Chapter 2
Preliminaries

M. Pivk

This chapter discusses basics necessary for the next chapters. All fields are skimped, because some areas would need more explanation, like quantum information theory, but this will be out of scope for this chapter.

2.1 Quantum Information Theory

In this section a short introduction to quantum information is given. For detailed explanation we refer to the book of Nielsen and Chuang [4], where this topic is amplified.

2.1.1 Quantum Bits

Since Shannon and the beginning of information theory, the *bit* has been the basic term in classical information. The states of a bit are either 0 or 1. In accordance with the classical concept in quantum information exists the *qubit* (short for quantum bit). Like for the classical bit two states are possible, $|0\rangle$ and $|1\rangle$. This special notation '$|\rangle$' is called the *Dirac notation* (or *ket*) and is the standard notation for states in quantum mechanics. The major difference to the classical bit, which accepts only 0 or 1, is that a qubit also allows states in between $|0\rangle$ and $|1\rangle$, which are called *superpositions*. Let us denote this by

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \tag{2.1}$$

where $\alpha, \beta \in \mathbb{C}$. Because these factors are complex numbers the state of a qubit can be described as a vector in a two-dimensional complex vector space $\mathbb{C}^2$, also called *Hilbert space*. The states $|0\rangle$ and $|1\rangle$ form the computational basis and are

M. Pivk (✉)
Safety & Security Department, Quantum Technologies, AIT Austrian Institute of Technology GmbH, Lakeside B01A, 9020 Klagenfurt, Austria, mario.pivk@ait.ac.at; http://www.ait.ac.at

Pivk, M.: *Preliminaries*. Lect. Notes Phys. **797**, 3–21 (2010)
DOI 10.1007/978-3-642-04831-9_2

orthonormal (see Definition 2.4) to each other, e.g., $|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $|1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Since a qubit state is a unit vector, meaning the length is normalized to 1, following equation must be fulfilled by the scalars α, β:

$$|\alpha|^2 + |\beta|^2 = 1. \tag{2.2}$$

Using this fact we can rewrite the state of a qubit

$$|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle, \tag{2.3}$$

where θ, φ are real numbers and define a point on a sphere called the *Bloch sphere* (see Fig. 2.1).

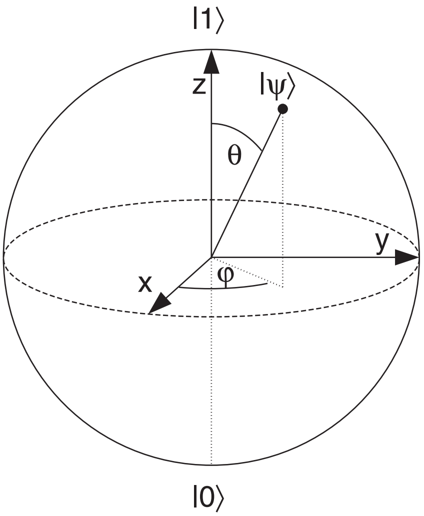

Fig. 2.1 Bloch sphere representation of a qubit

The measurement of qubits is a problem. In the special case when α or β is 0, the mapping to the classical bit will result in 1 or 0, respectively, as expected. But what happens if the qubit is in another superposition, i.e., $\alpha, \beta \neq 0$? Depending on the scalars the qubit will be measured as 1 with a certain probability or as 0 with the complementary probability. Since the scalars fulfil Eq. 2.2, the probability for a qubit to be measured as 0 is $|\alpha|^2$ and as 1 it is $|\beta|^2$. We see this in detail in Sect. 2.1.3.

Furthermore in quantum mechanics the scalars α and β are also called the *amplitudes* of the states $|0\rangle$ and $|1\rangle$, respectively. But there exists a second term describing a qubit, the *phase*. Consider the state $e^{i\varphi}|\psi\rangle$, where $|\psi\rangle$ is a state vector, and φ is a real number. We say that the state $e^{i\varphi}|\psi\rangle$ is equal to $|\psi\rangle$, up to the *global phase factor* $e^{i\varphi}$. The measurements for these two states are from the point of statistics the same as you will see in Sect. 2.1.2.

Another kind of phase is the *relative phase*. Consider these two states

$$|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \qquad \text{and} \qquad |-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle). \tag{2.4}$$

In the state $|+\rangle$ the amplitude of $|1\rangle$ is $\frac{1}{\sqrt{2}}$. In state $|-\rangle$ the amplitude has the same magnitude but a different sign. We define that two amplitudes α_1, α_2 for some states differ by a *relative phase* if there is a real φ such that the $\alpha_1 = e^{i\varphi}\alpha_2$. In contrast to the *global phase*, where both amplitudes of the state are different by the factor $e^{i\varphi}$, the *relative phase* differs only in one amplitude by the factor $e^{i\varphi}$.

2.1.2 Linear Operators

The state change of qubits is done by linear operators. Therefore a function A is used, taking vectors from $\mathcal{V}$ to $\mathcal{W}$ ($\mathcal{V}$ and $\mathcal{W}$ are vector spaces of $\mathbb{C}^*$). The most convenient way to describe such a function is the *matrix representation*. If matrix A has m columns and n rows and this matrix is multiplied with the vector $|v\rangle \in \mathbb{C}^n$ we get a new vector $|w\rangle \in \mathbb{C}^m$ as result. The claim for such a matrix A is to fulfill the linearity equation [4]

$$A\left(\sum_i a_i |v_i\rangle\right) = \sum_i a_i A|v_i\rangle. \tag{2.5}$$

Let $A : \mathcal{V} \longrightarrow \mathcal{W}$ be a linear operator and $|v_1\rangle, \ldots, |v_n\rangle$ be a basis of $\mathcal{V}$ and $|w_1\rangle, \ldots, |w_m\rangle$ a basis of $\mathcal{W}$. There exist complex numbers $A_{1j}, \ldots, A_{mj}$,

$$A|v_j\rangle = \sum_i A_{ij}|w_i\rangle \qquad \text{with} \qquad 1 \leq i \leq m, 1 \leq j \leq n, \tag{2.6}$$

which form the matrix representation of the operator A.

Contrarily, a $n \times m$ matrix can be understood as the opposite linear operator sending vectors out of the vector space $\mathcal{W}$ to the vector space $\mathcal{V}$ by performing the matrix multiplication with those vectors.

We use a notation which is different to the usual notation in linear algebra. Table 2.1 lists some frequently used symbols in quantum mechanics. As we know, a vector can be represented by the sum of the vectors out of the computational basis. For simplification we take the computational basis $v_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, v_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots,$

Table 2.1 Summary of some standard quantum mechanical notation

Notation	Description
z^*	Complex conjugate of the complex number z. e.g., $(1+i)^* = 1-i$
$\vert v\rangle$	Vector. Also known as a *ket*. $\vert v\rangle = \sum_i a_i \vert v_i\rangle$
$\langle v\vert$	Vector dual to $\vert v\rangle$. Also known as a *bra*. $\langle v\vert = \sum_i a_i^* \vert v_i\rangle^T$
$\lambda\vert v\rangle$	Multiplication by a scalar λ. $\lambda\vert v\rangle = \sum_i \lambda a_i \vert v_i\rangle$
$\langle v\vert w\rangle$	Inner product between the vectors $\vert v\rangle$ and $\vert w\rangle$
$\vert v\rangle\langle w\vert$	Outer product of $\vert v\rangle$ and $\vert w\rangle$
$\vert v\rangle \otimes \vert w\rangle$	Tensor product of $\vert v\rangle$ and $\vert w\rangle$
A^*	Complex conjugate of the A matrix
A^T	Transpose of the A matrix
$A^\dagger$	Hermitian conjugate or ad-joint of the A matrix, $A^\dagger = (A^T)^*$
$\langle\varphi\vert A\vert\psi\rangle$	Inner product between $\vert\varphi\rangle$ and $A\vert\psi\rangle$

$v_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$ such that a vector $|v\rangle = \sum_i a_i |v_i\rangle$ can also be written as $|v\rangle = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$. If an other computational basis is used, it is written explicitly.

2.1.2.1 The Pauli Matrices

Four extremely useful matrices are the *Pauli matrices*. These are 2 by 2 matrices and represent some needed effects on qubits. The matrices are

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.7}$$

The fourth matrix is the identity matrix (I). The Pauli operators X and Z are also known as *bit flip* and *phase flip* operators. If we apply the X operation on a qubit we see that $|0\rangle$ changes to $|1\rangle$ and vice versa, i.e.,

$$X|0\rangle = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{and} \quad X|1\rangle = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The Z operator is called phase flip operator because it changes the phase of $|1\rangle$ by the sign, i.e.,

$$Z|+\rangle = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} \sqrt{2} \\ \sqrt{2} \end{pmatrix} = \begin{pmatrix} \sqrt{2} \\ -\sqrt{2} \end{pmatrix} \quad \text{and}$$
$$Z|-\rangle = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} \sqrt{2} \\ -\sqrt{2} \end{pmatrix} = \begin{pmatrix} \sqrt{2} \\ \sqrt{2} \end{pmatrix}$$

A possibility to illustrate Y is to multiply the matrix with the imaginary unit i so we deal only with natural-numbered matrices. Thus, the reformulated version of Y is

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

The iY operator performs both flips, a bit flip and a phase flip, since

$$iY = ZX = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = i \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$

Therefore, when the iY operator is applied on the states $|0\rangle$ and $|1\rangle$ we get

$$iY|0\rangle = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \quad \text{and} \quad iY|1\rangle = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

2.1.2.2 Inner Products

An *inner product* (or *scalar product*) $\langle v|w\rangle$ (usually notation in linear algebra $(|v\rangle, |w\rangle)$) is a function which takes as input two vectors $|v\rangle$ and $|w\rangle$ from vector space $\mathcal{V}$ and produces a complex number as output. For example, the inner product of two n-dimensional vectors over the field of complex numbers is defined as

$$\langle v|w\rangle = \sum_i a_i^* b_i = \begin{pmatrix} a_1^* \cdots a_n^* \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}. \tag{2.8}$$

The inner product satisfies the following requirements:

1. It is linear in the second argument

$$\left(|v\rangle, \sum_i \lambda_i |w_i\rangle\right) = \sum_i \lambda_i \left(|v\rangle, |w_i\rangle\right).$$

2. $\langle v|w\rangle = \langle w|v\rangle^*$.
3. $\langle v|v\rangle \geq 0$ with equality if and only if $|v\rangle = 0$.

In the following some definitions in connection with the inner product are given.

Definition 2.1 Let $\mathcal{V}$ be a set of vectors over $\mathbb{C}^n$. Then two vectors $|v\rangle, |w\rangle \in \mathcal{V}$ are *orthogonal*, if their inner product is 0.

Definition 2.2 Let $\mathcal{V}$ be a set of vectors over $\mathbb{C}^n$. The norm of a vector $|v\rangle \in \mathcal{V}$ is defined by $\||v\rangle\| = \sqrt{\langle v|v\rangle}$. The norm of a vector is often understood as its length or size.

Definition 2.3 Let $\mathcal{V}$ be a set of vectors over $\mathbb{C}^n$. Then a vector $|v\rangle \in \mathbb{V}$ is called a unit vector or the vector is normalized if $\||v\rangle\| = 1$. Normalizing a vector means dividing it by its norm:

$$\left\| \frac{|v\rangle}{\||v\rangle\|} \right\| = 1.$$

Definition 2.4 Let $\mathcal{V}$ be a set of vectors over $\mathbb{C}^n$. Then a subset of vectors $|v_i\rangle \in \mathbb{V}$ is called orthonormal if each vector $|v_i\rangle$ is a unit vector, and distinct vectors are orthogonal $\langle v_i|w_j\rangle = 0$, $i, j = 1...n$, $i \neq j$.

For the computational basis of a vector space the last definition of orthonormal must hold. So those vectors form the spanning set for the vector space and any vector out of this space can be written as a linear combination.

2.1.2.3 Outer Products

The *outer product* of two vectors is the contrary multiplication to the inner product. In opposite to the inner product resulting in a single complex value, the outer product yields to a matrix:

$$|v\rangle\langle w| = A_{i,j} = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} \cdot \left(b_1^* \cdots b_n^* \right) = \begin{pmatrix} a_1 b_1^* & \cdots & a_1 b_n^* \\ \vdots & \ddots & \vdots \\ a_m b_1^* & \cdots & a_m b_n^* \end{pmatrix}. \tag{2.9}$$

The outer product representation is a useful way of representing linear operators which makes use of the inner product. Let $|v\rangle$ be a vector in an inner product space $\mathcal{V}$ and $|w\rangle$ be a vector in an inner product space $\mathcal{W}$. Define $|w\rangle\langle v|$ to be the linear operator from $\mathcal{V}$ to $\mathcal{W}$ like

$$(|w\rangle\langle v|)(|v'\rangle) = |w\rangle\langle v|v'\rangle = \langle v|v'\rangle|w\rangle. \tag{2.10}$$

The equation imposes at least two interpretations. On the one hand the vector $|v'\rangle$ is mapped by the matrix to a vector which lies in $\mathcal{W}$; on the other hand, it is only the representation of the vector $|w\rangle$ multiplied by a complex value.

One application of the outer product notation can be discerned from an important result known as the completeness relation for orthonormal vectors. Let $|v_i\rangle$ be orthonormal basis for the vector space $\mathcal{V}$. Then following equation must be fulfilled

$$\sum_i |v_i\rangle\langle v_i| = I. \tag{2.11}$$

2.1.2.4 Tensor Products

The *tensor product* is an operation to create a larger vector space from two smaller vector spaces. We have two vector spaces V and W of dimensions m and n, respectively. Then $V \otimes W$ is an mn dimensional vector space, whose elements are linear combinations of tensor products of elements $|v\rangle \in V$ and $|w\rangle \in W$.

For example, the tensor product of vectors (1, 2) and (3, 4) is the vector

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} \otimes \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 \times 3 \\ 1 \times 4 \\ 2 \times 3 \\ 2 \times 4 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ 6 \\ 8 \end{pmatrix}.$$

In the previous example, we perform the operation on two vector spaces, but the tensor product can also be applied on the linear operators of vector spaces. Assume $A : V \rightarrow V'$ and $B : W \rightarrow W'$ then $A \otimes B : V \otimes W \rightarrow V' \otimes W'$. Suppose A is a m by n matrix, and B is a p by q matrix. Then we have the matrix representation:

$$A \otimes B \equiv \underbrace{\left.\begin{pmatrix} A_{11}B & A_{12}B & \cdots & A_{1n}B \\ A_{21}B & A_{22}B & \cdots & A_{2n}B \\ \vdots & \vdots & \vdots & \vdots \\ A_{m1}B & A_{m2}B & \cdots & A_{mn}B \end{pmatrix}\right\}}_{nq} mp, \tag{2.12}$$

where $A_{11}B$ denotes a p by q submatrix. For example, the tensor product of the Pauli matrices X and Y is

$$X \otimes Y = \begin{pmatrix} 0 \cdot Y & 1 \cdot Y \\ 1 \cdot Y & 0 \cdot Y \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix}.$$

2.1.3 Quantum Measurement

This section provides ways for describing the effects of measurements on quantum systems with reference to the four postulates in [4]. Before we can measure a quantum we have to set up the area in which quantum mechanics takes place.

Postulate 1 of [4]: A complex vector space with inner product (also called Hilbert space) is related with any isolated physical system. This is also known as the state space of the system. With the unit vectors of the system's state space (state vectors) we can span the complete system.

To get more information of a particular system we would measure the state space. But not in quantum mechanics, here we cannot measure what the state space of the system is, nor we can tell what the state vector of that system is. The simplest and most important system is the *qubit*, described in Sect. 2.1.1. The next postulate gives the description how states change with time.

Postulate 2 of [4]: The evolution of a *closed* quantum system is described by a *unitary transformation*. That is, the state $|\psi\rangle$ of the system at time t_1 is related to the state $|\psi'\rangle$ of the system at time t_2 by a unitary operator U which depends only on the times t_1 and t_2,

$$|\psi'\rangle = U|\psi\rangle. \tag{2.13}$$

After quantum mechanics does not tell us the state space and quantum state of a system, it only assures us which unitary operators U describe the change in any closed quantum system. Such operators we have already seen in Sect. 2.1.2. Setting up the base, we can continue with the measurement.

Postulate 3 of [4]: Quantum measurements are described by a collection $\{M_m\}$ of *measurement operators*. These are operators acting on the state space of the system being measured. The index m refers to the measurement outcomes that may occur in the experiment. If the state of the quantum system is $|\psi\rangle$ immediately before the measurement then the probability that result m occurs is given by

$$p(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle, \tag{2.14}$$

and the state of the system after the measurement is

$$\frac{M_m|\psi\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}}. \tag{2.15}$$

The measurement operators satisfy the completeness equation

$$\sum_m M_m^\dagger M_m = I. \tag{2.16}$$

The completeness equation expresses the fact that the probabilities sum to 1:

$$\sum_m p\,(m) = \sum_m \langle\psi|M_m^\dagger M_m|\psi\rangle = 1. \tag{2.17}$$

There are different types of measurements, but the most important in our case is the measurement in the computational basis. Hence, we know that $|0\rangle$ and $|1\rangle$ form a computational basis for the two-dimensional complex vector space (space of qubits). As said in a previous section we can map these states onto the states 0 and 1 of a classical bit during measurement. Now we define two measurement operators M_0, M_1:

$$M_0 = |0\rangle\langle 0| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{and} \qquad M_1 = |1\rangle\langle 1| = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{2.18}$$

Observe that for the operators apply $M_0^\dagger = M_0$, $M_1^\dagger = M_1$ and $M_0^2 = M_0$, $M_1^2 = M_1$. A measurement on a qubit with state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$. If we use the operator M_0 for the qubit we obtain a probability that the result is 0

$$\begin{aligned} p(0) &= \langle\psi|M_0^\dagger M_0|\psi\rangle \\ &= \langle\psi|M_0|\psi\rangle \\ &= \begin{pmatrix}\alpha^* & \beta^*\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix} \\ &= \begin{pmatrix}\alpha^* & \beta^*\end{pmatrix}\begin{pmatrix}\alpha\\ 0\end{pmatrix} \\ &= \alpha^*\alpha + 0 = |\alpha|^2. \end{aligned} \tag{2.19}$$

Similarly, we get the probability $p(1) = |\beta|^2$ for the measurement result 1.

The state of the system after the measurement is

$$\begin{aligned} |\psi'\rangle = \frac{M_0|\psi\rangle}{\sqrt{\langle\psi|M_0^\dagger M_0|\psi\rangle}} &= \frac{\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}}{\sqrt{|\alpha|^2}} \\ = \frac{\begin{pmatrix}\alpha\\ 0\end{pmatrix}}{|\alpha|} &= \frac{\alpha|0\rangle}{|\alpha|} = \frac{\alpha}{|\alpha|}|0\rangle \end{aligned} \tag{2.20}$$

if the result was 0. Analog the state

$$|\psi'\rangle = \frac{\beta}{|\beta|}|1\rangle \tag{2.21}$$

for the measurement result 1.

Based on Eq. 2.4 if the qubit is in the specific state $|+\rangle$ we have $\alpha = \beta = \frac{1}{\sqrt{2}}$ and using Eq. 2.14 we get 0 as well as 1 with probability

$$p(0) = p(1) = |\alpha|^2 = |\beta|^2 = \left|\frac{1}{\sqrt{2}}\right|^2 = \frac{1}{2}. \tag{2.22}$$

Due to Eq. 2.15 after the measurement the system's state is

$$\frac{\alpha}{|\alpha|}|0\rangle = \frac{\frac{1}{\sqrt{2}}}{\left|\frac{1}{\sqrt{2}}\right|}|0\rangle = |0\rangle \qquad \text{or} \qquad \frac{\beta}{|\beta|}|1\rangle = \frac{\frac{1}{\sqrt{2}}}{\left|\frac{1}{\sqrt{2}}\right|}|1\rangle = |1\rangle, \tag{2.23}$$

respectively. We get the same results for the state $|-\rangle$.

To perform a correct measurement for such states, we cannot use the computational basis $\{|0\rangle, |1\rangle\}$. Therefore we have to use the basis $\{|+\rangle, |-\rangle\}$ (defined in Eq. 2.4). We have seen that these states are orthonormal to each other, since $\langle+|+\rangle = \langle-|-\rangle = 1$ and $\langle+|-\rangle = 0$, so we can use them as computational basis.

As before, we map the two states $|+\rangle$, $|-\rangle$ onto classical bits, e.g., $|+\rangle \rightarrow 0$ and $|-\rangle \rightarrow 1$. Thus, the two operators are

$$M_0 = |+\rangle\langle+| = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \qquad \text{and} \qquad M_1 = |-\rangle\langle-| = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \tag{2.24}$$

Using these measurement operators and the states $|+\rangle$ and $|-\rangle$ of Eq. 2.4 we get the probability for the result 0

$$p\,(0) = \langle+|M_0^\dagger M_0|+\rangle = \langle+|+\rangle\langle+|+\rangle = 1 \cdot 1 = 1. \tag{2.25}$$

So for the other case

$$p\,(1) = \langle+|M_1^\dagger M_1|+\rangle = \langle+|-\rangle\langle-|+\rangle = 0 \cdot 0 = 0, \tag{2.26}$$

and the state after the measurement is

$$\frac{M_0|+\rangle}{\sqrt{\langle+|M_0^\dagger M_0|+\rangle}} = \frac{|+\rangle\langle+|+\rangle}{\sqrt{\langle+|+\rangle\langle+|+\rangle}} = \frac{1}{1} \cdot |+\rangle = |+\rangle. \tag{2.27}$$

As the probability for the result 0 was 1 the state $|+\rangle$ is preserved even after the measurement.

With the computational basis $\{|0\rangle, |1\rangle\}$ and measurement states $|+\rangle$, $|-\rangle$, now the measurement of states $|0\rangle$, $|1\rangle$ in the computational basis $\{|+\rangle, |-\rangle\}$ yields the same probability results:

$$\begin{aligned} p\,(0) &= \langle 0|M_0^\dagger M_0|0\rangle = \langle 0|+\rangle\langle+|0\rangle = \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} = \frac{1}{2}, \\ p\,(1) &= \langle 0|M_1^\dagger M_1|0\rangle = \langle 0|-\rangle\langle-|0\rangle = \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} = \frac{1}{2}, \\ p\,(0) &= \langle 1|M_0^\dagger M_0|1\rangle = \langle 1|+\rangle\langle+|1\rangle = \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} = \frac{1}{2}, \\ p\,(1) &= \langle 1|M_1^\dagger M_1|1\rangle = \langle 1|-\rangle\langle-|1\rangle = -\frac{1}{\sqrt{2}} \cdot -\frac{1}{\sqrt{2}} = \frac{1}{2}, \end{aligned} \tag{2.28}$$

and the state after the measurement is

$$\begin{aligned} &\frac{M_0|0\rangle}{\sqrt{\langle 0|M_0^\dagger M_0|0\rangle}} = \frac{|+\rangle\langle +|0\rangle}{\sqrt{\langle 0|+\rangle\langle +|0\rangle}} = \frac{\frac{1}{\sqrt{2}}}{\frac{1}{\sqrt{2}}} \cdot |+\rangle = |+\rangle \\ \text{or} \quad &\frac{M_1|0\rangle}{\sqrt{\langle 0|M_1^\dagger M_1|0\rangle}} = \frac{|-\rangle\langle -|0\rangle}{\sqrt{\langle 0|-\rangle\langle -|0\rangle}} = \frac{\frac{1}{\sqrt{2}}}{\frac{1}{\sqrt{2}}} \cdot |-\rangle = |-\rangle \\ \text{and} \quad &\frac{M_0|1\rangle}{\sqrt{\langle 1|M_0^\dagger M_0|1\rangle}} = \frac{|+\rangle\langle +|1\rangle}{\sqrt{\langle 0|+\rangle\langle +|0\rangle}} = \frac{\frac{1}{\sqrt{2}}}{\frac{1}{\sqrt{2}}} \cdot |+\rangle = |+\rangle \\ \text{or} \quad &\frac{M_1|1\rangle}{\sqrt{\langle 1|M_1^\dagger M_1|1\rangle}} = \frac{|-\rangle\langle -|1\rangle}{\sqrt{\langle 1|-\rangle\langle -|1\rangle}} = \frac{-\frac{1}{\sqrt{2}}}{\frac{1}{\sqrt{2}}} \cdot |-\rangle = (-1)\cdot|-\rangle = |-\rangle. \end{aligned} \tag{2.29}$$

In Sect. 2.1.1 we said if qubits differ only by a global phase factor, they have the same statistical properties and so we consider them to be the same. Thus, in the last line in Eq. 2.29 we can neglect the factor -1.

2.1.4 The No-Cloning Theorem

Is it possible to make a copy of an unknown quantum state? The answer is no. In [10] the no-cloning theorem was presented the first time.

Suppose we have a quantum machine with two slots labeled A and B. Slot A is the *data slot* and starts out in a quantum state, $|\psi\rangle$. We do not know which state it has. The goal is to copy the state into slot B, the *target slot*. We assume that the target slot starts out in some independent state, $|s\rangle$. Thus the initial state of the copying machine is

$$|\psi\rangle \otimes |s\rangle. \tag{2.30}$$

Some unitary evolution U now effects the copying procedure, ideally,

$$|\psi\rangle \otimes |s\rangle \rightarrow_U U(|\psi\rangle \otimes |s\rangle) = |\psi\rangle \otimes |\psi\rangle. \tag{2.31}$$

Suppose this copying procedure works for two particular states, $|\psi\rangle$ and $|\varphi\rangle$. Then we have

$$U(|\psi\rangle \otimes |s\rangle) = |\psi\rangle \otimes |\psi\rangle, \tag{2.32}$$

$$U(|\varphi\rangle \otimes |s\rangle) = |\varphi\rangle \otimes |\varphi\rangle. \tag{2.33}$$

Taking the inner product of these two equations gives

$$\langle\psi|\varphi\rangle = (\langle\psi|\varphi\rangle)^2. \tag{2.34}$$

But $x = x^2$ has only two solution, $x = 0$ and $x = 1$, so either $|\psi\rangle = |\varphi\rangle$ or $|\psi\rangle$ and $|\varphi\rangle$ are orthogonal. If a cloning device would exist it can only clone states which are orthogonal to one another, and therefore a general quantum cloning device cannot exist. For example, we have qubits with states $|\psi\rangle$ and $|0\rangle$ where ψ is not $|1\rangle$, so it is impossible for a quantum cloner to copy them, since these states are not orthogonal.

2.2 Unconditional Secure Authentication

QKD communication is divided into two channels: the quantum channel and the classical public channel. On the public channel we have the problem that the adversary Eve can start a man-in-the-middle attack. So the need of authentication for the communication is essential. Usually authentication is done by public key methods, e.g., RSA [6] or DSS [3]. But this security is only computational. QKD's security is an unconditional secure, by using such an authentication scheme we will reduce it also to computational security. Therefore unconditional secure authentication schemes are necessary for QKD.

Next we present universal hashing, which can be used as an authentication scheme, where the authentication parties hold a pre-shared secret. This was first presented by Wegman and Carter [9]. Nowadays, more effective symmetric authentication methods are known. We will use Wegman–Carter authentication, because it describes an upper bound for needed symmetric authentication key.

2.2.1 Universal Hashing

The fundamental idea of universal hashing is to choose a hash function at random, independent of the input, such that the probability that any two distinct inputs have the same hash values is sufficiently low. To render more precisely, let $\mathcal{A}$ and $\mathcal{B}$ be finite sets, characterized by the number of elements $a = |\mathcal{A}|$ and $b = |\mathcal{B}|$, where $a \geq b$. The hash function h maps an element $x \in \mathcal{A}$ into an element $h(x) \in \mathcal{B}$. For h and for $x, y \in \mathcal{A}$, $x \neq y$, we define $\delta_h(x, y) = 1$ if $h(x) = h(y)$ and $\delta_h(x, y) = 0$ otherwise. Meaning $\delta_h(x, y) = 1$ if and only if the hashed values of x and y collide. The collection of all hash functions h, mapping the set $\mathcal{A}$ to set $\mathcal{B}$, is given as $\mathcal{H}$. Hence, the number of collisions for the hash function set $\mathcal{H}$ and the values x, y is defined as

$$\delta_{\mathcal{H}}(x, y) = \sum_{h \in \mathcal{H}} \delta_h(x, y).$$

Thus, the probability for the two distinct values can be computed by $\delta_{\mathcal{H}}(x, y)/|\mathcal{H}|$. The goal is to make this probability small, which can be achieved by a large number of hash functions $|\mathcal{H}|$. But when increasing $|\mathcal{H}|$ the number of bits for specifying the function $\log_2|\mathcal{H}|$ increases as well. This case would be unpractical for applications since more random bits are needed. So we have to consider this when choosing $\mathcal{H}$.

Wegman and Carter [1, 9] were the first ones dealing with universal hashing and they defined some useful properties. Below, a definition of the term *universal* is given [8].

Definition 2.5 Let ε be a positive real number. $\mathcal{H}$ is ε-almost $universal_2$ (or ε-AU_2) if $\delta_{\mathcal{H}}(x, y) \leq \varepsilon|\mathcal{H}|$ for all $x, y \in \mathcal{A}, x \neq y$.

In other words the collision probability for any two inputs x and y is at most ε. Note that ε is bounded below by $1/b$. In the special case when $\varepsilon = 1/b$ we can skip the term *almost* and speak only about a $universal_2$ class of hash functions as Wegman and Carter [1] have done it. Next we define a stricter property for such classes [8].

Definition 2.6 Let ε be a positive real number. $\mathcal{H}$ is ε-almost strongly $universal_2$ (or ε-ASU_2) if

(a) for every $x_1 \in \mathcal{A}$ and for every $y_1 \in \mathcal{B}$, $|\{h \in \mathcal{H} : h(x_1) = y_1\}| = |\mathcal{H}|/|\mathcal{B}|$,
(b) for every $x_1, x_2 \in \mathcal{A}(x_1 \neq x_2)$ and for every $y_1, y_2 \in \mathcal{B}$,
$|\{h \in \mathcal{H} : h(x_1) = y_1, h(x_2) = y_2\}| \leq \varepsilon|\mathcal{H}|/|\mathcal{B}|$.

The first part of the definition says that any input x_1 has the probability $1/b$ to be mapped to a hashed value y_1 (see Fig. 2.2). The second part concerns the conditional probability that with given tuple (x_1, y_1) the probability that x_2 is mapped to y_2 is at most ε (see Fig. 2.3). As before in the special case we call it *strongly* $universal_2$ class of hash functions when $\varepsilon = 1/b$.

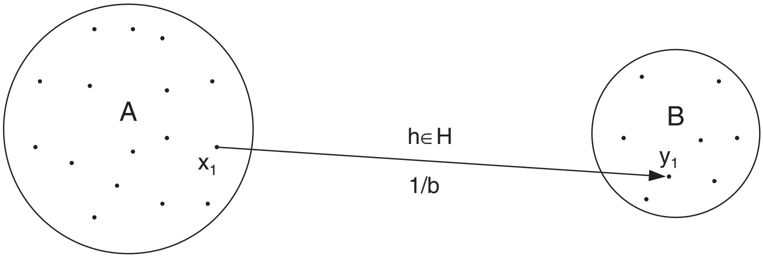

Fig. 2.2 Definition 2.6a. Any element of $\mathcal{A}$ is mapped to any element of $\mathcal{B}$ with probability $1/b$

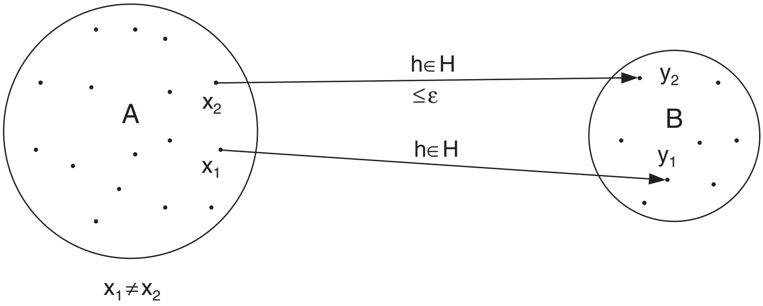

Fig. 2.3 Definition 2.6b. With given tuple (x_1, y_1) any x_2 is mapped to y_2 with probability $\leq \varepsilon$

2.2.2 Authentication

In the previous section we define what *ε-almost strongly* $universal_2$ families of hash function are. For authentication we can use them in the following manner: Alice and Bob share a secret random key k and have agreed on a ε-ASU set of hash function $\mathcal{H} = \{h|h : \mathcal{M} \rightarrow \mathcal{T}\}$ (not necessarily secret), where $\mathcal{M}$ is the set of possible messages, $\mathcal{T}$ the set of authentication tags, k has the length $\log_2(|\mathcal{H}|)$, and the hash function in $\mathcal{H}$ are ordered $\{h_0, h_1, ..., h_{|\mathcal{H}|-1}\}$. Alice sends Bob a message $m \in \mathcal{M}$ and the authentication tag $t = h_k(m)$. For Bob now it is easy to check if the message is really from Alice, he compares if the received tag t is equal to $h_k(m)$ and accepts the message as authentic if it does. The key k is then discarded. The authentication system's resistance against forgery has to be stated. A forger has two possibilities to attack the system: First, he can place a new message m' into the channel. Because the key k is random and secret, he does not know which hash function h_k he has to use and his only choice is a guess. Due to the Definition 2.62.6 the probability of success is $1/|\mathcal{T}|$. Second possibility for the forger is to wait for a message pair (m, t). With the knowledge of the tuple (m, t), the class of hash function $\mathcal{H}$ and enough computing power, the forger can find all $|\mathcal{H}|/|\mathcal{T}|$ matching keys or hash functions, respectively. Nevertheless, according to Definition 2.62.6 with this knowledge he has a probability of at most ε to guess the correct authentication tag for a modified message m', where $m' \neq m$.

So the aim is to make the ε very small concerning the security aspect. This means to bring it down to $1/|\mathcal{T}|$, which is the lower bound. The disadvantage when $\varepsilon = 1/|\mathcal{T}|$ is that the number of hash functions is very high. Since we need a string (key) to address the hash function, the size of it increases with $\log_2 |\mathcal{H}|$, which results in key length as long as the message length or higher ($|k| \in O(|m|)$). This is not practicable for applications, so we have to find an ε small enough but which increases the key length logarithmic in comparison to the message length.

The first solution for an *almost strongly* $universal_2$ class increasing the key logarithmic was presented by Wegman and Carter [9]. Their construction is shown in Fig. 2.4. Let $s = b + \log_2 \log_2 a$, where $b = \log_2 |\mathcal{T}|$ is the length of an authentication tags and $a = \log_2 |\mathcal{M}|$ is the length of the messages. Construct a *strongly*

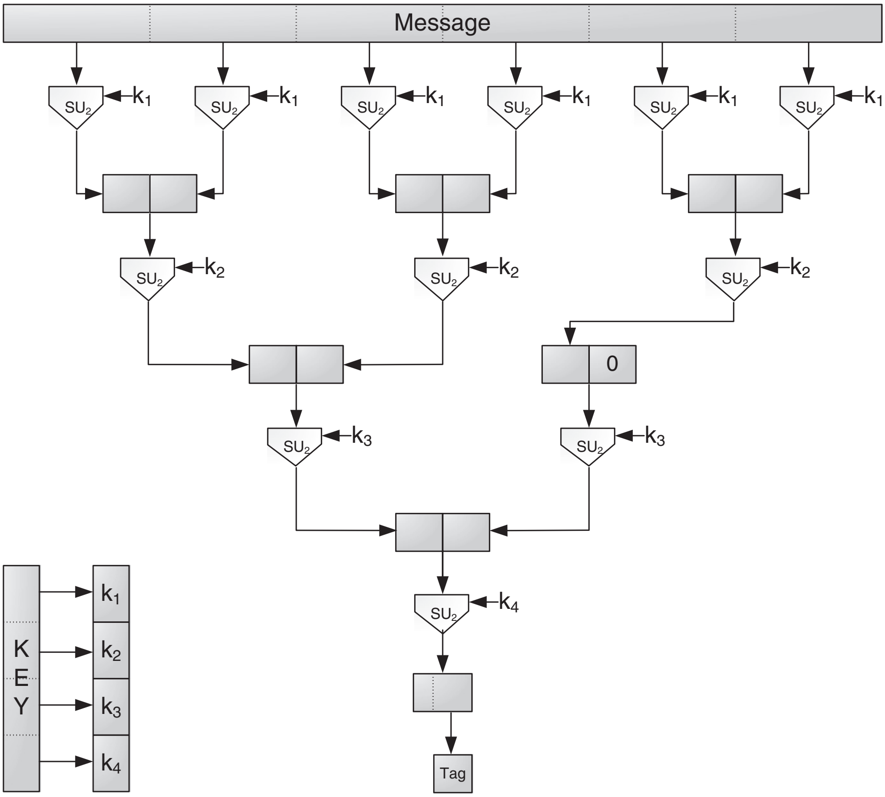

Fig. 2.4 Schematic of $2/|\mathcal{T}|$-*almost strongly universal*$_2$ class of hash function

universal$_2$ class of hash functions, which maps an input string of length $2s$ to a hash value of length s. The idea now is to split the message in blocks of length $2s$. If necessary, add zero padding at the end. Now pick a hash function $h \in \mathcal{H}$ and apply it on all blocks. The iteration ends with concatenating all outputs. Now start the new iteration again with splitting this string in blocks of length $2s$ and picking a hash function until only one block with length s remains. Finally the tag is the low-order $|\mathcal{T}|$ bits of this block. To generate a tag we need $\log_2 a - \log_2 b$ hash functions or keys, respectively. Before we mentioned that the key length in *strongly universal*$_2$ classes increases $O(|m|)$. Now the input for the hash function is a block of length $2s$. A key for one iteration would be $O(2s) = O(s) = O(b + \log_2 \log_2 a)$. We need $\log_2 a - \log_2 b$ iteration resulting in a key length of $O((b + \log_2 \log_2 a) \cdot \log_2 a)$. For further use we will use the suggestions in [9], where a *strongly universal*$_2$ class was used, needing a key roughly twice the size of the input. With this class the following equation is definitively an upper bound for the key size:

$$4 \cdot ((b + \log_2 \log_2 a) \cdot \log_2 a). \tag{2.35}$$

We achieve that the key increases only logarithmic in comparison to the message. To prove that this construction is an *ε-almost strongly universal$_2$* class of hash functions, we have to take a look on the iterations. After every iteration the probability for two distinct messages m_1, m_2 to be equal is $1/2^s$. Since we iterate $\log_2 a - \log_2 b$ times, the probability that the final tags match is at most $\log_2 a/2^s$, which equals to $1/2^b = 1/|\mathcal{T}|$. Because we use a *strongly universal$_2$* class of hash functions the last reduction will be taken m_1 to any tag t_1 with equal probability and fulfills the first part of Definition 2.6a. And as long as the penultimate blocks of m_1 and m_2 are different, m_2 will also be taken into any tag t_2 with probability less than $1/|\mathcal{T}|$, but if these blocks are the same, less than $2/|\mathcal{T}|$ hash functions will take m_2 to any t_2. So ε is $2/|\mathcal{T}|$ and it fulfills the second part of Definition 2.6b.

The second solution presented in [9] tries to reduce the key length needed per message. Therefore again a *strongly universal$_2$* class $\mathcal{H}$ is constructed, which maps $\mathcal{M}$ to $\mathcal{T}$. The two communicating parties (Alice and Bob) split the shared key in two parts. The first part has the length $\log_2 |\mathcal{H}|$ to specify a hash function and the second part is a sequence $(r_1, r_2, ...)$ of elements of $\mathcal{T}$, each with length $\log_2 |\mathcal{T}|$. Secure authentication requires unique indexed messages. The shared secret key indicates both parties which hash function h to choose. To create now the authentication tag t_i for a message m_i with a unique message number (e.g., i), it is hashed by the hash function and then exclusive-or's with r_i, so that $t_i = h(m_i) \oplus r_i$.

For the probability of guessing an authentication tag t for a message m, we define a set for the tag t as $S_t = \{(h, r)|h \in \mathcal{H}, r \in \mathcal{T}, h(m_1) \oplus r = t_1, \text{ and } h(m) \oplus r = t\}$, where m_1 and t_1 are the first message and its authentication tag. In words, S_t is the set of partial keys which map the new message m to the tag t with the knowledge that m_1 is mapped to t_1. Since the chosen class of hash function is *strongly universal$_2$*, the size of all sets S_t is the same. The only way to extend these partial keys of S_t is to append $r_i = h(m_i) \oplus t_i$ such that message m_i maps to tag t_i for $i = 2, 3, \ldots$. Thus, the sets have the same size and m will be assigned to any tag $t \in \mathcal{T}$ with same probability as any other tag. So the guessing probability of a forger depends on the tag length which results in $1/|\mathcal{T}|$.

Other than the first construction of Wegman and Carter, the needed key size per message depends on the tag length $\log_2 |\mathcal{T}|$ (neglecting the specifier for the hash function). The first construction returns one authentication tag for one large message, in contrast the second construction splits the message in smaller pieces, numbers them, and generate for each piece an authentication tag. In Fig. 2.5 we compare both constructions. Let $\mathcal{H}$ be a *strongly universal$_2$* class of hash functions, where $h : \mathcal{M} \to \mathcal{T}$ and $m_i, m_{j,k} \in \mathcal{M}$, $m' \notin \mathcal{M}$, $t_i \in \mathcal{T}$, $t' \in \mathcal{T}'$, $|m_i| = 2|t_i|$, $i = 1, \ldots, n$, $j = 1, \ldots, |m'|/|t|$, $k = 1, \ldots, \log_2(|m'|/|t|)$ and $lb_b(x) : b$ lower order bits of x. The additional needed space for the numeration of messages m_i is neglected.

The probability to insert a forge tuple (m'_i, t'_i), even with knowledge of a tuple (m_i, t_i) is $1/|T|$. Contrary to the first construction, the probability to insert a forge tuple (m'', t'') even with knowledge of a tuple (m', t') is $2/|T|$. For the second construction the input size (message size) of the *strongly universal$_2$* class is much larger than the output size (tag size).

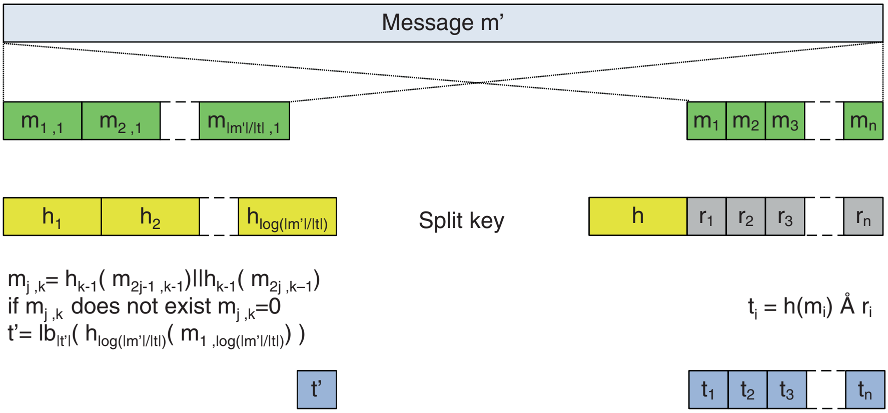

Fig. 2.5 The two constructs of Wegman and Carter [9]

2.3 Entropy

Entropy is a key concept in the field of information theory, which is a branch of applied mathematics concerned with the process to quantifying information. In the following a short part is presented, which we need in later chapters. In general it is recommendable to read up on information theory, hence we refer to the book of Cover and Thomas [2], where the subject is handled in detail.

2.3.1 Shannon Entropy

The Shannon entropy [7] is a basic measure in information theory. Let X be a discrete random variable on a finite set $\mathcal{X} = \{x_1, ..., x_n\}$ than the Shannon entropy (or information entropy) $H(X)$ is defined as

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x). \tag{2.36}$$

where $p(x)$ is the probability distribution function of X. In other words, $H(X)$ is the expected value of the amount of bits needed to specify a value $x \in \mathcal{X}$. Therefore it is easy to see that the entropy is upper bounded by

$$H(X) \leq \log_2 |\mathcal{X}|, \tag{2.37}$$

with equality if $p(x) = 1/|X|$ for each x. For probabilities $p(x) = 0$ we have the convention $0 \log_2 0 = 0$. Usually the base of the logarithm is 2, so the entropy is measured in *bit*. Other bases are e and 10.

If X and Y are random variables on $\mathcal{X}$ and $\mathcal{Y}$, respectively, the conditional Shannon entropy $H(X|Y)$ is defined as

$$H(X|Y) = -\sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} p(y)p(x|y)\log_2 p(x|y), \tag{2.38}$$

where $p(x|y) = \frac{p(x,y)}{p(y)}$ is the conditional probability distribution of X given Y.

Further, the mutual information $I(X;Y)$ is defined as

$$I(X;Y) = \sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} p(x,y)\log_2 \frac{p(x|y)}{p(x)} = \sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} p(x,y)\log_2 \frac{p(x,y)}{p(x)p(y)}. \tag{2.39}$$

Note that it is called mutual because

$$I(X;Y) = I(Y;X). \tag{2.40}$$

To understand the coherency between $H(X)$, $H(Y)$, $H(X|Y)$, $H(Y|X)$, and $I(X;Y)$ take a look on Fig. 2.6.

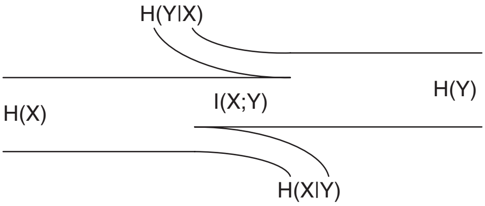

Fig. 2.6 Interpretation of $H(X)$, $H(Y)$, $H(X|Y)$, $H(Y|X)$, and $I(X;Y)$

2.3.2 Rényi Entropy

The Rényi entropy [5] of order α is a generalization of Shannon entropy. In the same manner as Shannon entropy, let X be a discrete random variable on a finite set $\mathcal{X} = \{x_1, ..., x_n\}$ than the Rényi entropy of order α $H_\alpha(X)$ with $\alpha \geq 0, \alpha \neq 1$ is defined as

$$H_\alpha(X) = \frac{1}{1-\alpha}\log_2 \sum_{x\in\mathcal{X}} p(x)^\alpha, \tag{2.41}$$

where $p(x)$ is the probability distribution function of X. In the case $\alpha \to 1$, $H_\alpha(X)$ converges to Shannon entropy $H(X)$. Again we use for the logarithm base 2 and the output of the entropy is *bit*. We pick out the Rényi entropy of order 2 ($R(X)$ for short):

$$R(X) = H_2(X) = -\log_2 \sum_{x\in\mathcal{X}} p(x)^2. \tag{2.42}$$

If we consider only the term in the logarithm, we observe that it equals the collision probability. The collision probability $p_c(X)$ of X is defined as the probability that X results in the same value or event twice in two independent executions. The connection to Shannon entropy is that it is upper bounded by Shannon

$$R(X) \leq H(X), \tag{2.43}$$

for every probability distribution $p(x)$.

If X and Y are random variables on $\mathcal{X}$ and $\mathcal{Y}$, respectively, the conditional Rényi entropy $R(X|Y)$ is defined as

$$R(X|Y) = -\sum_{y\in\mathcal{Y}} p(y) \log_2 \sum_{x\in\mathcal{X}} p(x|y)^2, \tag{2.44}$$

where $p(x|y)$ is the conditional probability distribution of X given Y. The upper bound can be derived following Eq. 2.43 as

$$R(X|Y) \leq H(X|Y). \tag{2.45}$$

References

1. Carter, L., Wegman, M.N.: Universal classes of hash functions. J. Comput. Syst. Sci. **18**(2), 143–154 (1979)
2. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley-Interscience (1991). URL http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09–20&amp;path=ASIN/0471062596
3. National Institute of Standards and Technology: FIPS PUB 186–2: Digital Signature Standard (DSS). National Institute for Standards and Technology, Gaithersburg, MD, USA (2000). URL http://www.itl.nist.gov/fipspubs/fip186–2.pdf
4. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information. Cambridge University Press, Cambridge (2000). URL http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09–20%&amp;path=ASIN/0521635039
5. Rényi, A.: On measures of information and entropy. Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability, pp. 547–561 (1961)
6. Rivest, R.L., Shamir, A., Adleman, L.M.: A method for obtaining digital signatures and public-key cryptosystems. Commun. ACM **21**(2), 120–126 (1978)
7. Shannon, C.E.: A mathematical theory of communication. The Bell System Technical Journal **27**, 379–423, 623–656 (1948). URL http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf
8. Stinson, D.R.: Universal hashing and authentication codes. In: J. Feigenbaum (ed.) CRYPTO, *Lecture Notes in Computer Science*, Vol. 576, pp. 74–85. Springer, Heidelberg (1991)
9. Wegman, M.N., Carter, L.: New hash functions and their use in authentication and set equality. J. Comput. Syst. Sci. **22**(3), 265–279 (1981)
10. Wootters, W.K., Zurek, W.H.: A single quantum cannot be cloned. Nature **299**(5886), 802–803 (1982). DOI 10.1038/299802a0. URL http://dx.doi.org/10.1038/299802a0