

## Chapter 2

# Quantum Information Theory Fundamentals

**Abstract** In this chapter, we provide the basic concepts of quantum information processing and quantum information theory. The following topics from quantum information processing will be covered: state vectors, operators, density operators, measurements, dynamics of a quantum system, superposition principle, quantum parallelism, no-cloning theorem, and entanglement. The following concepts from quantum information theory will be provided: Holevo information, accessible information, Holevo bound, Shannon entropy and von Neumann entropy, Schumacher's noiseless quantum coding theorem, and Holevo–Schumacher–Westmoreland theorem.

### 2.1 State Vectors, Operators, Projection Operators, and Density Operators

In quantum mechanics, the primitive undefined concepts are *physical system*, *observable*, and *state* [1–15]. A physical system is any sufficiently isolated quantum object, say an electron, a photon, or a molecule. An observable will be associated with a measurable property of a physical system, say energy or z-component of the spin. The state of a physical system is a trickier concept in quantum mechanics compared to classical mechanics. The problem arises when considering composite physical systems. In particular, states exist, known as *entangled states*, for bipartite physical systems in which neither of the subsystems is in a definite state. Even in cases where physical systems can be described as being in a state, two classes of states are possible: pure and mixed.

### 2.1.1 State Vectors and Operators

The condition of a quantum-mechanical system is completely specified by its *state vector*  $|\psi\rangle$  in a Hilbert space  $H$  (a vector space [16] on which a positive-definite scalar product is defined) over the field of complex numbers. Any state vector  $|\alpha\rangle$ , also known as a **ket**, can be expressed in terms of basis vectors  $|\phi_n\rangle$  by

$$|\alpha\rangle = \sum_{n=1}^{\infty} a_n |\phi_n\rangle. \quad (2.1)$$

An **observable**, such as momentum and spin, can be represented by an **operator**, say  $A$ , in the vector space of question. Quite generally, an operator acts on a ket from the left:  $(A) \cdot |\alpha\rangle = A |\alpha\rangle$ , which results in another ket. An operator  $A$  is said to be *Hermitian* if

$$A^\dagger = A, \quad A^\dagger = (A^T)^*. \quad (2.2)$$

Suppose that the Hermitian operator  $A$  has a discrete set of eigenvalues  $a^{(1)}, \dots, a^{(n)}, \dots$ . The associated eigenvectors (eigenkets)  $|a^{(1)}\rangle, \dots, |a^{(n)}\rangle, \dots$  can be obtained from

$$A|a^{(n)}\rangle = a^{(n)}|a^{(n)}\rangle. \quad (2.3)$$

The Hermitian conjugate of a ket  $|\alpha\rangle$  is denoted by  $\langle\alpha|$  and called the “bra.” The space dual to ket space is known as **bra** space. There exists a one-to-one correspondence, dual correspondence (D.C.), between a ket space and a bra space:

$$\begin{aligned} |\alpha\rangle &\overset{\text{D.C.}}{\leftrightarrow} \langle\alpha| \\ |a^{(1)}\rangle, |a^{(2)}\rangle, \dots &\overset{\text{D.C.}}{\leftrightarrow} \langle a^{(1)}|, \langle a^{(2)}|, \dots \\ |\alpha\rangle + |\beta\rangle &\overset{\text{D.C.}}{\leftrightarrow} \langle\alpha| + \langle\beta| \\ c_\alpha|\alpha\rangle + c_\beta|\beta\rangle &\overset{\text{D.C.}}{\leftrightarrow} c_\alpha^* \langle\alpha| + c_\beta^* \langle\beta|. \end{aligned} \quad (2.4)$$

The *scalar (inner) product* of two state vectors  $|\phi\rangle = \sum_n a_n |\phi_n\rangle$  and  $|\psi\rangle = \sum_n b_n |\phi_n\rangle$  is defined by

$$\langle\psi|\phi\rangle = \sum_{n=1}^{\infty} a_n b_n^*. \quad (2.5)$$

### 2.1.2 Projection Operators

The eigenkets  $\{|\xi^{(n)}\rangle\}$  of operator  $\Xi$  form the basis so that arbitrary ket  $|\psi\rangle$  can be expressed in terms of eigenkets by

$$|\psi\rangle = \sum_{n=1}^{\infty} c_n |\xi^{(n)}\rangle. \quad (2.6)$$

By multiplying (2.6) with  $\langle\xi^{(n)}|$  from the left, we obtain

$$\langle\xi^{(n)}|\psi\rangle = \sum_{j=1}^{\infty} c_j \langle\xi^{(n)}|\xi^{(j)}\rangle = c_n \langle\xi^{(n)}|\xi^{(n)}\rangle + \sum_{j=1, j \neq n}^{\infty} c_j \langle\xi^{(n)}|\xi^{(j)}\rangle. \quad (2.7)$$

Since the eigenkets  $\{|\xi^{(n)}\rangle\}$  form the basis, the principle of orthonormality is satisfied  $\langle\xi^{(n)}|\xi^{(j)}\rangle = \delta_{nj}$ ,  $\delta_{nj} = \begin{cases} 1, & n = j \\ 0, & n \neq j \end{cases}$  so that (2.7) becomes

$$c_n = \langle\xi^{(n)}|\psi\rangle. \quad (2.8)$$

By substituting (2.8) into (2.6), we obtain

$$|\psi\rangle = \sum_{n=1}^{\infty} \langle\xi^{(n)}|\psi\rangle |\xi^{(n)}\rangle = \sum_{n=1}^{\infty} |\xi^{(n)}\rangle \langle\xi^{(n)}|\psi\rangle. \quad (2.9)$$

Because  $|\psi\rangle = I|\psi\rangle$  from (2.9), it is clear that

$$\sum_{n=1}^{\infty} |\xi^{(n)}\rangle \langle\xi^{(n)}| = I, \quad (2.10)$$

and the relation above is known as the *completeness relation*. The operators under summation in (2.10) are known as *projection operators*  $P_n$ :

$$P_n = |\xi^{(n)}\rangle \langle\xi^{(n)}|, \quad (2.11)$$

which satisfy the relationship  $\sum_{n=1}^{\infty} P_n = I$ . It is easy to show that the ket (2.6) with  $c_n$  determined by (2.8) is of unit length:

$$\langle\psi|\psi\rangle = \sum_{n=1}^{\infty} \langle\psi|\xi^{(n)}\rangle \langle\xi^{(n)}|\psi\rangle = \sum_{n=1}^{\infty} |\langle\psi|\xi^{(n)}\rangle|^2 = 1. \quad (2.12)$$

The following theorem is an important theorem that will be used often throughout the chapter.

It can be shown that the eigenvalues of a Hermitian operator  $A$  are real, and the eigenkets are orthogonal:

$$\langle a^{(m)} | a^{(n)} \rangle = \delta_{nm}. \quad (2.13)$$

For the proof of this claim, an interested reader is referred to [1].

### 2.1.3 Photon, Spin- $\frac{1}{2}$ Systems, and Hadamard Gate

**Photon.** The  $x$ - and  $y$ -polarizations of the photon can be represented by

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

On the other hand, the right and left circular polarizations can be represented by

$$|E_R\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ j \end{pmatrix} \quad |E_L\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -j \end{pmatrix}.$$

The  $45^\circ$  polarization ket can be represented as follows:

$$|E_{45^\circ}\rangle = \cos\left(\frac{\pi}{4}\right)|E_x\rangle + \sin\left(\frac{\pi}{4}\right)|E_y\rangle = \frac{1}{\sqrt{2}}(|E_x\rangle + |E_y\rangle) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

The bras corresponding to the left and right polarization can be written as

$$\langle E_R| = \frac{1}{\sqrt{2}}(1 \quad -j) \quad \langle E_L| = \frac{1}{\sqrt{2}}(1 \quad j).$$

It can be easily verified that the left and right states are orthogonal and that the right polarization state is of unit length:

$$\langle E_R | E_L \rangle = \frac{1}{2}(1 \quad -j) \begin{pmatrix} 1 \\ -j \end{pmatrix} = 0 \quad \langle E_R | E_R \rangle = \frac{1}{2}(1 \quad -j) \begin{pmatrix} 1 \\ j \end{pmatrix} = 1.$$

The completeness relation is clearly satisfied because

$$|E_x\rangle\langle E_x| + |E_y\rangle\langle E_y| = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2.$$

An arbitrary polarization state can be represented by

$$|E\rangle = |E_R\rangle\langle E_R|E\rangle + |E_L\rangle\langle E_L|E\rangle.$$

For example, for  $E = E_x$ , we obtain

$$|E_x\rangle = |E_R\rangle\langle E_R|E_x\rangle + |E_L\rangle\langle E_L|E_x\rangle.$$

For the photon spin operator  $S$  matrix representation, we have to solve the following eigenvalue equation:

$$S|\psi\rangle = \lambda|\psi\rangle.$$

The photon spin operator satisfies  $S^2 = I$  so that we can write

$$|\psi\rangle = S^2|\psi\rangle = S(S|\psi\rangle) = S(\lambda|\psi\rangle) = \lambda S|\psi\rangle = \lambda^2|\psi\rangle.$$

It is clear from the previous equation that  $\lambda^2 = 1$  so that the corresponding eigenvalues are  $\lambda = \pm 1$ . By substituting the eigenvalues into the eigenvalue equation, we obtain that corresponding eigenkets are the left and the right polarization states:

$$S|E_R\rangle = |E_R\rangle \quad S|E_L\rangle = -|E_L\rangle.$$

The photon spin represented in  $\{|E_x\rangle, |E_y\rangle\}$  basis can be obtained by

$$S \doteq \begin{pmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{pmatrix} = \begin{pmatrix} \langle E_x|S|E_x\rangle & \langle E_x|S|E_y\rangle \\ \langle E_y|S|E_x\rangle & \langle E_y|S|E_y\rangle \end{pmatrix} = \begin{pmatrix} 0 & -j \\ j & 0 \end{pmatrix}.$$

**Spin-1/2 Systems.** The  $S_z$  basis in spin-1/2 systems can be written as  $\{|E_z; +\rangle, |E_z; -\rangle\}$ , where the corresponding basis kets represent the spin-up and spin-down states. The eigenvalues are  $\{\hbar/2, -\hbar/2\}$ , and the corresponding eigenket–eigenvalue relation is

$S_z|S_z; \pm\rangle = \pm\frac{\hbar}{2}|S_z; \pm\rangle$ , where  $S_z$  is the spin operator that can be represented in the basis above as follows:

$$S_z = \sum_{i=+,-} \sum_{j=+,-} |i\rangle\langle j| \underbrace{S_z|i\rangle}_{i\frac{\hbar}{2}|i\rangle} \langle j| = \sum_{i=+,-} i\frac{\hbar}{2}|i\rangle\langle i| = \frac{\hbar}{2}(|+\rangle\langle +| - |-\rangle\langle -|).$$

The matrix representation of spin- $1/2$  systems is obtained by

$$|S_z; +\rangle = \begin{pmatrix} \langle S_z; + | S_z; + \rangle \\ \langle S_z; - | S_z; + \rangle \end{pmatrix} \doteq \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad |S_z; -\rangle \doteq \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$S_z \doteq \begin{pmatrix} \langle S_z; + | S_z | S_z; + \rangle & \langle S_z; + | S_z | S_z; - \rangle \\ \langle S_z; - | S_z | S_z; + \rangle & \langle S_z; - | S_z | S_z; - \rangle \end{pmatrix} = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

**Hadamard Gate.** The matrix representation of Hadamard operator (gate) is given by

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

It can easily be shown that the Hadamard gate is Hermitian and unitary as follows:

$$H^\dagger = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = H$$

$$H^\dagger H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I.$$

The eigenvalues for Hadamard gate can be obtained from  $\det(H - \lambda I) = 0$  to be  $\lambda_{1,2} = \pm 1$ . By substituting the eigenvalues into the eigenvalue equation, namely,  $H|\Psi_{1,2}\rangle = \pm|\Psi_{1,2}\rangle$ , the corresponding eigenkets are obtained as follows:

$$|\Psi_1\rangle = \begin{bmatrix} \frac{1}{\sqrt{4-2\sqrt{2}}} \\ \frac{1}{\sqrt{2\sqrt{2}}} \end{bmatrix} \quad |\Psi_2\rangle = \begin{bmatrix} \frac{1}{\sqrt{4+2\sqrt{2}}} \\ -\frac{1}{\sqrt{2\sqrt{2}}} \end{bmatrix}.$$

### 2.1.4 Density Operators

Let the large number of quantum systems of the same kind be prepared, each in one of a set of states  $|\phi_n\rangle$ , and let the fraction of the system being in state  $|\phi_n\rangle$  be denoted by probability  $P_n$  ( $n = 1, 2, \dots$ ):

$$\langle \phi_m | \phi_n \rangle = \delta_{mn}, \quad \sum_n P_n = 1. \quad (2.14)$$

Therefore, this ensemble of quantum states represents a classical *statistical mixture* of kets. The probability of obtaining  $\xi_n$  from the measurement of  $\Xi$  will be

$$\Pr(\xi_k) = \sum_{n=1}^{\infty} P_n |\langle \xi_k | \phi_n \rangle|^2 = \sum_{n=1}^{\infty} P_n \langle \xi_k | \phi_n \rangle \langle \phi_n | \xi_k \rangle = \langle \xi_k | \rho | \xi_k \rangle, \quad (2.15)$$

where the operator  $\rho$  is known as a *density operator*, and it is defined by

$$\rho = \sum_{n=1}^{\infty} P_n |\phi_n \rangle \langle \phi_n|. \quad (2.16)$$

The expected value of density operator is given by

$$\langle \rho \rangle = \sum_{k=1}^{\infty} \xi_k \Pr(\xi_k) = \sum_{k=1}^{\infty} \xi_k \langle \xi_k | \rho | \xi_k \rangle = \sum_{k=1}^{\infty} \langle \xi_k | \rho \Xi | \xi_k \rangle = \text{Tr}(\rho \Xi). \quad (2.17)$$

The density operator *properties* can be summarized as follows:

1. The density operator is Hermitian ( $\rho^+ = \rho$ ), with the set of orthonormal eigenkets  $|\phi_n \rangle$  corresponding to the nonnegative eigenvalues  $P_n$  and  $\text{Tr}(\rho) = 1$ .
2. Any Hermitian operator with nonnegative eigenvalues and trace 1 may be considered as a density operator.
3. The density operator is positive definite:  $\langle \psi | \rho | \psi \rangle \geq 0$  for all  $|\psi \rangle$ .
4. The density operator has the property  $\text{Tr}(\rho^2) \leq 1$ , with equality iff one of the prior probabilities is 1, and all the rest 0:  $\rho = |\phi_n \rangle \langle \phi_n|$ , and the density operator is then a projection operator.
5. The eigenvalues of a density operator satisfy  $0 \leq \lambda_i \leq 1$ .

The proof of these properties is quite straightforward, and the proof is left as a homework problem. When  $\rho$  is the projection operator, we say that it represents the system in a *pure state*; otherwise, with  $\text{Tr}(\rho^2) < 1$ , it represents a *mixed state*. A mixed state in which all eigenkets occur with the same probability is known as a *completely mixed state* and can be represented by

$$\rho = \sum_{k=1}^{\infty} \frac{1}{n} |\phi_n \rangle \langle \phi_n| = \frac{1}{n} I \Rightarrow \text{Tr}(\rho^2) = \frac{1}{n} \Rightarrow \frac{1}{n} \leq \text{Tr}(\rho^2) \leq 1. \quad (2.18)$$

If the density matrix has off-diagonal elements different from zero, we say that it exhibits the *quantum interference*, which means that state term can interfere with each other. Let us observe the following pure state:

$$\begin{aligned} |\psi \rangle &= \sum_{i=1}^n \alpha_i |\xi_i \rangle \Rightarrow \rho = |\psi \rangle \langle \psi| = \sum_{i=1}^n |\alpha_i|^2 |a_i \rangle \langle a_i| + \sum_{i=1}^n \sum_{j=1, j \neq i}^n \alpha_i \alpha_j^* |\xi_i \rangle \langle \xi_j| \\ &= \sum_{i=1}^n \langle \xi_i | \rho | \xi_i \rangle |\xi_i \rangle \langle \xi_i| + \sum_{i=1}^n \sum_{j=1, j \neq i}^n \langle \xi_i | \rho | \xi_j \rangle |\xi_i \rangle \langle \xi_j|. \end{aligned} \quad (2.19)$$

The first term in (2.19) is related to the probability of the system being in state  $|a_i\rangle$ , and the second term is related to the quantum interference. It appears that the off-diagonal elements of a mixed state will be zero, while these of pure state will be nonzero. Notice that the existence of off-diagonal elements is base dependent; therefore, to check for purity, it is a good idea to compute  $\text{Tr}(\rho^2)$  instead.

In the quantum information theory, the density matrix can be used to determine the amount of information conveyed by the quantum state, i.e., to compute the von Neumann entropy:

$$S = \text{Tr}(\rho \log \rho) = -\sum_i \lambda_i \log_2 \lambda_i, \quad (2.20)$$

where  $\lambda_i$  are the eigenvalues of the density matrix. The corresponding Shannon entropy can be calculated by

$$H = -\sum_i p_i \log_2 p_i, \quad (2.21)$$

where  $p_i$  is the probability of selecting the  $i$ th vector from an ensemble of orthogonal vectors.

Suppose now that  $S$  is a *bipartite composite system* with component subsystems  $A$  and  $B$ . For example, the subsystem  $A$  can represent the quantum register  $Q$  and subsystem  $B$  the environment  $E$ . The composite system can be represented by  $AB = A \otimes B$ , where  $\otimes$  stands for the tensor product. If the dimensionality of Hilbert space  $H_A$  is  $m$  and the dimensionality of Hilbert space  $H_B$  is  $n$ , then the dimensionality of Hilbert space  $H_{AB}$  will be  $mn$ . If  $|\alpha\rangle \in A$  and  $|\beta\rangle \in B$ , then  $|\alpha\rangle|\beta\rangle = |\alpha\rangle \otimes |\beta\rangle \in AB$ . If the operator  $A$  acts on kets from  $H_A$  and the operator  $B$  on kets from  $H_B$ , then the action of  $AB$  on  $|\alpha\rangle|\beta\rangle$  can be described as follows:

$$(AB)|\alpha\rangle|\beta\rangle = (A|\alpha\rangle)(B|\beta\rangle). \quad (2.22)$$

The norm of state  $|\psi\rangle = |\alpha\rangle|\beta\rangle \in AB$  is determined by

$$\langle\psi|\psi\rangle = \langle\alpha|\alpha\rangle\langle\beta|\beta\rangle. \quad (2.23)$$

Let  $\{|\alpha_i\rangle\}$  ( $\{|\beta_i\rangle\}$ ) be a basis for the Hilbert space  $H_A$  ( $H_B$ ) and let  $E$  be an ensemble of physical systems  $S$  described by the density operator  $\rho$ . The *reduced density operator*  $\rho_A$  for subsystem  $A$  is defined to be the partial trace of  $\rho$  over  $B$ :

$$\rho_A = \text{Tr}_B(\rho) = \sum_j \langle\beta_j|\rho|\beta_j\rangle. \quad (2.24)$$

Similarly, the *reduced density operator*  $\rho_B$  for subsystems  $B$  is defined to be the partial trace of  $\rho$  over  $A$ :

$$\rho_B = \text{Tr}_A(\rho) = \sum_i \langle\alpha_i|\rho|\alpha_i\rangle. \quad (2.25)$$



## 2.2 Measurements, Uncertainty Relations, and Dynamics of a Quantum System

### 2.2.1 Measurements

Each measurable physical quantity—observable (such as position, momentum, or angular momentum) is associated with a Hermitian operator that has a complete set of eigenkets. According to P. A. Dirac, “A measurement always causes the system to jump into an eigenstate of the dynamical variable that is being measured [11].” The Dirac’s statement can be formulated as the following *postulate*: an exact measurement of an observable with operator  $A$  always yields as a result one of the eigenvalues  $a^{(n)}$  of  $A$ . Thus, the measurement changes the state, with the measurement system “thrown into” one of its eigenstates, which can be represented by:  $|\alpha\rangle \xrightarrow{A \text{ measurement}} |a^{(j)}\rangle$ . If before measurement the system was in state  $|\alpha\rangle$ , the probability that the result of a measurement will be the eigenvalue  $a^{(i)}$  is given by

$$\Pr(a^{(i)}) = \left| \langle a^{(i)} | \alpha \rangle \right|^2. \quad (2.26)$$

Since at least one of the eigenvalues must occur as the result of the measurements, these probabilities satisfy

$$\sum_i \Pr(a^{(i)}) = \sum_i \left| \langle a^{(i)} | \alpha \rangle \right|^2 = 1. \quad (2.27)$$

The expected value of the outcome of the measurement of  $A$  is given by

$$\langle A \rangle = \sum_i a^{(i)} \Pr(a^{(i)}) = \sum_i a^{(i)} \left| \langle a^{(i)} | \alpha \rangle \right|^2 = \sum_i a^{(i)} \langle \alpha | a^{(i)} \rangle \langle a^{(i)} | \alpha \rangle. \quad (2.28)$$

By applying the eigenvalue equation  $a^{(i)} |a^{(i)}\rangle = A |a^{(i)}\rangle$ , (2.28) becomes

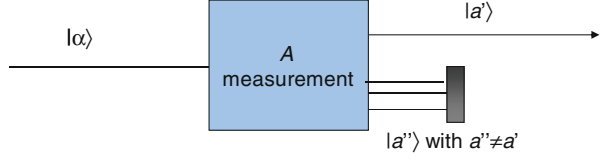
$$\langle A \rangle = \sum_i \langle \alpha | A | a^{(i)} \rangle \langle a^{(i)} | \alpha \rangle. \quad (2.29)$$

By using further the completeness relation  $\sum_i |a^{(i)}\rangle \langle a^{(i)}| = I$ , we obtain the expected value of the measurement of  $A$  to be simply

$$\langle A \rangle = \langle \alpha | A | \alpha \rangle. \quad (2.30)$$

In various situations, like initial state preparations for quantum information processing (QIP) applications, we need to select one particular outcome of the measurement. This procedure is known as the *selective measurement* (or filtration) and it can be conducted as shown in Fig. 2.1.

**Fig. 2.1** The illustration of the concept of a selective measurement (filtration)



The result of the selective measurement can be interpreted as applying the *projection operator*  $P_{a'}$  to  $|\alpha\rangle$  to obtain

$$P_{a'}|\alpha\rangle = |a'\rangle\langle a'|\alpha\rangle. \quad (2.31)$$

The probability that the outcome of the measurement of observable  $\Xi$  with eigenvalues  $\xi^{(n)}$  lies between  $(a, b)$  is given by

$$\begin{aligned} \Pr(\xi \in R(a, b)) &= \sum_{\xi^{(n)} \in R(a, b)} \left| \langle \xi^{(n)} | \alpha \rangle \right|^2 = \sum_{\xi^{(n)} \in R(a, b)} \langle \alpha | \xi^{(n)} \rangle \langle \xi^{(n)} | \alpha \rangle \\ &= \langle \alpha | P_{ab} | \alpha \rangle = \langle P_{ab} \rangle, \end{aligned} \quad (2.32)$$

where with  $P_{ab}$  we denoted the following projection operator:

$$P_{ab} = \sum_{\xi^{(n)} \in R(a, b)} |\xi^{(n)}\rangle\langle \xi^{(n)}|. \quad (2.33)$$

It can straightforwardly be shown that the projection operator  $P_{ab}$  satisfies

$$P_{ab}^2 = P_{ab} \Leftrightarrow P_{ab}(P_{ab} - I) = 0. \quad (2.34)$$

Therefore, the eigenvalues of projection operator  $P_{ab}$  are either 0 (corresponding to the “false proposition”) or 1 (corresponding to the “true proposition”), and it is of high importance in *quantum detection theory* [2].

In terms of projection operators, the state of the system after the measurement is given by

$$|\alpha\rangle \xrightarrow{\text{A measurement}} \frac{1}{\sqrt{\langle \alpha | P_j | \alpha \rangle}} P_j |\alpha\rangle, \quad P_j = |a^{(j)}\rangle\langle a^{(j)}|. \quad (2.35)$$

In case operator  $A$  has the same eigenvalue  $a_i$  for the following eigenkets  $\left\{ |a_i^{(j)}\rangle \right\}_{j=1}^{d_i}$ , with corresponding characteristic equation

$$A |a_i^{(j)}\rangle = a_i |a_i^{(j)}\rangle; \quad j = 1, \dots, d_i, \quad (2.36)$$

we say that eigenvalue  $a_i$  is *degenerate* of order  $d_i$ . The corresponding probability of obtaining the measurement result  $a_i$  can be found by

$$\Pr(a_i) = \sum_{j=1}^{d_i} \left| \langle a_i^{(j)} | \alpha \rangle \right|^2. \quad (2.37)$$

The projective measurements can be generalized as follows. Let the set of measurement operators be given by  $\{M_m\}$ , where index  $m$  stands for possible measurement result, satisfying the property  $\sum_m M_m^\dagger M_m = I$ . The probability of finding the measurement result  $m$ , given the state  $|\psi\rangle$ , is given by

$$\Pr(m) = \langle \psi | M_m^\dagger M_m | \psi \rangle. \quad (2.38)$$

After the measurement, the system will be left in following state:

$$|\psi_f\rangle = \frac{M_m |\psi\rangle}{\sqrt{\langle \psi | M_m^\dagger M_m | \psi \rangle}}. \quad (2.39)$$

For projective measurements, clearly  $M_m = P_m = |a^{(m)}\rangle \langle a^{(m)}|$ , and from the property above, we obtain

$$\sum_m M_m^\dagger M_m = \sum_m |a^{(m)}\rangle \underbrace{\langle a^{(m)} | a^{(m)} \rangle}_1 \langle a^{(m)}| = \sum_m |a^{(m)}\rangle \langle a^{(m)}| = \sum_m P_m = I, \quad (2.40)$$

which is the completeness relationship. The probability of obtaining the  $m$ , the result of the measurement, will be then

$$\Pr(m) = \text{Tr}(P_m^\dagger P_m \rho) = \text{Tr}(P_m \rho) = \text{Tr}(|a^{(m)}\rangle \langle a^{(m)}| \rho) = \langle a^{(m)} | \rho | a^{(m)} \rangle. \quad (2.41)$$

Another important type of measurement is known as a *positive operator-valued measure* (POVM). A POVM consists of the set of operators  $\{E_m\}$ , where each operator  $E_m$  is positive semidefinite, i.e.,  $\langle \psi | E_m | \psi \rangle \geq 0$ , satisfying the relationship

$$\sum_m E_m = I. \quad (2.42)$$

The POVM can be constructed from generalized measurement operators  $\{M_m\}$  by setting  $E_m = M_m^\dagger M_m$ . The probability of obtaining the  $m$ th result of measurements is given by  $\text{Tr}(E_m \rho)$ . The POVM concept is in particular suitable to situations when the measurements are not repeatable. For instance, by performing the measurement on a photon, it can be destroyed so that the repeated measurements are not possible.

### 2.2.2 Uncertainty Principle

Let  $A$  and  $B$  be two operators, which in general do not commute, i.e.,  $AB \neq BA$ . The quantity  $[A, B] = AB - BA$  is called the *commutator* of  $A$  and  $B$ , while the quantity  $\{A, B\} = AB + BA$  is called the *anticommutator*. Two observables  $A$  and  $B$  are said to be **compatible** when their corresponding operators commute:  $[A, B] = 0$ . Two observables  $A$  and  $B$  are said to be *incompatible* when  $[A, B] \neq 0$ . If in the set of operators  $\{A, B, C, \dots\}$  all operators commute in pairs, namely,  $[A, B] = [A, C] = [B, C] = \dots = 0$ , we say the set is a *complete set of commuting observables* (CSCO).

If two observables, say  $A$  and  $B$ , are to be measured simultaneously and exactly on the same system, the system after the measurement must be left in the state  $|a^{(n)}; b^{(n)}\rangle$  that is an eigenstate of both observables:

$$A|a^{(n)}; b^{(n)}\rangle = a^{(n)}|a^{(n)}; b^{(n)}\rangle, \quad B|a^{(n)}; b^{(n)}\rangle = b^{(n)}|a^{(n)}; b^{(n)}\rangle. \quad (2.43)$$

This will be true only if  $AB = BA$  or equivalently the commutator  $[A, B] = AB - BA = 0$ ; that is when two operators *commute* as shown below:

$$\begin{aligned} AB|a^{(n)}; b^{(n)}\rangle &= A(B|a^{(n)}; b^{(n)}\rangle) = Ab^{(n)}|a^{(n)}; b^{(n)}\rangle = b^{(n)} \cdot A|a^{(n)}; b^{(n)}\rangle \\ &= a^{(n)}b^{(n)}|a^{(n)}; b^{(n)}\rangle \end{aligned} \quad (2.44)$$

$$BA|a^{(n)}; b^{(n)}\rangle = a^{(n)}b^{(n)}|a^{(n)}; b^{(n)}\rangle \Rightarrow AB = BA.$$

When two operators do not commute, they cannot be simultaneously measured with the complete precision. Given an observable  $A$ , we define the operator  $\Delta A = A - \langle A \rangle$ , and the corresponding expectation value of  $(\Delta A)^2$  that is known as the *dispersion* of  $A$ :

$$\langle (\Delta A)^2 \rangle = \langle A^2 - 2A\langle A \rangle + \langle A \rangle^2 \rangle = \langle A^2 \rangle - \langle A \rangle^2. \quad (2.45)$$

Then, for any state, the following inequality is valid:

$$\langle (\Delta A)^2 \rangle \langle (\Delta B)^2 \rangle \geq \frac{1}{4} |\langle [\Delta A, \Delta B] \rangle|^2, \quad (2.46)$$

which is known as **the Heisenberg uncertainty principle** [1, 4, 17–26].

*Example* The commutation relation for coordinate  $X$  and momentum  $P$  observables is given by  $[X, P] = j\hbar$ . By substituting this commutation relation into (2.46), we obtain

$$\langle X^2 \rangle \langle P^2 \rangle \geq \frac{\hbar^2}{4}.$$

If we observe a large ensemble of  $N$  independent systems, all of them are in the state  $|\psi\rangle$ . On some systems,  $X$  is measured, and on some systems,  $P$  is measured. The uncertainty principle asserts that for none state the product of dispersions (variances) cannot be less than  $\hbar^2/4$ .

### 2.2.3 Time-Evolution Schrödinger Equation

Time-evolution operator  $U(t, t_0)$  transforms the initial ket at time instance  $t_0$ ,  $|\alpha, t_0\rangle$ , into the final ket at the time instance  $t$  by

$$|\alpha, t_0; t\rangle = U(t, t_0)|\alpha, t_0\rangle. \quad (2.47)$$

This time-evolution operators must satisfy the following two properties:

1. *Unitary property*:  $U^\dagger(t, t_0) U(t, t_0) = I$ .
2. *Composition property*:  $U(t_2, t_0) = U(t_2, t_1) U(t_1, t_0)$ ,  $t_2 > t_1 > t_0$ .

Following (2.47), the action of infinitesimal time-evolution operator  $U(t_0 + dt, t_0)$  can be described by

$$|\alpha, t_0; t_0 + dt\rangle = U(t_0 + dt, t_0)|\alpha, t_0\rangle. \quad (2.48)$$

The following operator satisfies all propositions above, when  $dt \rightarrow 0$ :

$$U(t_0 + dt, t_0) = 1 - j\Omega dt, \quad \Omega^\dagger = \Omega \quad (2.49)$$

where the operator  $\Omega$  is related to the Hamiltonian  $H$  by  $H = \hbar\Omega$  and the Hamiltonian eigenvalues correspond to the energy  $E = \hbar\omega$ . For the infinitesimal time-evolution operator  $U(t_0 + dt, t_0)$ , we can derive the time-evolution equation as follows. The starting point in derivation is the composition property:

$$U(t + dt, t_0) = U(t + dt, t)U(t, t_0) = \left(1 - \frac{j}{\hbar}H dt\right)U(t, t_0). \quad (2.50)$$

Equation (2.50) can be rewritten into the following form:

$$\lim_{dt \rightarrow 0} \frac{U(t + dt, t_0) - U(t, t_0)}{dt} = -\frac{j}{\hbar}HU(t, t_0), \quad (2.51)$$

and, by taking the partial derivative definition into account, (2.51) becomes

$$j\hbar \frac{\partial}{\partial t} U(t, t_0) = HU(t, t_0), \quad (2.52)$$

and this equation is known as **Schrödinger equation for time-evolution operator**.

The Schrödinger equation for a state ket [1, 4, 17–26] can be obtained by applying the time-evolution operator on initial ket:

$$j\hbar \frac{\partial}{\partial t} U(t, t_0) |\alpha, t_0\rangle = H U(t, t_0) |\alpha, t_0\rangle, \quad (2.53)$$

which based on (2.52) can be rewritten as

$$j\hbar \frac{\partial}{\partial t} |\alpha, t_0; t\rangle = H |\alpha, t_0; t\rangle. \quad (2.54)$$

For *conservative systems*, for which the Hamiltonian is time invariant, we can easily solve (2.54) to obtain

$$U(t, t_0) = e^{-\frac{j}{\hbar} H(t-t_0)}. \quad (2.55)$$

The time evolution of kets in conservative systems can therefore be described by applying (2.55) in (2.47), which yields to

$$|\alpha(t)\rangle = e^{-\frac{j}{\hbar} H(t-t_0)} |\alpha(t_0)\rangle. \quad (2.56)$$

Therefore, the operators do not explicitly depend on time and this concept is known as the *Schrödinger picture*.

In the *Heisenberg picture*, on the other hand, the state vector is independent of time, but operators depend on time:

$$A(t) = e^{\frac{j}{\hbar} H(t-t_0)} A e^{-\frac{j}{\hbar} H(t-t_0)}. \quad (2.57)$$

The time-evolution equation in Heisenberg picture is given by

$$j\hbar \frac{dA(t)}{dt} = [A(t), H] + j\hbar \frac{\partial A(t)}{\partial t}. \quad (2.58)$$

The density operator  $\rho$ , representing the statistical mixture of states, is independent on time in the Heisenberg picture. The expectation value of a measurement of an observable  $\Xi(t)$  at time instance  $t$  is given by

$$\begin{aligned} E_t[\Xi] &= \text{Tr}[\rho \Xi(t)] \\ E_t[\Xi] &= \text{Tr}[\rho(t) \Xi], \quad \rho(t) = e^{\frac{j}{\hbar} H(t-t_0)} \rho e^{-\frac{j}{\hbar} H(t-t_0)}. \end{aligned} \quad (2.59)$$

*Example* The Hamiltonian for a two-state system is given by

$$H = \begin{bmatrix} \omega_1 & \omega_2 \\ \omega_2 & \omega_1 \end{bmatrix}.$$

The basis for this system is given by  $\{|0\rangle = [1 \ 0]^T, |1\rangle = [0 \ 1]^T\}$ :

- (a) Determine the eigenvalues and eigenkets of  $H$ , and express the eigenkets in terms of basis.
- (b) Determine the time evolution of the system described by the Schrödinger equation

$$j\hbar \frac{\partial}{\partial t} |\psi\rangle = H|\psi\rangle, \quad |\psi(0)\rangle = |0\rangle.$$

To determine the eigenkets of  $H$ , we start from the characteristic equation  $\det(H - \lambda I) = 0$  and find that eigenvalues are  $\lambda_{1,2} = \omega_1 \pm \omega_2$ . The corresponding eigenvectors are

$$|\lambda_1\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \quad |\lambda_2\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle).$$

We now have to determine the time evolution of arbitrary ket  $|\psi(t)\rangle = [\alpha(t)\beta(t)]^T$ . The starting point is the Schrödinger equation:

$$\begin{aligned} j\hbar \frac{\partial}{\partial t} |\psi\rangle &= j\hbar \begin{bmatrix} \dot{\alpha}(t) \\ \dot{\beta}(t) \end{bmatrix}, \\ H|\psi\rangle &= \begin{bmatrix} \omega_1 & \omega_2 \\ \omega_2 & \omega_1 \end{bmatrix} \begin{bmatrix} \alpha(t) \\ \beta(t) \end{bmatrix} = \begin{bmatrix} \omega_1\alpha(t) + \omega_2\beta(t) \\ \omega_2\alpha(t) + \omega_1\beta(t) \end{bmatrix} \\ \Rightarrow j\hbar \begin{bmatrix} \dot{\alpha}(t) \\ \dot{\beta}(t) \end{bmatrix} &= \begin{bmatrix} \omega_1\alpha(t) + \omega_2\beta(t) \\ \omega_2\alpha(t) + \omega_1\beta(t) \end{bmatrix}. \end{aligned}$$

By substitution of  $\alpha(t) + \beta(t) = \gamma(t)$  and  $\alpha(t) - \beta(t) = \delta(t)$ , we obtain the ordinary set of differential equations

$$j\hbar \frac{d\gamma(t)}{dt} = (\omega_1 + \omega_2)\gamma(t) \quad j\hbar \frac{d\delta(t)}{dt} = (\omega_1 - \omega_2)\delta(t),$$

whose solution is  $\gamma(t) = C \exp\left(\frac{\omega_1 + \omega_2}{j\hbar} t\right)$  and  $\delta(t) = D \exp\left(\frac{\omega_1 - \omega_2}{j\hbar} t\right)$ . From the initial state  $|\psi(0)\rangle = |0\rangle = [1 \ 0]^T$ , we obtain the unknown constants  $C = D = 1$  so that state time evolution is given by

$$|\psi(t)\rangle = \exp\left(-\frac{j}{\hbar} \omega_1 t\right) \begin{bmatrix} \cos\left(\frac{\omega_2 t}{\hbar}\right) \\ -j \sin\left(\frac{\omega_2 t}{\hbar}\right) \end{bmatrix}.$$

## 2.3 Quantum Information Processing (QIP) Fundamentals

*Fundamental features* of QIP are different from that of classical computing and can be summarized into three: (1) linear superposition, (2) quantum parallelism, and (3) entanglement. Below we provide some basic details of these features:

1. *Linear superposition.* Contrary to the classical bit, a quantum bit or *qubit* can take not only two discrete values 0 and 1 but also *all* possible *linear combinations* of them. This is a consequence of a fundamental property of quantum states: it is possible to construct a *linear superposition* of quantum state  $|0\rangle$  and quantum state  $|1\rangle$ .
2. *Quantum parallelism.* The *quantum parallelism* is a possibility to perform a large number of operations in parallel, which represents the key difference from classical computing. Namely, in classical computing, it is possible to know what the internal status of the computer is. On the other hand, because of the no-cloning theorem, it is not possible to know the current state of the quantum computer. This property has led to the development of the Shor factorization algorithm, which can be used to crack the Rivest–Shamir–Adleman (RSA) encryption protocol. Some other important quantum algorithms include the Grover search algorithm, which is used to perform a search for an entry in an unstructured database; the quantum Fourier transform, which is a basis for a number of different algorithms; and Simon’s algorithm. The quantum computer is able to encode all input strings of length  $N$  simultaneously into a single computation step. In other words, the quantum computer is able simultaneously to pursue  $2^N$  classical paths, indicating that the quantum computer is significantly more powerful than the classical one.
3. *Entanglement.* At a quantum level, it appears that two quantum objects can form a single entity, even when they are well separated from each other. Any attempt to consider this entity as a combination of two independent quantum objects, given by tensor product of quantum states, fails, unless the possibility of signal propagation at superluminal speed is allowed. These quantum objects that cannot be decomposed into tensor product of individual independent quantum objects are called *entangled* quantum objects. Given the fact that arbitrary quantum states cannot be copied, which is the consequence of the no-cloning theorem, the communication at superluminal speed is not possible, and as consequence, the entangled quantum states cannot be written as the tensor product of independent quantum states. Moreover, it can be shown that the amount of information contained in an entangled state of  $N$  qubits grows exponentially instead of linearly, which is the case for classical bits.

In incoming subsections, we describe these fundamental features with more details.



### 2.3.1 Superposition Principle, Quantum Parallelism, Quantum Gates, and QIP Basics

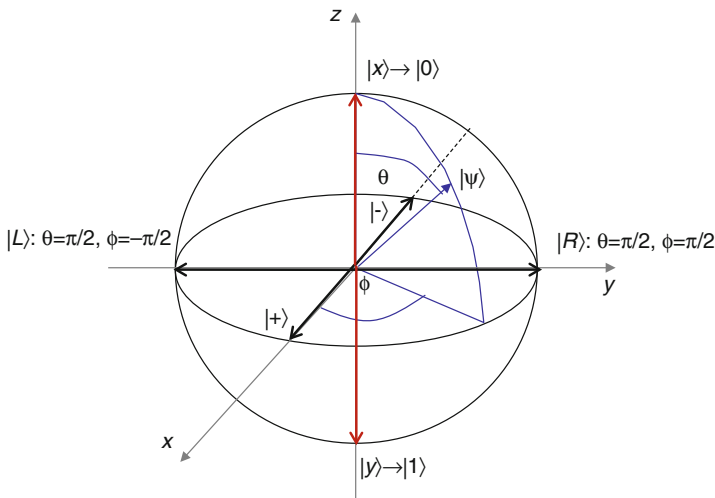
We say that the allowable states  $|\mu\rangle$  and  $|\nu\rangle$  of the quantum system satisfy the *superposition principle* if their linear superposition  $\alpha|\mu\rangle + \beta|\nu\rangle$ , where  $\alpha$  and  $\beta$  are the complex numbers ( $\alpha, \beta \in \mathbb{C}$ ), is also allowable quantum state. Without loss of generality, we typically observe the computational basis composed of the orthogonal canonical states  $|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ ,  $|1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ , so that the quantum bit, also known as *qubit*, lies in a two-dimensional Hilbert space  $H$ , isomorphic to the  $C^2$  space, and can be represented as

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}; \quad \alpha, \beta \in \mathbb{C}; \quad |\alpha|^2 + |\beta|^2 = 1. \quad (2.60)$$

If we perform the measurement of a qubit, we will get  $|0\rangle$  with probability  $|\alpha|^2$  and  $|1\rangle$  with probability of  $|\beta|^2$ . Measurement changes the state of a qubit from a superposition of  $|0\rangle$  and  $|1\rangle$  to the specific state consistent with the measurement result. If we parameterize the probability amplitudes  $\alpha$  and  $\beta$  as follows:

$$\alpha = \cos\left(\frac{\theta}{2}\right), \quad \beta = e^{j\phi} \sin\left(\frac{\theta}{2}\right), \quad (2.61)$$

where  $\theta$  is a polar angle and  $\phi$  is an azimuthal angle, we can geometrically represent the qubit by the Bloch sphere (or the Poincaré sphere for the photon) as illustrated in Fig. 2.2. Bloch vector coordinates are given by  $(\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta)$ . This Bloch vector representation is related to computational basis (CB) by



**Fig. 2.2** Bloch (Poincaré) sphere representation of the single qubit

$$|\psi(\theta, \phi)\rangle = \cos(\theta/2)|0\rangle + e^{j\phi} \sin(\theta/2)|1\rangle \doteq \begin{pmatrix} \cos(\theta/2) \\ e^{j\phi} \sin(\theta/2) \end{pmatrix}, \quad (2.62)$$

where  $0 \leq \theta \leq \pi$  and  $0 \leq \phi < 2\pi$ . The north and south poles correspond to computational  $|0\rangle$  ( $|x\rangle$ -polarization) and  $|1\rangle$  ( $|y\rangle$ -polarization) basis kets, respectively. Other important bases are the *diagonal basis*  $\{|+\rangle, |-\rangle\}$ , very often denoted as  $\{|\nearrow\rangle, |\searrow\rangle\}$ , related to CB by

$$|+\rangle = |\nearrow\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad |-\rangle = |\searrow\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle), \quad (2.63)$$

and the *circular basis*  $\{|R\rangle, |L\rangle\}$ , related to the CB as follows:

$$|R\rangle = \frac{1}{\sqrt{2}}(|0\rangle + j|1\rangle), \quad |L\rangle = \frac{1}{\sqrt{2}}(|0\rangle - j|1\rangle). \quad (2.64)$$

The pure qubit states lie on the Bloch sphere, while the mixed qubit states lie in the interior of the Bloch sphere. The maximally mixed state  $I/2$  ( $I$  denotes the identity operator) lies in the center of the Bloch sphere. The orthogonal states are antipodal. From Fig. 2.2, we see that CB, diagonal basis, and circular bases are  $90^\circ$  apart from each other, and we often say that these three bases are mutually *conjugate bases*. These bases are used as three pairs of signal states for the six-state quantum key distribution (QKD) protocol. Another important basis used in QKD and for eavesdropping is the *Breidbart basis* given by  $\{\cos(\pi/8)|0\rangle + \sin(\pi/8)|1\rangle, -\sin(\pi/8)|0\rangle + \cos(\pi/8)|1\rangle\}$ .

The superposition principle is the key property that makes quantum parallelism possible. To see this, let us juxtapose  $n$  qubits lying in  $n$  distinct two-dimensional Hilbert spaces  $H_0, H_1, \dots, H_{n-1}$  that are isomorphic to each other. In practice, this means the qubits have been prepared separately, without any interaction, which can be mathematically described by the tensor product

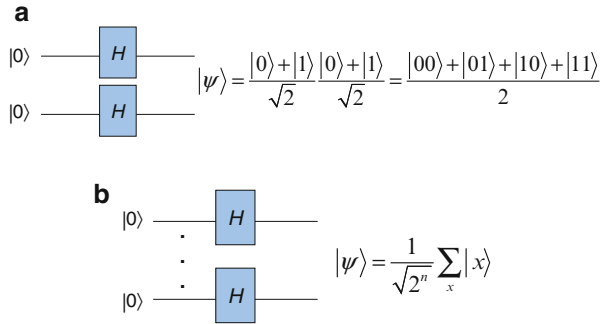
$$|\psi\rangle = |\psi_0\rangle \otimes |\psi_1\rangle \otimes \dots \otimes |\psi_{n-1}\rangle \in H_0 \otimes H_1 \otimes \dots \otimes H_{n-1}. \quad (2.65)$$

Any arbitrary basis can be selected as the computational basis for  $H_i$ ,  $i = 0, 1, \dots, n-1$ . However, for ease of exposition, we assume the computational basis to be  $|0_i\rangle$  and  $|1_i\rangle$ . Consequently, we can represent the  $i$ th qubit as  $|\psi_i\rangle = \alpha_i|0_i\rangle + \beta_i|1_i\rangle$ . Introducing a further assumption,  $\alpha_i = \beta_i = 2^{-1/2}$ , without loss of generality, we now have

$$|\psi\rangle = \prod_{i=0}^{n-1} \frac{1}{\sqrt{2}}(|0_i\rangle + |1_i\rangle) = 2^{-n/2} \sum_{\mathbf{x}} |\mathbf{x}\rangle, \quad \mathbf{x} = x_0 x_1 \dots x_{n-1}, x_j \in \{0, 1\}. \quad (2.66)$$

This composite quantum system is called the  **$n$ -qubit register**, and as can be seen from the equation above, it represents a superposition of  $2^n$  quantum states that exist simultaneously! This is an example of quantum parallelism. In the classical realm, a linear increase in size corresponds roughly to a linear increase in processing power.

**Fig. 2.3** The Walsh–Hadamard transform: (a) on two qubits and (b) on  $n$  qubits. The action of the Hadamard gate  $H$  on computational basis kets is given by:  
 $H|0\rangle = 2^{-1/2}(|0\rangle + |1\rangle)$  and  
 $H|1\rangle = 2^{-1/2}(|0\rangle - |1\rangle)$



In the quantum world, due to the power of quantum parallelism, a linear increase in size corresponds to an exponential increase in processing power. The downside, however, is the accessibility to this parallelism. Remember that superposition collapses the moment we attempt to measure it. The quantum circuit to create the superposition state above, in other words Walsh–Hadamard transform, is shown in Fig. 2.3. Therefore, the Walsh–Hadamard transform on  $n$  ancilla qubits in state  $|00 \dots 0\rangle$  can be implemented by applying the Hadamard operators (gates)  $H$ , whose action is described in Fig. 2.3, on ancillary qubits.

More generally, a linear operator (gate)  $B$  can be expressed in terms of eigenkets  $\{|a^{(m)}\rangle\}$  of a Hermitian operator  $A$ . The operator  $B$  is associated with a *square matrix* (albeit infinite in extent), whose elements are

$$B_{mn} = \langle a^{(m)} | B | a^{(n)} \rangle, \quad (2.67)$$

and can explicitly be written as

$$B \doteq \begin{pmatrix} \langle a^{(1)} | B | a^{(1)} \rangle & \langle a^{(1)} | B | a^{(2)} \rangle & \dots \\ \langle a^{(2)} | B | a^{(1)} \rangle & \langle a^{(2)} | B | a^{(2)} \rangle & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}, \quad (2.68)$$

where we use the notation  $\doteq$  to denote that operator  $B$  is represented by the matrix above. Very important single-qubit gates are Hadamard gate  $H$ , the phase shift gate  $S$ , the  $\pi/8$  (or  $T$ ) gate, controlled-NOT (or CNOT) gate, and Pauli operators  $X$ ,  $Y$ ,  $Z$ . The Hadamard gate  $H$ , phase shift gate,  $T$  gate, and CNOT gate have the following matrix representation in CB  $\{|0\rangle, |1\rangle\}$ :

$$H \doteq \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad S \doteq \begin{bmatrix} 1 & 0 \\ 0 & j \end{bmatrix}, \quad T \doteq \begin{bmatrix} 1 & 0 \\ 0 & e^{j\pi/4} \end{bmatrix}, \quad \text{CNOT} \doteq \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \quad (2.69)$$

The Pauli operators, on the other hand, have the following matrix representation in CB:

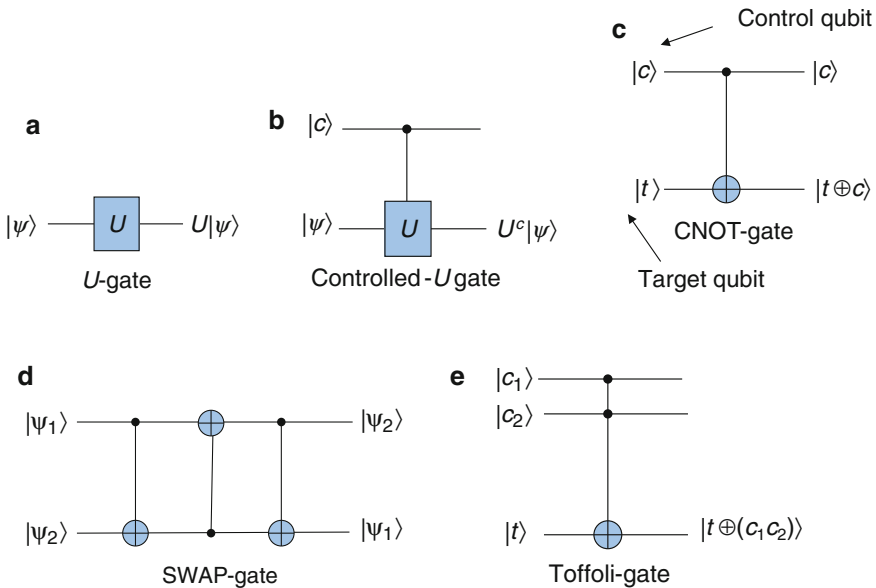
$$X \doteq \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad Y \doteq \begin{bmatrix} 0 & -j \\ j & 0 \end{bmatrix}, \quad Z \doteq \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (2.70)$$

The action of Pauli gates on an arbitrary qubit  $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$  is given as follows:

$$\begin{aligned} X(\alpha|0\rangle + \beta|1\rangle) &= \alpha|1\rangle + \beta|0\rangle, & Y(\alpha|0\rangle + \beta|1\rangle) &= j(\alpha|1\rangle - \beta|0\rangle), \\ Z(\alpha|0\rangle + \beta|1\rangle) &= \alpha|0\rangle - \beta|1\rangle. \end{aligned} \quad (2.71)$$

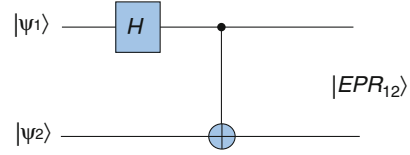
So the action of  $X$  gate is to introduce the bit flip, the action of  $Z$  gate is to introduce the phase flip, and the action of  $Y$  gate is to simultaneously introduce the bit and phase flips.

Several important single-, double-, and triple-qubit gates are shown in Fig. 2.4. The action of a single-qubit gate is to apply the operator  $U$  on qubit  $|\psi\rangle$ , which results in another qubit. The controlled- $U$  gate conditionally applies the operator  $U$  on target qubit  $|\psi\rangle$ , when the control qubit  $|c\rangle$  is in the  $|1\rangle$ -state. One particularly important controlled- $U$  gate is CNOT gate. This gate flips the content of target qubit  $|t\rangle$  when the control qubit  $|c\rangle$  is in  $|1\rangle$ -state. The purpose of SWAP gate is to interchange the positions of two qubits, and it can be implemented by using three CNOT gates as shown in Fig. 2.4d. Finally, the Toffoli gate represents the generalization of CNOT gate, where two control qubits are used.



**Fig. 2.4** Important quantum gates and their action: (a) single-qubit gate, (b) controlled- $U$  gate, (c) CNOT gate, (d) SWAP gate, and (e) Toffoli gate

**Fig. 2.5** Bell state (EPR pairs) preparation circuit



The minimum set of gates that can be used to perform arbitrary quantum computation algorithm is known as the *universal set of gates*. The most popular sets of universal quantum gates are  $\{H, S, \text{CNOT}, \text{Toffoli}\}$  gates,  $\{H, S, \pi/8 \text{ (T), CNOT}\}$  gates, Barenco gate, and Deutsch gate. By using these universal quantum gates, more complicated operations can be performed. As an illustration, in Fig. 2.5, the Bell state [Einstein–Podolsky–Rosen (EPR) pairs] preparation circuit is shown, which is of high importance in quantum teleportation and QKD applications.

So far, single-, double-, and triple-qubit quantum gates have been considered. An arbitrary quantum state of  $K$  qubits has the form  $\sum_s \alpha_s |s\rangle$ , where  $s$  runs over all binary strings of length  $K$ . Therefore, there are  $2^K$  complex coefficients, all independent except for the normalization constraint:

$$\sum_{s=00\dots 00}^{11\dots 11} |\alpha_s|^2 = 1. \quad (2.72)$$

For example, the state  $\alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_{11}|11\rangle$  (with  $|\alpha_{00}|^2 + |\alpha_{01}|^2 + |\alpha_{10}|^2 + |\alpha_{11}|^2 = 1$ ) is the general 2-qubit state (we use  $|00\rangle$  to denote the tensor product  $|0\rangle \otimes |0\rangle$ ). The multiple qubits can be **entangled** so that they cannot be decomposed into two separate states. For example, the Bell state or EPR pair  $(|00\rangle + |11\rangle)/\sqrt{2}$  cannot be written in terms of tensor product  $|\psi_1\rangle |\psi_2\rangle = (\alpha_1|0\rangle + \beta_1|1\rangle) \otimes (\alpha_2|0\rangle + \beta_2|1\rangle) = \alpha_1\alpha_2|00\rangle + \alpha_1\beta_2|01\rangle + \beta_1\alpha_2|10\rangle + \beta_1\beta_2|11\rangle$ , because it would require  $\alpha_1\alpha_2 = \beta_1\beta_2 = 1/\sqrt{2}$ , while  $\alpha_1\beta_2 = \beta_1\alpha_2 = 0$ , which a priori has no reason to be valid. This state can be obtained by using circuit shown in Fig. 2.5, for two-qubit input state  $|00\rangle$ . We will return to the concept of entanglement in Sect. 2.3.3.

The quantum parallelism can now be introduced more formally as follows. The QIP device, denoted as *QIP*, implemented on a quantum register maps the input string  $i_1, \dots, i_N$  to the output string  $O_1(i), \dots, O_N(i)$ :

$$\begin{pmatrix} O_1(i) \\ \vdots \\ O_N(i) \end{pmatrix} = U(\text{QIP}) \begin{pmatrix} i_1 \\ \vdots \\ i_N \end{pmatrix}; \quad (i)_{10} = (i_1, \dots, i_N)_2. \quad (2.73)$$

The CB states are denoted by

$$|i_1, \dots, i_N\rangle = |i_1\rangle \otimes \dots \otimes |i_N\rangle; \quad i_1, \dots, i_N \in \{0, 1\}. \quad (2.74)$$

The linear superposition allows us to form the following  $2N$ -qubit state:

$$|\psi_{\text{in}}\rangle = \left[ \frac{1}{\sqrt{2^N}} \sum_i |i_1, \dots, i_N\rangle \right] \otimes |0 \dots 0\rangle, \quad (2.75)$$

and upon the application of quantum operation  $U(QIP)$ , the output can be represented by

$$|\psi_{\text{out}}\rangle = U(QIP)|\psi_{\text{in}}\rangle = \frac{1}{\sqrt{2^N}} \sum_i |i_1, \dots, i_N\rangle \otimes |O_1(i), \dots, O_N(i)\rangle. \quad (2.76)$$

The QIP circuit (or a quantum computer) has been able to encode all input strings generated by  $QIP$  into  $|\psi_{\text{out}}\rangle$ ; in other words, it has simultaneously pursued  $2^N$  classical paths. This ability of a QIP circuit to encode multiple computational results into a quantum state in a single quantum computational step is known as **quantum parallelism**, as mentioned earlier.

### 2.3.2 No-Cloning Theorem and Distinguishing the Quantum States

Just like in quantum parallelism, the quantum superposition is also the key concept behind our inability to clone arbitrary quantum states. To see this, let us think of a quantum copier that takes as input an arbitrary quantum state and outputs two copies of that state, resulting in a clone of the original state. For example, if the input state is  $|\psi\rangle$ , then the output of the copier is  $|\psi\rangle|\psi\rangle$ . For an arbitrary quantum state, such a copier raises a fundamental contradiction. Consider two arbitrary states  $|\psi\rangle$  and  $|\chi\rangle$  that are inputted to the copier. When they are inputted individually, we expect to get  $|\psi\rangle|\psi\rangle$  and  $|\chi\rangle|\chi\rangle$ . Now consider a superposition of these two states given by

$$|\varphi\rangle = \alpha|\psi\rangle + \beta|\chi\rangle. \quad (2.77)$$

Based on the above description that the quantum copier clones the original state, we expect the output,

$$\begin{aligned} |\varphi\rangle|\varphi\rangle &= (\alpha|\psi\rangle + \beta|\chi\rangle)(\alpha|\psi\rangle + \beta|\chi\rangle) \\ &= \alpha^2|\psi\rangle|\psi\rangle + \alpha\beta|\psi\rangle|\chi\rangle + \alpha\beta|\chi\rangle|\psi\rangle + \beta^2|\chi\rangle|\chi\rangle. \end{aligned} \quad (2.78)$$

On the other hand, the linearity of quantum mechanics, as evidenced by the Schrodinger wave equation, tells us that the quantum copier can be represented by a unitary operator that performs the cloning. If such a unitary operator were to act on

the superposition state  $|\varphi\rangle$ , the output would be a superposition of  $|\psi\rangle|\psi\rangle$  and  $|\chi\rangle|\chi\rangle$ , that is,

$$|\varphi'\rangle = \alpha|\psi\rangle|\psi\rangle + \beta|\chi\rangle|\chi\rangle. \quad (2.79)$$

As is clearly evident, the difference between previous two equations leads to contradiction mentioned above. As a consequence, there is no unitary operator that can clone  $|\varphi\rangle$ . We therefore, formulate the no-cloning theorem as follows.

*No-cloning Theorem.* No quantum copier exists that can clone an arbitrary quantum state.

This result raises a related question: do there exist some specific states for which cloning is possible? The answer to this question is (surprisingly) yes. Remember, a key result of quantum mechanics is that unitary operators preserve probabilities. This implies that inner (dot) products  $\langle\varphi|\varphi\rangle$  and  $\langle\varphi'|\varphi'\rangle$  should be identical. The inner products  $\langle\varphi|\varphi\rangle$  and  $\langle\varphi'|\varphi'\rangle$  are, respectively, given by

$$\begin{aligned} \langle\varphi|\varphi\rangle &= (\langle\psi|\alpha^* + \langle\chi|\beta^*)(\alpha|\psi\rangle + \beta|\chi\rangle) \\ &= |\alpha|^2\langle\psi|\psi\rangle + |\beta|^2\langle\chi|\chi\rangle + \alpha^*\beta\langle\psi|\chi\rangle + \alpha\beta^*\langle\chi|\psi\rangle \\ \langle\varphi'|\varphi'\rangle &= (\langle\psi|\langle\psi|\alpha^* + \langle\chi|\langle\chi|\beta^*)(\alpha|\psi\rangle|\psi\rangle + \beta|\chi\rangle|\chi\rangle) \\ &= |\alpha|^2|\langle\psi|\psi\rangle|^2 + |\beta|^2|\langle\chi|\chi\rangle|^2 + \alpha^*\beta|\langle\psi|\chi\rangle|^2 + \alpha\beta^*|\langle\chi|\psi\rangle|^2. \end{aligned} \quad (2.80)$$

We know that  $\langle\psi|\psi\rangle = \langle\chi|\chi\rangle = 1$ . Therefore, the discrepancy lies in the cross terms. Specifically, to avoid the contradiction that resulted in the no-cloning theorem, we require that  $|\langle\psi|\chi\rangle|^2 = \langle\psi|\chi\rangle$ . This condition can only be satisfied when the states are orthogonal. Thus cloning is possible only for mutually orthogonal states. It is, however, important to remember a subtle point here. Even if we have a mutually orthogonal set of states, we need a quantum copier (or unitary operator) specifically for those states. If the unitary operator is specific to a different set of mutually orthogonal states, cloning would fail. It would seem that the no-cloning theorem would prevent us from exploiting the richness of quantum mechanics. It turns out that this is not the case. A key example is the QKD that with very high probability guarantees secure communication.

Not only that non-orthogonal quantum states cannot be cloned, they also cannot be reliably distinguished. There is no measurement device we can create that can reliably distinguish non-orthogonal states. This fundamental result plays an important role in quantum cryptography. Its proof is based on contradiction. Let us assume that the measurement operator  $M$  is the Hermitian operator (with corresponding eigenvalues  $m_i$  and corresponding projection operators  $P_i$ ) of an observable  $\mathfrak{M}$ , which allows unambiguously to distinguish between two non-orthogonal states  $|\psi_1\rangle$  and  $|\psi_2\rangle$ . The eigenvalue  $m_1$  ( $m_2$ ) unambiguously identifies the state  $|\psi_1\rangle$  ( $|\psi_2\rangle$ ) as the premeasurement state. We know that for projection operators the following properties are valid:

$$\begin{aligned}\langle\psi_1|P_1|\psi_1\rangle &= 1 & \langle\psi_2|P_2|\psi_2\rangle &= 1 \\ \langle\psi_1|P_2|\psi_1\rangle &= 0 & \langle\psi_2|P_1|\psi_2\rangle &= 0.\end{aligned}\tag{2.81}$$

Since  $|\psi_1\rangle$  and  $|\psi_2\rangle$  are non-orthogonal states,  $\langle\psi_1|\psi_2\rangle \neq 0$ , and  $|\psi_2\rangle$  can be represented in terms of  $|\psi_1\rangle$  and another state  $|\chi\rangle$  that is orthogonal to  $|\psi_1\rangle$  ( $\langle\psi_1|\chi\rangle = 0$ ) as follows:

$$|\psi_2\rangle = \alpha|\psi_1\rangle + \beta|\chi\rangle.\tag{2.82}$$

From the projection operator properties, listed above, we can conclude the following:

$$\begin{aligned}0 &= \langle\psi_1|P_2|\psi_1\rangle \stackrel{P_2^2=P_2}{=} \langle\psi_1|P_2P_2|\psi_1\rangle = \|P_2|\psi_1\rangle\|^2 \Rightarrow P_2|\psi_1\rangle = 0 \\ 1 &= \langle\psi_2|P_2|\psi_2\rangle = \langle\psi_2|P_2P_2|\psi_2\rangle = (\alpha^*\langle\psi_1| + \beta^*\langle\chi|)(\alpha P_2|\psi_1\rangle + \beta P_2|\chi\rangle) = |\beta|^2 \langle\chi|P_2|\chi\rangle.\end{aligned}\tag{2.83}$$

Now we use the completeness relationship

$$1 = \langle\chi|\chi\rangle = \langle\chi|\overbrace{\sum_i^I P_i}^I|\chi\rangle = \sum_i \langle\chi|P_i|\chi\rangle \stackrel{\langle\chi|P_i|\chi\rangle \geq 0}{\geq} \langle\chi|P_2|\chi\rangle.\tag{2.84}$$

By combining the previous two equations, we obtain

$$1 = \langle\psi_2|P_2|\psi_2\rangle \leq |\beta|^2 \Rightarrow |\beta|^2 = 1,\tag{2.85}$$

indicating that the probability of finding of  $|\psi_2\rangle$  in  $|\chi\rangle$  is 1. Therefore, we conclude that  $|\psi_2\rangle = |\chi\rangle$ , which is a contradiction. Therefore, indeed, *it is impossible to unambiguously distinguish non-orthogonal quantum states*.

### 2.3.3 Quantum Entanglement

Let  $|\psi_0\rangle, \dots, |\psi_{n-1}\rangle$  be  $n$  qubits lying in the Hilbert spaces  $H_0, \dots, H_{n-1}$ , respectively, and let the state of the joint quantum system lying in  $H_0 \otimes \dots \otimes H_{n-1}$  be denoted by  $|\psi\rangle$ . The qubit  $|\psi\rangle$  is then said to be entangled if it cannot be written in the product state form

$$|\psi\rangle = |\psi_0\rangle \otimes |\psi_1\rangle \otimes \dots \otimes |\psi_{n-1}\rangle.\tag{2.86}$$

Important examples of two-qubit states are *Bell states*, also known as *EPR* states (pairs):



$$\begin{aligned}
|B_{00}\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), & |B_{01}\rangle &= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \\
|B_{10}\rangle &= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle), & |B_{11}\rangle &= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle).
\end{aligned} \tag{2.87}$$

The  $n$ -qubit ( $n > 2$ ) analogs of Bell states will be now briefly reviewed. One popular family of entangled multiqubit states is Greenberger–Horne–Zeilinger (GHZ) states:

$$|\text{GHZ}\rangle = \frac{1}{\sqrt{2}}(|00 \cdots 0\rangle \pm |11 \cdots 1\rangle). \tag{2.88}$$

Another popular family of multiqubit entangled states is known as  $W$  states:

$$|W\rangle = \frac{1}{\sqrt{N}}(|00 \cdots 01\rangle + |00 \cdots 10\rangle + \cdots + |01 \cdots 00\rangle + |10 \cdots 00\rangle). \tag{2.89}$$

The  $W$  state of  $n$  qubits represents a superposition of single-weighted CB states, each occurring with probability amplitude of  $N^{-1/2}$ .

For a bipartite system, we can elegantly verify whether or not the qubit  $|\psi\rangle$  is a product state or an entangled one, by Schmidt decomposition [1, 7]. The *Schmidt decomposition theorem* states that a pure state  $|\psi\rangle$  of the composite system  $H_A \otimes H_B$  can be represented as

$$|\psi\rangle = \sum_i c_i |i_A\rangle |i_B\rangle, \tag{2.90}$$

where  $|i_A\rangle$  and  $|i_B\rangle$  are orthonormal basis of the subsystems  $H_A$  and  $H_B$ , respectively, and  $c_i \in \Re^+$  ( $\Re^+$  is the set of nonnegative real numbers) are *Schmidt coefficients* that satisfy the following condition  $\sum_i c_i^2 = 1$ . For the proof of the theorem, please refer to [1]. The Schmidt coefficients can be calculated from the partial density matrix  $\text{Trace}_B(|\psi\rangle\langle\psi|)$ . A corollary of the Schmidt decomposition theorem is that a pure state in a composite system is a product state if and only if the Schmidt rank is 1 and is an entangled state if and only if the Schmidt rank is greater than one.

As an illustration, let us verify if the Bell state  $|B_{11}\rangle$  is an entangled one. We first determine the density matrix:

$$\begin{aligned}
\rho &= |\psi\rangle\langle\psi| = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \frac{1}{\sqrt{2}}(\langle 01| - \langle 10|) \\
&= \frac{1}{2}(|01\rangle\langle 01| - |01\rangle\langle 10| - |10\rangle\langle 01| + |10\rangle\langle 10|).
\end{aligned}$$

By tracing out the subsystem  $B$ , we obtain

$$\begin{aligned}\rho_A &= \text{Trace}_B(|\psi\rangle\langle\psi|) = \langle 0_B|\psi\rangle\langle\psi|0_B\rangle + \langle 1_B|\psi\rangle\langle\psi|1_B\rangle \\ &= \frac{1}{2}(|1\rangle\langle 1| + |0\rangle\langle 0|) = \frac{1}{2}I.\end{aligned}$$

The eigenvalues are  $c_1 = c_2 = 1/2$ , and the Schmidt rank is 2 indicating that the Bell state  $|B_{11}\rangle$  is an entangled state.

### 2.3.4 Operator Sum Representation

Let the composite system  $C$  be composed of quantum register  $Q$  and environment  $E$ . This kind of system can be modeled as a closed quantum system. Because the composite system is closed, its dynamic is unitary, and final state is specified by a unitary operator  $U$  as follows:  $U(\rho \otimes \varepsilon_0)U^\dagger$ , where  $\rho$  is a density operator of initial state of quantum register  $Q$  and  $\varepsilon_0$  is the initial density operator of the environment  $E$ . The reduced density operator of  $Q$  upon interaction  $\rho_f$  can be obtained by tracing out the environment:

$$\rho_f = \text{Tr}_E[U(\rho \otimes \varepsilon_0)U^\dagger] \equiv \xi(\rho). \quad (2.91)$$

The transformation (mapping) of initial density operator  $\rho$  to the final density operator  $\rho_f$ , denoted as  $\xi : \rho \rightarrow \rho_f$ , given by (2.91), is often called the *superoperator* or *quantum operation*. The final density operator can be expressed in the so-called operator sum representation as follows:

$$\rho_f = \sum_k E_k \rho E_k^\dagger, \quad (2.92)$$

where  $E_k$  are the operation elements for the superoperator. Clearly, in the absence of environment, the superoperator becomes  $U\rho U^\dagger$ , which is nothing else but a conventional time-evolution quantum operation.

The operator sum representation can be used in the classification of quantum operations into two categories: (1) *trace preserving* when  $\text{Tr} \xi(\rho) = \text{Tr} \rho = 1$  and (2) *non-trace preserving* when  $\text{Tr} \xi(\rho) < 1$ . Starting from the trace-preserving condition

$$\text{Tr} \rho = \text{Tr} \xi(\rho) = \text{Tr} \left[ \sum_k E_k \rho E_k^\dagger \right] = \text{Tr} \left[ \rho \sum_k E_k E_k^\dagger \right] = 1, \quad (2.93)$$

we obtain

$$\sum_k E_k E_k^\dagger = I. \quad (2.94)$$

For non-trace-preserving quantum operation, (2.93) is not satisfied, and informally we can write  $\sum_k E_k E_k^\dagger < I$ .

If the *environment dimensionality is large enough*, it can be found in pure state,  $\varepsilon_0 = |\phi_0\rangle\langle\phi_0|$ , and the corresponding superoperator becomes

$$\begin{aligned} \xi(\rho) &= \text{Tr}_E [U(\rho \otimes \varepsilon)U^\dagger] = \sum_k \langle\phi_k| (U\rho \otimes \varepsilon U^\dagger) |\phi_k\rangle \\ &= \sum_k \langle\phi_k| \left( U\rho \otimes \left( \underbrace{|\phi_0\rangle\langle\phi_0|}_{\varepsilon} \right) U^\dagger \right) |\phi_k\rangle \\ &= \sum_k \underbrace{\langle\phi_k|U|\phi_0\rangle}_{E_k} \rho \underbrace{\langle\phi_0|U^\dagger|\phi_k\rangle}_{E_k^\dagger} \\ &= \sum_k E_k \rho E_k^\dagger, \quad E_k = \langle\phi_k|U|\phi_0\rangle. \end{aligned} \quad (2.95)$$

The  $E_k$  operators in operator sum representation are known as *Kraus operators*.

As an illustration, let us consider bit-flip and phase-flip channels. Let the composite system be given by  $|\phi_E\rangle| \psi_Q \rangle$ , wherein the initial state of environment is  $|\phi_E\rangle = |0_E\rangle$ . Let further the quantum subsystem  $Q$  interacts to the environment  $E$  by the Pauli- $X$  operator:

$$U = \sqrt{1-p}I \otimes I + \sqrt{p}X \otimes X, \quad 0 \leq p \leq 1. \quad (2.96)$$

Therefore, with probability  $1-p$ , we leave the quantum system untouched, while with probability  $p$  we apply Pauli- $X$  operator to both quantum subsystem and the environment. By applying the operator  $U$  on environment state, we obtain

$$U|\phi_E\rangle = \sqrt{1-p}I \otimes I|0_E\rangle + \sqrt{p}X \otimes X|0_E\rangle = \sqrt{1-p}|0_E\rangle I + \sqrt{p}|1_E\rangle X. \quad (2.97)$$

The corresponding Kraus operators are given by

$$E_0 = \langle 0_E|U|\phi_E\rangle = \sqrt{1-p}I, \quad E_1 = \langle 1_E|U|\phi_E\rangle = \sqrt{p}X. \quad (2.98)$$

Finally, the operator sum representation is given by

$$\xi(\rho) = E_0 \rho E_0^\dagger + E_1 \rho E_1^\dagger = (1-p)\rho + pX\rho X. \quad (2.99)$$

In similar fashion, the Kraus operators for the phase-flip channel are given by

$$E_0 = \langle 0_E|U|\phi_E\rangle = \sqrt{1-p}I, \quad E_1 = \langle 1_E|U|\phi_E\rangle = \sqrt{p}Z, \quad (2.100)$$

and the corresponding operator sum representation is

$$\xi(\rho) = E_0 \rho E_0^\dagger + E_1 \rho E_1^\dagger = (1 - p)\rho + pZ\rho Z. \quad (2.101)$$

### 2.3.5 *Decoherence Effects, Depolarization, and Amplitude Damping Channel Models*

Quantum computation works by manipulating quantum interference effect. The quantum interference, a manifestation of coherent superposition of quantum states, is the cornerstone behind all quantum information tasks such as quantum computation and quantum communication. A major source of problem is our inability to prevent our quantum system of interest from interacting with the surrounding environment. This interaction results in an entanglement between the quantum system and the environment leading to decoherence. To understand this system–environment entanglement and decoherence better, let us consider a qubit described by density state (matrix)  $\rho = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$  interacting with the environment, described by the following three states:  $|0_E\rangle$ ,  $|1_E\rangle$ , and  $|2_E\rangle$ . Without loss of generality, we assume that environment was initially in state  $|0_E\rangle$ . The unitary operator introducing the entanglement between the quantum system and the environment is defined as

$$\begin{aligned} U|0\rangle|0_E\rangle &= \sqrt{1-p}|0\rangle|0_E\rangle + \sqrt{p}|0\rangle|1_E\rangle \\ U|0\rangle|0_E\rangle &= \sqrt{1-p}|1\rangle|0_E\rangle + \sqrt{p}|1\rangle|2_E\rangle. \end{aligned} \quad (2.102)$$

The corresponding Kraus operators are given by

$$E_0 = \sqrt{1-p}I, \quad E_1 = \sqrt{p}|0\rangle\langle 0|, \quad E_2 = \sqrt{p}|1\rangle\langle 1|. \quad (2.103)$$

The operator sum representation is given by

$$\xi(\rho) = E_0 \rho E_0^\dagger + E_1 \rho E_1^\dagger + E_2 \rho E_2^\dagger = \begin{bmatrix} a & (1-p)b \\ (1-p)c & d \end{bmatrix}. \quad (2.104)$$

By applying these quantum operation  $n$ -times, the corresponding final state would be

$$\rho_f = \begin{bmatrix} a & (1-p)^n b \\ (1-p)^n c & d \end{bmatrix}. \quad (2.105)$$

If the probability  $p$  is expressed as  $p = \gamma\Delta t$ , we can write  $n = t/\Delta t$  and in limit we obtain

$$\lim_{\Delta t \rightarrow 0} (1 - p)^n = (1 - \gamma\Delta t)^{t/\Delta t} = e^{-\gamma t}. \quad (2.106)$$

Therefore, the corresponding operator sum representation as  $n$  tends to plus infinity is given by

$$\xi(\rho) = \begin{bmatrix} a & e^{-\gamma t} b \\ e^{-\gamma t} c & d \end{bmatrix}. \quad (2.107)$$

Clearly, the terms  $b$  and  $c$  go to zero as  $t$  increases, indicating that the relative phase in the original state of the quantum system is lost, and the corresponding channel model is known as the *phase damping* channel model.

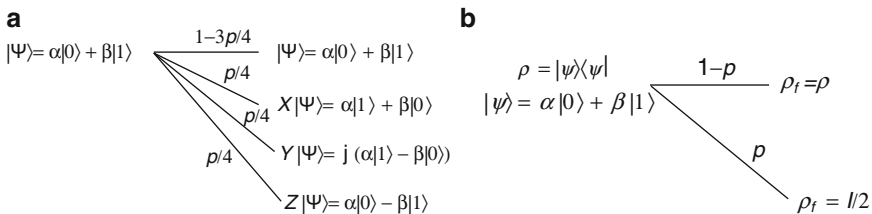
In the above example, we have considered the coupling between a single-qubit quantum system and the environment and discussed the resulting loss of interference or coherent superposition. In general, for multiple qubit systems, decoherence also results in loss of coupling between the qubits. In fact, with increasing complexity and size of the quantum computer, the decoherence effect becomes worse. Additionally, the quantum system lies in some complex Hilbert space where there are infinite variations of errors that can cause decoherence.

A more general example of dephasing is the depolarization. The *depolarizing channel*, as shown in Fig. 2.6, with probability  $1 - p$  leaves the qubit as it is, while with probability  $p$  moves the initial state into  $\rho_f = I/2$  that maximizes the von Neumann entropy  $S(\rho) = -\text{Tr } \rho \log \rho = 1$ . The properties describing the model can be summarized as follows:

1. Qubit errors are independent.
2. Single-qubit errors ( $X, Y, Z$ ) are equally likely.
3. All qubits have the same single-error probability  $p/4$ .

The Kraus operators  $E_i$  of the channel should be selected as follows:

$$E_0 = \sqrt{1 - 3p/4}I; \quad E_1 = \sqrt{p/4}X; \quad E_2 = \sqrt{p/4}Y; \quad E_3 = \sqrt{p/4}Z. \quad (2.108)$$



**Fig. 2.6** Depolarizing channel model: (a) Pauli operator description and (b) density operator description

The action of depolarizing channel is to perform the following mapping:  $\rho \rightarrow \xi(\rho) = \sum_i E_i \rho E_i^\dagger$ , where  $\rho$  is the initial density operator. Without loss of generality, we will assume that initial state was pure  $|\psi\rangle = a|0\rangle + b|1\rangle$  so that

$$\begin{aligned} \rho &= |\psi\rangle\langle\psi| = (a|0\rangle + b|1\rangle)(\langle 0|a^* + \langle 1|b^*) \\ &= |a|^2 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + ab^* \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + a^*b \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + |b|^2 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} |a|^2 & ab^* \\ a^*b & |b|^2 \end{bmatrix}. \end{aligned} \quad (2.109)$$

The resulting quantum operation can be represented using operator sum representation as follows:

$$\begin{aligned} \xi(\rho) &= \sum_i E_i \rho E_i^\dagger = \left(1 - \frac{3p}{4}\right)\rho + \frac{p}{4}(X\rho X + Y\rho Y + Z\rho Z) \\ &= \left(1 - \frac{3p}{4}\right) \begin{bmatrix} |a|^2 & ab^* \\ a^*b & |b|^2 \end{bmatrix} + \frac{p}{4} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} |a|^2 & ab^* \\ a^*b & |b|^2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \\ &\quad + \frac{p}{4} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} |a|^2 & ab^* \\ a^*b & |b|^2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + \frac{p}{4} \begin{bmatrix} 0 & -j \\ j & 0 \end{bmatrix} \begin{bmatrix} |a|^2 & ab^* \\ a^*b & |b|^2 \end{bmatrix} \begin{bmatrix} 0 & -j \\ j & 0 \end{bmatrix} \\ &= \left(1 - \frac{3p}{4}\right) \begin{bmatrix} |a|^2 & ab^* \\ a^*b & |b|^2 \end{bmatrix} + \frac{p}{4} \begin{bmatrix} |b|^2 & a^*b \\ ab^* & |a|^2 \end{bmatrix} + \frac{p}{4} \begin{bmatrix} |a|^2 & -ab^* \\ -a^*b & |b|^2 \end{bmatrix} + \frac{p}{4} \begin{bmatrix} |b|^2 & -a^*b \\ -ab^* & |a|^2 \end{bmatrix} \\ &= \begin{bmatrix} \left(1 - \frac{p}{2}\right)|a|^2 + \frac{p}{2}|b|^2 & (1-p)ab^* \\ (1-p)a^*b & \frac{p}{2}|a|^2 + \left(1 - \frac{p}{2}\right)|b|^2 \end{bmatrix} \\ &= \begin{bmatrix} (1-p)|a|^2 + \frac{p}{2}(|a|^2 + |b|^2) & (1-p)ab^* \\ (1-p)a^*b & \frac{p}{2}(|a|^2 + |b|^2) + (1-p)|b|^2 \end{bmatrix}. \end{aligned} \quad (2.110)$$

Since  $|a|^2 + |b|^2 = 1$ , the operator sum representation can be written as

$$\xi(\rho) = \sum_i E_i \rho E_i^\dagger = (1-p) \begin{bmatrix} |a|^2 & ab^* \\ a^*b & |b|^2 \end{bmatrix} + \frac{p}{2}I = (1-p)\rho + \frac{p}{2}I. \quad (2.111)$$

The first line in (2.110) corresponds to the model shown in Fig. 2.6a, and the last line (see also (2.111)) corresponds to the model shown in Fig. 2.6b. It is clear from (2.109) and (2.111) that  $\text{Tr } \xi(\rho) = \text{Tr } \rho = 1$  meaning that the superoperator is trace preserving. Notice that the depolarizing channel model in some other books/papers can slightly be differently defined.

In the rest of this subsection, we describe the amplitude damping channel model. In certain quantum channels, the errors  $X$ ,  $Y$ , and  $Z$  do not occur with the same probability. In amplitude damping channel, the operation elements are given by

$$E_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\epsilon^2} \end{pmatrix}, \quad E_1 = \begin{pmatrix} 0 & \epsilon \\ 0 & 0 \end{pmatrix}. \quad (2.112)$$

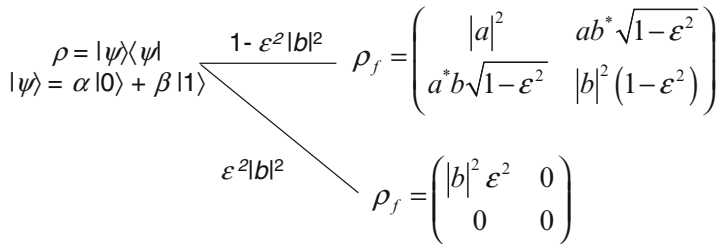
The *spontaneous emission* is an example of a physical process that can be modeled using the amplitude damping channel model. If  $|\psi\rangle = a|0\rangle + b|1\rangle$  is the initial qubit state  $\left(\rho = \begin{bmatrix} |a|^2 & ab^* \\ a^*b & |b|^2 \end{bmatrix}\right)$ , the effect of amplitude damping channel is to perform the following mapping:

$$\begin{aligned} \rho \rightarrow \xi(\rho) &= E_0 \rho E_0^\dagger + E_1 \rho E_1^\dagger = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\epsilon^2} \end{pmatrix} \begin{bmatrix} |a|^2 & ab^* \\ a^*b & |b|^2 \end{bmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\epsilon^2} \end{pmatrix} \\ &\quad + \begin{pmatrix} 0 & \epsilon \\ 0 & 0 \end{pmatrix} \begin{bmatrix} |a|^2 & ab^* \\ a^*b & |b|^2 \end{bmatrix} \begin{pmatrix} 0 & 0 \\ \epsilon & 0 \end{pmatrix} \\ &= \begin{pmatrix} |a|^2 & ab^* \sqrt{1-\epsilon^2} \\ a^*b \sqrt{1-\epsilon^2} & |b|^2 (1-\epsilon^2) \end{pmatrix} + \begin{pmatrix} |b|^2 \epsilon^2 & 0 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} |a|^2 + \epsilon^2 |b|^2 & ab^* \sqrt{1-\epsilon^2} \\ a^*b \sqrt{1-\epsilon^2} & |b|^2 (1-\epsilon^2) \end{pmatrix}. \end{aligned} \quad (2.113)$$

Probabilities  $P(0)$  and  $P(1)$  that  $E_0$  and  $E_1$  occur are given by

$$\begin{aligned} P(0) &= \text{Tr}(E_0 \rho E_0^\dagger) = \text{Tr} \begin{pmatrix} |a|^2 & ab^* \sqrt{1-\epsilon^2} \\ a^*b \sqrt{1-\epsilon^2} & |b|^2 (1-\epsilon^2) \end{pmatrix} = 1 - \epsilon^2 |b|^2 \\ P(1) &= \text{Tr}(E_1 \rho E_1^\dagger) = \text{Tr} \begin{pmatrix} |b|^2 \epsilon^2 & 0 \\ 0 & 0 \end{pmatrix} = \epsilon^2 |b|^2. \end{aligned} \quad (2.114)$$

The corresponding amplitude damping channel model is shown in Fig. 2.7.



**Fig. 2.7** Amplitude damping channel model

## 2.4 Classical (Shannon) and Quantum (von Neumann) Entropies

Let us observe a classical discrete memoryless source with the alphabet  $X = \{x_1, x_2, \dots, x_N\}$ . The symbols from the alphabet are emitted by the source with probabilities  $P(X = x_n) = p_n$ ,  $n = 1, 2, \dots, N$ . The amount of information carried by the  $k$ th symbol is related to the uncertainty that is resolved when this symbol occurs, and it is defined as  $I(x_n) = \log(1/p_n) = -\log(p_n)$ , where the logarithm is to the base 2. The classical (Shannon) entropy is defined as the measure of the average amount of information per source symbol [27, 28]:

$$H(X) = E[I(x_n)] = \sum_{n=1}^N p_n I(x_n) = \sum_{n=1}^N p_n \log_2 \left( \frac{1}{p_n} \right). \quad (2.115)$$

The Shannon entropy satisfies the following inequalities:

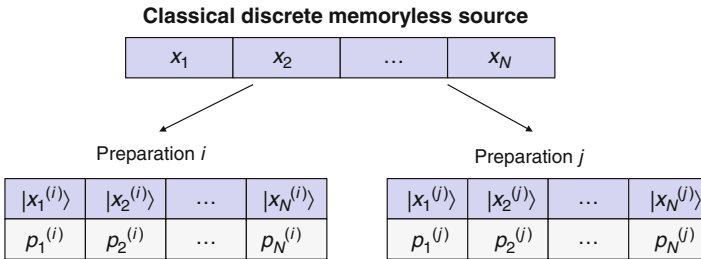
$$0 \leq H(X) \leq \log N = \log |X|. \quad (2.116)$$

Let  $Q$  be a quantum system with the state described by the density operator  $\rho_x$ . The probability that the output  $\rho_x$  is obtained is given by  $p_x = P(x)$ . The quantum source is, therefore, described by the ensemble  $\{\rho_x, p_x\}$ , characterized by the mixed density operator  $\rho = \sum_{x \in X} p_x \rho_x$ . Then the *quantum (von Neumann) entropy* is defined as

$$S(\rho) = -\text{Tr}(\rho \log \rho) = -\sum_{\lambda_i} \lambda_i \log \lambda_i, \quad (2.117)$$

where  $\lambda_i$  are eigenvalues of  $\rho$ . When all quantum states are pure and mutually orthogonal, then von Neumann entropy equals the Shannon entropy as in that case  $p_i = \lambda_i$ . As an illustration, in Fig. 2.8, we provide an interpretation of quantum representation of the classical information.

Two different preparations  $P^{(i)}$  and  $P^{(j)}$ , with  $H(X^{(i)}) \neq H(X^{(j)}) (i \neq j)$ , can generate the same  $\rho$  and hence have the same von Neumann entropy  $S(\rho)$  because the states of the two preparations may not be physically distinguishable from each other.



**Fig. 2.8** The illustration of quantum representation of the classical information



## 2.5 Holevo Information, Accessible Information, and Holevo Bound

A classical discrete memoryless channel (DMC) is described by the set of transition probabilities

$$p(y_k|x_j) = P(Y = y_k|X = x_j); \quad 0 \leq p(y_k|x_j) \leq 1 \quad (2.118)$$

satisfying the condition  $\sum_k p(y_k|x_j) = 1$ . The conditional entropy of  $X$ , given  $Y = y_k$ , is related to the uncertainty of the channel input  $X$  by observing the channel output  $y_k$ , and it is defined as

$$H(X|Y = y_k) = \sum_j p(x_j|y_k) \log \left[ \frac{1}{p(x_j|y_k)} \right], \quad (2.119)$$

where

$$p(x_j|y_k) = p(y_k|x_j) p(x_j) / p(y_k), \quad p(y_k) = \sum_j p(y_k|x_j) p(x_j). \quad (2.120)$$

The amount of uncertainty remaining about the channel input  $X$  after the observing channel output  $Y$  can be then determined by

$$\begin{aligned} H(X|Y) &= \sum_k H(X|Y = y_k) p(y_k) \\ &= \sum_k p(y_k) \sum_j p(x_j|y_k) \log \left[ \frac{1}{p(x_j|y_k)} \right]. \end{aligned} \quad (2.121)$$

Therefore, the amount of uncertainty about the channel input  $X$  that is resolved by observing the channel input would be the difference  $H(X) - H(X|Y)$ , and this difference is known as the mutual information  $I(X, Y)$  (also known as transinformation) [29, 30]. In other words,

$$I(X, Y) = H(X) - H(X|Y) = \sum_j p(x_j) \sum_k p(y_k|x_j) \log \left[ \frac{p(y_k|x_j)}{p(y_k)} \right]. \quad (2.122)$$

The mutual information can also be defined between any two random variables  $X$  and  $Y$ . In this case, the mutual information is related to the amount information that  $Y$  has about  $X$  in average, namely,  $I(X;Y) = H(X) - H(X|Y)$ . In other words, it represents the amount of uncertainty about  $X$  that is resolved given that we know  $Y$ .

In the quantum information theory, we can provide a similar interpretation of Holevo information. The *Holevo information*  $\chi$  for the ensemble of states  $\{\rho_x, p_x\}$

corresponds to the average reduction in quantum entropy given that we know how  $\rho$  was prepared, namely,  $\rho = \sum_x p_x \rho_x$ , and it is defined as

$$\chi = S(\rho) - \sum_x p_x S(\rho_x). \quad (2.123)$$

Holevo information is upper bounded by Shannon entropy, namely,  $\chi \leq H(X)$ .

Maximum of the mutual information over all generalized POVM measurement schemes  $M_y$ , denoted as  $H(X:Y)$ , is known as *accessible information*, and it is officially defined as

$$H(X : Y) = \max_{M_y} I(X; Y). \quad (2.124)$$

If the quantum states are pure and mutually orthogonal, instead of POVM we consider projective measurements such that  $p(y|x) = \text{Tr}(M_y \rho_x) = \text{Tr}(P_y \rho_x) = 1$ , iff  $x = y$  and zero otherwise. In that case,  $H(X:Y) = H(X)$  and Bob (receiver) is able accurately to estimate the information sent by Alice (transmitter). If the states are non-orthogonal, the accessible information is bounded by

$$H(X : Y) \leq S(\rho) \leq H(X). \quad (2.125)$$

When  $\rho_x$  states are mixed states, the accessible information is bounded by the Holevo information

$$H(X : Y) \leq \chi. \quad (2.126)$$

Since the Holevo information is upper bounded by the Shannon entropy, we can write

$$H(X : Y) \leq \chi \leq H(X). \quad (2.127)$$

Therefore, it is impossible for Bob to completely recover the classical information, characterized by  $H(X)$ , which Alice has sent him over the quantum channel!

## 2.6 Schumacher's Noiseless Quantum Coding Theorem and Holevo–Schumacher–Westmoreland Theorem

### 2.6.1 Schumacher's Noiseless Quantum Coding Theorem and Quantum Compression

Consider a classical source  $X$  that generates symbols 0 and 1 with probabilities  $p$  and  $1 - p$ , respectively. The probability of output sequence  $x_1, \dots, x_N$  is given by

$$p(x_1, \dots, x_N) = p \sum_i x_i (1-p) \sum_i (1-x_i) \xrightarrow{N \rightarrow \infty} p^{Np} (1-p)^{N(1-p)}. \quad (2.128)$$

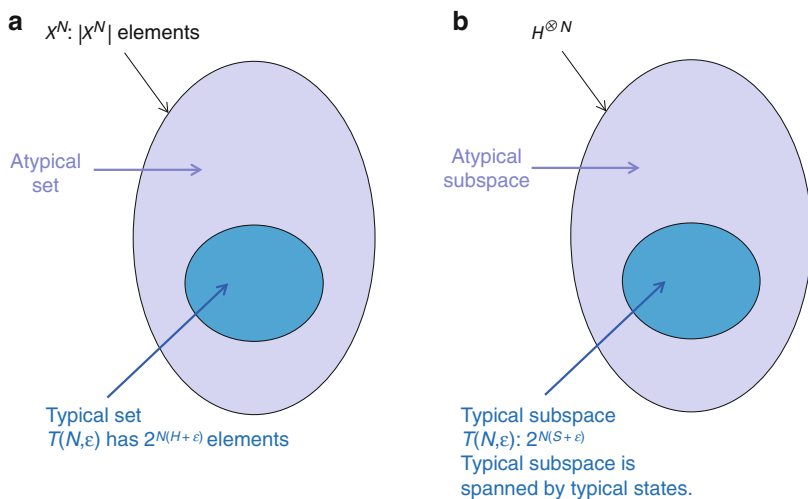
By taking the logarithm of this probability, we obtain

$$\log p(x_1, \dots, x_N) \approx Np \log p + N(1-p) \log(1-p) = -NH(X). \quad (2.129)$$

Therefore, the probability of occurrence of the so-called typical sequence is  $p(x_1, \dots, x_N) \approx 2^{-NH(X)}$ , and in average,  $NH(X)$  bits are needed to represent any typical sequence, which is illustrated in Fig. 2.9a. We denoted the typical set by  $T(N, \epsilon)$ . Clearly, not much of the information will be lost if we consider  $2^{NH(X)}$  typical sequences, instead of  $2^N$  possible sequences, in particular for large  $N$ . This observation can be used in data compression. Namely, with  $NH(X)$  bits, we can enumerate all typical sequences. If the source output is a typical sequence, it will be represented with  $NH(X)$  bits. On the other hand, when atypical sequence gets generated by the source, we will assign a fixed index to it resulting in compression loss. However, as  $N \rightarrow \infty$ , the probability of occurrence of atypical sequence will tend to zero resulting in arbitrary small information loss.

Let us now extend this concept to nonbinary sources. The law of large numbers [27] applied to a nonbinary source with i.i.d. outputs  $X_1, \dots, X_N$  claims that as  $N \rightarrow \infty$ , the expected value of the source approaches the true value  $E(X)$  in probability sense. In other words,

$$P\left(\left|\frac{1}{N} \sum_i X_i - E(X)\right| \leq \epsilon\right) > 1 - \delta \quad (2.130)$$



**Fig. 2.9** The illustration of typical set (a) and typical subspace (b)

for sufficiently large  $N$  and  $\varepsilon$ ,  $\delta > 0$ . Based on the discussion for the typical set, we can write

$$-N[H(X) + \varepsilon] \leq \log P(x_1, \dots, x_N) \leq -N[H(X) - \varepsilon]. \quad (2.131)$$

Since  $P[T(N, \varepsilon)] > 1 - \delta$ , the size of typical set, denoted as  $|T(N, \varepsilon)|$ , will be bounded by

$$(1 - \varepsilon)2^{N[H(X) - \varepsilon]} \leq |T(N, \varepsilon)| \leq 2^{N[H(X) + \varepsilon]}. \quad (2.132)$$

Suppose now that a compression rate  $R$  is larger than  $H(X)$ , say  $R > H(X) + \varepsilon$ . Then based on the discussion above, the total number of typical sequences is bounded by

$$|T(N, \varepsilon)| \leq 2^{N[H(X) + \varepsilon]} \leq 2^{NR}, \quad (2.133)$$

and the probability of occurrence of typical sequence is lower bounded by  $1 - \delta$ , where  $\delta$  is arbitrary small.

The source encoding strategy can be described as follows. Let us divide the set of all sequences generated by the source into typical and atypical sets as shown in Fig. 2.9a. From the equation above, it is clear that sequences in the typical set can be represented by at most  $NR$  bits. If an atypical sequence occurs, we assign to it a fixed index. As  $N$  tends to infinity, the probability for this event to occur tends to zero, indicating that we can encode and decode typical sequences reliably. On the other hand, when  $R < H(X)$ , we can encode maximum  $2^{NR}$  typical sequences, labeled as  $T_R(N, \varepsilon)$ . Since the probability of occurrence of typical sequence is upper bounded by  $2^{-N[H(X) - \varepsilon]}$ , the typical sequences in  $T_R(N, \varepsilon)$  will occur with probability  $2^{N[R - H(X) + \varepsilon]}$ , which tends to zero as  $N$  tends to infinity, indicating that the reliable compression scheme does not exist in this case. With this, we have just proved the Shannon's source coding theorem, which can be formulated as follows.

**Shannon's Source Coding Theorem.** For an i.i.d. source  $X$ , the reliable compression method exists for  $R > H(X)$ . If, on the other hand,  $R < H(X)$ , then no reliable compression scheme exists.

Consider now a quantum source emitting a pure state  $|\psi_x\rangle$  with probability  $p_x$ , described in terms of the mixed density operator

$$\rho = \sum_x p_x |\psi_x\rangle \langle \psi_x|. \quad (2.134)$$

The quantum message comprises  $N$  quantum source outputs, independent of each other, so that

$$\rho_{\otimes N} = \rho \otimes \cdots \otimes \rho. \quad (2.135)$$

The mixed density operator can be expressed in terms of eigenkets as

$$\rho = \sum_{\alpha} \lambda_{\alpha} |\lambda_{\alpha}\rangle \langle \lambda_{\alpha}|. \quad (2.136)$$

The *typical state* is a state  $|\lambda_1\rangle \dots |\lambda_N\rangle$  for which  $\lambda_1 \dots \lambda_N$  is a typical sequence satisfying the following inequality for sufficiently large  $N$ :

$$P\left(\left|\frac{1}{N}\log\left(\frac{1}{P(\lambda_1)P(\lambda_2)\dots P(\lambda_N)}\right) - S(\rho)\right| \leq \varepsilon\right) > 1 - \delta. \quad (2.137)$$

The projector on a typical subspace is given by

$$P_T = \sum_{\lambda} |\lambda_1\rangle \langle \lambda_1| \otimes \dots \otimes |\lambda_N\rangle \langle \lambda_N|. \quad (2.138)$$

The *projection* of the quantum message on the typical subspace is determined by

$$P_T \rho_{\otimes N}. \quad (2.139)$$

The probability of the projections is given by its trace. By using the projection operator, we can separate the total Hilbert space into typical and atypical subspaces, as illustrated in Fig. 2.9b. The probability that the state of the quantum message lies in the typical subspace is given by

$$P(P_T \rho_{\otimes N}) > 1 - \delta. \quad (2.140)$$

The bounds for the typical subspace  $T_{\lambda}(N, \varepsilon)$ , following a similar methodology as for typical sequences, can be determined as

$$(1 - \varepsilon)2^{N[S(\rho) - \varepsilon]} \leq |T_{\lambda}(N, \varepsilon)| \leq 2^{N[S(\rho) + \varepsilon]}. \quad (2.141)$$

Now we are in a position to formulate the corresponding *compression procedure*. Define the projector on the typical subspace  $P_T$  and its complement projecting on orthogonal subspace  $P_T^{\perp} = I - P_T$  ( $I$  is the identity operator) with corresponding outcomes 0 and 1. For compression purposes, we perform the measurements using two orthogonal operators with outputs denoted as 1 and 0, respectively. If the outcome is 1, we know that the message is in a typical state, and we do nothing further. If, on the other hand, the outcome is 0, we know that it belongs to atypical subspace and numerate it as a *fixed* state from the typical subspace. Given that the probability of this to happen can be made as small as possible for large  $N$ , we can compress the quantum message without the loss of information. We shall now formulate the Schumacher's source coding theorem, the quantum information theory equivalent to the Shannon's source coding theorem. For derivation, an interested reader is referred to [1].

**Schumacher's Source Coding Theorem.** Let  $\{|\psi_x\rangle, p_x\}$  be an i.i.d. quantum source. If  $R > S(\rho)$ , then there exists a reliable compression scheme of rate  $R$  for this source. Otherwise, if  $R < S(\rho)$ , then no reliable compression scheme of rate  $R$  exists.

In the formulation of the Schumacher's source coding theorem, we used the concept of reliable compression, without formally introducing it. We say that the compression is reliable if the corresponding *entanglement fidelity* tends to 1 for large  $N$ :

$$F(\rho_{\otimes N}; D^N \circ C^N) \xrightarrow{N \rightarrow \infty} 1; \quad \rho_{\otimes N} = |\psi_{\otimes N}\rangle\langle\psi_{\otimes N}|, \quad \langle\psi_{\otimes N}| \\ = |\psi_{x_1}\rangle \otimes \cdots \otimes |\psi_{x_N}\rangle \quad (2.142)$$

where we used  $D^N$  and  $C^N$  to denote the decompression and compression operations, respectively, defined as the following mappings:

$$D^N : H_c^N \rightarrow H^N, \quad C^N : H^N \rightarrow H_c^N, \quad (2.143)$$

where  $H_c^N$  is  $2^{NR}$ -dimensional subspace of  $H^N$ . The entanglement fidelity  $F$  represents the measure of the preservation of entanglement before and after performing the trace-preserving quantum operation. There exist different definitions of fidelity. Let  $\rho$  and  $\sigma$  be two density operators. Then the fidelity can be defined as

$$F(\rho, \sigma) = \text{Tr} \left( \sqrt{\rho^{1/2} \sigma \rho^{1/2}} \right). \quad (2.144)$$

For two pure states with density operators  $\rho = |\psi\rangle\langle\psi|$ ,  $\sigma = |\phi\rangle\langle\phi|$ , the corresponding fidelity will be

$$F(\rho, \sigma) = \text{Tr} \left( \sqrt{\rho^{1/2} \sigma \rho^{1/2}} \right) \stackrel{\rho^2 = \rho}{=} \text{Tr} \left( \sqrt{|\psi\rangle\langle\psi| (|\phi\rangle\langle\phi|) |\psi\rangle\langle\psi|} \right) \\ = \text{Tr} \left( \frac{|\psi\rangle\langle\psi| \underbrace{|\psi\rangle\langle\phi| \langle\phi|\psi\rangle}_{|\langle\phi|\psi\rangle|^2} |\psi\rangle\langle\psi|}{|\langle\phi|\psi\rangle|^2} \right)^{1/2} = |\langle\phi|\psi\rangle| \underbrace{\text{Tr}(|\psi\rangle\langle\psi|)}_1 = |\langle\phi|\psi\rangle| \quad (2.145)$$

where we used the property of the density operators for pure states  $\rho^2 = \rho$  (or equivalently  $\rho = \rho^{1/2}$ ). Clearly, for pure states, the fidelity corresponds to the square root of probability of finding the system in state  $|\phi\rangle$  if it is known to be prepared in state  $|\psi\rangle$  (and vice versa). Since fidelity is related to the probability, it ranges between 0 and 1, with 0 indicating that there is no overlap and 1 meaning that the states are identical. The following *properties* of fidelity hold:

1. The symmetry property:

$$F(\rho, \sigma) = F(\sigma, \rho). \quad (2.146)$$

2. The fidelity is invariant under unitary operations:

$$F(U\rho U^\dagger, U\sigma U^\dagger) = F(\rho, \sigma). \quad (2.147)$$

3. If  $\rho$  and  $\sigma$  commute, the fidelity can be expressed in terms of eigenvalues of  $\rho$ , denoted as  $r_i$ , and  $\sigma$ , denoted as  $s_i$ , as follows:

$$F(\rho, \sigma) = \sum_i (r_i s_i)^{1/2}. \quad (2.148)$$

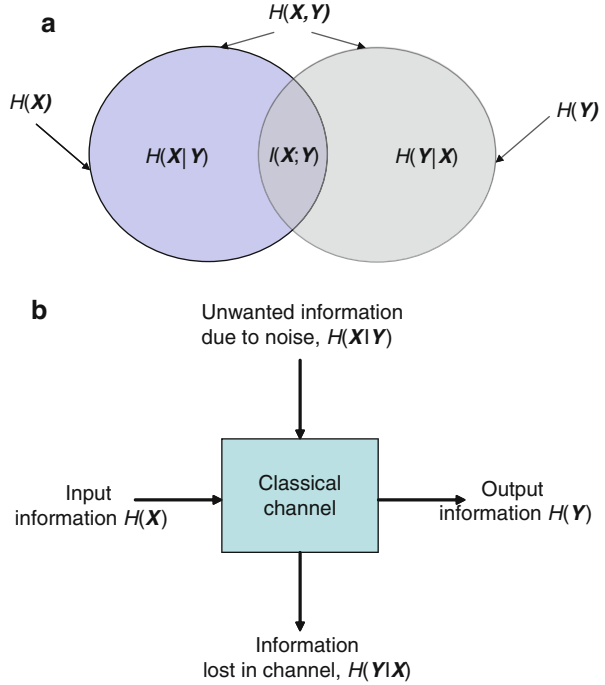
### 2.6.2 Holevo–Schumacher–Westmoreland Theorem and Channel Coding

We have already introduced the concept of mutual information  $I(X, Y)$  in Sect. 2.5, where we defined it as  $I(X, Y) = H(X) - H(Y|X)$ .  $H(X)$  represents the uncertainty about the channel input  $X$  before observing the channel output  $Y$ , while  $H(X|Y)$  denotes the conditional entropy or the amount of uncertainty remaining about the channel input after the channel output has been received. Therefore, the mutual information represents the amount of information (per symbol) that is conveyed by the channel. In other words, it represents the uncertainty about the channel input that is resolved by observing the channel output. The mutual information can be interpreted by means of a Venn diagram shown in Fig. 2.10a. The left circle represents the entropy of channel input, the right circle represents the entropy of channel output, and the mutual information is obtained in intersection of these two circles. Another interpretation due to Ingels [28] is shown in Fig. 2.10b. The mutual information, i.e., the information conveyed by the channel, is obtained as the output information minus information lost in the channel. One important figure of merit the classical channel is the *channel capacity*, which is obtained by maximization of mutual information  $I(X, Y)$  over all possible input distributions:

$$C = \max_{\{p(x_i)\}} I(X; Y). \quad (2.149)$$

The classical channel encoder accepts the message symbols and adds redundant symbols according to a corresponding prescribed rule. The channel coding is the act of transforming of a length- $K$  sequence into a length- $N$  codeword. The set of rules specifying this transformation are called the *channel code*, which can be represented as the following mapping:

**Fig. 2.10** Interpretation of the mutual information: (a) using Venn diagrams and (b) using the approach due to Ingels

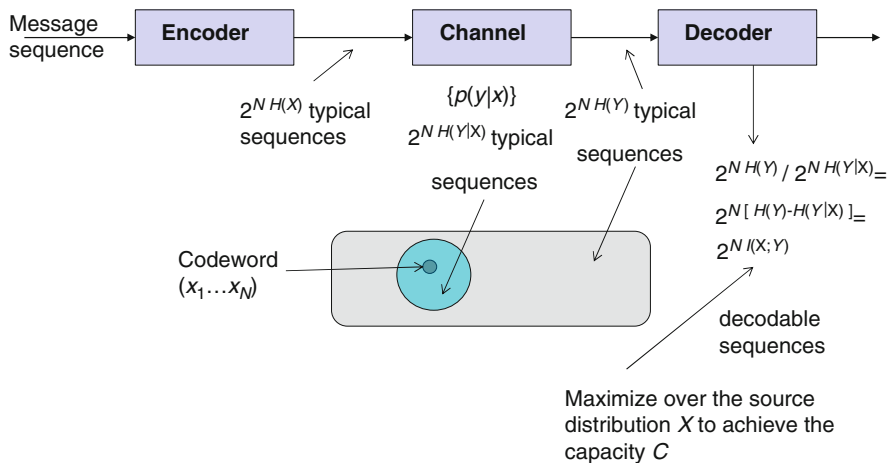


$$C : M \rightarrow X, \quad (2.150)$$

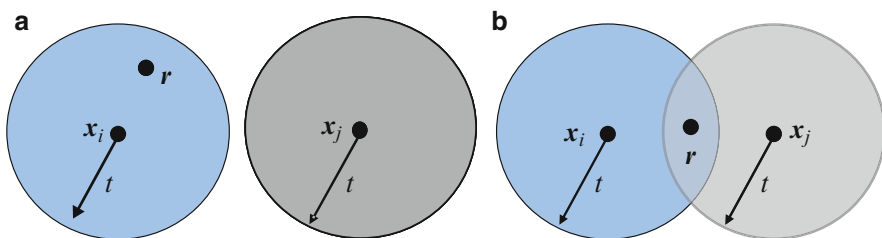
where  $C$  is the channel code,  $M$  is the set of information sequences of length  $K$ , and  $X$  is the set of codewords of length  $N$ . The decoder exploits these redundant symbols to determine which message symbol was actually transmitted. The concept of classical channel coding is introduced in Fig. 2.11. An important class of channel codes is the class of *block codes*. In an  $(N, K)$  *block code*, the channel encoder accepts information in successive  $K$ -symbol blocks and adds  $N - K$  redundant symbols that are algebraically related to the  $K$  message symbols, thus producing an overall encoded block of  $N$  symbols ( $N > K$ ), known as a *codeword*. If the block code is *systematic*, the information symbols stay unchanged during the encoding operation, and the encoding operation may be considered as adding the  $N - K$  generalized parity checks to  $k$  information symbols. The code rate of the code is defined as  $R = K/N$ .

In order to determine the error-correction capability of the linear block code, we have to introduce the concepts of Hamming distance and Hamming weight. The Hamming distance  $d(\mathbf{x}_1, \mathbf{x}_2)$  between two codewords  $\mathbf{x}_1$  and  $\mathbf{x}_2$  is defined as the number of locations in which these two vectors differ. The Hamming weight  $\text{wt}(\mathbf{x})$  of a codeword vector  $\mathbf{x}$  is defined as the number of nonzero elements in the vector. The minimum distance  $d_{\min}$  of a linear block code is defined as the smallest Hamming distance between any pair of code vectors in the code space. Since the





**Fig. 2.11** The illustration of classical channel coding and Shannon capacity theorem derivation



**Fig. 2.12** The illustration of Hamming distance: (a)  $d(x_i, x_j) \geq 2t + 1$  and (b)  $d(x_i, x_j) < 2t + 1$

zero vector is also a codeword, the minimum distance of a linear block code can be defined as the smallest Hamming weight of the nonzero code vectors in the code. The codewords can be represented as points in  $N$ -dimensional space, as shown in Fig. 2.12. Decoding process can be visualized by creating the spheres of radius  $t$  around codeword points. The received word vector  $r$  in Fig. 2.12a will be decoded as a codeword  $x_i$  because its Hamming distance  $d(x_i, r) \leq t$  is closest to the codeword  $x_i$ . On the other hand, in the example shown in Fig. 2.12b, the Hamming distance satisfies relation  $d(x_i, x_j) \leq 2t$ , and the received vector  $r$  that falls in intersection area of two spheres cannot be uniquely decoded.

Therefore,  $(N, K)$  linear block code of minimum distance  $d_{\min}$  can correct up to  $t$  errors if and only if  $t \leq \lfloor 1/2(d_{\min} - 1) \rfloor$  or  $d_{\min} \geq 2t + 1$  (where  $\lfloor \cdot \rfloor$  denotes the largest integer smaller or equal to the enclosed quantity). If we are only interested in detecting  $e_d$  errors, then the minimum distance should be  $d_{\min} \geq e_d + 1$ . However, if we are interested in detecting  $e_d$  errors and correcting  $e_c$  errors, then the minimum distance should be  $d_{\min} \geq e_d + e_c + 1$ .

Shannon has shown that if  $R < C$ , we can construct  $2^{NR}$  length- $N$  codewords that can be sent over the (classical) channel with maximum probability of error

approaching zero for large  $N$ . In order to prove this claim, we introduce the concept of jointly typical sequences. Two length- $N$  sequences  $\mathbf{x}$  and  $\mathbf{y}$  are jointly typical sequences if they satisfy the following set of inequalities:

$$P\left(\left|\frac{1}{N}\log\left(\frac{1}{P(\mathbf{x})}\right) - H(X)\right| \leq \varepsilon\right) > 1 - \delta \quad (2.151)$$

$$P\left(\left|\frac{1}{N}\log\left(\frac{1}{P(\mathbf{y})}\right) - H(Y)\right| \leq \varepsilon\right) > 1 - \delta \quad (2.152)$$

$$P\left(\left|\frac{1}{N}\log\left(\frac{1}{P(\mathbf{x}, \mathbf{y})}\right) - H(X, Y)\right| \leq \varepsilon\right) > 1 - \delta \quad (2.153)$$

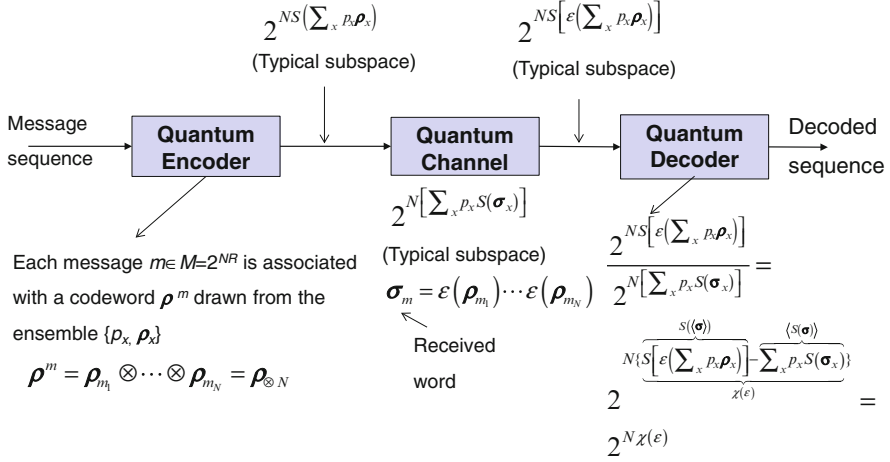
where  $P(\mathbf{x}, \mathbf{y})$  denotes the joint probability of the two sequences and  $H(X, Y)$  is their joint entropy.

For an length- $N$  input codeword randomly generated according to the probability distribution of a source  $X$ , the number of random input sequences is approximately  $2^{NH(X)}$ , and the number of output typical sequences is approximately  $2^{NH(Y)}$ . Furthermore, the total number of input and output sequences that are jointly typical is  $2^{NH(X, Y)}$ . Therefore, the total pairs of sequences that are simultaneously  $\mathbf{x}$ -typical,

$$2^{\underbrace{N[H(X) + H(Y) - H(X, Y)]}_{I(X, Y)}} = 2^{NI(X, Y)},$$

$\mathbf{y}$ -typical, and also jointly typical are  $2^{NI(X, Y)}$ , where  $I(X, Y)$  is the mutual information between  $X$  and  $Y$ , as illustrated in Fig. 2.11(bottom). These are the maximum number of codeword sequences that can be distinguished. One way of seeing this is to consider a single codeword. For this codeword, the action of the channel, characterized by the conditional probability  $P(\mathbf{y}|\mathbf{x})$ , defines the Hamming sphere in which this codeword can lie after the action of the channel. The size of this Hamming sphere is approximately  $2^{NH(\mathbf{y}|\mathbf{x})}$ . Given that the total number of output typical sequences is approximately  $2^{NH(Y)}$ , if we desire to have no overlap between two Hamming spheres, the maximum number of codewords we can consider is given by  $2^{NH(X)} / 2^{NH(\mathbf{y}|\mathbf{x})} = 2^{NI(X, Y)}$ . To increase this number, we need to maximize  $I(X, Y)$  over the distribution of  $X$ , as we do not have a control over the channel. This maximal mutual information is referred to as the capacity of the channel. If we have a rate  $R < C$ , then Shannon's noisy channel coding theorem tells us that we can construct  $2^{NR}$  length- $N$  codewords that can be sent over the channel with maximum probability of error approaching zero for large  $N$ . Now we can formally formulate Shannon channel coding theorem as follows [1].

**Shannon's Channel Coding Theorem.** Let us consider the transmission of  $2^{N(C - \varepsilon)}$  equiprobable messages. Then there exists a classical channel coding scheme of rate  $R < C$  in which the codewords are selected from all  $2^N$  possible words such that the decoding error probability can be made arbitrary small for sufficiently large  $N$ .



**Fig. 2.13** The illustration of quantum channel coding and HSW theorem derivation

Now we consider the communication over the quantum channel, as illustrated in Fig. 2.13. The quantum encoder for each message  $m$  out of  $M = 2^{NR}$  generates a product state codeword  $\rho^m$  drawn from the ensemble  $\{p_x, \rho_x\}$  as follows:

$$\rho^m = \rho_{m_1} \otimes \cdots \otimes \rho_{m_N} = \rho_{\otimes N}. \quad (2.154)$$

This quantum codeword is sent over the quantum channel described by the trace-preserving quantum operation  $\epsilon$ , resulting in the received quantum word:

$$\sigma_m = \epsilon(\rho_{m_1}) \otimes \epsilon(\rho_{m_2}) \otimes \cdots \otimes \epsilon(\rho_{m_N}). \quad (2.155)$$

Bob performs the measurement on received state using the POVM measurement operators  $\{M_m\}$  in order to decode Alice's message. The probability of successful decoding is given by  $p_m = \text{Trace}(\sigma_m M_m)$ . The goal is to maximize the transmission rate over the quantum channel so that the probability of decoding error is arbitrarily small. The von Neumann entropy associated with the quantum encoder would be  $S(\rho) = S(\sum_x p_x \rho_x)$ , while the dimensionality of the typical subspace of quantum

encoder is given by  $2^{NS(\sum_x p_x \rho_x)}$ . On the other hand, the dimensionality of the quantum subspace characterizing the quantum channel is given by  $2^{N[\sum_x p_x S(\sigma_x)]}$ . The quantum channel perturbs the quantum codeword transmitted over the quantum channel by performing the trace-preserving quantum operation so that the entropy at the channel output can be written as  $S[\epsilon(\sum_x p_x \rho_x)]$ , while the dimensionality of the corresponding subspace is given by

$2^{NS} \left[ \varepsilon \left( \sum_x p_x \rho_x \right) \right]$ . The number of decodable codewords would be then

$$\frac{2^{NS} \left[ \varepsilon \left( \sum_x p_x \rho_x \right) \right]}{2^N \left[ \sum_x p_x S(\rho_x) \right]} = 2^N \left\{ \underbrace{S \left[ \varepsilon \left( \sum_x p_x \rho_x \right) \right]}_{\chi(\varepsilon)} - \sum_x p_x S(\rho_x) \right\} = 2^{N\chi(\varepsilon)}. \quad (2.156)$$

In order to maximize the number of decodable codewords, we need to perform the optimization of  $\chi(\varepsilon)$  over  $p_x$  and  $\rho_x$ , which represents the simplified derivation of the Holevo–Schumacher–Westmoreland (HSW) theorem, which can be formulated as follows [1].

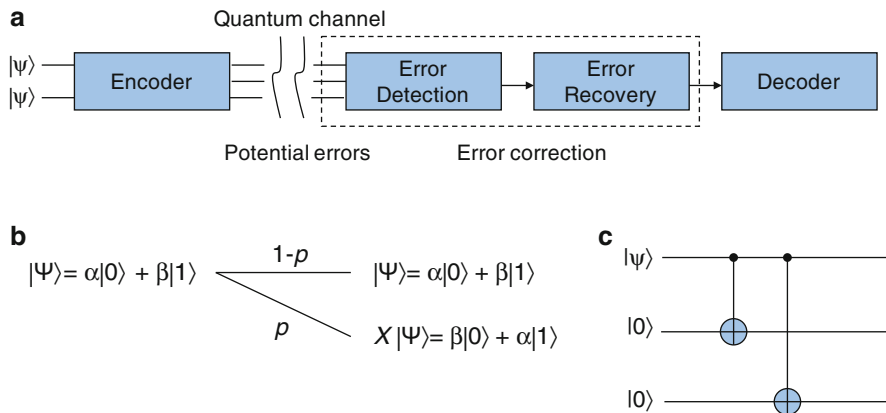
**Holevo–Schumacher–Westmoreland (HSW) Theorem.** Let us consider the transmission of a codeword  $\rho^m$  drawn from the ensemble  $\{p_x, \rho_x\}$  over the quantum channel, characterized by the trace-preserving quantum operation  $\varepsilon$ , with the rate  $R < C(\varepsilon)$ , where  $C(\varepsilon)$  is the product state capacity defined as

$$\begin{aligned} C(\varepsilon) &= \max_{\{p_x, \rho_x\}} \chi(\varepsilon) = \max_{\{p_x, \rho_x\}} [S(\langle \sigma \rangle) - \langle S(\rho_x) \rangle] \\ &= \max_{\{p_x, \rho_x\}} \left[ S \left[ \varepsilon \left( \sum_x p_x \rho_x \right) \right] - \sum_x p_x S(\rho_x) \right], \end{aligned} \quad (2.157)$$

where the maximization of  $\chi(\varepsilon)$  is performed over  $p_x$  and  $\rho_x$ . Then there exists a coding scheme that allows reliable error-free transmission over the quantum channel.

## 2.7 Quantum Error-Correction Concepts

The QIP relies on delicate superposition states, which are sensitive to interactions with environment, resulting in decoherence. Moreover, the quantum gates are imperfect and the use of quantum error-correction coding (QECC) is necessary to enable the fault-tolerant computing and to deal with quantum errors. QECC is also essential in quantum communication and quantum teleportation applications. The elements of quantum error-correction codes are shown in Fig. 2.14a. The  $(N, K)$  QECC code performs the encoding of the quantum state of  $K$  qubits, specified by  $2^K$  complex coefficients  $\alpha_s$ , into a quantum state of  $N$  qubits, in such a way that errors can be detected and corrected, and all  $2^K$  complex coefficients can be perfectly restored, up to the global phase shift. Namely, from quantum mechanics, we know that two states  $|\psi\rangle$  and  $e^{i\theta}|\psi\rangle$  are equal up to a *global phase shift* as the results of measurement on both states are the same. A quantum error correction consists of



**Fig. 2.14** (a) A quantum error-correction principle. (b) Bit-flipping channel model. (c) Three-qubit flip code encoder

four major steps: encoding, error detection, error recovery, and decoding, as shown in Fig. 2.14a. The sender (Alice) encodes quantum information in state  $|\psi\rangle$  with the help of local ancilla qubits  $|0\rangle$  and then sends the encoded qubits over a noisy quantum channel (say free-space optical channel or optical fiber). The receiver (Bob) performs multiqubit measurement on all qubits to diagnose the channel error and performs a recovery unitary operation  $R$  to reverse the action of the channel. The quantum error correction is essentially more complicated than classical error correction. Difficulties for quantum error correction can be summarized as follows: (1) the no-cloning theorem indicates that it is impossible to make a copy of an arbitrary quantum state, (2) quantum errors are continuous and a qubit can be in any superposition of the two bases states, and (3) the measurements destroy the quantum information. The quantum error-correction principles will be more evident after a simple example given below.

Assume we want to send a single qubit  $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$  through the quantum channel in which during transmission the transmitted qubit can be flipped to  $X|\psi\rangle = \beta|0\rangle + \alpha|1\rangle$  with probability  $p$ . Such a quantum channel is called a *bit-flip channel* and it can be described as shown in Fig. 2.14b. A three-qubit flip code sends the same qubit three times and therefore represents the *repetition code* equivalent. The corresponding codewords in this code are  $|\bar{0}\rangle = |000\rangle$  and  $|\bar{1}\rangle = |111\rangle$ . The three-qubit flip code encoder is shown in Fig. 2.14c. One input qubit and two ancillas are used at the input encoder, which can be represented by  $|\psi_{123}\rangle = \alpha|000\rangle + \beta|100\rangle$ . The first ancilla qubit (the second qubit at the encoder input) is controlled by the information qubit (the first qubit at encoder input) so that its output can be represented by  $\text{CNOT}_{12}(\alpha|000\rangle + \beta|100\rangle) = \alpha|000\rangle + \beta|110\rangle$  (if the control qubit is  $|1\rangle$  the target qubit gets flipped; otherwise it stays unchanged). The output of the first CNOT gate is used as input to the second CNOT gate in which the second ancilla qubit (the third qubit) is controlled by the information qubit (the first qubit) so that the corresponding encoder output is

obtained as  $\text{CNOT}_{13}(\alpha|000\rangle + \beta|110\rangle) = \alpha|000\rangle + \beta|111\rangle$ , which indicates that basis codewords are indeed  $|\bar{0}\rangle$  and  $|\bar{1}\rangle$ . With this code, we are capable to correct a single-qubit flip error, which occurs with probability  $(1-p)^3 + 3p(1-p)^2 = 1 - 3p^2 + 2p^3$ . Therefore, the probability of an error remaining uncorrected or wrongly corrected with this code is  $3p^2 - 2p^3$ . It is clear from Fig. 2.14c that three-qubit bit-flip encoder is a *systematic encoder* in which the information qubit is unchanged, and the ancilla qubits are used to impose the encoding operation and create the parity qubits (the output qubits 2 and 3).

Let us assume that a qubit flip occurred on the first qubit leading to received quantum word  $|\psi_r\rangle = \alpha|100\rangle + \beta|011\rangle$ . In order to identify the error, it is needed to perform the measurements on the following observables:  $Z_1Z_2$  and  $Z_2Z_3$ , where the subscript denotes the index of a qubit on which a given Pauli gate is applied. The result of the measurement is the eigenvalue  $\pm 1$ , and corresponding eigenvectors are two valid codewords, namely,  $|000\rangle$  and  $|111\rangle$ . The observables can be represented as follows:

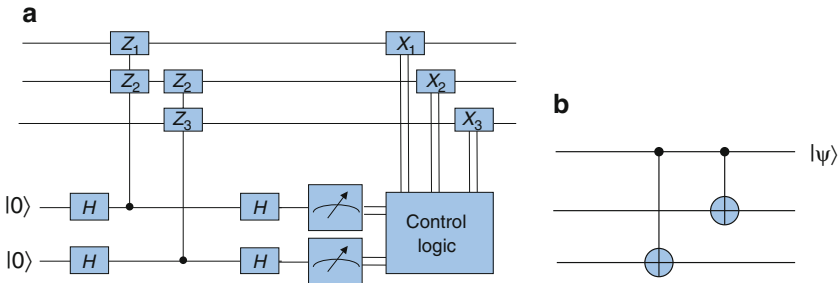
$$\begin{aligned} Z_1Z_2 &= (|00\rangle\langle 11| + |11\rangle\langle 11|) \otimes I - (|01\rangle\langle 01| + |10\rangle\langle 10|) \otimes I \\ Z_2Z_3 &= I \otimes (|00\rangle\langle 11| + |11\rangle\langle 11|) - I \otimes (|01\rangle\langle 01| + |10\rangle\langle 10|). \end{aligned} \quad (2.158)$$

It can be showed that  $\langle \psi_r | Z_1Z_2 | \psi_r \rangle = -1$ ,  $\langle \psi_r | Z_2Z_3 | \psi_r \rangle = +1$ , indicating that an error occurred on either the first or second qubit but neither on the second nor third qubit. The intersection reveals that the first qubit was in error. By using this approach, we can create the three-qubit lookup table (LUT), given as Table 2.1.

Three-qubit flip code error detection and error-correction circuits are shown in Fig. 2.15. The results of measurements on ancillas (see Fig. 2.15a) will determine

**Table 2.1** The three-qubit flip code LUT

$Z_1Z_2$	$Z_2Z_3$	Error
+1	+1	/
+1	-1	$X_3$
-1	+1	$X_1$
-1	-1	$X_2$



**Fig. 2.15** (a) Three-qubit flip code error detection and error-correction circuit. (b) Decoder circuit configuration

the error syndrome  $[\pm 1 \pm 1]$ , and based on LUT given by Table 2.1, we identify the error event and apply corresponding  $X_i$  gate on the  $i$ th qubit being in error, and the error gets corrected since  $X^2 = I$ . The control logic operation is described in Table 2.1. For example, if both outputs at the measurements circuits are  $-1$ , the operator  $X_2$  is activated. The last step is to perform decoding as shown in Fig. 2.15b by simply reversing the order of elements in corresponding encoder.

## 2.8 Hydrogen-Like Atoms and Beyond

At this point, it is convenient to establish *connection* between *wave quantum mechanics* and *matrix quantum mechanics*, as we will use the concept of the wave function in the rest of the section. In wave mechanics, the information about the state of a particle is described by the corresponding **wave function**  $\psi(x, t) = \langle x|\psi \rangle$ . The wave function gives the information about the location of the particle, namely, the magnitude squared of the wave functions  $|\psi(x, t)|^2$  is related to the *probability density function*. The probability of finding the particle within the interval  $x$  and  $x + dx$  is given by

$$dP(x, t) = |\psi(x, t)|^2 dx. \quad (2.159)$$

In wave quantum mechanics, the actions of the position  $X$  and the momentum  $P$  operators are defined by

$$X\psi(x, t) = x\psi(x, t) \quad P\psi(x, t) = -j\hbar \frac{\partial}{\partial x} \psi(x, t). \quad (2.160)$$

When we apply the commutator to the test wave function  $\psi(x, t)$ , we obtain

$$\begin{aligned} [X, P]\psi(x, t) &= (XP - PX)\psi(x, t) = XP\psi(x, t) - PX\psi(x, t) \\ &= -j\hbar x \frac{\partial}{\partial x} \psi(x, t) + j\hbar \frac{\partial}{\partial x} (X\psi(x, t)) \\ &= -j\hbar x \frac{\partial}{\partial x} \psi(x, t) + j\hbar \frac{\partial}{\partial x} (x\psi(x, t)) \\ &= -j\hbar x \frac{\partial}{\partial x} \psi(x, t) + j\hbar \left( \psi(x, t) + x \frac{\partial}{\partial x} \psi(x, t) \right) = j\hbar \psi(x, t), \end{aligned} \quad (2.161)$$

indicating, therefore, that  $[X, P] = j\hbar$ , which was used in the section on uncertainty principle.

A hydrogen atom is a bound system, consisting of a proton and a neutron, with potential given by

$$V(r) = -\frac{1}{4\pi\epsilon_0} \frac{e^2}{r}, \quad (2.162)$$

where  $e$  is an electron charge. Therefore, the potential is only function of radial coordinate, and because of spherical symmetry, it is convenient to use the spherical coordinate system in which the Laplacian is defined by

$$\nabla^2 = \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2}. \quad (2.163)$$

The angular momentum operator  $L^2$  in spherical coordinates is given by

$$L^2 = -\hbar^2 \left[ \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right]. \quad (2.164)$$

The Hamiltonian can be written by

$$H = -\frac{\hbar^2}{2m} \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial}{\partial r} \right) + \frac{1}{2mr^2} L^2 + V(r). \quad (2.165)$$

Because the operators  $L^2$  and  $L_z$  have the common eigenkets, the Hamiltonian leads to the following three equations:

$$\begin{aligned} H\Psi(r, \theta, \phi) &= E\Psi(r, \theta, \phi) \\ L^2\Psi(r, \theta, \phi) &= \hbar^2 l(l+1)\Psi(r, \theta, \phi) \\ L_z\Psi(r, \theta, \phi) &= m\hbar\Psi(r, \theta, \phi). \end{aligned} \quad (2.166)$$

With this problem, we can associate three quantum numbers: (1) the *principal quantum number*,  $n$ , corresponding to the energy (originating from Hamiltonian  $H$ ); (2) the *azimuthal quantum number*,  $l$ , representing the angular momentum (originating from  $L^2$ ); and (3) the *magnetic quantum number*,  $m$ , originating from  $L_z$ . In order to solve (2.166), we can use the method of separation of variables:  $\Psi(r, \theta, \phi) = R(r)\Theta(\theta)\Phi(\phi)$ . Since  $\Theta(\theta)\Phi(\phi)$  is related to the spherical harmonics  $\Theta(\theta)\Phi(\phi) = Y_l^m(\theta, \phi)$ , we are left with the radial equation to solve

$$-\frac{\hbar^2}{2mr} \frac{d^2}{dr^2} [rR_{nl}(r)] + \left[ \frac{l(l+1)\hbar^2}{2mr^2} + V(r) \right] R_{nl}(r) = ER_{nl}(r). \quad (2.167)$$



The spherical harmonics  $Y_l^m(\theta, \phi)$  are defined by

$$Y_l^m(\theta, \phi) = \begin{cases} (-1)^m \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} P_l^m(\cos \theta) e^{jm\phi}, & m > 0 \\ (-1)^{|m|} \sqrt{\frac{(2l+1)(l-|m|)!}{4\pi(l+m)!}} P_l^{|m|}(\cos \theta) e^{jm\phi}, & m < 0 \end{cases}. \quad (2.168)$$

With  $P_l^m(x)$ , we denoted the associated Legendre polynomials

$$P_l^m(x) = (1-x^2)^{m/2} \frac{d^m}{dx^m} P_l(x), \quad (2.169)$$

where  $P_l(x)$  are the Legendre polynomials, defined by

$$P_l(x) = \frac{1}{2^l l!} \frac{d^l}{dx^l} [(x^2-1)^l]. \quad (2.170)$$

The Legendre polynomial can be determined recursively as follows:

$$(l+1)P_{l+1}(x) = (2l+1)xP_l(x) - lP_{l-1}(x); \quad P_0(x) = 1, \quad P_1(x) = x. \quad (2.171)$$

By substituting the Coulomb potential into the radial equation, we obtain

$$-\frac{\hbar^2}{2mr} \frac{d^2}{dr^2} [rR_{nl}(r)] + \left[ \frac{l(l+1)\hbar^2}{2mr^2} - \frac{1}{4\pi\epsilon_0} \frac{e^2}{r} \right] R_{nl}(r) = ER_{nl}(r). \quad (2.172)$$

The solution of radial equation can be written as

$$R_{nl}(r) = \sqrt{\left(\frac{2}{na_H}\right) \frac{(n-l-1)!}{2n[(n-1)!]^2}} e^{-r/2a_H} \left(\frac{r}{a_H}\right)^l L_{n+1}^{2l+1}\left(\frac{r}{a_H}\right), \quad (2.173)$$

where  $a_H$  is the first Bohr radius (the lowest energy orbit radius) ( $a_H = 0.0529$  nm) and  $L_{n+1}^{2l+1}(r/a_H)$  are the corresponding associated Laguerre polynomials, defined by

$$L_n^\alpha(x) = \frac{x^{-\alpha} e^x}{n!} \frac{d^n}{dx^n} (e^{-x} x^{n+\alpha}). \quad (2.174)$$

The radial portion of the wave function is typically normalized as follows:

$$\int_0^\infty r^2 |R(r)|^2 dr = 1. \quad (2.175)$$

The complete wave function is the product of radial wave function and spherical harmonics:

$$\Psi(r, \theta, \phi) = \sqrt{\left(\frac{2}{na_H}\right) \frac{(n-l-1)!}{2n[(n-1)!]^2}} e^{-r/2a_H} \left(\frac{r}{a_H}\right)^l L_{n+1}^{2l+1}\left(\frac{r}{a_H}\right) Y_l^m(\theta, \phi). \quad (2.176)$$

By substituting (2.168) and (2.173) into the wave function expression  $\Psi(r, \theta, \phi) = R(r)\Theta(\theta)\Phi(\phi)$ , we obtain

$$\Psi(\rho, \theta, \phi) = C_{n,l,m} e^{-\rho/2} (\rho)^l L_{n+1}^{2l+1}(\rho) P_l^{|m|}(\cos \theta) e^{jm\phi}, \quad (2.177)$$

where  $\rho = r/a_H$  and  $C_{n,l,m}$  is the normalization constant obtained as product of normalization constants in (2.168) and (2.176). The energy levels in hydrogen atom are only functions on  $n$  and are given by

$$E_n = \frac{E_1}{n^2}, \quad E_1 = -\frac{m^2 e^4}{32\pi^2 \epsilon_0^2 \hbar^2} = -13.6 \text{ eV}. \quad (2.178)$$

Because the values that  $l$  can take are  $\{0, 1, \dots, n-1\}$ , while the values that  $m$  can take are  $\{-l, -l+1, \dots, l-1, l\}$ , and since the radial component is not a function of  $m$ , the number of states with the same energy (the total *degeneracy* of the energy level  $E_n$ ) is

$$2 \sum_{l=0}^{n-1} (2l+1) = 2n^2. \quad (2.179)$$

The states in which  $l=0, 1, 2, 3, 4$  are traditionally called *s*, *p*, *d*, *f*, and *g*, respectively. The simpler eigenfunctions of the hydrogen atom, by ignoring the normalization constant, are given in Table 2.2, which is obtained based on (2.177).

The results above applicable to many two-particle systems with attraction energy are reversely proportional to the distance between them, provided that parameters are properly chosen. For instance, if the charge of nucleus is  $Z$ , then in the calculations above, we need to substitute  $e^2$  by  $Ze^2$ . Examples include deuterium, tritium, ions that contain only one electron, positronium, and muonic atoms.

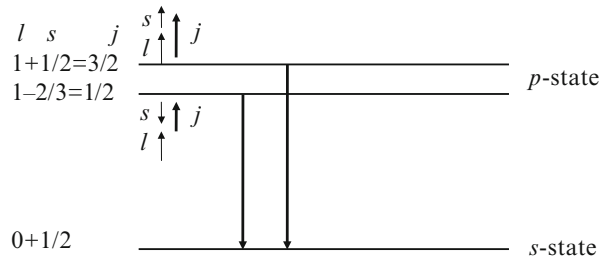
The total angular momentum of an electron  $\mathbf{j}$  in an atom can be found as the vector sum of orbital angular momentum  $\mathbf{l}$  and spin  $\mathbf{s}$  as by

$$\mathbf{j} = \mathbf{l} + \mathbf{s}. \quad (2.180)$$

For a given value of azimuthal quantum number  $l$ , there exist two values of total angular momentum quantum number of an electron:  $j = l + 1/2$  and  $j = l - 1/2$ . Namely, as the electron undergoes orbital motion around the nucleus, it experiences the magnetic field, and this interaction is known as the spin-orbit interaction.

**Table 2.2** Hydrogen atom eigenfunctions

State	$n$	$l$	$m$	Eigenfunction
$1s$	1	0	0	$e^{-\rho/2}$
$2s$	2	0	0	$e^{-\rho/2}(1 - \rho)$
$2p$	2	1	-1	$e^{-\rho/2}\rho \begin{cases} \sin \theta e^{-j\phi} \\ \cos \theta \\ \sin \theta e^{j\phi} \end{cases}$
			0	
			1	
$3s$	3	0	0	$e^{-\rho/2}(\rho^2 - 4\rho + 2)$
$3p$	3	1	-1	$e^{-\rho/2}(\rho^2 - 2\rho) \begin{cases} \sin \theta e^{-j\phi} \\ \cos \theta \\ \sin \theta e^{j\phi} \end{cases}$
			0	
			1	
$3d$	3	2	-2	$e^{-\rho/2}\rho^2 \begin{cases} \sin^2 \theta e^{-j2\phi} \\ \sin \theta \cos \theta e^{-j\phi} \\ 1 - 3 \cos^2 \theta \\ \sin \theta \cos \theta e^{j\phi} \\ \sin^2 \theta e^{j2\phi} \end{cases}$
			-1	
			0	
			1	
			2	

**Fig. 2.16** Illustration of spin-orbit interaction

The result of this interaction is two states  $j = l + s$  and  $j = l - s$  with slightly different energies as shown in Fig. 2.16.

For atoms containing more than one electron, the total angular momentum  $\mathbf{J}$  is given by the sum of individual orbital momenta  $\mathbf{L} = \mathbf{l}_1 + \mathbf{l}_2 + \cdots$  and spins  $\mathbf{S} = \mathbf{s}_1 + \mathbf{s}_2 + \cdots$ , so that we can write

$$\mathbf{J} = \mathbf{L} + \mathbf{S}; \quad \mathbf{L} = \mathbf{l}_1 + \mathbf{l}_2 + \cdots, \quad \mathbf{S} = \mathbf{s}_1 + \mathbf{s}_2 + \cdots. \quad (2.181)$$

This type of coupling is known as  $LS$  coupling. Another type of coupling, namely,  $JJ$  coupling, occurs when individual  $j$ s add together to produce the resulting  $\mathbf{J}$ . The  $LS$  coupling typically occurs in the lighter elements, while  $JJ$  coupling typically occurs in heavy elements. In  $LS$  coupling, the magnitudes of  $\mathbf{L}$ ,  $\mathbf{S}$ , and  $\mathbf{J}$  are given by

$$|\mathbf{L}| = \hbar\sqrt{L(L+1)}, \quad |\mathbf{S}| = \hbar\sqrt{S(S+1)}, \quad |\mathbf{J}| = \hbar\sqrt{J(J+1)}, \quad (2.182)$$

where  $L$ ,  $S$ , and  $J$  are quantum numbers satisfying the following properties: (1)  $L$  is always a nonnegative integer, (2) the spin quantum number  $S$  is either integral or

half-integral depending whether the number of electrons is even or odd, and (3) the total angular momentum quantum number  $J$  is either integral or half-integral depending whether the number of electrons is even or odd, respectively. The *spectroscopic notation* of a state characterized by the quantum numbers  $L$ ,  $S$ , and  $J$  is as follows:

$$^{2S+1}L_J, \quad (2.183)$$

where the quantity  $2S + 1$  is known as the *multiplicity* and determines the numbers of different  $J$ s for a given value of  $L$ . If  $L \ll S$ , different values of  $J$  are  $L + S$ ,  $L + S - 1, \dots, L - S$ , meaning that there are  $2S + 1$  possible values for  $J$ . If, on the other hand,  $L < S$ , then the possible values of  $J$  are  $L + S$ ,  $L + S - 1, \dots, |L - S|$ , meaning only  $2L + 1$  different values for  $J$  exist. The states in which  $L = 0, 1, 2, 3, 4, 5, 6, 7, 8, \dots$  are traditionally called  $S, P, D, F, G, H, I, K, M, \dots$ , respectively. The states with multiplicity  $2S + 1 = 0, 1, 2, 3, 4, 5$ , and  $6$  are typically called singlet, doublet, triplet, quartet, quintet, and sextet states, respectively.

## 2.9 Concluding Remarks

This chapter has provided an overview of the basic concepts of QIP and quantum information theory. The following topics from QIP have been described in Sects. 2.1–2.3: state vectors, operators, density operators, measurements, dynamics of a quantum system, superposition principle, quantum parallelism, no-cloning theorem, and entanglement. The following concepts from quantum information theory have been described in Sects. 2.4–2.6: Holevo information, accessible information, Holevo bound, Shannon entropy and von Neumann entropy, Schumacher’s noiseless quantum coding theorem, and Holevo–Schumacher–Westmoreland theorem. Section 2.7 has been related to the quantum error-correction concepts. Finally, Sect. 2.8 is devoted to the hydrogen-like atoms (and beyond).

## References

1. Djordjevic IB (2012) Quantum information processing and quantum error correction: an engineering approach. Elsevier/Academic Press, Amsterdam
2. Helstrom CW (1976) Quantum detection and estimation theory. Academic, New York
3. Helstrom CW, Liu JWS, Gordon JP (1970) Quantum mechanical communication theory. Proc IEEE 58:1578–1598
4. Sakurai JJ (1994) Modern quantum mechanics. Addison-Wesley, Reading
5. Gaitan F (2008) Quantum error correction and fault tolerant quantum computing. CRC, Boca Raton
6. Baym G (1990) Lectures on quantum mechanics. Westview Press, New York
7. McMahon D (2007) Quantum computing explained. Wiley, Hoboken

8. McMahon D (2006) Quantum mechanics demystified. McGraw-Hill, New York
9. Nielsen MA, Chuang IL (2000) Quantum computation and quantum information. Cambridge University Press, Cambridge
10. Fleming S (2008) PHYS 570: quantum mechanics (Lecture notes). University of Arizona, Tucson
11. Dirac PAM (1958) Quantum mechanics, 4th edn. Oxford University Press, London
12. Peleg Y, Pnini R, Zaarur E (1998) Schaum's outline of theory and problems of quantum mechanics. McGraw-Hill, New York
13. Fowles GR (1989) Introduction to modern optics. Dover, New York
14. Scully MO, Zubairy MS (1997) Quantum optics. Cambridge University Press, Cambridge
15. Le Bellac M (2006) A short introduction to quantum information and quantum computation. Cambridge University Press, Cambridge
16. Pinter CC (2010) A book of abstract algebra. Dover, New York
17. Griffiths D (1995) Introduction to quantum mechanics. Prentice-Hall, Englewood Cliffs
18. Zettili N (2001) Quantum mechanics, concepts and applications. Wiley, New York
19. Von Neumann J (1955) Mathematical foundations of quantum mechanics. Princeton University Press, Princeton
20. Jauch JM (1968) Foundations of quantum mechanics. Addison-Wesley, Reading
21. Peres A (1995) Quantum theory: concepts and methods. Kluwer, Boston
22. Goswami A (1996) Quantum mechanics. McGraw-Hill, New York
23. Jammer M (1974) Philosophy of quantum mechanics. Wiley, New York
24. McWeeny R (2003) Quantum mechanics, principles and formalism. Dover, Mineola
25. Weyl H (1950) Theory of groups and quantum mechanics. Dover, New York
26. Liboff RL (1997) Introductory quantum mechanics. Addison-Wesley, Reading
27. Cover TM, Thomas JA (1991) Elements of information theory. Wiley, New York
28. Ingels FM (1971) Information and coding theory. Intext Educational Publishers, Scranton
29. Cvijetic M, Djordjevic IB (2013) Advanced optical communications and networks. Artech House, Boston
30. Djordjevic IB, Ryan W, Vasic B (2010) Coding for optical channels. Springer, New York



<http://www.springer.com/978-3-319-22815-0>

Quantum Biological Information Theory

Djordjevic, I.B.

2016, XI, 269 p., Hardcover

ISBN: 978-3-319-22815-0