bound)**.** *Let X be a discrete random variable with distribution $\Pr[X = i] = p_i$. Let $\rho = \sum_{i=1}^{n} p_i \rho_i$ be a mixed state generated by encoding the value of X using the states of ρ_i. Let Y be the random variable of the results of some measurement of the state ρ. Then*

$$I(X : Y) \leq S(\rho) - \sum_{i=1}^{n} p_i S(\rho_i)\,.$$

Holevo bound says that the von Neumann entropy is an upper estimate for the information available by any measurement of the random variable X. If all states of ρ_i are pure, then the inequality has the form $I(X : Y) \leq S(\rho)$. Moreover, we know that $S(\rho) \leq H(X)$, which gives the classical $I(X : Y) \leq H(X)$.

The equality $S(\rho) = H(X)$ occurs precisely when the states are pure and distinguishable. Then it is a classical random variable, it is not important that we understand its values as quantum states. Then also $I(X : Y) = H(X)$ for measurements in the basis containing the chosen states, when $Y = X$.

For the proof of the Holevo bound, consider, in addition to the prepared and measured system denoted by Q, two additional systems. A system P, containing information about the value of the random variable X encoded in basis states, and a system M containing in turn the similarly encoded result Y of the measurement given by the operators (M_j). After preparation, the density matrix of this composite system is

$$\rho_0^{PQM} = \sum_i p_i |i\rangle\langle i| \otimes \rho_i \otimes |0\rangle\langle 0|\,,$$

and after the measurement it is

$$\rho_1^{PQM} = \sum_{i,j} p_i |i\rangle\langle i| \otimes M_j \rho_i M_j^\dagger \otimes |j\rangle\langle j|\,.$$

It turns out that the Holevo bound is actually the inequality

$$S(\rho_1^P : \rho_1^M) \leq S(\rho_0^P : \rho_0^Q)\,.$$

For the right hand side we can verify the following:

$$\rho_0^P = \sum_i p_i |i\rangle\langle i| \qquad S(\rho_0^P) = H(X)$$

$$\rho_0^Q = \rho = \sum_i p_i \rho_i \qquad S(\rho_0^Q) = S(\rho)$$

$$\rho_0^{PQ} = \sum_i p_i |i\rangle\langle i| \otimes \rho_i \qquad S(\rho_0^{PQ}) = H(X) + \sum_i p_i S(\rho_i)$$

For the left hand side, first note that $\mathrm{tr}(M_j \rho_i M_j)$ is the probability that the measurement result is j, under the condition that the measured state is ρ_i, let us denote it by $p_{j|i}$. Since the measurement result is independent of the choice of X, $p_i p_{j|i}$ is the joint probability of i and j, let us denote it by p_{ij}. Thus:

$$\begin{aligned} \rho_1^P &= \sum_{i,j} p_{ij} |i\rangle\langle i| = \sum_i p_i |i\rangle\langle i| & S(\rho_1^P) &= H(X) \\ \rho_1^M &= \sum_{i,j} p_{ij} |j\rangle\langle j| = \sum_j p_j |j\rangle\langle j| & S(\rho_1^M) &= H(Y) \\ \rho_0^{PM} &= \sum_{i,j} p_{ij} |i\rangle\langle i| \otimes |j\rangle\langle j| & S(\rho_1^{PM}) &= H(X,Y) \end{aligned}$$

Holevo bound is now obtained as follows:

$$S(\rho_0^P : \rho_0^Q) = S(\rho_0^P : \rho_0^{QM}) \geq S(\rho_1^P : \rho_1^{QM}) \geq S(\rho_1^P : \rho_1^M).$$

The derivation follows from three intuitive (and provable) principles:

- mutual information is not changed by adding an additional (uncorrelated) system;
- the mutual information of two systems cannot be increased by any measurement (or any unitary operations);
- the mutual information cannot be increased by removing part of one of the systems.

The first principle simply follows from the relation of entropy of decomposable states $S(\sigma \otimes \rho) = S(\sigma) + S(\rho)$, which we get directly from the definition.

The third principle follows from strong subadditivity. In our case, we have

$$S(\rho_1^{PQM}) + S(\rho_1^M) \leq S(\rho_1^{PM}) + S(\rho_1^{QM}),$$

where we get the required

$$S(\rho_1^P) + S(\rho_1^M) - S(\rho_1^{PM}) \leq S(\rho_1^P) + S(\rho_1^{QM}) - S(\rho_1^{PQM}).$$

Regarding the second principle, let us first note that the matrices $U\rho U^\dagger$ are similar, i.e. they have the same diagonal form, i.e. $S(\rho) = S(U\rho U^\dagger)$. It is natural that the entropy does not change by choosing a different basis. For measurements, we can reduce the principle to the second one by showing that each measurement can be viewed as a unitary transformation of our system along with some external, additional system, which we again remove after the measurement. This elegant and useful construction proceeds as follows.

Let us denote the system to be measured by Q and consider measurements using the operators (M_j). The additional system M will have base elements $|j\rangle$. Then the "indexing" mapping

$$U : |\varphi\rangle \otimes |0\rangle \mapsto \sum_j M_j |\varphi\rangle \otimes |j\rangle$$

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION, YOUTH AND SPORTS

is unitary. More precisely, this mapping preserves the scalar product (as can be straightforwardly verified using the completeness relation $\sum_j M_j^\dagger M_j = E$), and can thus be extended to the unitary mapping of the system $Q \otimes M$. The resulting density matrix is thus

$$\rho^{QM} = U(|\varphi\rangle\langle\varphi| \otimes |0\rangle\langle 0|)U^\dagger = \sum_{j,j'} M_j|\varphi\rangle\langle\varphi|M_{j'}^\dagger \otimes |j\rangle\langle j'|$$

and the reduced matrix for the original system is

$$\rho^Q = \sum_j M_j|\varphi\rangle\langle\varphi|M_j^\dagger\,,$$

as we wanted.

10 Bibliography

Hirvensalo, M. (2003), *Quantum Computing*, Natural Computing Series, Springer Berlin Heidelberg.

Nielsen, M. A. and Chuang, I. L. (2010), *Quantum Computation and Quantum Information: 10th Anniversary Edition*, Cambridge University Press.