12/05/25

Lecture-35

11/12/20

Revanth Reddy Pannala
EBIT, Unibo
రేవంత్ రెడ్డి పన్నాల

# Chapter 1
# M. Rudan — Analytical Mechanics — An Introduction

- We will explore the similarities b/w Analytical Mechanics & Quantum Mechanics.
- Analytical mechanics has different forms

## 1.1 Lagrangian Function

The motion of a point-like particle of mass $m$ is described by Newton's law

$$\mathbf{F} = m\,\mathbf{a}\,, \tag{1.1}$$

where the components of the acceleration $\mathbf{a}$ will be indicated with $\ddot{x}_i(t)$, $i = 1,2,3$. The force $\mathbf{F}$ depends in general on the components $x_i(t)$ of position $\mathbf{r}$ of the particle, on the components $\dot{x}_i(t)$ of its velocity $\mathbf{u} = \dot{\mathbf{r}}$, and on the time $t$. It follows that the three scalar components of the dynamical law (1.1) take the form

$$\ddot{x}_i = \frac{1}{m} F_i(\mathbf{r}, \dot{\mathbf{r}}, t)\,, \qquad i = 1,2,3\,. \tag{1.2}$$

Lagrangian:

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}} = \frac{\partial L}{\partial x}$$

$L(r, \dot{r}, t)$

Eqs. (1.2) may be recast in a different form by introducing the Lagrangian $L$, that is, a scalar function $L = L(\mathbf{r}, \dot{\mathbf{r}}, t)$ such that the three differential equations (the *Lagrange equations*)

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{x}_i} = \frac{\partial L}{\partial x_i}\,, \qquad i = 1,2,3\,, \tag{1.3}$$

explicit forces are not mentioned

are respectively identical to (1.2). The Lagrangian does not necessarily exist for arbitrarily-prescribed $F_i$. However, if the forces derive from a potential energy $V = V(\mathbf{r})$, that is,

$$F_i = F_i(\mathbf{r}) = -\frac{\partial V}{\partial x_i}\,, \tag{1.4}$$

for Conservative forces i.e forces dependent on the position and are derivable! it is possible

then the Lagrangian exists and reads

$$L = T - V\,, \qquad T = \frac{1}{2}mu^2 = \frac{1}{2}m\left(\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2\right), \tag{1.5}$$

where $T$ is the kinetic energy of the particle. In fact, observing that $V$ does not depend on the velocity of the particle, one finds



* If Lagrangian is Invariant of the position of origin, then the Momentum of the S/S is conserved. if it is also Invariant after Rotation then Angular Momentum is conserved.

If lagrangian is invariant of Mirroring wrt time Then the Total Energy of the S/S is conserved

$$\frac{\partial L}{\partial \dot{x}_i} = \frac{\partial T}{\partial \dot{x}_i} = m\dot{x}_i\,, \tag{1.6}$$

whence

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{x}_i} = m\ddot{x}_i\,. \tag{1.7}$$

As $T$ does not depend on the position components,

$$\frac{\partial L}{\partial x_i} = -\frac{\partial V}{\partial x_i} = F_i\,. \tag{1.8}$$

- Equating (1.8) and (1.7) yields (1.2) for every index $i$. In the special case considered here ($F_i = -\partial V/\partial x_i$) the components of the force do not depend on the velocity and have no explicit dependence on the time.

- Function $p_i = p_i(t)$, defined for every index $i$ by

$$p_i = \frac{\partial L}{\partial \dot{x}_i}\,, \tag{1.9}$$

is the *momentum* conjugate to the coordinate $x_i$. The set of three canonically-conjugate pairs $x_i(t)$, $p_i(t)$ is also called the *state* of the particle at time $t$.

- The same concept is generalized to the case where, instead of a single particle, one considers a systems of $N$ interacting particles (if the particles were not interacting, they would actually form $N$ independent systems). Assuming for the sake of simplicity that the particles are not subjected to constraints, the degrees of freedom of the system are $s = 3N$. Letting $\mathbf{r}_j$ be the position vector of the $j$th particle, $j = 1,\ldots,N$, the component of the force and the Lagrangian read

$$F_i = F_i(\mathbf{r}_1, \dot{\mathbf{r}}_1, \ldots, \mathbf{r}_N, \dot{\mathbf{r}}_N, t)\,, \qquad L = L(\mathbf{r}_1, \dot{\mathbf{r}}_1, \ldots, \mathbf{r}_N, \dot{\mathbf{r}}_N, t)\,, \tag{1.10}$$

- where the index $i$ of the component of the force ranges over all the degrees of freedom of the system, $i = 1,\ldots,s$. The definition of the momentum $p_i$ conjugate to the coordinate $x_i$ is still given by (1.9), and the Lagrange equations are identical to (1.3) with the exception that the Lagrangian depends on the position and velocity of all particles, and index $i$ varies over all degrees of freedom:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{x}_i} = \frac{\partial L}{\partial x_i}\,, \qquad i = 1,\ldots,s\,. \tag{1.11}$$

The component $F_i$ of the force in (1.10) is the force affecting the $i$th degree of freedom; for instance, if index $i$ is related to the motion of the $j$th particle ($j = 1,\ldots,N$) along the $k$th axis of the rectangular reference ($k = 1,2,3$), then $F_i$ will determine the component of the acceleration of the $j$th particle along the $k$th axis. Such a force is in general due to the influence of the other particles of the system, that appear in the expression of $F_i$ through their positions and velocities, and also to the influence of other causes external to the system. Such causes appear in the

expression of $F_i$ through the position and velocity of the $i$th degree of freedom itself. If the force derives from a potential energy $V(\mathbf{r}_1,\ldots,\mathbf{r}_N)$, it follows

$$F_i = F_i(\mathbf{r}_1,\ldots,\mathbf{r}_N) = -\frac{\partial V}{\partial x_i}, \tag{1.12}$$

and (1.5) still holds with

$$T = \sum_{j=1}^{N} \frac{1}{2} m_j u_j^2, \qquad u_j = |\dot{\mathbf{r}}_j|. \tag{1.13}$$

## 1.2 Generalized Coordinates

It is useful to determine how the relations of section 1.1 change due to a coordinate transformation. Let the new coordinates be $q_i$, $i = 1,\ldots,s$ (*generalized coordinates*), and let the corresponding generalized velocities be $\dot{q}_i$. As the dimensions of the $q_i$ are not necessarily those of a length, those of the $\dot{q}_i$ are not necessarily the ratio of a length to a time. The transformation laws from the rectangular to the generalized coordinates have the form

$$q_i = q_i(x_1,\ldots,x_s,t), \qquad i = 1,\ldots,s. \tag{1.14}$$

The time $t$ may appear explicitly in (1.14) because the new reference may possess a relative motion with respect to the old one. One also assumes that Eqs. (1.14) are invertible, namely, that the following exist:

$$x_i = x_i(q_1,\ldots,q_s,t), \qquad i = 1,\ldots,s. \tag{1.15}$$

After (1.14) have been prescribed, one calculates

$$\dot{q}_i = \dot{q}_i(x_1,\dot{x}_1,\ldots,x_s,\dot{x}_s,t), \qquad i = 1,\ldots,s \tag{1.16}$$

by taking the total derivative with respect to time. In the same manner one calculates from (1.15)

$$\dot{x}_i = \dot{x}_i(q_1,\dot{q}_1,\ldots,q_s,\dot{q}_s,t), \qquad i = 1,\ldots,s. \tag{1.17}$$

Then, introducing (1.15) and (1.17) into (1.11) yields the Lagrange equations in terms of the generalized coordinates. A noteworthy results is that the form of the Lagrange equations is invariant:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i} = \frac{\partial L}{\partial q_i}, \qquad i = 1,\ldots,s, \tag{1.18}$$

with

$$L = L(q_1,\dot{q}_1,\ldots,q_s,\dot{q}_s,t). \tag{1.19}$$

For the momentum conjugate to $q_i$ the same symbol is used as in the case of the rectangular coordinates:

$$p_i = \frac{\partial L}{\partial \dot{q}_i}, \qquad i = 1, \ldots, s. \tag{1.20}$$

Combining (1.20) with (1.18) yields

$$\dot{p}_i = \frac{\partial L}{\partial q_i}, \qquad i = 1, \ldots, s. \tag{1.21}$$

As $L$ depends on the $q_i$, $\dot{q}_i$, and time, Eqs. (1.20,1.21) provide $p_i$ and $\dot{p}_i$ as functions of the $q_i$, $\dot{q}_i$, and time. Moreover, from (1.18) one finds

$$[q_i]\,[p_i] = [L]\,[t], \tag{1.22}$$

where the brackets indicate the dimensions. In the rectangular coordinates the Lagrangian has the dimension of an energy; as the coordinate transformation leaves the dimensions of the transformed function unchanged, from (1.22) it follows that the product of two canonically-conjugate variables has always the dimension of an action (energy × time). In the example of section 1.1 it is $p_i = m\dot{x}_i$ and the property (1.22) is easily checked.

## 1.3 Hamiltonian Function

Eq. (1.20) provides $p_i$ in terms of the $q_i$, $\dot{q}_i$, and time:

$$p_i = p_i(q_1, \dot{q}_1, \ldots, q_s, \dot{q}_s, t), \qquad i = 1, \ldots, s. \tag{1.23}$$

If Eqs. (1.23) are invertible, from them one expresses the $\dot{q}_i$ in terms of the $q_i$, $p_i$, and time:

$$\dot{q}_i = \dot{q}_i(q_1, p_1, \ldots, q_s, p_s, t), \qquad i = 1, \ldots, s. \tag{1.24}$$

Define

$$H = \sum_{i=1}^{s} p_i \dot{q}_i - L \tag{1.25}$$

where, due to (1.24), the generalized velocities are expressed in terms of the generalized coordinates $q_i$ and conjugate momenta $p_i$. It follows

$$H = H(q_1, p_1, \ldots, q_s, p_s, t). \tag{1.26}$$

Function $H$ is the *Hamiltonian function* of the system of particles. It is easily found that the following hold (*Hamilton equations*):

Newtons method ⇔ Lagrangian ⇔ Hamiltonian

This can be extended to cases where we don't have particles i.e Fields ex: EMF

Hamiltonian operator in QM is the counter part of the Hamiltonian function of QM.



Force

velocity ⇐ $\dot{q}_i = \frac{\partial H}{\partial p_i}$, $\dot{p}_i = -\frac{\partial H}{\partial q_i}$, $i = 1, \ldots, s$, (1.27)

that are equivalent to the Lagrange equations (1.18) for the description of the dynamics of the system.

They are 1st order ODE And have $2s$ degrees of freedom ∴ we have $2s$ no. of eqn

The number of initial conditions necessary to solve the problem is $2s$. In the Lagrangian formulation one must solve $s$ second-order equations, whose initial conditions are typically $q_i(t=0)$ and $\dot{q}_i(t=0)$, $i = 1, \ldots, s$. In the Hamiltonian formulation one must solve $2s$ first-order equations, whose initial conditions are typically $q_i(t=0)$ and $p_i(t=0)$, $i = 1, \ldots, s$.

In lagrangian can we had 2nd order ODE.

If the forces derive from a potential energy $V$, using the rectangular coordinates one finds

$H = T + V$, for a conservative system (1.28)

where, in case of a single particle,

$$T = \frac{1}{2m} p^2 = \frac{1}{2m} (p_1^2 + p_2^2 + p_3^2), \qquad V = V(\mathbf{r}). \tag{1.29}$$

although T & V depend on time but their sum total energy equal to Hamiltonian is constant.

In order to calculate Hamiltonian the Lagrangian function must exist

Instead, in case of a system of $N$ particles one has

$$T = \sum_{j=1}^{N} \frac{1}{2m_j} p_j^2, \qquad V = V(\mathbf{r}_1, \ldots, \mathbf{r}_N). \tag{1.30}$$

For instance, applying the Hamilton equations (1.27) to (1.29) yields, from the first one,

$$\dot{q}_i = \frac{\partial H}{\partial p_i} = \frac{\partial T}{\partial p_i} = \dot{x}_i, \tag{1.31}$$

and from the second one,

$$\dot{p}_i = -\frac{\partial H}{\partial q_i} = -\frac{\partial V}{\partial q_i} = F_i. \tag{1.32}$$

This again is Newton's Law

On the other hand, as remarked in section 1.2, in this case one also has $p_i = m\dot{x}_i$. Inserting the latter relation into the first of (1.31) and combining the result with (1.32) yields Newton's law $m\ddot{x}_i = F_i$. The same result is found when a system of $N$ particles is considered instead of a single particle.

Note that, in the example where the force derives from a potential energy, the expression of the kinetic energy $T$ in the Hamiltonian formulation in rectangular coordinates (the first of (1.29) in the case of a single particle) is deduced from that of the Lagrangian formulation, namely, from the first of (1.5), simply by replacing $u^2$ with $p^2/m^2$. Similarly, for a system of $N$ particles one goes from (1.13) to (1.30) by replacing $u_j^2$ with $p_j^2/m_j^2$.

## 1.4 Constants of Motion

Another property that holds when the forces derive from a potential energy is that, if the latter has no explicit dependence on time, the Hamiltonian function has no explicit dependence on time as well. In addition, the Hamiltonian function is a constant of motion. Namely, using the rectangular coordinates, the following holds:

$$H = H(x_1, p_1, \ldots, x_s, p_s) = \text{cost}. \tag{1.33}$$

In other terms, regardless of the fact that in general all coordinates and momenta vary with time, they combine with each other within the Hamiltonian function in such a way that the latter remains constant.

Taking the case of a single particle by way of example, the property above is easily found by considering the work exerted by the force $\mathbf{F}$ during an elementary interval of time $\mathrm{d}t$ starting at the time $t$. In such an interval the particle moves by an elementary displacement $\mathrm{d}\mathbf{r} = \mathbf{u}\,\mathrm{d}t$, where $\mathbf{u}$ is the particle velocity at time $t$. The work is given by

$$\mathbf{F} \cdot \mathrm{d}\mathbf{r} = \mathbf{F} \cdot \mathbf{u}\,\mathrm{d}t = m\,\dot{\mathbf{u}} \cdot \mathbf{u}\,\mathrm{d}t\,, \tag{1.34}$$

where the last form derives from $\mathbf{F} = m\,\mathbf{a}$ after expressing the acceleration as the derivative of the velocity. On the other hand, the term of the right of (1.34) is given by

$$m\,\dot{\mathbf{u}} \cdot \mathbf{u}\,\mathrm{d}t = \frac{m}{2}\,\frac{\mathrm{d}(\mathbf{u}\cdot\mathbf{u})}{\mathrm{d}t}\,\mathrm{d}t = \frac{1}{2}\,m\,\mathrm{d}(u^2) = \frac{1}{2\,m}\,\mathrm{d}(p^2)\,. \tag{1.35}$$

In turn, remembering that the components of the force are the negative derivatives of the potential energy, the term on the left of (1.34) reads

$$\mathbf{F} \cdot \mathrm{d}\mathbf{r} = -\,\mathrm{grad}\,V \cdot \mathrm{d}\mathbf{r} = -\mathrm{d}V\,. \tag{1.36}$$

namely, it represents the variation of the potential energy of the particle due to the displacement $\mathrm{d}\mathbf{r}$. As a consequence, (1.34) implies that the variation of the quantity $p^2/(2\,m) + V$ in the time interval $\mathrm{d}t$ is zero, namely, the sum $p^2/(2\,m) + V = E$ is a constant of motion. Such a constant is the total energy.

Still with reference to a force derived from a potential energy, the following observation is noteworthy. The particle motion is determined by the force, whereas the potential energy is an auxiliary function by which the law of motion is expressed in a more general form. As the force is the gradient of the potential energy, the latter has an arbitrary additive constant in it (on the contrary, there is no additive constant in the kinetic energy). As a consequence, also the total energy $H = T + V = E$ has an arbitrary additive constant in it. Such a constant has no effect upon the description of the dynamics, because the latter is obtained from the Hamilton equations (1.27), whence the constant is eliminated by the calculation of the partial derivatives. In any case, when a dynamical problem is investigated, the additive constant is intrisically fixed when the functional dependence of the potential energy $V$ on the coordinates

is prescribed. This, in turn, fixes the value of the total energy $E$ which, as shown above, is a constant of motion.

For a system possessing $s$ degrees of freedom the total number of constants of motion is $2s$, that is, the number of initial conditions that must be prescribed to solve the dynamical equations. In turn, the initial conditions are the values of the generalized coordinates and momenta at $t = 0$. One may combine the initial conditions into expressions $\alpha = \alpha(q_1, q_2, \ldots, p_{s-1}, p_s)$ that keep the initial value also for $t \neq 0$ as the invidual values of $q_1, \ldots, p_s$ evolve with time. Obviously there are more than $2s$ combinations with such a property. Out of them, however, $2s$ at most may be independent from each other, and are the constants of motion of the system. Examples of constants of motion are the total momentum and the total angular momentum of an isolated system, and the total energy of a conservative system.

## 1.5 Consequences of the Hamilton Equations

Consider a system of particles with $s$ degrees of freedom, described by $s$ pairs of conjugate canonical variables $q_i$, $p_i$. Then, let $\mathbf{c}$ be a vector made of $2s$ components, such that the first $s$ components are the generalized coordinates $q_1, \ldots, q_s$ and the remaining $s$ components are the momenta $p_1, \ldots, p_s$:

$$\mathbf{c} = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ p_{s-1} \\ p_s \end{bmatrix}. \tag{1.37}$$

This vector depends on time provides information about state of the sys in time

Vector $\mathbf{c}$ represents the state of the system at time $t$. It may be thought of as a point belonging to a $2s$-dimensional space. In such a space a single point represents the system as a whole and, as times evolves, the point representing the system follows a trajectory belonging to that space.

The space of $\mathbf{c}$ is called *phase space*; in order to distinguish it from other types of phase spaces, it is also called *$\gamma$ space* or *Gibbs space*. Due to the Hamilton equations (1.27), the time derivative of $\mathbf{c}$ reads

$$\dot{\mathbf{c}} = \begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \\ \vdots \\ \dot{p}_{s-1} \\ \dot{p}_s \end{bmatrix} = \begin{bmatrix} \partial H/\partial p_1 \\ \partial H/\partial p_2 \\ \vdots \\ -\partial H/\partial q_{s-1} \\ -\partial H/\partial q_s \end{bmatrix}. \tag{1.38}$$

Besides a possible explicit dependence on time, the Hamiltonian function depends on the canonical variables $q_i$, $p_i$. For this reason, vector $\dot{\mathbf{c}}$ depends on the canonical

variables as well. Its divergence in the $\gamma$ space then reads

$$\mathrm{div}_\gamma \dot{\mathbf{c}} = \frac{\partial \dot{c}_1}{\partial c_1} + \frac{\partial \dot{c}_2}{\partial c_2} + \cdots \tag{1.39}$$

Using (1.37,1.38) yields

$$\mathrm{div}_\gamma \dot{\mathbf{c}} = \sum_{i=1}^{s} \left( \frac{\partial \dot{q}_i}{\partial q_i} + \frac{\partial \dot{p}_i}{\partial p_i} \right) = \sum_{i=1}^{s} \left( \frac{\partial^2 H}{\partial p_i \partial q_i} - \frac{\partial^2 H}{\partial q_i \partial p_i} \right) = 0 \,. \tag{1.40}$$

In fluid dynamics, Eq. (1.40) states the incompressibility condition of the fluid. From the above result it also follows that, given a scalar function defined in the $\gamma$ space, $\sigma = \sigma(\mathbf{c}, t)$, the application of a known vector identity combined with (1.40) yields

$$\mathrm{div}_\gamma(\sigma \dot{\mathbf{c}}) = \sigma \, \mathrm{div}_\gamma \dot{\mathbf{c}} + \dot{\mathbf{c}} \cdot \mathrm{grad}_\gamma \sigma = \dot{\mathbf{c}} \cdot \mathrm{grad}_\gamma \sigma \,. \tag{1.41}$$

The total derivative of $\sigma$ with respect to time then reads

$$\frac{\mathrm{d}\sigma}{\mathrm{d}t} = \frac{\partial \sigma}{\partial t} + \sum_{i=1}^{s} \frac{\partial \sigma}{\partial c_i} \frac{\mathrm{d}c_i}{\mathrm{d}t} = \frac{\partial \sigma}{\partial t} + \dot{\mathbf{c}} \cdot \mathrm{grad}_\gamma \sigma \,. \tag{1.42}$$

Combining (1.41) and (1.42) yields

$$\frac{\mathrm{d}\sigma}{\mathrm{d}t} = \frac{\partial \sigma}{\partial t} + \mathrm{div}_\gamma(\sigma \dot{\mathbf{c}}) \,. \tag{1.43}$$

The scalar product of (1.41) of may be recast in a different form using the definition of the *Poisson parentheses* of two functions $a$, $b$ that depend on the conjugate variables $q_i$, $p_i$ and on time:

$$\{a, b\} = \sum_{i=1}^{s} \left( \frac{\partial a}{\partial q_i} \frac{\partial b}{\partial p_i} - \frac{\partial a}{\partial p_i} \frac{\partial b}{\partial q_i} \right) . \tag{1.44}$$

In fact, using (1.27),

$$\dot{\mathbf{c}} \cdot \mathrm{grad}_\gamma \sigma = \sum_{i=1}^{s} \left( \dot{q}_i \frac{\partial \sigma}{\partial q_i} + \dot{p}_i \frac{\partial \sigma}{\partial p_i} \right) = \sum_{i=1}^{s} \left( \frac{\partial H}{\partial p_i} \frac{\partial \sigma}{\partial q_i} - \frac{\partial H}{\partial q_i} \frac{\partial \sigma}{\partial p_i} \right) . \tag{1.45}$$

Combining (1.45) with (1.41), (1.42), (1.43), and (1.44) provides

$$\frac{\mathrm{d}\sigma}{\mathrm{d}t} = \frac{\partial \sigma}{\partial t} + \{\sigma, H\} \,. \tag{1.46}$$

By way of example, using (1.46) with $\sigma = H$ yields

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \frac{\partial H}{\partial t} \,, \tag{1.47}$$

as obviously it is $\{H,H\}=0$. This shows that the Hamiltonian function is a constant of motion if it does not depend explicitly on time.

## 1.6 Hamilton-Jacobi Equation

Besides the Newton law (1.1), the Lagrange equations (1.11,1.18), and the Hamilton equations (1.27), the dynamics of a system having $s$ degrees of freedom may also be described by means of the *Hamilton-Jacobi equation*, that is, a partial-differential equation whose unknown $S$ is called the *Hamilton principal function*.

A short-hand derivation of the Hamilton-Jacobi equation starts from the definition (1.25) of the Hamiltonian function, which can be recast as $L=\sum_{i=1}^{s} p_i\dot{q}_i - H$; then, one defines $S$ such that $\mathrm{d}S/\mathrm{d}t = L$, whence

$$\mathrm{d}S = \sum_{i=1}^{s} p_i\dot{q}_i\,\mathrm{d}t - H\,\mathrm{d}t = \sum_{i=1}^{s} p_i\,\mathrm{d}q_i - H\,\mathrm{d}t \tag{1.48}$$

The form of (1.48) is that of a total differential, such that

$$p_i = \frac{\partial S}{\partial q_i}\,, \qquad H = -\frac{\partial S}{\partial t}\,. \tag{1.49}$$

In conclusion, function $S$ depends on the $s$ generalized coordinates $q_1,\ldots,q_s$ and on time $t$. As shown by (1.49), the equation for $S$ is constructed starting from the Hamiltonian function $H$ of the system, which is transformed into a new function $H_S$ by replacing each momentum $p_i$ with the derivative $\partial S/\partial q_i$:

$$H(q_i,p_i,t) \quad \Leftarrow \quad H_S\left(q_i,\frac{\partial S}{\partial q_i},t\right). \tag{1.50}$$

By construction, function $H_S$ depends on the coordinates and time both explicitly and implicitly ($S$ may depend on time as well). Due to the second relation in (1.49), the Hamilton-Jacobi equation then reads

$$\frac{\partial S}{\partial t} + H_S\left(q_1,\ldots,q_s,\frac{\partial S}{\partial q_1},\ldots,\frac{\partial S}{\partial q_s},t\right) = 0\,. \tag{1.51}$$

As (1.51) is a first-order differential equation with $s+1$ variables, its general solution has $s+1$ integration constants in it, say, $\alpha_1,\ldots,\alpha_{s+1}$. On the other hand only the derivatives of $S$ are present in (1.51) whence, if $S$ is a solution, then $S+\text{const}$ is also a solution; it follows that one of the integration constants, say, $\alpha_{s+1}$, must be an additive constant on $S$. As shown below, the calculations based on $S$ never involve $S$ itself, but only its derivatives; as a consequence one may let $\alpha_{s+1}=0$, and $S$ may be considered a function of the generalized coordinates, time, and $s$ integration constants:

$$S = S(q_1, \ldots, q_s, \alpha_1, \ldots, \alpha_s, t) \,. \tag{1.52}$$

The Hamilton-Jacobi theory shows that, once $S$ has been determined, the following hold:

$$p_i = \frac{\partial S}{\partial q_i} \,, \qquad \beta_i = \frac{\partial S}{\partial \alpha_i} \,, \qquad i = 1, \ldots, s \,, \tag{1.53}$$

where $\beta_1, \ldots, \beta_s$ turn out to be constants of motion. Eqs. (1.53) may now be used to determine the time evolution of each $q_i$, $p_i$. To begin, one uses the $2s$ initial conditions in the first of (1.53) to obtain a set of $s$ equations of the form

$$p_i(0) = \frac{\partial S(q_1(0), \ldots, q_s(0), \alpha_1, \ldots, \alpha_s, 0)}{\partial q_i} \,. \tag{1.54}$$

Inverting (1.54) provides the values of $\alpha_1, \ldots, \alpha_s$. Such values are now inserted into the second of (1.53), which are again calculated by letting $t = 0$. This provides the values of $\beta_1, \ldots, \beta_s$.

After the calculation of the constants has been completed, the second of (1.53) may be considered at any time $t$. They provide a set of $s$ relations intrinsically relating $q_1, \ldots, q_s$ with time. Inverting such relations provides the time evolution of each coordinate $q_i$. Finally, inserting the time dependence of all $q_i$ into the first of (1.53) provides the time evolution of $p_1, \ldots, p_s$.

The above shows that the knowledge of the Hamilton principal function provides the dynamics of the conjugate coordinates of the system. It follows that the Hamilton-Jacobi equation is equivalent to the other theories for the description of the system dynamics. One may also observe that the method of solution involves a set of $2s$ constants that are a combination of the $2s$ initial conditions $q_1(0), \ldots, p_s(0)$.

## 1.7 Statistical Ensemble

If the Hamiltonian function and the initial conditions of the system are known, the trajectory of the point $\mathbf{c}$ representing the state of the system in the $\gamma$ space is completely determined. On the other hand, it may happen that the information about the system is insufficient for a complete specification of a precise state; in this case, there are several points $\mathbf{c}$ that describe the state of the system in a way that is compatible with the available information: the set of such points is called *ensemble*.

By way of example, let the system be made of a single particle constrained to a one-dimensional motion, so that the canonical variables reduce to the single pair $q, p$. Assuming that the $q$ coordinate indicates the distance of the particle from the origin, and that the force derives from a potential energy of the linear harmonic-oscillator type, namely, $V = m\omega^2 q^2/2$ with $\omega$ the angular frequency, the Hamiltonian function takes the form

The state vector is very huge and we don't know how to calculate Trajectory of the s/s in the Phase space!

eg: LHO in the classical case (linear harmonic oscillator)

$$H(q,p) = \frac{1}{2m}p^2 + \frac{m\omega^2}{2}q^2 = E\,, \qquad \omega = \sqrt{c/m}\,. \tag{1.55}$$

The trajectory in the $\gamma$ space $q, p$ is the ellipsis

$$\frac{p^2}{p_E^2} + \frac{q^2}{q_E^2} = 1\,, \qquad p_E = \sqrt{2mE}\,, \quad q_E = \frac{1}{\omega}\sqrt{\frac{2E}{m}}\,, \tag{1.56}$$

whose area is

$$\pi\, p_E\, q_E = 2\pi\frac{E}{\omega} = \frac{E}{\nu}\,, \tag{1.57}$$

with $\nu = \omega/(2\pi)$ the frequency.

If parameters $m, \omega$ are prescribed, the dynamics of the particle is uniquely determined by the initial conditions $p(0)$, $q(0)$. The total energy is determined as well, and is calculated by simply inserting the initial conditions into (1.55).

If, on the contrary, the initial conditions are not known, whereas the total energy is prescribed, the representative point of the particle in the $\gamma$ space at $t = 0$ may be any point compatible with the given set $m, \omega, E$; that is, any point of the ellipsis (1.55) may represent equally well the state of the particle at $t = 0$. This implies that at any other time the only possible information about the state of the particle is that the representative point belongs to the ellipsis (1.55). In this case the ensemble is the set of points of the ellipsis.

Finally, given $m$ and $\omega$, consider the case where the initial conditions are not known, while it is known that the total energy belongs to the interval $E_1 < E < E_2$, with $E_1 \geq 0$. This constraint on the total energy defines a strip of the $\gamma$ space; such a strip is limited by the ellipses of the form (1.55) determined by the values $E_1$ and $E_2$. Any point internal to the strip may represent equally well the state of the particle at $t = 0$. This implies that at any other time the only possible information about the state of the particle is that the representative point belongs to the strip. In this case the ensemble is the set of points of the strip.

It is important to note that the points of the ensemble do not interact with each other; in fact, each point is a replica of the same system. Also, each point of the ensemble evolves in time according to the dynamical laws; there is no possibility for a point to disappear or to be generated at some instant of time, because this would imply the annihilation or creation of the system under consideration. Finally, the trajectories of two different points can not cross each other at the same instant of time; in fact, if that happened, at the instant when the crossing occurs the two points would be described by the same Hamiltonian function, and the initial conditions for the subsequent motion would also be the same. From that instant on, the two trajectories would then be identical. On the other hand, as the dynamical laws work in the same way also when time is reversed, the two trajectories would also be identical backward in time, which contradicts the hypothesis that the two points of the ensemble were originally different.

## 1.8 Density of Distribution in the $\gamma$ Space

Let $\mathrm{d}c = \mathrm{d}q_1\,\mathrm{d}q_2\,\ldots\,\mathrm{d}p_{s-1}\,\mathrm{d}p_s$ be an elementary volume of the $\gamma$ space. The number $\mathrm{d}M$ of systems belonging to the elementary volume at time $t$ may be expressed by means of the density $\rho(\mathbf{c},t)$, such that

$$\mathrm{d}M = \rho(\mathbf{c},t)\,\mathrm{d}c\,, \qquad \rho \geq 0\,. \tag{1.58}$$

The integral of $\rho$ over the whole $\gamma$ space provides at all times the total number of elements of the ensemble:

$$M = \int_{\infty} \rho(\mathbf{c},t)\,\mathrm{d}c\,. \tag{1.59}$$

Only densities whose integral (1.59) converges will be considered; as there is no possibility of annihilation or generation of points, the density must fulfill the continuity equation

$$\frac{\partial \rho}{\partial t} + \mathrm{div}_{\gamma}(\dot{\mathbf{c}}\,\rho) = 0\,. \tag{1.60}$$

Using (1.43) provides

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = 0\,, \tag{1.61}$$

that expresses the so-called *conservation of density in phase* or, also, the *Liouville theorem*. Using (1.41) and (1.46) one finds other forms of (1.60)

$$\frac{\partial \rho}{\partial t} + \dot{\mathbf{c}} \cdot \mathrm{grad}_{\gamma}\rho = 0\,, \qquad \frac{\partial \rho}{\partial t} + \{\rho, H\} = 0\,. \tag{1.62}$$

The integral in (1.59) may be calculated by performing first the integration over $p_1, \ldots, p_s$, so that

$$M = \int_{\infty} D(\mathbf{q},t)\,\mathrm{d}q\,, \qquad D = \int_{\infty} \rho(\mathbf{q},\mathbf{p},t)\,\mathrm{d}p\,. \tag{1.63}$$

The function $D$ defined by (1.63) has the property that its integral over $\mathbf{q}$ provides the number of elements of the ensemble. As a consequence, $D$ is the density of the elements in the $q$ space. Conversely, from

$$M = \int_{\infty} Q(\mathbf{p},t)\,\mathrm{d}p\,, \qquad Q = \int_{\infty} \rho(\mathbf{q},\mathbf{p},t)\,\mathrm{d}q \tag{1.64}$$

one derives the density $Q$ of the elements in the $p$ space.

From the definitions (1.58,1.59) one may define the normalized density in the $\gamma$ space

$$P(\mathbf{c},t) = \frac{1}{M}\,\rho(\mathbf{c},t)\,, \tag{1.65}$$

that fulfills at all times the normalization condition

$$\int_{\infty} P(\mathbf{c},t)\,\mathrm{d}c = 1\,. \tag{1.66}$$

The statistical average of any dynamical function $A(\mathbf{c})$ over the $\gamma$ space is defined as

$$\mathrm{Av}_{\gamma}[A](t) = \int_{\infty} A(\mathbf{c})P(\mathbf{c},t)\,\mathrm{d}c\,. \tag{1.67}$$

Only functions $A$ such that the integral in (1.67) converges are considered.

## 1.9 Statistical Equilibrium

The condition of the ensemble of being in statistical equilibrium is defined as the condition where the density does not depend explicitly on time at all points of the $\gamma$ space.

A simple case where the equilibrium condition is fulfilled is $\rho = \text{const}$. Another more general case is when the density has the form $\rho = \rho(\alpha)$, where $\alpha = \alpha(\mathbf{c})$ is a constant of motion of the system. In fact, remembering that by construction a constant of motion has no explicit dependence on time in it, one finds from (1.46)

$$\frac{\mathrm{d}\alpha}{\mathrm{d}t} = 0 \iff \{\alpha, H\} = 0\,, \tag{1.68}$$

whence

$$\{\rho, H\} = \frac{\mathrm{d}\rho}{\mathrm{d}\alpha}\{\alpha, H\} = 0\,. \tag{1.69}$$

Combining (1.69) with the second of (1.62) shows that, if the density depends on the canonical coordinates in the form $\rho = \rho(\alpha)$, then it does not depend explicitly on time. According to the definition it is an equilibrium density.

Some of the expressions shown in section 1.8 may be recast in a different form, which is more useful for the application of the statistical concepts to the quantum case. For this, it is sufficient to assume that the vector $\mathbf{c}$ that describes the members of the ensemble spans a countable set of values instead of varying continuously in the $\gamma$ space. The members of the ensemble will then be identified with the vectors $\mathbf{c}_1, \mathbf{c}_2, \ldots$.

The use of discrete values for the state of the system is related to the uncertainty principle. For each pair of conjugate variables $q_i, p_i$ the uncertainty about the simultaneous knowledge of such variables is of the order of the Planck constant $h$. It is then sensible to introduce in the $q_i, p_i$ plane a tessellation; this is accomplished by subdividing the $q_i$ axis into equal intervals $\Delta q$, and the $p_i$ axis into equal intervals $\Delta p$, such that the plane is tessellated into equal cells of area $\Delta q\,\Delta p = h$.

For a system with $s$ degrees of freedom the procedure must be repeated $s$ times, to eventually yield a tessellation of the $\gamma$ space into equal cells of volume $h^s$. As the

In Quantum Mechanics can we have a problem

i.e. the definition $dC = dq_1 dq_2 \ldots dp_{s-1} dp_s$ does not apply to QM

uncertainty principle prevents one from considering the states belonging to the same cell as different, it is sufficient to identify the state with the cell, and to label the cell by the index of the **c** vector pointing to the cell's center. Letting

$$\rho_e = \rho(\mathbf{c}_e, t)\,, \qquad e = 1, 2, \ldots \tag{1.70}$$

Eqs. (1.58,1.59) must be replaced with

$$M_e(t) = \rho_e h^s\,, \qquad M = \sum_{e=1}^{\infty} M_e = h^s \sum_{e=1}^{\infty} \rho_e\,, \tag{1.71}$$

while definition (1.65) of the normalized density becomes

$$P_e(t) = \frac{M_e}{M} = \frac{\rho_e}{\sum_e \rho_e}\,, \qquad 0 \le P_e \le 1\,, \tag{1.72}$$

that fulfills at all times the normalization condition

$$\sum_{e=1}^{\infty} P_e(t) = 1\,. \tag{1.73}$$

In the discrete case considered in this section, and in contrast to the continuous case of section 1.8, the normalized density is dimensionless; also, due to the second of (1.72), it is a probability. The statistical average of any dynamical function $A_e = A(\mathbf{c}_e)$ is defined as

$$\mathrm{Av}_\gamma[A](t) = \sum_{e=1}^{\infty} A_e P_e(t)\,. \tag{1.74}$$

lecture - 36

## 1.10 Distribution in the $\mu$ Space — Entropy

The $\gamma$ space defined in Sect. 1.8 is a $(2s)$-dimensional space, with $s$ the total number of degrees of freedom of the system under consideration; as a consequence, a point **c** of the $\gamma$ space represents the system as a whole. In parallel to the $\gamma$ space it is often convenient to consider another space, called *$\mu$ space*,[1] whose dimension is twice the number of degrees of freedom of the single particle of the system; it follows that for a system made of point-like particles the dimension of the $\mu$ space is $2 \times 3 = 6$, and so on. Considering this case, one repeats for the $\mu$ space the tessellation shown above in the case of the $\gamma$ space; the difference is that here the volume of each cell is $h^3$. Remembering that the total number of particles of the system is $N$, let $N_r(t)$ be the number of particles that at time $t$ belong to the $r$-th cell, $r = 1, 2, \ldots$; then, one can introduce another type of average, different from the ensemble average given by (1.74); still considering a dynamic function $A$, the average over the particles of a

[1] Notation $\mu$ derives from the initial of "molecule"; in turn, notation $\gamma$ derives from the initial of "gas".

single system is defined as

$$\mathrm{Av}_\mu[A] = \frac{1}{N}\sum_{r=1}^{\infty} A_r N_r\,. \tag{1.75}$$

In the sum of (1.75), the function to be averaged is marked with the cell index $r$ because, in general, such a function depends on the canonical coordinates; thus, it is calculated using the values of such coordinates associated to the center of the cell. The ratio $0 \le N_r/N \le 1$ is the fraction of particles that belongs to the $r$-th cell; the fraction will be indicated here with $P_r(t)$ (note that the meaning of $P_r$ is different from that of $P_e$ given by (1.72)).

An important example of statistical average is

$$H_B(t) = \mathrm{Av}_\mu[\log P_r] = \sum_{r=1}^{\infty} P_r \log P_r \le 0\,, \qquad S = -k_B N H_B\,, \tag{1.76}$$

where $k_B = 1.38 \times 10^{-23}\ \mathrm{J\,K^{-1}}$ is the Boltzmann constant. Quantity $S \ge 0$ defined by (1.76) is the *entropy* of the system.[2] Note that in some cells it is $P_r = 0$; however, in (1.76) the logarithm of $P_r$ appears only in the product $P_r \log P_r$, which vanishes when $P_r \to 0$. Similarly, the product vanishes for $P_r = 1$, whereas it is negative for $0 < P_r < 1$. It follows that $S$ is by construction a non-negative quantity. Observing that $N > 0$, so that the condition $P_r = 0$ for all cells is ruled out, the only condition for $S$ to vanish occurs when all particles belong to the same cell. The graph of $P_r \log P_r$ is shown in Fig. 1.1.

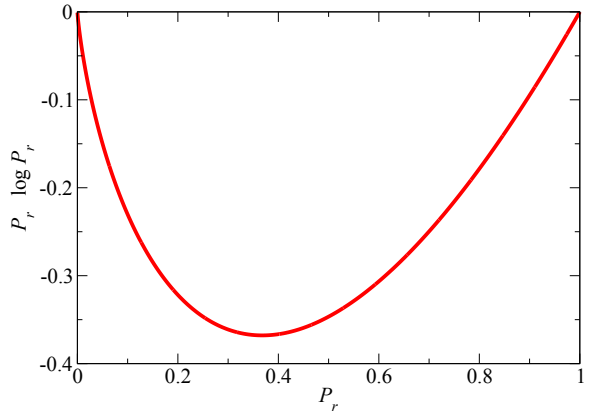



**Fig. 1.1** Form of the $P_r \log P_r$ function used in (1.76) to define the entropy of a system of particles.

To examine the time evolution of $H_B$ one calculates the derivative

[2] In turn, $H_B$ is called *Boltzmann's* H *function*.

$$\frac{\mathrm{d}H_B}{\mathrm{d}t} = \sum_r (1 + \log P_r)\, \frac{\mathrm{d}P_r}{\mathrm{d}t}. \tag{1.77}$$

The above relation merely shifts the problem of calculating $\mathrm{d}H_B/\mathrm{d}t$ to that of calculating $\mathrm{d}P_r/\mathrm{d}t$; in fact, the latter derivative can in principle be calculated by solving the equations of motion of all particles of the system. The difficulty of the problem is unsurmountable for systems made of a large number of particles. To avoid the difficulty one resorts to a probabilistic approach, that derives $\mathrm{d}P_r/\mathrm{d}t$ from a balance equation: let $U_{sr} \geq 0$ be the unconditional probability per unit time that a particle makes a transition from the $s$th to the $r$th cell, $s \neq r$. Thus, the number of particles that in the unit time make a transition from the $s$th to the $r$th cell is $U_{sr} N_s$; adding up the above over all cells different from $r$ provides the increase per unit time in the population of the $r$th cell. In parallel to this, there are particles that make a transition from the $r$th cell to any different cell, thus contributing to the decrease of $N_r$; letting $U_{rs}$ be the unconditional probability per unit time of a transition from the $r$th to the $s$th cell, the balance equation at time $t$ reads

$$\frac{\mathrm{d}N_r}{\mathrm{d}t} = \sum_s U_{sr} N_s - \sum_s U_{rs} N_r\,. \tag{1.78}$$

In the sums at the right hand side of (1.78) the prescription $s \neq r$ is not necessary because the two summands corresponding to $s = r$ cancel each other. The quantity $U_{sr}$ depends on the interactions of the particles among each other and with external perturbing agents. If the system is assumed to be isolated, $U_{sr}$ is determined only by the mutual interactions of the particles. The calculation shows that in this case $U_{sr}$ is independent of time; also, it can be shown that $U_{sr}$ is invariant upon the exchange of the indices,[3] namely, $U_{rs} = U_{sr}$. Using this result in (1.78) and dividing both sides by $N$ yields, due to $P_r = N_r/N$,

$$\frac{\mathrm{d}P_r}{\mathrm{d}t} = \sum_s U_{sr}\,(P_s - P_r)\,. \tag{1.79}$$

Inserting the above into (1.77),

$$\frac{\mathrm{d}H_B}{\mathrm{d}t} = \sum_{rs} (1 + \log P_r)\, U_{sr}\,(P_s - P_r)\,. \tag{1.80}$$

Now, considering that both indices in (1.80) span over all cells of the $\mu$ space, the derivative $\mathrm{d}H_B/\mathrm{d}t$ is equally well represented by an expression derived from (1.80) after exchanging $r$ and $s$. Observing that the order of the indices in the sum and in $U_{sr}$ is immaterial, one finds

$$\frac{\mathrm{d}H_B}{\mathrm{d}t} = \sum_{rs} (1 + \log P_s)\, U_{sr}\,(P_r - P_s)\,. \tag{1.81}$$

[3] The demonstration is carried out in the quantum case using the first-order perturbation method.

Finally, $\mathrm{d}H_B/\mathrm{d}t$ is given another expression, obtained as half the sum of (1.80) and (1.81); such a procedure has the advantage of eliminating the unity at the right hand side, and yields

$$\frac{\mathrm{d}H_B}{\mathrm{d}t} = \frac{1}{2}\sum_{rs}(\log P_r - \log P_s)\, U_{sr}\,(P_s - P_r)\,. \tag{1.82}$$

Remembering that $U_{sr}$ is non negative, and that the logarithm is a monotonic function of the argument, one draws the conclusion $\mathrm{d}H_B/\mathrm{d}t \leq 0$, whence $\mathrm{d}S/\mathrm{d}t \geq 0$. This result, called *Boltzmann's* H *theorem*, shows that the entropy of an isolated system may only increase. This implies that if a system is initially set in a condition described by a non-equilibrium distribution function, and the external forces are removed, then the initial distribution can not be stationary: an equilibration process occurs, that brings the distribution to the equilibrium one.

## 1.11 Entropy Paradox

It is known that the $H$-theorem of Boltzmann brings about two apparent paradoxes: the first one is that the theorem contains irreversibility, because any initial distribution function, different from the equilibrium one, evolves towards the equilibrium distribution when the external forces are removed, whereas an opposite evolution never occurs. This outcome is in contrast with the symmetry of the laws of mechanics with respect to time reversal. The second paradox is the violation of Poincaré's time recurrence, which states that every finite mechanical system returns to a state arbitrarily close to the initial one after a sufficiently long time (called *Poincaré cycle*); this is forbidden by the $H$-theorem, that prevents entropy from decreasing back to the initial value.

A qualitative insight into the question is given by a simple model, called *Kac's ring model*, reported in [68] and taken from [31] (Fig. 1.2). In the model, $N$ objects are uniformly distributed over a circle, so that at time $t = 0$ each object is ascribed to a specific arc. The objects have two possible states, say, either "0" or "1". The time variable is discrete so that, when time evolves from $k\Delta t$ to $(k+1)\Delta t$, $k = 0, 1, 2, \ldots$, each object moves clockwise from the arc it occupied at time $k\Delta t$ to the next arc. A number $n < N$ of markers is present along the circle: specifically, the markers positions are at the junctions between two neighboring arcs. The objects that cross the position of a marker change the state from "0" to "1" or vice versa; those that do not cross the position of a marker keep their state.

Given the number of objects and markers, the initial state of each object, and the markers' positions along the circle, one wants to investigate the time evolution of the states. Such an evolution is obviously time reversible and fulfills Poincaré's time recurrence; in fact, the set of objects goes back into the initial condition after $N$ time steps if $n$ is even, and after $2N$ time steps if $n$ is odd.

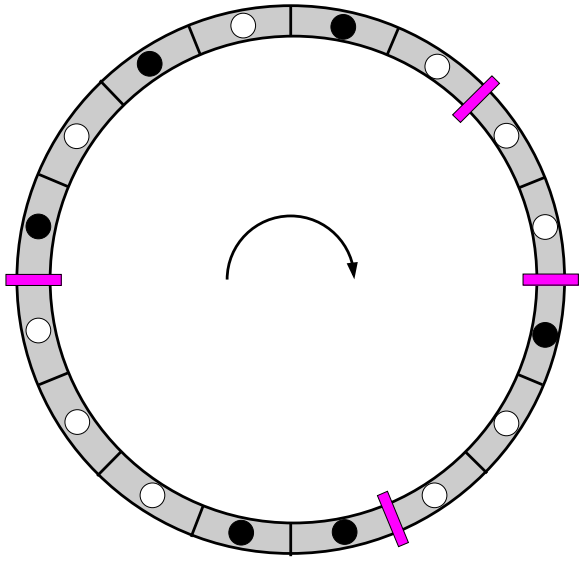


**Fig. 1.2** Illustration of the Kac-ring model. Here $N = 16$ objects are distributed over a circle. The state of an object is indicated by the black or white color. At each time step the objects move clockwise by a $2\pi/16$ arc. Four markers are present, indicated by the magenta blocks.

Providing the time evolution of the individual object's state is in fact a microscopic description of the system; as remarked above, such a description becomes impossible when the number of objects in the system is large. A less detailed, macroscopic description of the Kac ring consists, for instance, in providing the time evolution of the number of "0" states. However, the outcome of the latter analysis seems to indicate that an irreversible process takes place; for instance, Fig. 1.3 shows a computer calculation of the time evolution of the number of "0" states in a sample made of $N = 3{,}600$ objects, which at time $t = 0$ were all set to "0". The markers of the sample are $n = 12$, and the number of time steps is smaller than $N$. The curve tends to decrease and, after some fluctuations, stabilizes at a constant value.

On the other hand, a similar calculation using a number of time steps larger than the number of objects shows that the stabilization at or around a constant value is eventually lost: the system fulfills Poincaré's time recurrence and recovers the initial condition (Fig. 1.4). Such an outcome is not detectable in real many-body systems, because the Poincaré cycle is enormously long with respect to the typical time scales of experiments.[4]

A more detailed analysis of the dilemma shows that a key point in the demonstration of the Boltzmann $H$ theorem is the renunciation of using the equations of Mechanics

[4] A crude estimate of the Poincaré cycle yields $\sim \exp(N)$, with $N$ the total number of molecules in the system [29, Sect. 4.5]. In typical situations such a time is longer than the age of the universe.

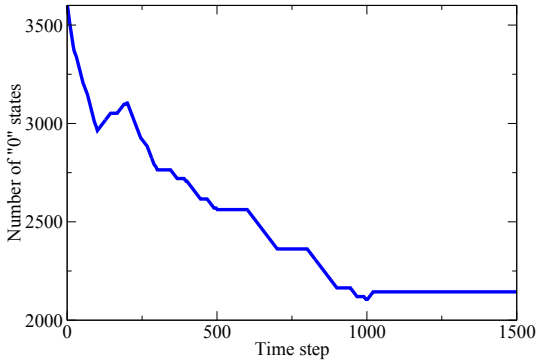



**Fig. 1.3** Kac-ring model: computer calculation of the time evolution of the number of "0" states in a sample made of $N = 3{,}600$ objects, which at time $t = 0$ were all set to "0". The markers of the sample are $n = 12$, and the number of time steps is smaller than $N$.

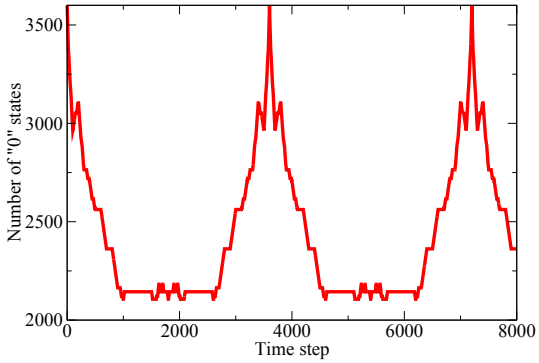



**Fig. 1.4** Kac-ring model: computer calculation of the time evolution of the number of "0" states in a sample made of $N = 3{,}600$ objects, which at time $t = 0$ were all set to "0". The markers of the sample are $n = 12$, and the number of time steps is larger than $N$.

and the introduction of the concept of probability through $U_{sr}$. More discussion on this issue is in [23].

## 1.12 Connection with Thermodynamics

The concepts introduced in Sect. 1.10 are also useful for determining the equilibrium distribution of a system of particles. Here, the hypothesis that the particles are point-like is dispensed with, and a system of $N$ identical molecules is considered, each molecule having $R$ degrees of freedom. The tessellation of the $\mu$ space yields, for the volume and units of each cell,

$$\Delta M = (\Delta q_1 \, \Delta p_1) \ldots (\Delta q_R \, \Delta p_R) , \qquad [\Delta M] = (\mathrm{J\,s})^R . \tag{1.83}$$

Letting $f_\mu$ be the distribution function, the population of the $i$th cell (whose center has coordinates $\mathbf{q}_i, \mathbf{p}_i$) is

$$N_i = f_i \, \Delta M , \qquad f_i = f_\mu(\mathbf{q}_i, \mathbf{p}_i, t) . \tag{1.84}$$

Remembering (1.75), the average with respect to $f_\mu$ of any dynamic quantity $A$ is given by

$$\mathrm{Av}[A](t) = \frac{\sum_i N_i A_i}{\sum_i N_i} = \frac{1}{N} \sum_i N_i A_i , \tag{1.85}$$

with $A_i$ the value of $A$ at the center of the $i$th cell and $N = \sum_i N_i$ the total number of molecules of the system. Letting $A = \log P$, and using $0 \leq P_i = N_i / N \leq 1$, yields an alternative expression for the Boltzmann $H$ function (1.76):

$$H_B(t) = \frac{1}{N} \sum_i N_i (\log N_i - \log N) = -\frac{1}{N} \left( N \log N - \sum_i N_i \log N_i \right) . \tag{1.86}$$

As mentioned in Sect. 1.10, apart from the special case in which all molecules belong to a single cell, say, $P_j = 1$ and $P_i = 0$ for $i \neq j$, function $H_B$ is strictly negative;[5] therefore, the quantity in parentheses at the right hand side of (1.86) vanishes only if $N_j = N$, otherwise it is strictly positive. Now, comparing (1.86) with [50, Eq. (6.13)] shows that

$$-N H_B(t) = N \log N - \sum_i N_i \log N_i = \log W , \tag{1.87}$$

where $W$ is the total number of possible ways of placing $N_1$ molecules in cell 1, $N_2$ molecules in cell 2, and so on, accounting for the fact that molecules belonging to the same cell are indistinguishable from each other.[6] The equilibrium distribution of the system of molecules is given by the set of numbers $N_1, N_2, \ldots$ that maximize $\log W$ and, at the same time, fulfill the constraint that the total energy of the system is fixed; other constraints (for example, the conservation of the number of particles) may enter the calculation depending on the system in hand. Since multiplicative or additive constants do not influence the maximization process, it follows that determining the equilibrium distribution using $-H_B$ or $\log W$ as a starting point yields the same result. One also remembers that in Thermodynamics the equilibrium state of a system corresponds to the condition in which entropy $S$ is a maximum; therefore, in order to make the above procedure consistent with the thermodynamical definition $\mathrm{d}S = \mathrm{d}Q / T$, with $\mathrm{d}Q$ the heat exchanged by the system during and infinitesimal transformation at temperature $T$, one lets

$$S = k_B \log W . \tag{1.88}$$

[5] The case $P_1 = P_2 = \cdots = 0$ is impossible because $N > 0$, whence $\sum_i P_i = 1$.

[6] As the calculation carried out in [50] aims at determining the equilibrium distribution, in that case the populations $N_i$ do not depend on time. However, the method by which the molecules are placed in the cells holds in general, given the constraints to which the molecules are subjected.

Before proceeding, one observes that the expressions derived above depend on the cell size $\Delta M$. If each side $\Delta q$ of the tessellation is multiplied by a factor $\alpha$, and each side $\Delta p$ is multiplied by a factor $\beta$, then $\Delta M$ becomes $\Delta M' = \tau \Delta M$ with $\tau = (\alpha\beta)^R$. It can be shown that scaling the cell volume alters $S$ by an additive (and arbitrary) constant; therefore, the equilibrium distribution is not altered, whereas entropy is altered. Further considerations show that the arbitrariness is removed by setting the value of entropy to zero in the $T \to 0$ limit [18, Sect. 30].

## 1.13 Entropy Variation in a Binary System

As an interesting example of application of the concepts illustrated above, consider a space made of $n$ blocks, each block being divided into two cells, say, the upper cell and the lower cell. Then, $n$ identical particles must be placed into the space, following these rules: (1) each block must contain one and only one particle, (2) the particle of a block must belong to either the upper or the lower cell. As the total number of possible ways of placing the particles is $W = 2^n$, the entropy of this system is

$$S = k_B \log 2^n = n k_B \log 2 . \tag{1.89}$$

Assume now that a transformation occurs such that one of the blocks, along with the corresponding particle, is erased; the entropy of the new system will be $S' = (n-1) k_B \log 2$, corresponding to an entropy variation $S - S' = k_B \log 2$. If the transformation occurs reversibly at some temperature $T$, the loss of energy of the system to the heat reservoir is

$$Q - Q' = k_B T \log 2 . \tag{1.90}$$

Since it corresponds to a reversible transformation, difference (1.90) is also the minimum loss of energy corresponding to the elimination of a block [18].

# Chapter 5
# M. Rudan — Classical Computation

Before considering the aspects of quantum computation it is necessary to provide a brief outline of the classical theory of computation. Traditionally, the beginning of the modern theory of computation is associated with the so-called *Hilbert tenth problem*, or *problem of decision*.[1] The first name derives from the fact that the problem was the tenth in a list of mathematical problems proposed in 1900 by David Hilbert; the problem asks whether it is possible to find a general algorithm able to decide, in a finite number of steps, whether a Diophantine equation has a solution such that all unknowns take integer values (a Diophantine equation is a polynomial equation having a finite number of unknowns and integer coefficients).[2]

At the time of Hilbert, the concept that an algorithm could be a "mechanical" procedure carried out by a machine was not clear; a number of years later (1936–1937), Alan Turing proposed a formalization of the working scheme of algorithms, in which the calculation procedure is decomposed into a set of fundamental blocks (*Turing Machine*, Sect. 5.9). Conversely, his result provides an axiomatic and fully abstract description of a mechanical procedure [63, 64].

## 5.1 Number Representations

Consider some examples of integer numbers expressed in *decimal* notation, namely, as linear combinations of the powers of 10 with integer coefficients:

$$19 = 10 + 9 = 1 \times 10^1 + 9 \times 10^0, \tag{5.1}$$

$$6903 = 6000 + 900 + 3 = 6 \times 10^3 + 9 \times 10^2 + 0 \times 10^1 + 3 \times 10^0. \tag{5.2}$$

The rightmost coefficient is obtained by dividing the original number by 10 and taking the remainder; the second coefficient from the right is obtained by dividing

[1] In German, *Entscheidungsproblem*.

[2] The answer to the Hilbert tenth problem has been found in 1970, and is negative.

by 10 the result of the previous division, and taking the new remainder, and so on; this operations can be arranged in columns as follows (the figure in parenthesis indicates the power of 10 to be multiplied by the coefficient thus found):

$$\begin{array}{r|ll} 19 & 9 & (0) \\ 1 & 1 & (1) \\ 0 & 0 & (2) \\ \vdots & \vdots & \end{array} \qquad \begin{array}{r|ll} 6903 & 3 & (0) \\ 690 & 0 & (1) \\ 69 & 9 & (2) \\ 6 & 6 & (3) \\ 0 & 0 & (4) \\ \vdots & \vdots & \end{array} \tag{5.3}$$

In summary, a number $N$ is expressed[3] as the sum of powers of a given *base*, indicated here with $b$ (in the example above, $b = 10$), each power being multiplied by a coefficient in the range $0, \ldots, b-1$.

By the same token, one can use the decompositions based on the powers of 7, namely, $19 = 14 + 5$ and $6903 = 4802 + 2058 + 42 + 1$; letting $b = 7$ and using the coefficients $0, \ldots, 6$ provides the *septimal* representation of 19 and 6903, that reads, respectively,

$$2 \times 7^1 + 5 \times 7^0 = 25\,, \qquad 2 \times 7^4 + 6 \times 7^3 + 0 \times 7^2 + 6 \times 7^1 + 1 \times 7^0 = 26061\,. \tag{5.4}$$

The standard way of decomposing a number in computer architectures is the *binary* one, which uses a sum made of the powers of 2, so that $b = 2$ with coefficients $0, 1$; considering the same numbers of the previous examples, the decompositions in decimal notation are $19 = 16 + 2 + 1$ and $6903 = 4096 + 2048 + 512 + 128 + 64 + 32 + 16 + 4 + 2 + 1$, this yielding

$$\begin{array}{r|ll} 19 & 1 & (0) \\ 9 & 1 & (1) \\ 4 & 0 & (2) \\ 2 & 0 & (3) \\ 1 & 1 & (4) \\ 0 & 0 & (5) \\ \vdots & \vdots & \end{array} \tag{5.5}$$

and the like for 6903. It follows that the binary notation for the number that, in decimal notation, is represented by 19, is

$$10011 = 1 \times 2^4 + 0 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0\,. \tag{5.6}$$

In this notation, the decomposition $19 = 16 + 2 + 1$ reads $10011 = 10000 + 10 + 1$.

[3] In the example just shown, $N$ is an integer. As one of the next examples demonstrates, the procedure is applicable also to non-integer numbers.

The same procedure can be carried out for numbers smaller than unity, e.g., using again the decimal notation, $0.8125 = .8 + .01 + .002 + .0005$, or

$$0.8125 = 8 \times 10^{-1} + 1 \times 10^{-2} + 2 \times 10^{-3} + 5 \times 10^{-4}. \tag{5.7}$$

In this case, the leftmost coefficient is obtained by multiplying the original number by 10 and taking the integer part of the result; the second coefficient from the left is obtained by multiplying by 10 the decimal part resulting from the previous multiplication, and taking the new integer part, and so on; this operations can be arranged in columns as shown in the left part of (5.8) where, again, the figure in parenthesis indicates the power of 10 to be multiplied by the coefficient thus found.

$$\begin{array}{r|ll} .8125 & 8 & (-1) \\ .125 & 1 & (-2) \\ .25 & 2 & (-3) \\ .5 & 5 & (-4) \\ 0 & 0 & (-5) \\ \vdots & \vdots & \end{array} \qquad \begin{array}{r|ll} .8125 & 1 & (-1) \\ .6250 & 1 & (-2) \\ .250 & 0 & (-3) \\ .50 & 1 & (-4) \\ 0 & 0 & (-5) \\ \vdots & \vdots & \end{array} \tag{5.8}$$

The binary decomposition of the same number is a sum made of the negative powers of 2, namely, $0.8125 = 1/2 + 1/4 + 1/16$; in this case the procedure is the one shown in the right part of (5.8). It follows that the binary notation[4] for the number that, in decimal notation, is represented by 0.8125, is

$$.1101 = 1 \times 2^{-1} + 1 \times 2^{-2} + 0 \times 2^{-3} + 1 \times 2^{-4}. \tag{5.9}$$

In this notation, the decomposition $0.8125 = 1/2 + 1/4 + 1/16$ becomes $.1101 = .1 + .01 + .0001$; finally, combining the two procedures above yields the expression of numbers where both the integer and decimal parts are different from zero:

$$19 + .8125 = 19.8125\,, \qquad 10011 + .1101 = 10011.1101\,. \tag{5.10}$$

Each coefficient (0 or 1) of the binary notation is called *bit* (from *binary digit*).

## 5.2 Elementary Operations

Given an algorithm able to provide the sum of two numbers, the other elementary operations (subtraction, multiplication, division) are easily accomplished; in fact, using, e.g., the decimal notation yields

[4] Note that the binary form of .8125 is made of a finite number of figures because the rightmost digit is 5. The same happens when the rightmost digit is 0.

$$\begin{cases} 12-3 = & 12+(-3) & = 9 \\ 13\times 7 = 0+\underbrace{13+13+13+13+13+13+13}_{7 \text{ times}} & & = 91 \\ 20\div 5 = & 20\underbrace{-5-5-5-5}_{4 \text{ times}} & = 4 \end{cases} \tag{5.11}$$

The examples used in (5.11) are deliberately simple: they use integer numbers and, in the division, the dividend is a multiple of the divisor; as for the subtraction, it is implied that a method is available for complementing the subtrahend: this can be carried out differently, depending on the computer architecture. In any case it is always assumed that the decimal part of the numbers under considerations is made of a finite number of digits.

The counterpart of (5.11) in binary notation is

$$\begin{cases} 1100-11 = & 1100+(-11) & = 1001 \\ 1101\times 111 = 0+\underbrace{1101+1101+\ldots+1101+1101}_{111 \text{ times}} & & = 1011011 \\ 10100\div 101 = & 10100\underbrace{-101-101-101-101}_{100 \text{ times}} & = 100 \end{cases} \tag{5.12}$$

The above examples show that, given an algorithm able to provide the sum of two numbers, the other elementary operations are easily accomplished, and all other calculations ensue. In other terms, combinations of elementary operations provide, to some degree of approximation, more complicate functions like, e.g., $\exp(x) \simeq \sum_{k=0}^{N} x^k/k!$ and so on; therefore, in order to implement the possibility that a machine performs calculations, the first step is the implementation of the sum.

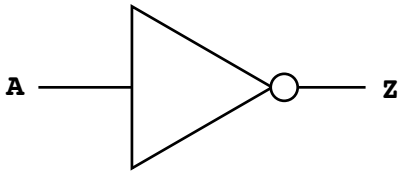
To proceed, one must observe that computers do not use numbers but (at least in the classical implementation) two different voltage levels, which are not related to numbers and must be viewed as two *logically-opposite conditions*; actually, once this point of view is accepted, several equivalent representations of opposite conditions are possible other than "High voltage-Low voltage": e.g., "Full tank-Empty tank", "True proposition-False proposition", "Set $A$-Complement of $A$". In several cases, in fact, initial examples of textbooks on the subject are based on the behavior of tanks connected by pipes, on propositional calculus, or on set theory.

A short consideration shows that two logically-opposite conditions can be condensed into two symbols, like $F$ ("False") and $T$ ("True"), or "0" and "1"; the latter symbols, "0" and "1" (not numbers!) are those commonly used, and are termed *logic values*. Any parameter $A, B, C, \ldots$ that can take the logic value 0 or 1 is a *binary variable*; in turn, a combination $Z = Z(A, B, C, \ldots)$ of binary variables is a *logic function*. The possible values of a logic function $Z$ are still 0 or 1; in order to determine $Z$ one must list all combinations of logic values of $A, B, C, \ldots$ and associate a logic value (0 or 1) to each combination: such an association provides a table, called *truth table* (examples of truth tables are given in Sect. 5.3). Finally, an abstract object implementing a truth table is a *logic operator* or *logic gate*; in the next sections, a number

of important logic operators are shown, along with suitable combinations that make it possible to obtain the sum of two binary numbers.

## 5.3 Classical Logic Gates

It is convenient to start with a short summary of Boolean algebra, that is, the framework used to describe the logic functions. Given the binary variables $A, B, C, \ldots$, one considers binary functions $Z = Z(A, B, C, \ldots)$, of which elementary examples are the NOT, AND, and OR operators. The symbols and truth tables of such operators are shown in Figs. 5.1, 5.2, and 5.3, respectively. The implementation of the NOT operator using the CMOS technology (also called *inverter*) is shown in Fig. 5.4. Other important operators are the NAND, NOR, and XOR (Exclusive OR)




| A | Z |
|---|---|
| 1 | 0 |
| 0 | 1 |

**Fig. 5.1** Symbol and truth table of the NOT operator $Z = \overline{A}$.

operators, whose symbols and truth tables are shown in Figs. 5.5, 5.6, and 5.7, respectively. The implementations of the NAND and NOR operators, still using the CMOS technology, are shown in Fig. 5.8. The AND, OR, and XOR operations are commutative and associative:

$$AB = BA, \qquad A(BC) = (AB)C, \tag{5.13}$$

$$A + B = B + A, \qquad A + (B + C) = (A + B) + C, \tag{5.14}$$

$$A \oplus B = B \oplus A, \qquad A \oplus (B \oplus C) = (A \oplus B) \oplus C. \tag{5.15}$$

In contrast, the NAND and NOR operations are commutative, but not associative:

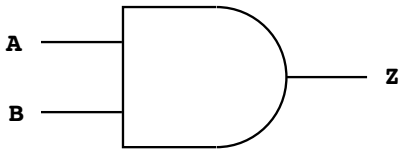
$$\overline{AB} = \overline{BA} \qquad \overline{A\left(\overline{BC}\right)} \neq \overline{\left(\overline{AB}\right)C}, \tag{5.16}$$

$$\overline{A + B} = \overline{B + A}, \qquad \overline{A + \left(\overline{B + C}\right)} \neq \overline{\left(\overline{A + B}\right) + C}. \tag{5.17}$$

The *De Morgan theorems* hold:

$$\overline{AB} = \overline{A} + \overline{B}, \qquad \overline{A + B} = \overline{A}\,\overline{B}. \tag{5.18}$$

Other useful theorems are:




| A | B | Z |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 1 | 1 | 1 |

**Fig. 5.2** Symbol and truth table of the AND operator $Z = AB$.

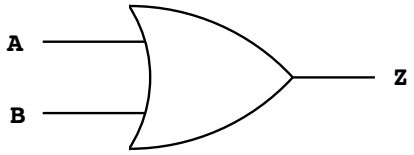
$$A+0=A\,,\qquad A+1=1\,,\qquad A+A=A\,,\qquad A+\overline{A}=1\,, \tag{5.19}$$

$$A\cdot 1=A\,,\qquad A\cdot 0=0\,,\qquad A\cdot A=A\,,\qquad A\cdot\overline{A}=0\,; \tag{5.20}$$

also, the following *contraction rules* are easily verified:

$$A+AB=A\,(\overline{B}+B)+AB=A\,(\overline{B}+B+B)=A\,, \tag{5.21}$$

$$A+\overline{A}B=(A+AB)+\overline{A}B=A+(\overline{A}+A)\,B=A+B\,. \tag{5.22}$$




| A | B | Z |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 1 |

**Fig. 5.3** Symbol and truth table of the OR operator $Z = A + B$.

## 5.4 Logic Implication and Logic Expressions

Another interesting truth table, shown in Fig. 5.9, defines the *logic implication*. The meaning of the denomination of $Z = A \sqsupset B$ is better understood if, in the truth tables shown here, symbol 1 is made to correspond to the logic value "True" (T), and symbol 0 is made to correspond to the logic value "False" (F). Thus, the meaning of, e.g., the AND operator $Z = AB$ is rendered as

$$(Z = \text{True})\ \text{if}\ [(A = \text{True})\ \text{and}\ (B = \text{True})]\,; \tag{5.23}$$

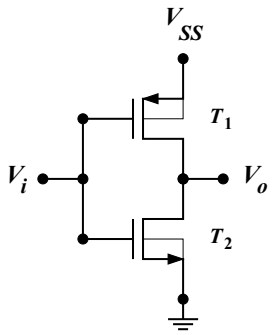
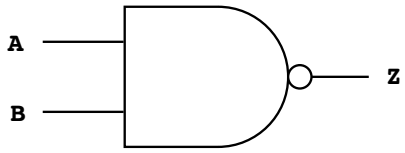
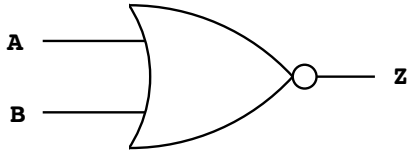



**Fig. 5.4** Implementation of the NOT operator in the CMOS technology (*inverter*).



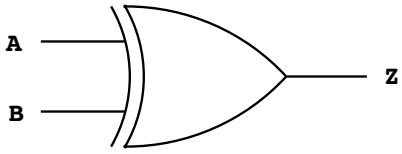

| A | B | Z |
|---|---|---|
| 0 | 0 | 1 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 0 |

**Fig. 5.5** Symbol and truth table of the NAND operator $Z = \overline{AB}$.




| A | B | Z |
|---|---|---|
| 0 | 0 | 1 |
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 1 | 1 | 0 |

**Fig. 5.6** Symbol and truth table of the NOR operator $Z = \overline{A + B}$.




| A | B | Z |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 0 |

**Fig. 5.7** Symbol and truth table of the XOR operator $Z = A \oplus B$.

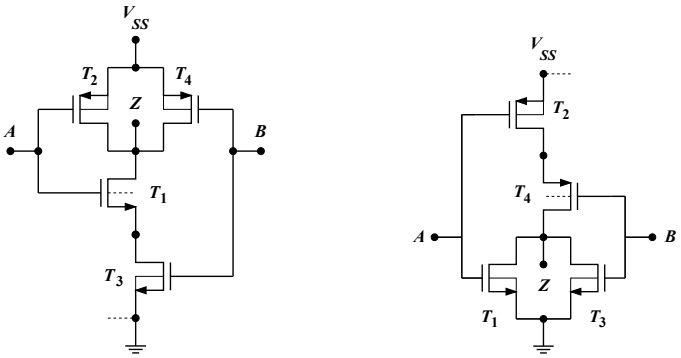



**Fig. 5.8** Implementation of the NAND (left) and NOR (right) operators in the CMOS technology. The dashed lines in the left circuit indicate that the bulk contact of $T_1$ and $T_3$ is connected to ground; it follows that the source and bulk contacts of $T_1$ are not shorted, namely, the body effect is present in $T_1$. Similarly, the dashed lines in the right circuit indicate that the *n*-type well within which $T_2$ and $T_4$ are fabricated is connected to the $V_{SS}$ bias; it follows that the source and bulk contacts of $T_4$ are not shorted, namely, the body effect is present in $T_4$.

in turn, the meaning of the OR operator $Z = A + B$ is

$$(Z = \text{True}) \text{ if } [(A = \text{True}) \text{ or } (B = \text{True}) \text{ or } (A \text{ and } B = \text{True})]. \tag{5.24}$$

Basing on this, one describes the logic implication as

$$\begin{cases} (A = \text{True}) \text{ and } (B = \text{True}) & \Rightarrow & (Z = \text{True}) \\ (A = \text{False}) \text{ and } (B = \text{False}) & \Rightarrow & (Z = \text{True}) \\ (A = \text{True}) \text{ and } (B = \text{False}) & \Rightarrow & (Z = \text{False}) \\ (A = \text{False}) \text{ and } (B = \text{True}) & \Rightarrow & (Z = \text{True}) \end{cases} \tag{5.25}$$

which is equivalent to

$$Z = A \sqsupset B = \overline{A} + B. \tag{5.26}$$

As an example of evaluation of logic expressions, let $Z = Z_2 Z_5 Z_3 Z_4 Z_1$, where

$$Z_1 = A \oplus B = A\overline{B} + \overline{A}B, \qquad Z_2 = C + E, \tag{5.27}$$

$$Z_3 = D \sqsupset B = \overline{D} + B, \quad Z_4 = AC + \overline{A}\,\overline{C}, \quad Z_5 = E \sqsupset (CD) = \overline{E} + CD. \tag{5.28}$$

One finds:

$$Z_2 Z_5 = (C+E)(\overline{E} + CD) = C\overline{E} + 0 + CD + CDE = C\overline{E} + CD,$$

$$Z_2 Z_5 Z_3 = (C\overline{E} + CD)(B + \overline{D}) = BCD + BC\overline{E} + 0 + C\overline{D}\,\overline{E},$$

$$Z_2 Z_5 Z_3 Z_4 = (BCD + BC\overline{E} + C\overline{D}\,\overline{E})(AC + \overline{A}\,\overline{C}) = ABCD + ABC\overline{E} + AC\overline{D}\,\overline{E},$$

whence

$$Z = Z_2 Z_5 Z_3 Z_4 Z_1 = (ABCD + ABC\overline{E} + AC\overline{D}\,\overline{E})\,(A\overline{B} + \overline{A}B) = A\overline{B}C\overline{D}\,\overline{E}\,. \quad (5.29)$$

In other terms, $Z = 1$ iff $A = C = 1$ and $B = D = E = 0$.

| A | B | Z |
|---|---|---|
| 0 | 0 | 1 |
| 0 | 1 | 1 |
| 1 | 0 | 0 |
| 1 | 1 | 1 |

**Fig. 5.9** Truth table of the logic implication $Z = A \sqsupset B = \overline{A} + B$.

## 5.5 Canonical Forms

For applications to digital circuits it is convenient to present the logic functions in either one of the *canonical forms*; the latter are the *sum of minterms* and the *product of maxterms*.[5] The minterms are the logic AND of a set of variables, and the maxterms are the logic OR of a set of variables; more specifically, in a minterm or in a maxterm, *all* logic variables that form the logic function must appear once, either in the complemented or uncomplemented form. By way of example, consider the logic function $Z = (\overline{A} + BC)\,(B + CD)$, whose form is non-canonical because it is made of the product of sums that in turn embed other products; expanding the product yields

$$Z = \overline{A}B + BC + \overline{A}CD + BCD, \quad (5.30)$$

which is a sum of products, but not of the canonical form yet: in fact, the products do not contain all variables. As a second example, consider the logic function $Z = (A + \overline{BC})\,(\overline{D + BE})$ which, using the De Morgan theorems (5.18), becomes

$$Z = \left(A + \overline{B} + \overline{C}\right)\overline{D}\left(\overline{BE}\right) = \left(A + \overline{B} + \overline{C}\right)\overline{D}\left(\overline{B} + \overline{E}\right) =$$

$$= A\,\overline{B}\,\overline{D} + \overline{B}\,\overline{D} + \overline{B}\,\overline{C}\,\overline{D} + A\,\overline{D}\,\overline{E} + \overline{B}\,\overline{D}\,\overline{E} + \overline{C}\,\overline{D}\,\overline{E}\,. \quad (5.31)$$

Again, this is a sum of products, but not of the canonical form. As a third example, consider the function $Z = A + \overline{B}C$; its canonical form is obtained by the following procedure:

$$Z = A\,\left(B + \overline{B}\right)\,\left(C + \overline{C}\right) + \left(A + \overline{A}\right)\,\overline{B}C = ABC + A\overline{B}C + AB\overline{C} + A\overline{B}\,\overline{C} + \overline{A}\,\overline{B}C\,. \quad (5.32)$$

The construction of the minterms for the logic function 5.32 is shown in the table of Fig. (5.10). The eight combinations of bits $A$, $B$, and $C$ are ordered according

[5] The two canonical forms are also indicated with *sum of products* and *product of sums*.

to the *Gray code*;[6] then, each minterm is formed by taking the uncomplemented (complemented) value of the logic variable if the logic value of the latter in the same line is 1 (0):

| Gray code | A | B | C | Z | Minterm |
|---|---|---|---|---|---|
| $m_0$ | 0 | 0 | 0 | 0 | |
| $m_1$ | 0 | 0 | 1 | 1 | $\Rightarrow \overline{A}\,\overline{B}\,C$ |
| $m_2$ | 0 | 1 | 1 | 0 | |
| $m_3$ | 0 | 1 | 0 | 0 | |
| $m_4$ | 1 | 1 | 0 | 1 | $\Rightarrow A\,B\,\overline{C}$ |
| $m_5$ | 1 | 1 | 1 | 1 | $\Rightarrow A\,B\,C$ |
| $m_6$ | 1 | 0 | 1 | 1 | $\Rightarrow A\,\overline{B}\,C$ |
| $m_7$ | 1 | 0 | 0 | 1 | $\Rightarrow A\,\overline{B}\,\overline{C}$ |

**Fig. 5.10** Definition of the minterms to be used in (5.33).

$$Z = A + \overline{B}C = m_1 + m_4 + m_5 + m_6 + m_7\,. \tag{5.33}$$

## 5.6 Universal Gates

In principle, the implementation of a logic function entails a combination of different gates like the NOT, AND, and OR gates. By way of example, a combination of the above provides the XOR operator (Fig. 5.11). More generally, by a suitable combination of the NOT, AND, and OR gates it is possible to implement *any* logic function; to show this it is sufficient to observe that any logic function can be reduced to a canonical form.

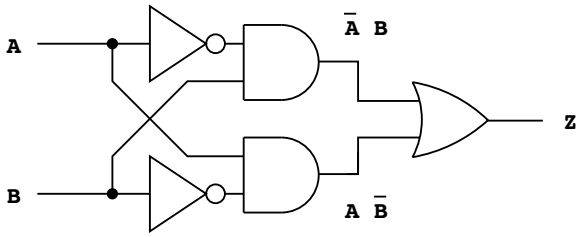



**Fig. 5.11** Implementation of the XOR operator using the NOT, AND, and OR operators. The output $Z = \overline{A}B + A\overline{B}$ fulfills the truth table shown in Fig. 5.7.

[6] The Gray code is an ordering such that only one bit is changed from one combination to the next. It has the advantage that, in the implementation with electronic circuits, only one gate at the time is switched. This overcomes the practical impossibility of having several gates switching simultaneously.

From the practical standpoint, the implementation of a logic function by means of a combination of different types of gates is not efficient: in view of the actual fabrication of a circuit implementing a logic function, it would be preferable to use one type of logic gate only; this would in fact allow the designer to optimize, and then replicate, the same scheme. It is therefore important to identify a single logic gate able to reproduce the NOT, AND, and OR operators; if such a gate exists, it is termed *universal*.

It is easily found that the NAND operator is universal; the implementation of the NOT, AND, and OR operators using suitable combinations of the NAND operator is shown in Fig. 5.12. The NOR operator is universal as well: still with reference to Fig. 5.12, one could replace each NAND operator with a NOR operator, while keeping the same connections; in this way, the top, middle, and bottom part of the figure would correspond to the NOT, OR, and AND operators, respectively.

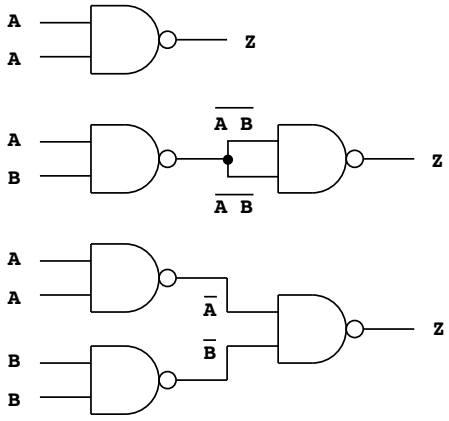
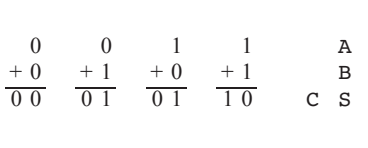



**Fig. 5.12** Implementation of the NOT, AND, and OR operators using the NAND operator only. In fact, the output of the operations is (top) $Z = \overline{A\,A} = \overline{A}$, (middle) $Z = \overline{\overline{A\,B}} = A\,B$, (bottom) $Z = \overline{\overline{A}\,\overline{B}} = A + B$.

## 5.7 Half Adder and Full Adder

In the above sections it has been shown that, by suitably combining a number of logic gates, it is possible to implement any logic function; this provides the ingredients necessary to solve the problem stated at the end of Sect. 5.2, namely, the implementation of the sum of two binary numbers. The task is accomplished by identifying, first, the truth table corresponding to addition, then by constructing the corresponding logic function and, finally, by implementing the function by means of logic gates.

To proceed, it is convenient to consider the case where each of the two numbers to be added is made of a single digit; the operator that, given two one-digit summands $A$ and $B$ in binary notation, provides the sum ($S$) and the carry ($C$), is called *half adder*. The four possible combinations of the summands are shown in the left part




| A | B | S | C |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 0 | 1 | 1 | 0 |
| 1 | 0 | 1 | 0 |
| 1 | 1 | 0 | 1 |

**Fig. 5.13** Operation and truth table of the half adder.

of Fig. 5.13, and the corresponding truth table is shown in the right part of the same figure; comparing this truth table with those of Figs. 5.2 and 5.7 shows that the logic expressions describing the half adder are

$$S = A \oplus B\,, \qquad C = A\,B\,. \tag{5.34}$$

Clearly, the half adder is in itself insufficient, because it does not account for a possible carry produced by a less significant digit. To perform the addition of numbers made of $n$ digits one must refer to the table shown in Fig. 5.14, that provides the functioning of the *full adder* operator. Adding the least significant digits $A_0$ and $B_0$

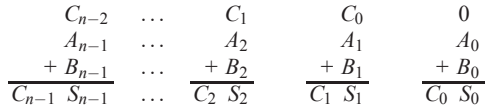



**Fig. 5.14** Operation of the full adder.

of the summands, provides the least significant digit $S_0$ of the sum, along with the carry of order zero, $C_0$. The latter is added to the next summands $A_1$ and $B_1$, and so on; the result of the calculation is the string

$$S = [C_{n-1}\,, S_{n-1}\,, \ldots S_2\,, S_1\,, S_0]\,. \tag{5.35}$$

The truth table corresponding to the $i$th digit of the above calculation is shown in Fig. 5.15, and the logic expression describing the full adder is:

$$S_i = C_{i-1} \oplus A_i \oplus B_i\,, \qquad C_i = A_i B_i + (A_i \oplus B_i)\, C_{i-1}\,. \tag{5.36}$$

| $C_{i-1}$ | $A_i$ | $B_i$ | $S_i$ | $C_i$ |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 1 | 0 |
| 0 | 1 | 0 | 1 | 0 |
| 0 | 1 | 1 | 0 | 1 |
| 1 | 0 | 0 | 1 | 0 |
| 1 | 0 | 1 | 0 | 1 |
| 1 | 1 | 0 | 0 | 1 |
| 1 | 1 | 1 | 1 | 1 |

**Fig. 5.15** Truth table of the $i$th digit of the full adder.

## 5.8 Logic and Physical Irreversibility

Most of the logic gates described so far have two properties in common, *physical irreversibility* and *logic irreversibility*; as shown below, the two properties are related. A process is termed *physically reversible* if it does not produce an increase in the entropy of the system under consideration. In practice, no dynamic process is physically reversible; however, if one succeeds in isolating the physical system from the environment, in order to prevent the unknown effects of the interactions with the latter, the laws describing the evolution of the system are precisely known. In this limiting case, one may approach the reversibility of the process.

Achieving physical reversibility would be of enormous advantage in the field of computation, thanks to the improvement in computational efficiency from the viewpoint of energy consumption. When the classical computer architecture is considered, there is a minimum theoretical limit for the energy dissipated in an irreversible bit operation; as shown in [67] its value, called the *Landauer limit* or *von Neumann-Landauer limit*,[7] is $L = k_B T \log 2$; the derivation of $L$ is also shown in Sect. 1.13.

As observed first in [35], a computational process must be logically reversible in order to be physically reversible.[8]

[7] The Landauer limit is enormously smaller than the actual energy consumption of present-day computers [6].

[8] This means that logic reversibility is necessary, but not sufficient, for physical reversibility. Consider for instance the implementation of the NOT operator shown in Fig. 5.4.

## 5.9 The Turing Machine

The (deterministic) Turing machine consists in an infinitely-long tape parted into equal cells. In each cell a "zero" (0), a "one" (1), or a blank may be written (Fig. 5.16). The tape is supplemented with a head that can move in either direction along the tape, and is able to read or write the content of each cell. The head has an internal state, which belongs to a finite set of possible states; also, the head contains a list of instructions, indicated with the term *program*; the program specifies, given the internal state of the head and the bit (that is, the binary digit 0 or 1) present at the current location of the head and being read by the latter, whether the bit should be changed and in which direction the head should move.

The *internal state*, or *configuration*, of a system is the condition of the different components of the system at a given instant $t$. The components to be considered in a Turing machine are the number of the observed cell, the content of the cell, and the instruction to be executed. Among the possible states it is convenient to distinguish *a)* the initial configuration (corresponding to $t = t_0$), prior to the starting of the program, *b)* the final configuration (corresponding to $t = t_n$), at the end of the program, and *c)* the intermediate configuration (corresponding to $t = t_i$), prior to the execution of operation $\omega_i$. Implementing an algorithm in the Turing machine consist in performing one of the following operations:

- The head moves to the cell on the right of the present one.
- The head moves to the cell on the left of the present one.
- The head writes into the present cell a symbol taken from a list of symbols.
- The cell erases the symbol that appears in the present cell.
- The program stops.

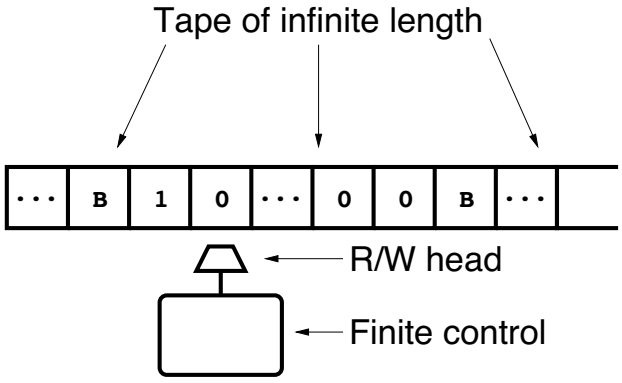



**Fig. 5.16** Schematic view of the Turing machine.

The execution of operation $\omega_1$ between the instants $t_1$ and $t_2$ is equivalent to pass from the internal state $s_1$ to the internal state $s_2$. This is indicated with the symbol

$\{s_1, a_1, \omega_1, s_2\}$, meaning that, starting from state $s_1$, the machine reads symbol $a_1$, executes operation $\omega_1$, and ends up in the internal state $s_2$. Given enough time, such a machine is able to perform any calculation; in fact, one can demonstrate that the Turing machine is able to carry out all elaborations that are possible with the existing models of calculation.[9] An advantage of the Turing machine is that it is described by simple rules; this property entails the possibility of describing the machine by elementary mechanisms, easily illustrated. In this respect it is postulated that, for each computable problem, there exists a Turing machine able to solve it (*Church-Turing conjecture*); the algorithms than can be implemented using a Turing machine are called "Turing-computable algorithms".

## 5.10 Computational Complexity

The Turing machine provides a model for real sequential computers; in particular, the computer memory is an approximation of the infinitely-long tape of the Turing machine. An issue still left open is the practical possibility of bringing a computational problem to an end; it may in fact happen that the power required to solve a given problem, in terms of memory and time, is so large as to make the task unaffordable. To provide a first example of how to estimate the computational power necessary to solve a given task, one may consider the following problem:

**Change in the interest rate**

The interest rate on the 15-year loans is changed. A bank must recalculate, for each monthly installment of $n$ loans using the French amortization, the part related to the new interest.

The number of calculations necessary to solve the problem is of the order of $K = 15 \times 12 \times n$; if $n = 1{,}000{,}000$, then $K \approx 180{,}000{,}000$.

Although the number of calculations is huge, one notes that it is proportional to the size $n$ of the problem. More generally, if the size of a problem is $n$, and the number of calculations necessary to solve it is some function of $n$ of a polynomial form (like in the example above), the complexity of the problem is of the *polynomial type* ($P$).

As a second example one may consider a combinatorial problem, of which a typical example is that of the *Travelling Salesman Problem*; the problem is a benchmark in operations research, and is formulated as follows:

**The Travelling Salesman Problem**

Given a list of $n$ cities and the distances between each pair of cities, find the shortest route that visits each city exactly once and returns to the city whence the route started.

[9] In other terms, the Turing machine can perform the same calculations that can be accomplished by a super-computer, in a longer time.

The solution is conceptually easy: let $K = (n-1)!$ be the total number of routes. Calculate the length of routes 1 and 2, and keep the shorter one; then, calculate the length of route 3, compare it with the one selected at the previous step, and keep the shorter one. Continue down to route $K$.

However, when the size $n$ of the problem is large,[10] it is (Stirling approximation)

$$n! \sim \sqrt{2\pi n}\,(n/e)^n\,, \tag{5.37}$$

showing that the time necessary for solving the Travelling Salesman Problem grows exponentially with the size.

For the discussion it is necessary to clarify the difference between a *deterministic* and a *non-deterministic* algorithm: a deterministic algorithm constructs the demonstration of the result by considering all possibilities, in other terms, it *demonstrates* the theorem under consideration; in the example of the Travelling Salesman Problem, the algorithm calculates the length of all possible $(n-1)!$ routes, and selects the shortest. A non-deterministic algorithm, instead, *verifies* the validity of a theorem.

The discussion above leads to a classification of the complexity of problems; specifically, the problems whose deterministic solution is polynomial are grouped in a class denoted with $P$, while those whose verification is polynomial are grouped in a class denoted $NP$. If a problem is solvable in polynomial time, then a solution is also verifiable in polynomial time by solving the same problem; it follows that a deterministic algorithm is a particular case of a non-deterministic algorithm, whence $P \subset NP$. The problems of the combinatorial type, like the Travelling Salesman Problem, are called "$NP$-complete", and the corresponding class is indicated with $NPC$.[11]

The example of the Travelling Salesman Problem, along with other problems (like, e.g., that of the weather forecast), shows that a classical computer, that performs the calculations one by one, may not be suited for some types of problems. The possibility of performing calculations in parallel would be of help; this issue is addressed in Sect. 24.1.1.

[10] The Stirling approximation is already quite accurate for $n = 10$.

[11] Following [58], a gross estimate of the time necessary for solving the Travelling Salesman Problem with $n = 26$ cities could be made by assuming that computers are reduced to the size of an atom, $r_a \simeq 3 \times 10^{-8}$ cm, and that the time necessary to calculate the length of a route equals the time $\Delta t$ taken by light, whose speed is $c \simeq 3 \times 10^{10}$ cm/s, to cross the atom. It turns out $\Delta t = r_a/c \simeq 10^{-18}$ s. On the other hand, from the Stirling approximation one finds $\log[(n-1)!] = \log(25!) \simeq 58$, whence $25! \simeq \exp(58) = 10^{58/\log(10)} \simeq 10^{25}$. Thus, the time necessary to calculate all routes would be $10^{25} \times 10^{-18} = 10^7$ s, equivalent to about 116 days.

Lecture-39

# Chapter 23
# M. Rudan — Matrix Formulation of Quantum Mechanics

The formal structure of Quantum Mechanics, based on the concepts of operators and eigenvalue equations, may be recast in a language that is the analogue of that of vectors and matrices.

## 23.1 Bra and Ket Vectors

As a starting point, one may consider the example of a vector $\mathbf{a}$ in a $k$-dimensional space, expressed as a linear combination of the components along the mutually orthogonal axes $x_1, \ldots, x_k$. Letting $\mathbf{i}_1, \ldots, \mathbf{i}_k$ be the unit vectors of the axes, it is

$$\mathbf{a} = \sum_{j=1}^{k} a_j \mathbf{i}_j = a_1 \mathbf{i}_1 + \cdots + a_k \mathbf{i}_k = a_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + a_k \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix} . \qquad (23.1)$$

In the quantum-mechanical case, consider a Hermitean [1] operator $\mathscr{A}$ having discrete eigenvalues $A_n$ and a complete set of mutually-orthogonal eigenfunctions $u_n$, normalized to unity. The state of a particle or of a system of particles is described by a wave function $\psi$ which, thanks to the completeness of set $u_n$, and to the orthonormality property $\langle u_m | u_n \rangle = \delta_{nm}$, is expressible as $\psi = \sum_n c_n u_n$, where the coefficients of the expansion are given by the scalar products $\langle u_n | \psi \rangle$. It is customary to adopt the Dirac notation, so that the expansion reads $|\psi\rangle = \sum_n c_n |u_n\rangle$. If $|\psi\rangle$ is thought of as a vector made of infinite components (*ket vector*), the above expansion reads[2]

[1] The definition of Hermitean operator is given in Sec. 23.2.

[2] Obviously, the values of coefficients $c_n$ depend on the choice of the complete set $u_n$. The reasoning leading to (23.2) is stated more appropriately as: "*in a given basis* $u_n$, function $\psi$ is thought of as a vector made of the coefficients of the expansion".

$$|\psi\rangle = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \end{bmatrix} + \cdots \tag{23.2}$$

It is apparent from the above that the ket vectors are considered as column vectors. The *bra vector* corresponding to $\psi$ is defined as $\langle\psi| = [c_1^*, c_2^*, \ldots]$, namely, it is a row vector whose components are the conjugate of those of the ket vector.[3] Using this notation, the *scalar product* of $|\psi\rangle = \sum_n c_n |u_n\rangle$ and $|\phi\rangle = \sum_k b_k |u_k\rangle$ is defined as

$$\langle\psi|\phi\rangle = \sum_n c_n^* \langle u_n| \sum_k b_k |u_k\rangle = \sum_{nk} c_n^* b_k \langle u_n|u_k\rangle = [c_1^*, c_2^*, \ldots] \begin{bmatrix} b_1 \\ b_2 \\ \vdots \end{bmatrix}, \tag{23.3}$$

where the mutual orthonormality of vectors $u_n$ has been exploited; in a more compact notation, it is $\langle\psi|\phi\rangle = \sum_n c_n^* b_n$.

## 23.2 Matrix Associated to an Operator

Given an operator $\mathscr{B}$, not necessarily Hermitean, let $|\phi\rangle = \mathscr{B}|\psi\rangle$. Expanding $|\psi\rangle$ and $|\phi\rangle$ into the same set $|u_n\rangle$ yields

$$|\phi\rangle = \sum_k b_k |u_k\rangle, \qquad b_k = \langle u_k|\phi\rangle = \langle u_k|(\mathscr{B} \sum_n c_n |u_n\rangle)\rangle = \sum_n B_{kn} c_n, \tag{23.4}$$

where $B_{kn} = \langle u_k| (\mathscr{B}|u_n\rangle)\rangle$. It follows[4]

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} & \cdots \\ B_{21} & B_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \end{bmatrix}, \tag{23.5}$$

namely, in this formalism operators are represented by matrices. Given two vectors

$$|f\rangle = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \end{bmatrix}, \qquad |g\rangle = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \end{bmatrix}, \tag{23.6}$$

and repeating the procedure leading to the definition of $B_{kn}$, it is easily found that

[3] The terms *bra* and *ket* are the two parts of the word *bracket*.

[4] Similarly to what specified in the note at p. 349, a matrix is associated to an operator *after specifying the basis used for this association*.

$$\langle f|\,(\mathscr{B}g)\rangle = \sum_{kn} f_k^*\, B_{kn}\, g_n\,, \qquad \langle(\mathscr{B}f)\,|g\rangle = \sum_{kn} f_k^*\, B_{nk}^*\, g_n\,. \tag{23.7}$$

Letting $\mathbf{B}$ indicate the matrix of elements $B_{kn}$, then the matrix of elements $B_{nk}^*$ is its conjugate transpose. The operator associated to the conjugate transpose of $\mathbf{B}$ is called *adjoint* of $\mathscr{B}$ and is indicated with symbol $\mathscr{B}^\dagger$. From (23.7) it is apparent that

$$\langle\left(\mathscr{B}^\dagger f\right)|g\rangle = \langle f|\,(\mathscr{B}g)\rangle\,, \qquad \left(\mathscr{B}^\dagger\right)^\dagger = \mathscr{B}\,. \tag{23.8}$$

In general it is $\mathscr{B}^\dagger \neq \mathscr{B}$; however, there are operators such that $\mathscr{B}^\dagger = \mathscr{B}$: such operators are called *self-adjoint* or *Hermitean*. The matrix corresponding to a Hermitean operator is equal to its conjugate transpose, $\mathbf{B}^* = \mathbf{B}^{\mathrm{T}}$. From (23.8) it follows that a Hermitean operator fulfills the relation $\langle f|\,(\mathscr{B}g)\rangle = \langle(\mathscr{B}f)\,|g\rangle$; in other terms, it is irrelevant whether the operator is applied to the right or left function, so that the parentheses are not necessary any more: the expression used in this case is $\langle f|\mathscr{B}|g\rangle$.

Definition (23.4) of $B_{kn}$ is based on a set of vectors $|u_n\rangle$ that are not necessarily eigenvectors of $\mathscr{B}$. Consider now the case where vectors $|u_n\rangle$ are the eigenvectors of a Hermitean operator $\mathscr{A}$, so that $\mathscr{A}|u_n\rangle = A_n\,|u_n\rangle$, with $A_n$ the eigenvalue. The definition of the matrix associated to $\mathscr{A}$ yields, thanks to the orthonormality of vectors $|u_n\rangle$,

$$\langle u_k|\mathscr{A}|u_n\rangle = A_n\,\langle u_k|u_n\rangle = A_n\,\delta_{kn}\,, \tag{23.9}$$

with $\delta_{kn}$ the Kronecker delta. It follows that the matrix associated to $\mathscr{A}$ by the set $|u_n\rangle$ is diagonal,

$$\mathbf{A} = \begin{bmatrix} A_1 & 0 & \cdots \\ 0 & A_2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}, \tag{23.10}$$

and its diagonal entries are the eigenvalues of $\mathscr{A}$. This shows that finding the eigenvectors of an operator is equivalent to finding the reference frame in which the corresponding matrix is diagonal. In fact, given a Hermitean operator $\mathscr{B}$, its eigenvalue equation reads $\mathscr{B}|v\rangle = b\,|v\rangle$; letting $|v\rangle = \sum_n v_n\,|u_n\rangle$ yields

$$\langle u_m|\mathscr{B}\sum_n v_n\,u_n\rangle = \langle u_m|b\sum_n v_n\,u_n\rangle\,, \qquad \sum_n B_{mn}\,v_n = b\,v_m\,, \tag{23.11}$$

so that the matrix form of $\mathscr{B}|v\rangle = b\,|v\rangle$ is

$$\begin{bmatrix} B_{11} & B_{12} & \cdots \\ B_{21} & B_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \end{bmatrix} = b \begin{bmatrix} v_1 \\ v_2 \\ \vdots \end{bmatrix}, \tag{23.12}$$

where the column vector $\mathbf{v}$ of entries $v_i$ represents $|v\rangle$. Letting $\mathbf{I}$ be the identity matrix, the eigenvalue equation takes the form $\mathbf{B}\,\mathbf{v} = b\,\mathbf{I}\,\mathbf{v}$, corresponding to an infinite set of homogeneous equations in the unknowns $\mathbf{v}$. The eigenvalues are the solutions

of equation $\det(\mathbf{B} - b\,\mathbf{I}) = 0$; after the eigenvalues are found, replacing them one by one in (23.12) provides the corresponding eigenvectors $\mathbf{v}$.

## 23.3 Unitary Transformations

Let $\mathscr{A}$ be a Hermitean operator with eigenvectors $|u_n\rangle$ and eigenvalues $A_n$. The latter are assumed to be simple. Next, consider another operator $\mathscr{S}$, to which a matrix $\mathbf{S}$ is associated by the eigenvectors $|u_n\rangle$, $S_{kn} = \langle u_k | (\mathscr{S}|u_n\rangle)\rangle$. Given a vector $|f\rangle$ represented in terms of the same eigenvectors, $|f\rangle = \sum_n f_n |u_n\rangle$, it follows

$$\mathscr{S}|f\rangle = \sum_n f_n |v_n\rangle\,, \qquad |v_n\rangle = \mathscr{S}|u_n\rangle\,. \tag{23.13}$$

It is of interest to identify the conditions under which vectors $|v_n\rangle$ are mutually orthogonal. Remembering (23.8) one finds that this condition is equivalent to

$$\langle v_k|v_n\rangle = \langle (\mathscr{S}u_k)\,|\,(\mathscr{S}u_n)\rangle = \langle u_k|\left(\mathscr{S}^\dagger \mathscr{S} u_n\right)\rangle = \delta_{kn}\,, \tag{23.14}$$

which is fulfilled if $\mathscr{S}^\dagger\mathscr{S} = \mathscr{S}\mathscr{S}^\dagger = \mathscr{I}$, where $\mathscr{I}$ is the identity operator. The above relation is recast as

$$\mathscr{S}^\dagger = \mathscr{S}^{-1}\,, \tag{23.15}$$

namely, the adjoint operator equals the inverse operator. An operator fulfilling (23.15) is called *unitary*; the transformation $|v_n\rangle = \mathscr{S}|u_n\rangle$ that brings from set $|u_n\rangle$ to set $|v_n\rangle$ is called *unitary transformation*. Besides orthonormality, a unitary transformation preserves completeness; as a consequence, a unitary transformation allows one to obtain any reference (e.g., $|v_n\rangle$) starting from any other reference (e.g., $|u_n\rangle$).

In general, a unitary operator is not Hermitean; if it is such, it must fulfill $\mathscr{S}^\dagger = \mathscr{S}$ besides (23.15); it follows $\mathscr{S}^{-1} = \mathscr{S}$, whence $\mathscr{S} = \mathscr{I}$. A unitary operator leaves the norm of a function unchanged; letting $|g\rangle = \mathscr{S}|f\rangle$, it is in fact

$$\langle g|g\rangle = \langle (\mathscr{S}f)\,|\,(\mathscr{S}f)\rangle = \langle f|\left(\mathscr{S}^\dagger\mathscr{S}f\right)\rangle = \langle f|f\rangle\,. \tag{23.16}$$

The norm of the eigenvalues of a unitary operator is equal to unity;[5] letting $\mathscr{S}|s\rangle = \lambda\,|s\rangle$, with $\lambda$ an eigenvalue of $\mathscr{S}$ and $|s\rangle$ an eigenfunction corresponding to it, the following equalities hold together:

$$\langle (\mathscr{S}s)\,|\,(\mathscr{S}s)\rangle = \langle \lambda\, s|\lambda\, s\rangle = |\lambda|^2\,\langle s|s\rangle\,, \qquad \langle (\mathscr{S}s)\,|\,(\mathscr{S}s)\rangle = \langle s|s\rangle\,, \tag{23.17}$$

the first of which derives from the eigenvalue equation, the second one from the norm-conservation property. Since an eigenfunction can not vanish identically, it is $|\lambda|^2 = 1$, whence $\lambda = \exp(\mathrm{i}\,b)$, with $b$ a real number.

[5] Whence the term *unitary*.

## 23.4 Similarity Transformations — Functions of Operators

Remembering (23.4), one can transform an operator $\mathscr{B}$ into a matrix using a given reference $|u_n\rangle$: one finds in fact $B^u_{kn} = \langle u_k|\,(\mathscr{B}u_n)\rangle$. If, instead, another reference $|v_n\rangle$ is used, the entries of the corresponding matrix are $B^v_{kn} = \langle v_k|\,(\mathscr{B}v_n)\rangle$. It is of interest to determine the relation between $B^u_{kn}$ and $B^v_{kn}$. For this, one observes that a unitary operator $\mathscr{S}$ exists, that transforms vectors $|u_n\rangle$ into vectors $|v_n\rangle$, so that

$$B^v_{kn} = \langle v_k|\,(\mathscr{B}v_n)\rangle = \langle(\mathscr{S}u_k)\,|\,(\mathscr{B}\mathscr{S}u_n)\rangle = \langle u_k|\left(\mathscr{S}^\dagger\mathscr{B}\mathscr{S}u_n\right)\rangle\,. \tag{23.18}$$

From (23.15) it follows $B^v_{kn} = \langle u_k|\left(\mathscr{S}^{-1}\mathscr{B}\mathscr{S}u_n\right)\rangle$. The corresponding relations among matrices are found after associating a matrix to $\mathscr{S}$, namely,

$$\mathbf{B}^v = \mathbf{S}^\dagger\,\mathbf{B}^u\,\mathbf{S} = \mathbf{S}^{-1}\,\mathbf{B}^u\,\mathbf{S}\,, \qquad S_{kn} = \langle u_k|\,(\mathscr{S}u_n)\rangle\,. \tag{23.19}$$

If operator $\mathscr{B}$ is Hermitean, then matrices $\mathbf{B}^u$ and $\mathbf{B}^v$ are also Hermitean.

Given a square, non-singular matrix $\mathbf{G}$ (non necessarily associated to a unitary operator) the operation that brings from $\mathbf{B}$ to $\mathbf{G}^{-1}\,\mathbf{B}\,\mathbf{G}$ is called *similarity transformation*; matrices $\mathbf{B}$ and $\mathbf{G}^{-1}\,\mathbf{B}\,\mathbf{G}$ are called *similar*. Relation (23.19) above is an example of this. Similar matrices have the same determinant, the same eigenvalues, and the same trace; the first property is demonstrated by observing that

$$\det(\mathbf{G}^{-1}\,\mathbf{B}\,\mathbf{G}) = \det(\mathbf{G}^{-1})\,\det(\mathbf{B})\,\det(\mathbf{G}) = \det(\mathbf{G}^{-1}\,\mathbf{G})\,\det(\mathbf{B}) = \det(\mathbf{B})\,; \tag{23.20}$$

the second property is demonstrated by observing that the eigenvalues $\lambda$ of $\mathbf{G}^{-1}\,\mathbf{B}\,\mathbf{G}$ are found from the algebraic equation $\det(\mathbf{G}^{-1}\,\mathbf{B}\,\mathbf{G} - \lambda\,\mathbf{I}) = 0$, and that $\mathbf{I} = \mathbf{G}^{-1}\,\mathbf{I}\,\mathbf{G}$. It follows

$$\det(\mathbf{G}^{-1}\,\mathbf{B}\,\mathbf{G} - \lambda\,\mathbf{I}) = \det[\mathbf{G}^{-1}\,(\mathbf{B} - \lambda\,\mathbf{I})\,\mathbf{G}] = \det(\mathbf{G}^{-1}\,\mathbf{G})\,\det(\mathbf{B} - \lambda\,\mathbf{I})\,, \tag{23.21}$$

namely, the equations that provide the eigenvalues of similar matrices are identical to each other. As for the eigenvectors, let $\mathbf{e}$ be an eigenvector of $\mathbf{B}$ corresponding to $\lambda$, namely, $\mathbf{B}\,\mathbf{e} = \lambda\,\mathbf{e}$. Using $\mathbf{G}\,\mathbf{G}^{-1} = \mathbf{I}$, the latter becomes $\mathbf{B}\,\mathbf{G}\,\mathbf{G}^{-1}\,\mathbf{e} = \lambda\,\mathbf{G}\,\mathbf{G}^{-1}\,\mathbf{e}$; left multiplying by $\mathbf{G}^{-1}$ shows that $\mathbf{G}^{-1}\,\mathbf{e}$ is an eigenvector of $\mathbf{G}^{-1}\,\mathbf{B}\,\mathbf{G}$ belonging to the same eigenvalue $\lambda$.

As for the trace, for any pair of square matrices $\mathbf{A}$, $\mathbf{B}$ one finds $\mathrm{Tr}\,(\mathbf{A}\,\mathbf{B}) = \sum_{ir} A_{ir}\,B_{ri} = \sum_{ri} B_{ir}\,A_{ri} = \mathrm{Tr}\,(\mathbf{B}\,\mathbf{A})$; it follows

$$\mathrm{Tr}\,(\mathbf{G}^{-1}\,\mathbf{B}\,\mathbf{G}) = \mathrm{Tr}\,(\mathbf{B}\,\mathbf{G}\,\mathbf{G}^{-1}) = \mathrm{Tr}\,(\mathbf{B})\,. \tag{23.22}$$

In Sect. 23.2 the case has been considered where vectors $|u_n\rangle$ are the eigenvectors of a Hermitean operator $\mathscr{A}$, so that $\mathscr{A}\,|u_n\rangle = A_n\,|u_n\rangle$, with $A_n$ the eigenvalue. The definition of the matrix associated to $\mathscr{A}$ has yielded the diagonal matrix $\mathbf{A}$ of (23.10). This case is useful as a starting point for defining the function of an operator; to this purpose, given a function $F(\xi)$ of the variable $\xi$, one constructs a matrix of the

form

$$\mathbf{F}(\mathbf{A}) = \begin{bmatrix} F(A_1) & 0 & \cdots \\ 0 & F(A_2) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}, \tag{23.23}$$

whose $j$th diagonal entry is the function $F$ calculated at $A_j$. In the same reference frame $|u_n\rangle$ in which matrix (23.10) has been constructed, $\mathbf{F}(\mathbf{A})$ is associated to an operator, which is indicated with $F(\mathscr{A})$; the association has the usual form $F_{kn} = \langle u_k | (F(\mathscr{A})\, u_n)\rangle$. Operator $F(\mathscr{A})$, and the corresponding matrix, are not necessary Hermitean (as a simple example one may consider the case where one or more diagonal entries of (23.23) are not real). On the other hand, both $\mathbf{A}$ and $\mathbf{F}(\mathbf{A})$ are diagonal, whence they commute; as a consequence, the same property holds for the corresponding operators:

$$\mathbf{A}\,\mathbf{F}(\mathbf{A}) = \mathbf{F}(\mathbf{A})\,\mathbf{A}\,, \qquad \mathscr{A}\,F(\mathscr{A}) = F(\mathscr{A})\,\mathscr{A}\,. \tag{23.24}$$

In another reference, say, $|v_n\rangle = \mathscr{S}|u_n\rangle$, operator $F(\mathscr{A})$ is represented by the matrix

$$\mathbf{G}(\mathbf{A}) = \mathbf{S}^{\dagger}\,\mathbf{F}(\mathbf{A})\,\mathbf{S} = \mathbf{S}^{-1}\,\mathbf{F}(\mathbf{A})\,\mathbf{S}\,. \tag{23.25}$$

The derivation of (23.25) is similar to that of (23.19).

## 23.5 The Schrödinger Representation

Among the matrices that are associated to operators, it is of interest to consider that associated to the Hamiltonian. To begin, consider a Hermitean operator $\mathscr{A}$ and let $|u_n\rangle$ be the complete set of its eigenvectors. A wave function $|\psi\rangle$ describing the state of a physical system can be expanded in terms of vectors $|u_n\rangle$, this yielding

$$|\psi\rangle = \sum_n a_n\,|u_n\rangle\,, \qquad a_n(t) = \langle u_n|\psi\rangle\,. \tag{23.26}$$

The wave function is the solution of the Schrödinger equation $\mathrm{i}\hbar\,\partial|\psi\rangle/\partial t = \mathscr{H}|\psi\rangle$; a left-scalar multiplication of the latter by $|u_k\rangle$ yields[6]

$$\mathrm{i}\hbar\,\frac{\mathrm{d}}{\mathrm{d}t}\langle u_k|\psi\rangle = \langle u_k|\mathscr{H}|\psi\rangle\,, \qquad \mathrm{i}\hbar\,\frac{\mathrm{d}a_n}{\mathrm{d}t} = \sum_n H_{kn}\,a_n\,, \quad H_{kn} = \langle u_k|\mathscr{H}|u_n\rangle\,. \tag{23.27}$$

The matrix form of (23.27) reads

$$\mathrm{i}\hbar\,\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} a_1 \\ a_2 \\ \vdots \end{bmatrix} = \mathbf{H}_A \begin{bmatrix} a_1 \\ a_2 \\ \vdots \end{bmatrix}, \qquad \mathbf{H}_A = \begin{bmatrix} H_{11} & H_{12} & \cdots \\ H_{21} & H_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}; \tag{23.28}$$

[6] As indicated in the notes at p. 349 and 350, coefficients $a_n$ depend on the choice of the basis $|u_n\rangle$.

it provides a system of infinite differential equations of the first order with respect to time, whose solution, given the initial condition $a_{n0} = a_n(t=0)$, $n = 1,2,\ldots$, provides the time evolution of the system.

The derivation of (23.28) has been carried out without specifying the choice of set $|u_n\rangle$; a sensible choice is using the set $|w_n\rangle$ made of the mutually-orthogonal and normalized eigenvectors of $\mathscr{H}$: in this case it follows

$$|\psi\rangle = \sum_n c_n |w_n\rangle\,, \qquad H_{kn} = \langle w_k|\mathscr{H}|w_n\rangle = \langle w_k|E_n w_n\rangle = E_n\,, \tag{23.29}$$

so that (23.28) takes the form

$$\mathrm{i}\hbar\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} c_1 \\ c_2 \\ \vdots \end{bmatrix} = \mathbf{H}\begin{bmatrix} c_1 \\ c_2 \\ \vdots \end{bmatrix}, \qquad \mathbf{H} = \begin{bmatrix} E_1 & 0 & \cdots \\ 0 & E_2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}. \tag{23.30}$$

From (23.30) it follows that in the $|w_n\rangle$ reference the equations are independent, and their solution is $c_n(t) = c_{n0}\exp(-\mathrm{i}E_n t/\hbar)$ with $c_{n0} = c_n(t=0)$; in matrix form,

$$|\psi\rangle = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} \exp(-\mathrm{i}E_1 t/\hbar) & 0 & \cdots \\ 0 & \exp(-\mathrm{i}E_2 t/\hbar) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}\begin{bmatrix} c_{10} \\ c_{20} \\ \vdots \end{bmatrix}. \tag{23.31}$$

Remembering definition (23.23) of a matrix whose entries are functions of the entries of another matrix, one defines

$$\mathbf{S}(\mathbf{H}) = \begin{bmatrix} \exp(-\mathrm{i}E_1 t/\hbar) & 0 & \cdots \\ 0 & \exp(-\mathrm{i}E_2 t/\hbar) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}. \tag{23.32}$$

As the diagonal entries of $\mathbf{S}$ are complex, $\mathbf{S}$ is not Hermitean; also, it is

$$\mathbf{S}^\dagger = \begin{bmatrix} \exp(\mathrm{i}E_1 t/\hbar) & 0 & \cdots \\ 0 & \exp(\mathrm{i}E_2 t/\hbar) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}, \qquad \mathbf{S}^\dagger\mathbf{S} = \mathbf{S}\mathbf{S}^\dagger = \mathbf{I}, \tag{23.33}$$

namely, $\mathbf{S}$ is unitary. Assume that at $t = 0$ the wave function is normalized to unity,

$$\langle\psi(t=0)|\psi(t=0)\rangle = \sum_n |c_{n0}|^2 = 1\,; \tag{23.34}$$

thanks to (23.31), the vector $c_n$, that defines $|\psi\rangle$ at any instant of time, is obtained by multiplying the initial vector $c_{n0}$ by $\mathbf{S}$; as the latter is unitary, the norm of $|\psi\rangle$ is conserved in time,

$$\langle\psi|\psi\rangle = \sum_n |c_n|^2 = \sum_n |c_{n0}|^2 = 1\,. \tag{23.35}$$

Therefore, the time variation induced by $\mathbf{S}$ on $|\psi\rangle$ is a rotation; this type of description of the state of a system, in which the reference (namely, $|w_n\rangle$) and the operator are fixed, while the wave function evolves with respect to time, is called *Schrödinger representation* or *Schrödinger picture*.

The short-hand notation[7] for (23.32) is $\mathbf{S} = \exp(-\mathrm{i}\mathbf{H}t/\hbar)$; the time derivative of $\mathbf{S}$ is

$$\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}t} = -\frac{\mathrm{i}}{\hbar}\begin{bmatrix} E_1\exp(-\mathrm{i}E_1 t/\hbar) & 0 & \cdots \\ 0 & E_2\exp(-\mathrm{i}E_2 t/\hbar) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} = \frac{\mathbf{H}\mathbf{S}}{\mathrm{i}\hbar} = \frac{\mathbf{S}\mathbf{H}}{\mathrm{i}\hbar}\,. \tag{23.36}$$

Remembering the association between operators and matrices outlined in Sect. 23.2, one associates to $\mathbf{S}$ a unitary operator $\mathscr{S}$ such that

$$\mathscr{S} = \exp(-\mathrm{i}\mathscr{H}t/\hbar)\,, \qquad |\psi\rangle = \mathscr{S}\,|\psi(t=0)\rangle\,. \tag{23.37}$$

## 23.6 Outer Product and Tensor Product

Consider the case where the matrices under consideration are not square; the multiplication of two such matrices is possible if the number of columns of the left matrix equals the number of rows of the right matrix: specifically, if $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{B}$ is an $n \times q$ matrix, then $\mathbf{C} = \mathbf{A}\mathbf{B}$ is an $m \times q$ matrix. Considering the simple example where $m = 3$, $n = 2$, $q = 2$, and assuming that the entries are complex,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}\,, \qquad \mathbf{B} = \begin{bmatrix} b_{11}^* & b_{12}^* \\ b_{21}^* & b_{22}^* \end{bmatrix}\,, \tag{23.38}$$

yields

$$\mathbf{C} = \mathbf{A}\mathbf{B} = \begin{bmatrix} a_{11}b_{11}^* + a_{12}b_{21}^* & a_{11}b_{12}^* + a_{12}b_{22}^* \\ a_{21}b_{11}^* + a_{22}b_{21}^* & a_{21}b_{12}^* + a_{22}b_{22}^* \\ a_{31}b_{11}^* + a_{32}b_{21}^* & a_{31}b_{12}^* + a_{32}b_{22}^* \end{bmatrix}\,. \tag{23.39}$$

A special case occurs when $n = 1$, which results in

$$\mathbf{A} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}\,, \quad \mathbf{B} = \begin{bmatrix} b_1^* & b_2^* \cdots b_p^* \end{bmatrix}\,, \quad \mathbf{C} = \begin{bmatrix} a_1 b_1^* & a_1 b_2^* & \cdots & a_1 b_p^* \\ a_2 b_1^* & a_2 b_2^* & \cdots & a_2 b_p^* \\ \vdots & \vdots & \ddots & \vdots \\ a_m b_1^* & a_m b_2^* & \cdots & a_m b_p^* \end{bmatrix}\,. \tag{23.40}$$

Comparing with the definitions introduced in Sect. 23.1, one finds that an $n = 1$ case like (23.40) can be thought of as the product of a bra and a ket vector, in

[7] More details about the exponential of a matrix are given in Sect. 23.13.1 of the Complements.

which the ket vector is placed on the left of the bra vector. This type of product is also called *outer product* of the two vectors;[8] remembering the symbols of Sect. 23.1, and assuming that vector **A** represents $|\psi\rangle$ and vector **B** represents $\langle\phi|$, such a product is also indicated with the notation

$$\mathbf{A}\,\mathbf{B} = |\psi\rangle\,\langle\phi|\,. \tag{23.41}$$

Consider now the more general case where **A** is an $m \times n$ matrix and **B** is a $p \times q$ matrix, with no relations among the numbers $m$, $n$, $p$, $q$; one defines the *tensor product* of the two matrices as

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\,\mathbf{B} & a_{12}\,\mathbf{B} & \cdots & a_{1n}\,\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\,\mathbf{B} & a_{m2}\,\mathbf{B} & \cdots & a_{mn}\,\mathbf{B} \end{bmatrix}. \tag{23.42}$$

it is easily found that (23.42) is an $(m\,p) \times (n\,q)$ matrix. The tensor product is also called *Kronecker product*. Given four matrices **A**, **B**, **C**, **D** such that the products **AC** and **BD** exist, the following holds [39]

$$(\mathbf{A} \otimes \mathbf{B})\,(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{A}\,\mathbf{C}) \otimes (\mathbf{B}\,\mathbf{D})\,. \tag{23.43}$$

Comparing (23.42) with (23.40) shows that the tensor product coincides with the outer product when $p = n = 1$. Another interesting case occurs when $n = q = 1$, namely, when **A** and **B** are column vectors; one finds

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} \otimes \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix} = \begin{bmatrix} a_1\,b_1 \\ a_1\,b_2 \\ \vdots \\ a_m\,b_p \end{bmatrix}. \tag{23.44}$$

Assuming that vector **A** represents $|\psi\rangle$ and vector **B** represents $|\phi\rangle$, their tensor product is also indicated with the notation

$$|\psi\rangle \otimes |\phi\rangle = |\psi\,\phi\rangle\,. \tag{23.45}$$

## 23.7 Quantum-Mechanical Ensemble

Given the initial condition $|\psi(t=0)\rangle$, the wave function is determined by the Schrödinger equation $\mathrm{i}\hbar\,\partial\psi/\partial t = \mathscr{H}\psi$. If the Hamiltonian operator and the initial condition are known, the time evolution of the wave function at all spatial points, that is, the quantum state of the system, is completely determined. On the other hand, it may happen that the information about the system is insufficient for a complete

[8] By the same token, the scalar product defined in (23.3) is also termed *inner product*.

specification of a precise state. In this case, there are several wave functions that describe the state of the system in a way that is compatible with the available information. As in the classical case outlined in Sect. 1.7, the set of such wave functions is called *ensemble*. They provide a set of non-interacting replicas of the system, and will be labelled by a discrete index $e$.

It is useful to introduce a function similar to the ensemble density of the classical case. For this, one may start by considering the expansion of a wave function of the ensemble into a complete set of orthonormal functions $|u(\mathbf{k})\rangle$, where $\mathbf{k}$ instead of $n$ is used as label.[9] Remembering (23.26),

$$|\psi_e\rangle = \sum_{\mathbf{k}} a_e(\mathbf{k},t)\,|u(\mathbf{k})\rangle\,, \qquad a_e(\mathbf{k},t) = \langle u(\mathbf{k})|\psi_e\rangle\,. \tag{23.46}$$

In (23.46), the discrete vector $\mathbf{k}$ that labels the functions $u$ has $s$ components $k_1, \ldots k_s$, and the sum is performed over the $s$-fold set of such components. Functions $u$ and $a_e$ fulfill the relations[10]

$$\langle u(\mathbf{h})|u(\mathbf{k})\rangle = \delta[\mathbf{h}-\mathbf{k}]\,, \qquad \sum_{\mathbf{k}} |a_e(\mathbf{k},t)|^2 = 1\,, \tag{23.47}$$

the second of which derives from the normalization condition $\langle\psi_e|\psi_e\rangle = 1$. Due to such a condition, the quantity $0 \leq |a_e(\mathbf{k},t)|^2 \leq 1$ is the probability that the system described by $\psi_e$ is found at the time $t$ in the state labelled by $\mathbf{k}$. Note that the ensemble index $e$ is ascribed only to the coefficients of the expansion (23.46). In fact, the same set of functions $u$ is used to expand all members $\psi_e$ of the ensemble.

## 23.8 Expectation Value

Let $\mathscr{A}$ be a Hermitean operator, to which a matrix $\mathbf{A}$ is associated like in (23.4), namely, $A_{\mathbf{kh}} = \langle u(\mathbf{h})|\mathscr{A}|u(\mathbf{k})\rangle$. The *expectation value* of $\mathscr{A}$ with respect to the wave function $\psi_e$ is given by

$$\langle A\rangle_e = \langle\psi_e|\mathscr{A}|\psi_e\rangle = \sum_{\mathbf{kh}} a_e^*(\mathbf{h},t)\,a_e(\mathbf{k},t)\,A_{\mathbf{kh}}\,, \tag{23.48}$$

where (23.46) has been used. If functions $u$ happen to be the eigenfunctions of $\mathscr{A}$, matrix $\mathbf{A}$ becomes diagonal due to the orthonormality condition (23.47), and its diagonal entries coincide with the eigenvalues of $\mathscr{A}$. Moreover, if the system described by the wave function $\psi_e$ is in the state labelled, say, by $\mathbf{h}$, it follows

[9] The number of dimensions of $\mathbf{k}$ equals the number of degrees of freedom of the problem in hand.

[10] The Kronecker delta whose argument is the difference between to vectors is equivalent to $\delta[\mathbf{h}-\mathbf{k},\mathbf{0}]$, where $\mathbf{0}$ is the null vector of the space of $\mathbf{h}$ and $\mathbf{k}$. With this provision, it is $\delta[\mathbf{h}-\mathbf{k}] = 1$ if $\mathbf{k} = \mathbf{h}$, while it is $\delta[\mathbf{h}-\mathbf{k}] = 0$ otherwise.

$|\psi_e\rangle = a_e(\mathbf{h},t)\,|u(\mathbf{h})\rangle$ whence $|a_e(\mathbf{k},t)|^2 = \delta[\mathbf{h}-\mathbf{k}]$. As a consequence, the expectation value $\langle A\rangle_e$ coincides with the eigenvalue of $\mathscr{A}$ labelled by $\mathbf{h}$.

Assuming that $\mathscr{A}$ is independent of time, a standard calculation yields[11]

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle A\rangle_e = \frac{\mathrm{i}}{\hbar}\langle\psi_e|\,[\mathscr{H},\mathscr{A}]\,|\psi_e\rangle\,, \qquad [\mathscr{H},\mathscr{A}] = \mathscr{H}\,\mathscr{A} - \mathscr{A}\,\mathscr{H}\,. \tag{23.49}$$

Eq. (23.49) is the quantum-mechanical analogue of the classical rate of change of a dynamical quantity expressed in terms of the Poisson parentheses. In fact, consider a function $\sigma$ defined in the $\gamma$ space, with no explicit dependence on time (this condition corresponds to the time independence of the operator $\mathscr{A}$); Eq. (1.46) then yields $\mathrm{d}\sigma/\mathrm{d}t = \{\sigma, H\}$, which has the same form as (23.49).

[11] The definition (23.48) of the expectation value can be generalized by using two different wave functions of the ensemble; observing that the time evolution for two wave functions belonging to the same ensemble is given by the same Hamiltonian operator, one finds $\langle A\rangle_{ef} = \langle\psi_e|\mathscr{A}|\psi_f\rangle$, with $\mathrm{d}\langle A\rangle_{ef}/\mathrm{d}t = (\mathrm{i}/\hbar)\,\langle\psi_e|\,[\mathscr{H},\mathscr{A}]\,|\psi_f\rangle$.

lecture -36 (continued)

# Chapter 24
# M. Rudan — Quantum Computing

Following [5], the main aspects of quantum mechanics relevant in the field of information theory are *superposition of states*, *interference*, *correlation*, *non-clonability*, and *uncertainty*.[1]

Extending to the case of $s$ bits the result found in Sect. 1.13, the energy cost of the unmindful erasure of $s$ bits of known information is $sL = s k_B T \log 2$ (compare with (1.89)); it follows that a computational process, able to perform a one-to-one mapping of the old states into the new ones, would not produce any erasure and could therefore constitute the basis for an isoentropic process. On the other hand, such a mapping uniquely determines the input logic states starting from the output logic states, that is, it is logically reversible.

The possibility of physically-reversible processes is inherent in the time-reversible nature of the Hamiltonian formulation of Classical Mechanics, and in the unitary time-evolution operators of Quantum Mechanics. In practice, implementing a reversible computation requires a precise control of the physical mechanisms by which the computation is implemented, so that the uncertainty of the physical state of the mechanisms becomes negligible. In so doing, the energy involved in a computation step can be recovered and reused in the next step, instead of being dissipated in the form of heat.

In the following it will be shown that logically-reversible, universal gates can in theory be implemented. To achieve the practical implementation, the following steps are necessary:

1. Associate to each gate one or more quantum-mechanical operators.
2. Seek for physical systems that are described by the same operators.
3. Use such systems to implement the quantum gates.
4. Identify problems whose solution is made easier by quantum calculations.
5. For each implementation, check the conditions that make thermodynamic reversibility possible.

[1] The outline of Sects. 5.9, 5.10, and 24.1.1 is based on [60], that of Sects. 5.8, 24.7 is based on [2].

## 24.1 Quantum Parallelism

The main characteristic of quantum computation is given by the quantum parallelism, that is, the ability of carrying out a superposition of computations. As shown by the example of the algorithms described in Sects. 3.18 and 3.19, this characteristic makes such algorithms much superior to the classical ones. It must be remarked that, to date, there is no general recipe that shows how to exploit the quantum parallelism to tackle a general class of problems: for each problem, an *ad hoc* implementation must be found.

### *24.1.1 An Example of Quantum Parallelism*

As an example of parallelism, one may consider a computation as an operation that transforms the content of an input register into the content of an output register. Assume the the length of both the input and output registers is $n$: if one associates the input and output registers to vectors, the computation may be viewed as an operator that transforms one vector into another; if the transformation is unitary, the length of the vector during the transformation is left unchanged. This is equivalent to associating to each element $x_i$ of the input register, $i = 1,\ldots,n$, the corresponding element $f(x_i)$ of the output register; in other terms, *all values* of a function are calculated by applying the transformation *only once*. This marvelous outcome has some limitations: in fact, from the standpoint of quantum mechanics it is impossible to carry out a measurement able to extract all values $f(x_1), f(x_2), \ldots$. However, as shown by the example below, it is possible to carry out some types of measurements that allow one to extract *global properties* of the $f(x_1), f(x_2), \ldots$.

Assume that $x_1, x_2, \ldots$ are a finite set of integers, and that $f(x_1), f(x_2), \ldots$ belong to the same set; then, one wants to recognize whether $f$ is constant or not. For solving the problem using a classical computer, it is necessary to calculate all outputs $f(x_i)$ and compare them (remembering that the input and output data belong to the same set, one may assume that the possible values of $x_i$ and $f$ are 0 and 1); in the simple case $n = 2$, with $x_1 = 0$ and $x_2 = 1$, the calculation amounts to determining $f(0)$ and $f(1)$, that is, a classical computer must be used twice. It is also obvious that, if the classical computer were used only once, the outcome would be insufficient to determine whether $f$ is constant or not. To proceed, one observes that for the problem in hand there are four possible combinations of values:

| $A$ | $B$ | $C$ | $D$ |
|---|---|---|---|
| $f(0) = 0$<br>$f(1) = 0$ | $f(0) = 1$<br>$f(1) = 1$ | $f(0) = 0$<br>$f(1) = 1$ | $f(0) = 1$<br>$f(1) = 0$ |

(24.1)

Combinations $A$ and $B$ in (24.1) correspond to the outcome "equal", whereas combinations $C$ and $D$ correspond to the outcome "different". If the same problem is tack-

led using a quantum algorithm, it can be demonstrated that a single measurement carried out on the combination of the input and output registers has a probability equal to $1/2$ to yield the result "equal" or "different"; the other half of the cases correspond to a non-significant outcome, that indicates that algorithm has failed and that the computation must be repeated [14]. The corresponding table is (24.2):

| $A$ | $B$ | $C$ | $D$ |
|---|---|---|---|
| equal | failed | different | failed |

(24.2)

In the example above, the quantum parallelism allows one to use a single computation to determine, with probability $1/2$, whether the values of $f$ are equal or different; however, it does not allow to determine $f(0)$ and $f(1)$ independently. This conclusion seems unsatisfactory: after all, it is true that the classical solution requires two calculations instead of one, but, on the other hand, the quantum calculation provides a useful answer only in 50% of cases; due to this, the time required in the average by the quantum solution is the same as that of the classical solution.

Consider, however, another example, taken from [60]: one must perform a calculation that is crucial with respect to some decision to be taken, like, e.g., investing in the stock market on a day-by-day basis; the decision must be taken within 24 hours, and a single calculation of $f$ takes almost 24 hours, because $n$ is very large. Obviously, a classical computer would in this case be useless, whereas a quantum computer would, at least, provide a sensible answer one day out of two.

## 24.2 Evaluating a Function as a Whole

The idea of evaluating a function as a whole, introduced in Sect. 24.1, naturally connects this analysis with typical features of quantum mechanics. Remembering 23.2, a wave functions describing the state of a particle or of a system can be expressed as

$$\psi = c_1 w_1 + c_2 w_2 + \cdots , \tag{24.3}$$

where functions $w_k$ are the (mutually-orthogonal) eigenfunctions of the operator $\mathscr{A}$ associated to a dynamic variable $A$. Assuming that the wave function is normalizable, namely, that $||\psi||^2 = \langle\psi|\psi\rangle < \infty$, it follows that $|c_k|^2$ is proportional to the probability that a measurement finds the particle or the system in state $A_k$. In computation it suffices to consider systems with two states $w_1$ and $w_2$:

$$\psi = \alpha w_1 + \beta w_2 , \qquad |\alpha|^2 + |\beta|^2 = ||\psi||^2 . \tag{24.4}$$

As shown in Sect. 24.5, examples of the two states are the polarization directions of a photon (vertical or horizontal polarization) and the orientations of an electron spin ("spin up" or "spin down"). As indicated still in Sect. 24.5, a *qubit* is similar to a classical bit in that it can take on the states 0 or 1, but it differs from a bit in that it can also take on a continuous range of values representing a superposition of states

like (24.4). A *quantum logic gate* (or *quantum gate*) is a physical object performing logical operations on a qubit or on a small number of qubits; examples of quantum gates are those given in Sect. 24.4. Connected quantum gates form *quantum circuits*; when the superposition of states is exploited to carry out calculations, it is also called *quantum parallelism*.

An important point, already noted at the beginning of this chapter, is that the equations of quantum mechanics are reversible with respect to time; it follows that, when dealing with quantum gates, one must consider the issue of logical reversibility and thermodynamic reversibility (this point is also addressed in Sects. 5.8 and 24.4). Classical gates consume energy: extending the Landauer limit $L$ (1.89) to the case of $s$ bits shows that the minimum energy consumed to erase $s$ bits is $sL = s k_B T \log(2)$, with $T$ the temperature of the heat sink surrounding the device. The classical AND, OR, NAND, NOR gates (Sect. 5.3), and their combinations, erase bits because these gates are logically irreversible; in fact, some of their output values are such that the input values can not be reconstructed.[2] As outlined in Sect. 24.4, it is possible to realize logically-reversible gates by preventing the bit erasure, namely, by obtaining gates in which the number of output variables is equal to that of the input variables, and reconstruction of inputs is possible.

## 24.3 A Useful Representation of the Qubit

Starting from the expression (24.4) of the qubit, one lets

$$\alpha = a \exp(\mathrm{i}\, p)\,, \quad \beta = b \exp(\mathrm{i}\, q)\,, \quad \varphi = q - p\,, \quad 0 \leq \varphi < 2\pi\,, \tag{24.5}$$

with $a$, $b$, $p$, $q$ real quantities, $a$, $b$ non-negative. It follows, thanks to the orthonormality of $w_1$, $w_2$,

$$\psi \exp(-\mathrm{i}\, p) = a\, w_1 + b \exp(\mathrm{i}\, \varphi)\, w_2\,, \qquad ||\psi||^2 = a^2 + b^2\,. \tag{24.6}$$

Then, one defines

$$\Phi = \frac{\psi}{||\psi||} \exp(-\mathrm{i}\, p) = \frac{a}{||\psi||}\, w_1 + \frac{b}{||\psi||} \exp(\mathrm{i}\, \varphi)\, w_2\,, \tag{24.7}$$

and introduces the symbols, with $0 \leq \theta \leq \pi$,

$$\cos\left(\frac{\theta}{2}\right) = \frac{a}{||\psi||}\,, \qquad \sin\left(\frac{\theta}{2}\right) = \frac{b}{||\psi||}\,. \tag{24.8}$$

In conclusion, the new representation of (24.4) becomes

[2] The NOT operator is logically reversible, but its standard implementation consumes energy.

$$\Phi = \cos\left(\frac{\theta}{2}\right) w_1 + \sin\left(\frac{\theta}{2}\right) \exp(\mathrm{i}\,\varphi)\, w_2\,, \tag{24.9}$$

where $w_1$ and $w_2$ are given. It follows that $\Phi$ is a vector of unit length whose orientation is defined by the $\theta$ and $\varphi$ angles. By changing such angles within their full range, the tip of the $\Phi$ vector spans a sphere called *Bloch sphere* (the latter is shown in Fig. 24.12 after introducing new symbols).

## 24.4 Logically-Reversible Gates

Coming back to the logic gates illustrated so far it is found that, in contrast to the other operators, the NOT gate is logically reversible. For this reason, it is represented with a new symbol to remind one of this property (Fig. 24.1). To proceed, it is necessary to consider two additional logic functions, namely, the *Fan-out* operator (Fig. 24.2) and the *Exchange* operator (shown in Fig. 24.3 along with the corresponding truth table). Both the Fan-out and Exchange functions are reversible.


**Fig. 24.1** Symbol of the NOT operator reminding of logic reversibility

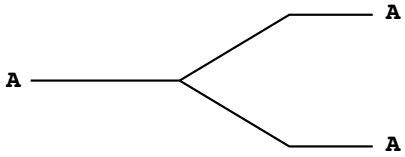
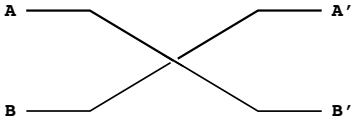



**Fig. 24.2** Symbol of the Fan-out operator

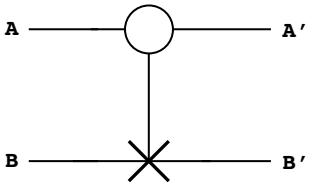



| A | B | A′ | B′ |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 0 | 1 | 1 | 0 |
| 1 | 0 | 0 | 1 |
| 1 | 1 | 1 | 1 |

**Fig. 24.3** Symbol and truth table of the Exchange operator

In standard circuits, the Fan-out and Exchange operators are simply implemented by wires. This is not possible in quantum computing due to dissipation; therefore, special gates performing these logic operations must be sought. As a simple example where the coeherence of the wave function is kept, one may consider the system illustrated in Sect. 3.17.1 of the Complements.

To proceed, one observes that a logically-reversible operation between a pair $A,B$ of input bits and a pair $A',B'$ of output bits is achievable by the logic operator, called *Controlled NOT* (CNOT), whose symbol is shown in Fig. 24.4 along with the corresponding truth table. The latter shows that the input bit $A$ is passed unchanged to the output bit $A'$; instead, $B$ is complemented, namely, $B'=\overline{B}$, if $A=1$, whereas it is left unchanged if $A=0$. Logic reversibility is apparent from the one-to-one correspondence of the input and output pairs; also, it is $B'=A\oplus B$.




| A | B | A′ | B′ |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 1 |
| 1 | 0 | 1 | 1 |
| 1 | 1 | 1 | 0 |

**Fig. 24.4** Symbol and truth table of the CNOT operator

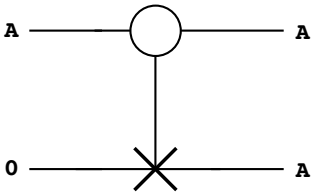



**Fig. 24.5** Use of the CNOT operator to implement the Fan-out function

The CNOT operator lends itself to implementing the Fan-out function (Fig. 24.5) and the Exchange function (Fig. 24.6). The former is intuitive: its functioning is described by the first and third lines of the truth table of Fig. 24.4; the latter is better understood with the aid of the truth table of Fig. 24.7.

Another logically-reversible operator, whose input and output are triads of logic variables ($A,B,C$ and $A',B',C'$, respectively) is shown in Fig. 24.8 along with its truth table. Its functioning is such that $A'=A$, $B'=B$, namely, $A$ and $B$ are passed

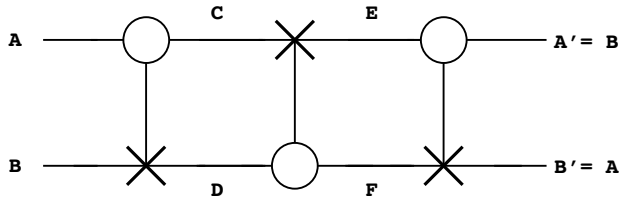
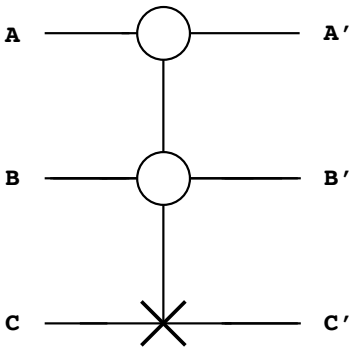



**Fig. 24.6** Use of the CNOT operator to implement the Exchange function

| A | B | C | D | E | F | A′ | B′ |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 |
| 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 |
| 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 |

**Fig. 24.7** Truth table of the Exchange operator implemented with the CNOT operator.

unchanged to the output; instead, $C$ is complemented ($C' = \overline{C}$) if $A = B = 1$, otherwise $C$ is left unchanged. This operator is called *Controlled controlled NOT* (CCNOT), or also *Toffoli gate*.




| A | B | C | A′ | B′ | C′ |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 | 1 |
| 0 | 1 | 0 | 0 | 1 | 0 |
| 0 | 1 | 1 | 0 | 1 | 1 |
| 1 | 0 | 0 | 1 | 0 | 0 |
| 1 | 0 | 1 | 1 | 0 | 1 |
| 1 | 1 | 0 | 1 | 1 | 1 |
| 1 | 1 | 1 | 1 | 1 | 0 |

**Fig. 24.8** Symbol and truth table of the CCNOT operator

The truth table of Fig. 24.8 shows that $C' = AB$ when $C = 0$, whereas it is $C' = \overline{AB}$ when $C = 1$. It follows that the CCNOT operator provides both the AND and NAND functions. Remembering that the NAND operator is universal, every logic function can be sinthesized, in a reversible manner, using the CCNOT operator.

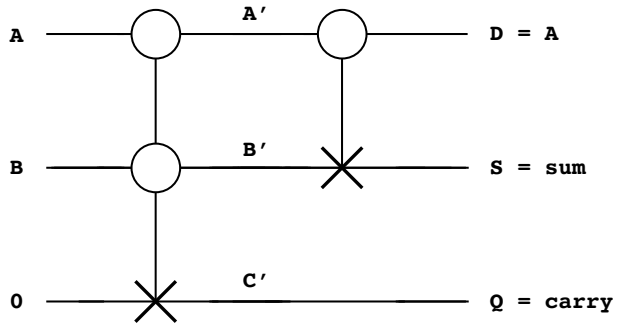



**Fig. 24.9** Implementation of the reversible half adder using a combination of the CCNOT and CNOT operators

### *24.4.1 Reversible Half Adder and Full Adder*

A suitable combination of the CCNOT and CNOT operators provides a reversible half adder (Fig. 24.9); its functioning is better understood by observing the truth table of Fig. 24.10, in which the lines of interest are those corresponding to $C = 0$. The relations among the variables are

$$A' = A\,, \qquad S = B' = A \oplus B\,, \qquad Q = C' = AB\,. \tag{24.10}$$

| A | B | C | A′ | B′ | C′ | D | S | Q |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 |
| 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 |
| 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 |
| 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 |

**Fig. 24.10** Truth table of the reversible half adder of Fig. 24.9.

The reversible full adder is obtained by replicating the combinations of the CCNOT and CNOT operators (Fig. 24.11). The relations among the variables are

$$A' = A\,, \qquad S = A \oplus B \oplus C = P \oplus C\,, \tag{24.11}$$

$$C = \overline{G}PC + G\overline{PC} = G \oplus PC = G + PC\,. \tag{24.12}$$

Like in the case of the half adder, one may think of associating to the full adder an operator $\mathscr{M}$ made of the successive applications of five unitary operators,

$$\mathscr{M} = \mathscr{A}_5\, \mathscr{A}_4\, \mathscr{A}_3\, \mathscr{A}_2\, \mathscr{A}_1\,, \tag{24.13}$$

each associated to the elementary gates of Fig. 24.11. Again, the operators at the right hand side operate in backwards order ($\mathscr{A}_1$ first).

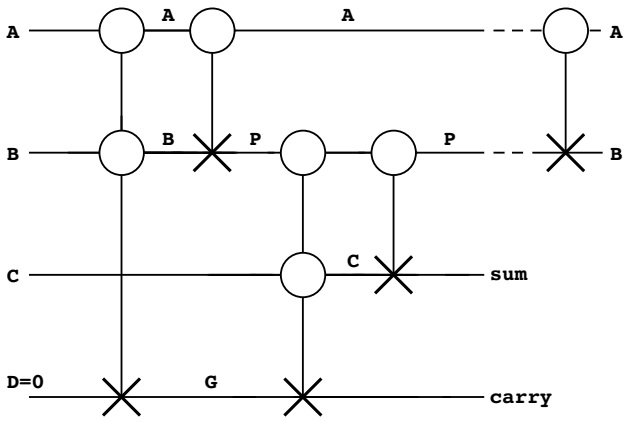



**Fig. 24.11** Implementation of the reversible full adder using a combination of the CCNOT and CNOT operators

## 24.5 Qubits and their Representation

In principle, any quantum-mechanical system with two quantum states can be used for quantum computing; one may in fact associate the logic values 0 and 1 to the two quantum states. These two physical states are called *quantum bits* or *qubits*, and can be represented as two-component vectors,

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \langle 0|0\rangle = \langle 1|1\rangle = 1\,, \quad \langle 0|1\rangle = \langle 1|0\rangle = 0\,. \tag{24.14}$$

With respect to a classical system, two major differences are present:

1. The state of a quantum physical system can be represented by a superposition; in the case of a two-qubit system, remembering (24.4),

$$\psi = \alpha\, w_1 + \beta\, w_2\,, \qquad |\alpha|^2 + |\beta|^2 = ||\psi||^2\,. \tag{24.15}$$

2. A measurement carried out on the quantum system will force it to collapse to one of its eigenstates; in the case of a two-qubit system, either $w_1$ or $w_2$; besides a measurement, also an unwanted interaction with the environment can destroy the coherence of the quantum state.

The representation of the qubit introduced in Sect. 24.3 is easily adapted to the new nomenclature introduced here by defining the symbols

$$|\Phi\rangle = \Phi\,, \qquad |0\rangle = w_1\,, \qquad |1\rangle = w_2\,. \tag{24.16}$$

This yields a representation of the Bloch sphere in the form

$$|\Phi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + \sin\left(\frac{\theta}{2}\right)\exp(\mathrm{i}\,\varphi)\,|1\rangle\,, \tag{24.17}$$

where $|0\rangle$ and $|1\rangle$ are given. One notes that the condition $\theta = 0$ provides $|\Phi\rangle = |0\rangle$ for any $\varphi$, while the condition $\theta = \pi$, $\varphi = 0$ provides $|\Phi\rangle = |1\rangle$; these two conditions correspond, respectively, to the "north pole" and "south pole" of the Bloch sphere (Fig. 24.12). Similarly, the condition $\theta = \pi/2$, $\varphi = 0$ provides $|\Phi_p\rangle =$

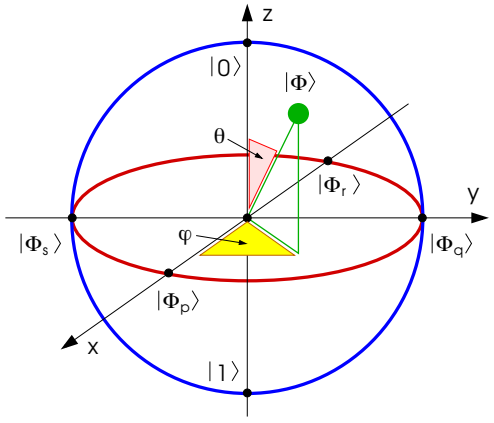



**Fig. 24.12** The Bloch sphere

$(|0\rangle + |1\rangle)/\sqrt{2}$, and the condition $\theta = 3\pi/2$, $\varphi = 0$ provides $|\Phi_r\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$. Finally, the condition $\theta = \pi/2$, $\varphi = \pi/2$ provides $|\Phi_q\rangle = (|0\rangle + \mathrm{i}\,|1\rangle)/\sqrt{2}$, and the condition $\theta = \pi/2$, $\varphi = 3\pi/2$ provides $|\Phi_s\rangle = (|0\rangle - \mathrm{i}\,|1\rangle)/\sqrt{2}$. These points are also marked in the figure.

It is useful to identify the operator $\mathscr{B}$ such that its eigenstates are (24.14). Letting **B** the matrix corresponding to $\mathscr{B}$, it is easily found that[3]

[3] It is found by inspection that **B** and $\mathscr{B}$ are Hermitean.

$$\mathbf{B} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} . \tag{24.18}$$

In fact,

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0 \begin{bmatrix} 1 \\ 0 \end{bmatrix} , \qquad \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 1 \begin{bmatrix} 0 \\ 1 \end{bmatrix} , \tag{24.19}$$

namely, $\mathscr{B}|0\rangle = 0\,|0\rangle$, $\mathscr{B}|1\rangle = 1\,|1\rangle$; in other terms, 0 and 1 are the eigenvalues of $\mathscr{B}$ corresponding to the eigenstates $|0\rangle$ and $|1\rangle$, respectively. If the system is in one of the two eigenstates, its value is the corresponding qubit; if, instead, the system is in a superposition of states like (24.17), a measurement of the qubit yields 0 with probability $\cos^2(\theta/2)$ and 1 with probability $\sin^2(\theta/2)$.

The expectation value of $\mathscr{B}$ is $\langle B\rangle = \langle \Phi|\mathscr{B}|\Phi\rangle$, namely, letting $a = \cos(\theta/2)$, $b = \sin(\theta/2)\exp(\mathrm{i}\varphi)$, and using (24.9) to calculate $\mathscr{B}|\psi\rangle = \mathscr{B}(a\,|0\rangle + b\,|1\rangle) = a\,\mathscr{B}|0\rangle + b\,\mathscr{B}|1\rangle = b\,|1\rangle$,

$$\langle B\rangle = a^*\,\langle 0|b|1\rangle + b^*\,\langle 1|b|1\rangle = a^*\,b\,\langle 0|1\rangle + b^*\,b\,\langle 1|1\rangle = |b|^2 = \sin^2(\theta/2)\,. \tag{24.20}$$

### *24.5.1 Photon Polarization*

As an example of a quantum-mechanical system with two quantum states, consider a monochromatic radiation travelling along the $z$ direction; the matrix representations of the $x$- and $y$-polarized states of the photons are

$$|\phi_x\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} , \qquad |\phi_y\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix} . \tag{24.21}$$

Any linear combination of the above, $|\phi\rangle = a\,|\phi_x\rangle + b\,|\phi_y\rangle$ can be written as

$$\begin{bmatrix} a \\ b \end{bmatrix} = a \begin{bmatrix} 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \end{bmatrix} , \qquad |a|^2 + |b|^2 = 1\,. \tag{24.22}$$

The density matrix associated to this example is found from the definition (23.73) of the density matrix of a pure state, and reads

$$\mathbf{R} = \begin{bmatrix} a\,a^* & a\,b^* \\ b\,a^* & b\,b^* \end{bmatrix} . \tag{24.23}$$

If the photon is in the $x$-polarized state, it is $a = 1$, $b = 0$; if, conversely, the photon is in the $y$-polarized state, it is $a = 0$, $b = 1$. The density matrices corresponding to these two cases are

$$\mathbf{R}_x = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} , \qquad \mathbf{R}_y = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} ; \tag{24.24}$$

they represent pure states, as shown by the property $\mathbf{R}_x^2 = \mathbf{R}_x$, $\mathbf{R}_y^2 = \mathbf{R}_y$. If the photon is in the 45°-polarized state, it is $a = b = 1/\sqrt{2}$; if, conversely, the photon is in the 135°-polarized state, it is $-a = b = 1/\sqrt{2}$. The density matrices corresponding to these two cases are

$$\mathbf{R}_{45} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \qquad \mathbf{R}_{135} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \tag{24.25}$$

again fulfilling the property $\mathbf{R}_{45}^2 = \mathbf{R}_{45}$, $\mathbf{R}_{135}^2 = \mathbf{R}_{135}$. An equal-weight mixture of $x$- and $y$-polarized states, and an equal-weight mixture of 45°- and 135°-polarized states have, respectively, a density matrix of the form

$$\mathbf{R}' = \frac{1}{2}\mathbf{R}_x + \frac{1}{2}\mathbf{R}_y = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \mathbf{R}'' = \frac{1}{2}\mathbf{R}_{45} + \frac{1}{2}\mathbf{R}_{135} = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{24.26}$$

As $\mathbf{R}'' = \mathbf{R}'$, the physical effects are the same.

### *24.5.2 Electron Spin*

In contrast with other dynamic quantities, there is no classical counterpart of spin. Therefore, the latter can not be derived from the expression of a dynamic variable by replacing conjugate coordinates with suitable operators. It can be shown that the eigenvalues of spin are derived in a manner similar to that of angular momentum: this leads to determining the square modulus of spin, and its component along one of the coordinate axes, say, $z$. Their values are given by

$$S^2 = \hbar^2 s(s+1), \qquad S_z = \hbar s_z. \tag{24.27}$$

The important difference with respect to the case of angular momentum is that $s$, instead of being a non-negative integer, is a non-negative half integer: $s = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \ldots$; in turn, $s_z$ can take the $2s+1$ values $-s, -s+1, \ldots, s-1, s$.

The introduction of spin must be accounted for in the expression of the wave function: the latter, in the case of a single particle, must be indicated with $\psi(\mathbf{r}, s_z, t)$, and its normalization to unity, if existing, is expressed by

$$\sum_{s_z} \int_\Omega |\psi(\mathbf{r}, s_z, t)|^2 \, \mathrm{d}^3 r = 1. \tag{24.28}$$

If (24.28) holds, the product $|\psi(\mathbf{r}, s_z, t)|^2 \, \mathrm{d}^3 r$ is the probability that at time $t$ the particle is in the elementary volume $\mathrm{d}^3 r$ centered on $\mathbf{r}$, and the component of its spin along the $z$ axis is $S_z = \hbar s_z$.

The connection between spin and boson-like or fermion-like behavior is the following: the quantum number $s$ is integer for bosons, half integer for fermions [47]. It is

then meaningful to use the terms "boson" or "fermion" for an individual particle. All known fermions have $s = 1/2$, whence $2s + 1 = 2$. It follows that for fermions the $z$-component of spin has two possible values, $\hbar/2$ (*spin up*) and $-\hbar/2$ (*spin down*). Electrons are fermions, while photons are bosons with $s = 1$.

The similarity between the expressions of the quantum numbers for spin and those of the angular momentum is the origin of the qualitative visualization of spin in classical terms: spin is described as an intrinsic angular momentum of the particle, as if the particle were a sphere spinning on its axis.

It is possible to associate operators to spin in the following manner:

$$\mathscr{S}^2|\psi\rangle = s\,(s+1)\,\hbar^2\,|\psi\rangle\,, \qquad \mathscr{S}_z|\psi\rangle = s_z\,\hbar\,|\psi\rangle\,. \tag{24.29}$$

The corresponding matrices are

$$\mathbf{S}^2 = \frac{3}{4}\,\hbar^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \mathbf{S}_z = \frac{\hbar}{2} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{24.30}$$

The above implies that the eigenvectors are such that $\mathbf{S}_z$ is diagonal. It is possible to associate matrices also to the other spin components:

$$\mathbf{S}_x = \frac{\hbar}{2} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \mathbf{S}_y = \frac{\hbar}{2} \begin{bmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{bmatrix}. \tag{24.31}$$

It is found

$$\mathbf{S}^2\,\mathbf{S}_x - \mathbf{S}_x\,\mathbf{S}^2 = 0\,, \quad \mathbf{S}^2\,\mathbf{S}_y - \mathbf{S}_y\,\mathbf{S}^2 = 0\,, \quad \mathbf{S}^2\,\mathbf{S}_z - \mathbf{S}_z\,\mathbf{S}^2 = 0\,, \tag{24.32}$$

showing that each component of spin is simultaneously observable with the square modulus. In contrast,

$$\mathbf{S}_y\,\mathbf{S}_z - \mathbf{S}_z\,\mathbf{S}_y = \mathrm{i}\,\hbar\,\mathbf{S}_x\,, \quad \mathbf{S}_z\,\mathbf{S}_x - \mathbf{S}_x\,\mathbf{S}_z = \mathrm{i}\,\hbar\,\mathbf{S}_y\,, \quad \mathbf{S}_x\,\mathbf{S}_y - \mathbf{S}_y\,\mathbf{S}_x = \mathrm{i}\,\hbar\,\mathbf{S}_z\,, \tag{24.33}$$

showing that the components of spin are not simultaneously observable.

## 24.6 Entanglement

Two quantum systems together can be regarded as one combined system described by a state in a space larger than the invidual spaces. Such a larger space is also called the *tensor product* of the Hilbert spaces of the individual systems. As an easy example, assume that the combined system is made of two interacting particles, so that the wave function (avoiding for the moment the Dirac notation) reads $\psi_{12}(\mathbf{r}_1, \mathbf{r}_2, t)$. Assuming that $\psi_{12}$ is normalized to unity, the probability that particle 1 is found at time $t$ within the elementary volume centered at $\mathbf{r}_1$, and particle 2 is found at time $t$ within the elementary volume centered at $\mathbf{r}_2$, is given by $\mathrm{d}P_{12} = |\psi_{12}(\mathbf{r}_1, \mathbf{r}_2, t)|^2\,\mathrm{d}^3 r_1\,\mathrm{d}^3 r_2$;

the expectation value of a Hermitean operator $\mathscr{B}$ is given by

$$\int \psi_{12}^{*}\, \mathscr{B} \psi_{12}\, \mathrm{d}^3 r_1\, \mathrm{d}^3 r_2 \,. \tag{24.34}$$

The above applies irrespective of the fact that $\mathscr{B}$ operates on both sets of coordinates $\mathbf{r}_1$ and $\mathbf{r}_2$, or on one of them only. Letting $u_k(\mathbf{r}_1)$ be a complete set in the $\mathbf{r}_1$ space, one finds

$$\psi_{12}(\mathbf{r}_1, \mathbf{r}_2, t) = \sum_k b_k(\mathbf{r}_2, t)\, u_k(\mathbf{r}_1)\,, \qquad b_k = \langle u_k | \psi_{12} \rangle\,; \tag{24.35}$$

then, letting $w_n(\mathbf{r}_2)$ be a complete set in the $\mathbf{r}_2$ space,

$$\psi_{12}(\mathbf{r}_1, \mathbf{r}_2, t) = \sum_k \left( \sum_n a_{kn} w_n \right) u_k = \sum_{kn} a_{kn} u_k w_n\,, \qquad a_{kn}(t) = \langle w_n | b_k \rangle\,. \tag{24.36}$$

If it happens that the two particles have no interaction, then the probability density $|\psi_{12}|^2$ becomes the product of two densities, $|\psi_{12}(\mathbf{r}_1, \mathbf{r}_2, t)|^2 = |\psi_1(\mathbf{r}_1, t)|^2 \times |\psi_2(\mathbf{r}_2, t)|^2$; remembering that the phase of the wave function has no effect on the probability, one may let $\psi_{12} = \psi_1\, \psi_2$: in other terms the combined state is the tensor product of the individual states. Returning to the Dirac notation, if system 1 is in state $|\psi\rangle_1$ and system 2 is in state $|\psi\rangle_2$, the combined state of two non-interacting systems reads

$$|\psi\rangle_{12} = |\psi\rangle_1\, |\psi\rangle_2\,. \tag{24.37}$$

When operators are considered, for the sake of clarity one attaches a label to them to distinguish the subspace where they operate. For instance, operator $\mathscr{B}_1$ operates only on $|\psi\rangle_1$, so that (still assuming that the wave functions are normalized to unity),

$$\langle \psi | \mathscr{B}_1 | \psi \rangle_{12} = \langle \psi | \mathscr{B}_1 | \psi \rangle_1\, \langle \psi | \psi \rangle_2 = \langle \psi | \mathscr{B}_1 | \psi \rangle_1\,. \tag{24.38}$$

It follows that in this situation the expectation value of $\mathscr{B}_1$ is independent of the state of system 2; thus, each qubit can be regarded as being in a well-defined state independent of his partner. In the situation described by (24.37), the wave function is separable into the wave functions of the individual systems, whereas in the more general situation described by (24.36), the wave function is not separable.

As a special example of a non-separable wave function, consider the case in which each of the two systems has two quantum states $|0\rangle$ and $|1\rangle$, so that the following combinations are possible: $|0\rangle_1\, |0\rangle_2$, $|0\rangle_1\, |1\rangle_2$, $|1\rangle_1\, |0\rangle_2$, and $|1\rangle_1\, |1\rangle_2$. The linear combination is a special case of (24.36) and reads

$$|\psi\rangle_{12} = \sum_{k,n=0}^{1} a_{kn}\, |k\rangle_1\, |n\rangle_2\,, \tag{24.39}$$

and neither subsystem can be said to be in some definite state independent of his partner: the two subsystems form an *entangled* system.

Rather surprisingly, the entanglement is kept also if the two subsystems are far apart, like in the case of two photons generated together with some related polarization states, that fly apart at the speed of light. Despite separation, the two photons retain the entanglement until some irreversible interaction, e.g., a measurement, occurs. For instance, if the original state of the combined system is

$$|\psi\rangle_{12} = \frac{1}{\sqrt{2}}\,|0\rangle_1\,|0\rangle_2 + \frac{1}{\sqrt{2}}\,|1\rangle_1\,|1\rangle_2\,, \tag{24.40}$$

and a measurement carried out on photon 1 gives as a result the polarization state $|0\rangle_1$, then a similar measurement carried out using the same basis on photon 2 at a large distance will provide as a result the polarization state $|0\rangle_2$ with probability equal to 1.

## 24.7 Formal Implementation of Quantum Operators

The description of the operation of a number of reversible logic gates has been given in Sect. 24.4 by means of the truth tables. It is then necessary to associate operators and matrices to such gates; this is done in the subsections below.

### *24.7.1 The NOT operator*

Remembering the association of qubits with vectors introduced in Sec. 24.5, it is easily found that the NOT operation is achieved by the real, unitary operator $\mathscr{A}_0$ represented by the matrix

$$\mathbf{A}_0 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{24.41}$$

In fact, it is

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{24.42}$$

corresponding to $\mathscr{A}_0|0\rangle = |1\rangle$ and, respectively, to $\mathscr{A}_0|1\rangle = |0\rangle$. The eigenvalues of $\mathbf{A}_0$ are $\lambda_1 = 1$, $\lambda_2 = -1$; to find them, after indicating with $v_1$, $v_2$ the components of an eigenvector $\mathbf{v}$, one writes the secular equation $\mathbf{A}_0\mathbf{v} = \lambda\,\mathbf{v}$, corresponding to the algebraic system $v_2 = \lambda\,v_1$, $v_1 = \lambda\,v_2$, whence $v_1 = \lambda^2\,v_1$, $v_1 \neq 0$. The eigenvectors $\mathbf{v}^{(1)}$, $\mathbf{v}^{(2)}$ fulfill the relation $v_2^{(1)} = v_1^{(1)}$, $v_2^{(2)} = -v_1^{(2)}$.

One notes that $\mathbf{A}_0$ is a *permutation matrix*; in particular, it differs from the identity by the exchange of the rows (columns): as a consequence, it is $\det(\mathbf{A}_0) = -1$. Also, it is easily checked that $\mathbf{A}_0^T = \mathbf{A}_0$ and that $\mathbf{A}_0^T\,\mathbf{A}_0 = \mathbf{I}$. The latter relation shows that the real matrix $\mathbf{A}_0$ is *orthogonal*; as the orthogonality property of real matrices corresponds to the unitarity property of complex matrices, it follows that the

transformations produced by $\mathbf{A}_0$ are unitary. Combining the unitarity property with $\mathbf{A}_0^T = \mathbf{A}_0$ shows that it is also $\mathbf{A}_0^2 = \mathbf{I}$, namely, $\mathbf{A}_0$ is *involutory*.

In the following, other matrices will be considered which, like $\mathbf{A}_0$ above, are permutation matrices. All such matrices are orthogonal; in fact, they can be obtained by successive permutations of columns of the identity matrix: as the latter is orthogonal, the permutation matrices are orthogonal as well (see, e.g., [48, Sect. 9-5] and Prob. 3.2); in addition, permutation matrices are also involutory (Prob. 3.3).

### *24.7.2 The Hadamard Operator*

With reference to the representation (24.9) of the qubit, it has been shown in Sect. 24.5 that, if the condition $\theta = 0$, $\varphi = 0$ becomes $\theta = \pi/2$, $\varphi = 0$, the qubit transforms from $|0\rangle$ into $|\Phi_p\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$; similarly, if the condition $\theta = \pi$, $\varphi = 0$ becomes $\theta = 3\pi/2$, $\varphi = 0$, the qubit transforms from $|1\rangle$ into $|\Phi_r\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$. Inspecting Fig. 24.12 shows that such transformations are equivalent to a rotation of $|\Phi\rangle$, starting from $|0\rangle$ or $|1\rangle$, in which $\theta$ is increased by $\pi/2$ while $\varphi$ is always kept to zero. These transformations may be thought of as the action of an operator $\mathscr{R}$ that converts each of the states of the basis into a linear combination of them, according to the rules:

$$\mathscr{R}|0\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}}, \qquad \mathscr{R}|1\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}}. \tag{24.43}$$

The matrix associated to $\mathscr{R}$ is easily found to be

$$\mathbf{R} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}; \tag{24.44}$$

in fact,

$$\sqrt{2}\mathbf{R} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \sqrt{2}\mathbf{R} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{24.45}$$

Operator $\mathscr{R}$ is called *Hadamard operator*; it belongs to a class of operators that induce rotations of the qubit.

### *24.7.3 The CNOT and CCNOT Operators*

The relation expressed by the truth table of Fig. 24.4 is recast in matrix form as

$$\begin{bmatrix} 0\ 0 \\ 0\ 1 \\ 1\ 1 \\ 1\ 0 \end{bmatrix} = \mathbf{A}_2 \begin{bmatrix} 0\ 0 \\ 0\ 1 \\ 1\ 0 \\ 1\ 1 \end{bmatrix}, \qquad \mathbf{A}_2 = \begin{bmatrix} 1\ 0\ 0\ 0 \\ 0\ 1\ 0\ 0 \\ 0\ 0\ 0\ 1 \\ 0\ 0\ 1\ 0 \end{bmatrix}, \tag{24.46}$$

where the $4 \times 2$ matrix on the right hand side represents the possible combinations of the input bits $A$, $B$, the $4 \times 4$ real matrix $\mathbf{A}_2$ represents the CNOT operator, and the $4 \times 2$ matrix on the left hand side shows the corresponding values of the output bits $A'$, $B'$. Like $\mathbf{A}_0$ of (24.41), one notes that $\mathbf{A}_2$ is a permutation matrix; as indicated in Sect. 24.7.1, it follows that the real matrix $\mathbf{A}_2$ is orthogonal and the transformation described by (24.46) is unitary.

By the same token, the relation expressed by the truth table of Fig. 24.8 is recast in matrix form as

$$\begin{bmatrix} 0\ 0\ 0 \\ 0\ 0\ 1 \\ 0\ 1\ 0 \\ 0\ 1\ 1 \\ 1\ 0\ 0 \\ 1\ 0\ 1 \\ 1\ 1\ 1 \\ 1\ 1\ 0 \end{bmatrix} = \mathbf{A}_1 \begin{bmatrix} 0\ 0\ 0 \\ 0\ 0\ 1 \\ 0\ 1\ 0 \\ 0\ 1\ 1 \\ 1\ 0\ 0 \\ 1\ 0\ 1 \\ 1\ 1\ 0 \\ 1\ 1\ 1 \end{bmatrix}, \qquad \mathbf{A}_1 = \begin{bmatrix} 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0 \\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0 \\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0 \\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0 \\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0 \\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1 \\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0 \end{bmatrix}. \tag{24.47}$$

Like $\mathbf{A}_2$ of (24.46), $\mathbf{A}_1$ is a permutation matrix, namely, as indicated in Sect. 24.7.1, the real matrix $\mathbf{A}_1$ is orthogonal and the transformation described by (24.47) is unitary.

### *24.7.4 The Half Adder operator*

The operation of the half adder is described in term of matrices starting from (24.47); the latter relates, through the $8 \times 8$ matrix $\mathbf{A}_1$, the $8 \times 3$ matrix at the right hand side, which represents the triad $A$, $B$, $C$, with the $8 \times 3$ matrix on the left hand side, which represents the triad $A'$, $B'$, $C'$ (compare with the truth table of Fig. 24.10). It is then necessary to relate the triad $A'$, $B'$, $C'$ with triad $D$, $S$, $Q$ visible in the same table. This is readily accomplished by

$$\begin{bmatrix} 0\ 0\ 0 \\ 0\ 0\ 1 \\ 0\ 1\ 0 \\ 0\ 1\ 1 \\ 1\ 1\ 0 \\ 1\ 1\ 1 \\ 1\ 0\ 1 \\ 1\ 0\ 0 \end{bmatrix} = \mathbf{A}_2' \begin{bmatrix} 0\ 0\ 0 \\ 0\ 0\ 1 \\ 0\ 1\ 0 \\ 0\ 1\ 1 \\ 1\ 0\ 0 \\ 1\ 0\ 1 \\ 1\ 1\ 1 \\ 1\ 1\ 0 \end{bmatrix}, \qquad \mathbf{A}_2' = \begin{bmatrix} 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0 \\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0 \\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0 \\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1 \\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0 \\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0 \\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0 \end{bmatrix} \tag{24.48}$$

where, like $\mathbf{A}_0$, $\mathbf{A}_1$, and $\mathbf{A}_2$ above, $\mathbf{A}'_2$ is a permutation matrix, whence it is orthogonal. It follows that the transformation from triad $A$, $B$, $C$ to triad $D$, $S$, $Q$ is provided by the product $\mathbf{A}'_2\,\mathbf{A}_1$ of the two matrices; as latter are orthogonal, the product is orthogonal as well.

One may think of associating to the half adder an operator $\mathscr{L}$ made of the successive application of two unitary operators, $\mathscr{A}_1$ and $\mathscr{A}'_2$,

$$\mathscr{L} = \mathscr{A}'_2\,\mathscr{A}_1\,, \tag{24.49}$$

where the rightmost operator at the right hand side ($\mathscr{A}_1$) operates first.

### *24.7.5 The Creation and Annihilation Operators*

To proceed, it is convenient to introduce the *annihilation* operator $\hat{a}$ and the *creation* operator $\hat{a}^\dagger$ represented, respectively, by the matrices

$$\mathbf{a} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad \mathbf{a}^\dagger = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \tag{24.50}$$

and fulfilling the relations

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \tag{24.51}$$

in short, $\hat{a}|1\rangle = |0\rangle$ and $\hat{a}^\dagger|0\rangle = |1\rangle$; also,

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{24.52}$$

in short, $\hat{a}|0\rangle = 0$, $\hat{a}^\dagger|1\rangle = 0$, where symbols "0" at the right hand sides indicate the null vector. The above operators and the corresponding matrices are real; one finds that the NOT operator is expressible as the sum of the annihilation and creations operators,

$$\mathbf{A}_0 = \mathbf{a} + \mathbf{a}^\dagger, \qquad \mathscr{A}_0 = \hat{a} + \hat{a}^\dagger, \tag{24.53}$$

and

$$\mathbf{a}\,\mathbf{a}^\dagger = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \qquad \mathbf{a}^\dagger\,\mathbf{a} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \tag{24.54}$$

whence

$$\mathbf{a}^\dagger\,\mathbf{a} + \mathbf{a}\,\mathbf{a}^\dagger = \mathbf{I}, \qquad \hat{a}^\dagger\,\hat{a} + \hat{a}\,\hat{a}^\dagger = \mathscr{I}\,. \tag{24.55}$$

Also, observing that

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{24.56}$$

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{24.57}$$

it follows

$$\hat{a}\,\hat{a}^\dagger|0\rangle = |0\rangle\,, \quad \hat{a}\,\hat{a}^\dagger|1\rangle = 0\,, \quad \hat{a}^\dagger\,\hat{a}|1\rangle = |1\rangle\,, \quad \hat{a}^\dagger\,\hat{a}|0\rangle = 0\,. \tag{24.58}$$

In conclusion, operator $\hat{a}\,\hat{a}^\dagger$ confirms the qubit if the latter is $|0\rangle$, whereas it annihilates the qubit if the latter is $|1\rangle$. In contrast, operator $\hat{a}^\dagger\,\hat{a}$ confirms the qubit if the latter is $|1\rangle$, whereas it annihilates the qubit if the latter is $|0\rangle$.

## 24.8 Combining the $\hat{a}$ and $\hat{a}^\dagger$ Operators

For consistency with the description of logic operators of Sect. 24.4, letters instead of numbers will be used here as suffixes: e.g., $\mathscr{A}_A$ indicates that the operator operates on qubit $A$; multiple suffixes (e.g., $\mathscr{A}_{AB}$) indicate that the operator operates on qubits $A$, $B$. Consider now the operator $\mathscr{A}_{AB}$ defined as

$$\mathscr{A}_{AB} = (\hat{a}^\dagger\,\hat{a})_A\,(\hat{a}^\dagger + \hat{a})_B + (\hat{a}\,\hat{a}^\dagger)_A\,\mathscr{I}_B\,, \tag{24.59}$$

and consider the application of it to $|0\rangle_A\,|0\rangle_B$; in matrix form, using (24.41), (24.53), and (24.55), one obtains

$$\left( \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}_A \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}_B + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}_A \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_B \right) \begin{bmatrix} 1 \\ 0 \end{bmatrix}_A \begin{bmatrix} 1 \\ 0 \end{bmatrix}_B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}_A \begin{bmatrix} 1 \\ 0 \end{bmatrix}_B, \tag{24.60}$$

corresponding to $\mathscr{A}_{AB}\,|0\rangle_A\,|0\rangle_B = |0\rangle_A\,|0\rangle_B$. In a similar manner, using the other combinations of states yields

$$\left( \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}_A \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}_B + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}_A \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_B \right) \begin{bmatrix} 1 \\ 0 \end{bmatrix}_A \begin{bmatrix} 0 \\ 1 \end{bmatrix}_B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}_A \begin{bmatrix} 0 \\ 1 \end{bmatrix}_B, \tag{24.61}$$

corresponding to $\mathscr{A}_{AB}\,|0\rangle_A\,|1\rangle_B = |0\rangle_A\,|1\rangle_B$; then,

$$\left( \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}_A \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}_B + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}_A \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_B \right) \begin{bmatrix} 0 \\ 1 \end{bmatrix}_A \begin{bmatrix} 1 \\ 0 \end{bmatrix}_B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}_A \begin{bmatrix} 0 \\ 1 \end{bmatrix}_B. \tag{24.62}$$

corresponding to $\mathscr{A}_{AB}\,|1\rangle_A\,|0\rangle_B = |1\rangle_A\,|1\rangle_B$; finally,

$$\left( \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}_A \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}_B + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}_A \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_B \right) \begin{bmatrix} 0 \\ 1 \end{bmatrix}_A \begin{bmatrix} 0 \\ 1 \end{bmatrix}_B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}_A \begin{bmatrix} 1 \\ 0 \end{bmatrix}_B. \tag{24.63}$$

corresponding to $\mathscr{A}_{AB}\,|1\rangle_A\,|1\rangle_B = |1\rangle_A\,|0\rangle_B$. A comparison with the truth table shown in Fig. 24.8 demonstrates that operator $\mathscr{A}_{AB}$ of (24.4) implements the CNOT logic gate. Using $\hat{a}^\dagger\,\hat{a} + \hat{a}\,\hat{a}^\dagger = \mathscr{I}$ gives (24.59) another form, namely,

$$\mathscr{A}_{AB} = \mathscr{I}_{AB} + (\hat{a}^\dagger \hat{a})_A \, (\hat{a}^\dagger + \hat{a})_B - (\hat{a}^\dagger \hat{a})_A \, \mathscr{I}_B \,. \tag{24.64}$$

The CCNOT operator is described in a similar manner:

$$\mathscr{B}_{ABC} = \mathscr{I}_{ABC} + (\hat{a}^\dagger \hat{a})_A \, (\hat{a}^\dagger \hat{a})_B \, (\hat{a}^\dagger + \hat{a})_C - (\hat{a}^\dagger \hat{a})_A \, (\hat{a}^\dagger \hat{a})_B \, \mathscr{I}_C \,. \tag{24.65}$$

The interpretation of (24.65) is as follows: if $A = |0\rangle$ or $B = |0\rangle$, the second and third terms of (24.65) contain a factor of the form $\hat{a}^\dagger \hat{a} \, |0\rangle$ and are annihilated; the only remaining term is the first one, whence $A' = A$, $B' = B$, and $C' = C$. If, instead, $A = B = |1\rangle$, the third term cancels the first, and $C' = \overline{C}$.

## 24.9 Complements

### *24.9.1 The DiVincenzo Criteria*

The *DiVincenzo Criteria*, listed below,[4] are a formalization of what a quantum computer consists of:

1. A scalable physical system with *i*) qubits that are distinct from one another and *ii*) the ability to count exactly how many qubits there are in it.
2. The ability to initialize the state of any qubit to a definite state in the computational basis (in the examples above, the computational basis is $|0\rangle$, $|1\rangle$).
3. The system's qubits must hold their state: the system must be isolated from the outside world, otherwise the qubits will decohere. In practice, the qubits must hold their state long enough to apply the next operator with assurance that the qubits have not changed state due to outside influences between operations.
4. The system must be able to apply a sequence of unitary operators to the qubit states. The system must also be able to apply a unitary operator to two qubits at once: this entails entanglement between those qubits. Let

$$g = s_{11} \, u_1 \, w_1 + s_{12} \, u_1 \, w_2 + s_{21} \, u_2 \, w_1 + s_{22} \, u_2 \, w_2 \,, \tag{24.66}$$

with

$$||g||^2 = \sum_{ij=1}^{2} |s_{ij}|^2 \, ||u_i||^2 \, ||w_j||^2 \,. \tag{24.67}$$

If $s_{11} \, s_{22} = s_{12} \, s_{21}$, then

$$g = (s_{11} \, u_1 + s_{21} \, u_2) \, (w_1 + s_{12} \, w_2 / s_{11}) = u \, w \,, \tag{24.68}$$

namely, $g$ is separable; otherwise, $g$ is entangled. Quoting DiVincenzo (from [28]): "*... entanglement between different parts of the quantum computer is good;*

[4] The list of criteria is taken from [28], that in turn summarizes the contents of [15].

*entanglement between the quantum computer and its environment is bad, since it corresponds to decoherence*."

5. The system must be capable of making "strong" measurements of each qubit. That is, the measuring technique in the system actually does measure the state of the qubit for the property being measured and leaves the qubit in that state. For example, assume that index 1 (2) means "spin up (down)" and that initially the total spin of a two-electron system equals zero. Thus,

$$g = s_{12}\, u_1\, w_2 + s_{21}\, u_2\, w_1\,. \tag{24.69}$$

Assume that Alice (sitting on Earth) measures $u$ and finds "spin up"; this is equivalent to forcing $s_{12} = 1$ and $s_{21} = 0$. As a consequence, when Bob (sitting on Anacreon)[5] measures $w$, he must necessarily find "spin down".

[5] In the jargon of cryptography, *Alice* and *Bob* are the customary names used to designate the two individuals who exchange secret messages. Anacreon (ca 570–485 BC) was a Greek lyric poet. Planet Anacreon does not exist, it is a fictional planet from Isaac Asimov's *Foundation Series*.

$$\mathscr{A}_{AB} = \mathscr{I}_{AB} + (\hat{a}^\dagger \hat{a})_A \, (\hat{a}^\dagger + \hat{a})_B - (\hat{a}^\dagger \hat{a})_A \, \mathscr{I}_B \, . \tag{26.64}$$

The CCNOT operator is described in a similar manner:

$$\mathscr{B}_{ABC} = \mathscr{I}_{ABC} + (\hat{a}^\dagger \hat{a})_A \, (\hat{a}^\dagger \hat{a})_B \, (\hat{a}^\dagger + \hat{a})_C - (\hat{a}^\dagger \hat{a})_A \, (\hat{a}^\dagger \hat{a})_B \, \mathscr{I}_C \, . \tag{26.65}$$

The interpretation of (26.65) is as follows: if $A = |0\rangle$ or $B = |0\rangle$, the second and third terms of (26.65) contain a factor of the form $\hat{a}^\dagger \hat{a} \, |0\rangle$ and are annihilated; the only remaining term is the first one, whence $A' = A$, $B' = B$, and $C' = C$. If, instead, $A = B = |1\rangle$, the third term cancels the first, and $C' = \overline{C}$.

## 26.9 Complements

### *26.9.1 The DiVincenzo Criteria*

The *DiVincenzo Criteria*, listed below,[5] are a formalization of what a quantum computer consists of:

1. A scalable physical system with *i*) qubits that are distinct from one another and *ii*) the ability to count exactly how many qubits there are in it.
2. The ability to initialize the state of any qubit to a definite state in the computational basis (in the examples above, the computational basis is $|0\rangle$, $|1\rangle$).
3. The system's qubits must hold their state: the system must be isolated from the outside world, otherwise the qubits will decohere. In practice, the qubits must hold their state long enough to apply the next operator with assurance that the qubits have not changed state due to outside influences between operations.
4. The system must be able to apply a sequence of unitary operators to the qubit states. The system must also be able to apply a unitary operator to two qubits at once: this entails entanglement between those qubits. Let

$$g = s_{11} \, u_1 \, w_1 + s_{12} \, u_1 \, w_2 + s_{21} \, u_2 \, w_1 + s_{22} \, u_2 \, w_2 \, , \tag{26.66}$$

with

$$||g||^2 = \sum_{ij=1}^{2} |s_{ij}|^2 \, ||u_i||^2 \, ||w_j||^2 \, . \tag{26.67}$$

If $s_{11} \, s_{22} = s_{12} \, s_{21}$, then

$$g = (s_{11} \, u_1 + s_{21} \, u_2) \, (w_1 + s_{12} \, w_2 / s_{11}) = u \, w \, , \tag{26.68}$$

namely, $g$ is separable; otherwise, $g$ is entangled. Quoting DiVincenzo (from [34]): "*... entanglement between different parts of the quantum computer is good;*

[5] The list of criteria is taken from [34], that in turn summarizes the contents of [20].

*entanglement between the quantum computer and its environment is bad, since it corresponds to decoherence*."

5. The system must be capable of making "strong" measurements of each qubit. That is, the measuring technique in the system actually does measure the state of the qubit for the property being measured and leaves the qubit in that state. For example, assume that index 1 (2) means "spin up (down)" and that initially the total spin of a two-electron system equals zero. Thus,

$$g = s_{12}\, u_1\, w_2 + s_{21}\, u_2\, w_1\,. \tag{26.69}$$

Assume that Alice (sitting on Earth) measures $u$ and finds "spin up"; this is equivalent to forcing $s_{12} = 1$ and $s_{21} = 0$. As a consequence, when Bob (sitting on Anacreon)[6] measures $w$, he must necessarily find "spin down".

### 26.9.2 Quantum Transport in a One-Dimensional Channel

With reference to Fig. 26.13, consider a particle subjected to a potential energy of the form

$$V = 0 \quad \text{in} \quad 0 \le x_2 \le d_2\,, \quad 0 \le x_3 \le d_3\,, \qquad V = V_0 \to \infty \quad \text{elsewhere.} \tag{26.70}$$

The Schrödinger equation $-[\hbar^2/(2\,m)]\,\nabla^2 w = E\,w$ is separable, so that the solution is expressible as a product, $w = w_1(x_1)\,w_2(x_2)\,w_3(x_3)$. Using primes to indicate the derivatives of the factors with respect to the corresponding variables, one finds

$$-\frac{2\,m\,E}{\hbar^2} = \frac{w_1''}{w_1} + \frac{w_2''}{w_2} + \frac{w_3''}{w_3}\,, \tag{26.71}$$

where each fraction at the right hand side is equal to a constant. Equation (26.71) then splits into three eigenvalue equations in the unknowns $w_i$, of the form $w_i'' = \text{const} \times w_i$; for $i = 2,3$, the boundary conditions are $w_i(0) = w_i(d_i) = 0$. Excluding the trivial cases $w_2 = 0$ and $w_3 = 0$, one finds that the only solutions compatible with such boundary conditions correspond to a negative value of the constant; it follows $w_i = 2\,\mathrm{i}\,w_{i0}\,\sin(n_i\,\pi\,x_i/d_i)$, with $n_i = 1,2,\ldots$ and $w_{i0}$ a complex constant (compare, e.g., with [58, Sect. 8.2.2]). The constant is specified by imposing the normalization condition of $w_i$, namely,[7]

$$\frac{1}{4\,|w_{i0}|^2} = \int_0^{d_i} \sin^2\left(\frac{n_i\,\pi\,x_i}{d_i}\right) \mathrm{d}x_i = \frac{d_i}{2}\,, \qquad 2\,\mathrm{i}\,w_{i0} = \sqrt{\frac{2}{d_i}}\,. \tag{26.72}$$

[6] In the jargon of cryptography, *Alice* and *Bob* are the customary names used to designate the two individuals who exchange secret messages. Anacreon (ca 570–485 BC) was a Greek lyric poet. Planet Anacreon does not exist, it is a fictional planet from Isaac Asimov's *Foundation Series*.

[7] The integral in (26.72) is readily obtained from an integration by parts.

As for the $i = 1$ case, excluding again the trivial solution $w_1 = 0$, the constant in $w_1'' = \text{const} \times w_1$ must be negative as well, to prevent $w_1$ from diverging at infinity; from (26.71) it then follows $E > 0$, as should be.[8] Function $w_1$ is not normalizable; letting $\kappa$ be a real positive quantity, the form of $w_1$ is

$$w_1 = w_1^+ \exp(\mathrm{i}\kappa x_1) + w_1^- \exp(-\mathrm{i}\kappa x_1)\,, \tag{26.73}$$

so that the relation between the total energy and the other parameters reads

$$\frac{2m}{\hbar^2} E = \kappa^2 + \frac{\pi^2}{d_2^2} n_2^2 + \frac{\pi^2}{d_3^2} n_3^2\,. \tag{26.74}$$

The time-dependent wave function corresponding to energy $E$ has the form

$$\phi = \left[w_1^+ \exp(\mathrm{i}\kappa x_1 - \mathrm{i}\omega t) + w_1^- \exp(-\mathrm{i}\kappa x_1 - \mathrm{i}\omega t)\right] w_2(x_2)\, w_3(x_3)\,, \tag{26.75}$$

with $\omega = E/\hbar$ and $w_1^+$, $w_1^-$ two complex constants, that is, a superposition of two monochromatic plane waves propagating in the opposite directions of the $x_1$ axis, and modulated by the $w_2(x_2)\, w_3(x_3)$ factor. Assuming that the total energy $E$ is

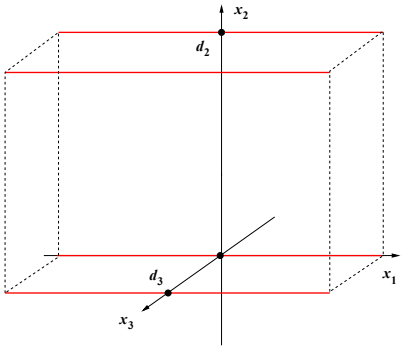



**Fig. 26.13** Schematic picture of a one-dimensional channel. The wave packet propagates along the $x_1$ axis

prescribed, relation (26.74) introduces constraints among the different terms at the right hand side; specifically, the following inequalities must be fulfilled:

$$\frac{\pi^2}{d_2^2} n_2^2 + \frac{\pi^2}{d_3^2} n_3^2 < \frac{2m}{\hbar^2} E\,, \qquad 0 < \kappa^2 \le \kappa_{\max}^2 = \frac{2m}{\hbar^2} E - \frac{\pi^2}{d_2^2} - \frac{\pi^2}{d_3^2}\,. \tag{26.76}$$

For the considerations that follow it is sufficient to assume a two-dimensional case, e.g., in the $x_1 x_3$ plane, whence (26.74, 26.75) become

[8] Compare, e.g., with the discussion in [58, Sect. 8.2.3].

$$\frac{2m}{\hbar^2}E = \kappa^2 + \frac{\pi^2}{d_3^2}n_3^2, \tag{26.77}$$

$$\phi(x_1,x_3,t) = \left[w_1^+ \exp(\mathrm{i}\kappa x_1 - \mathrm{i}\omega t) + w_1^- \exp(-\mathrm{i}\kappa x_1 - \mathrm{i}\omega t)\right] w_3(x_3), \tag{26.78}$$

where $\phi$ depends also on $\kappa$ and $n_3$, and inequalities (26.76) reduce to

$$\frac{\pi^2}{d_3^2}n_3^2 < \frac{2m}{\hbar^2}E, \qquad 0 < \kappa^2 \le \kappa_{\max}^2 = \frac{2m}{\hbar^2}E - \frac{\pi^2}{d_3^2}. \tag{26.79}$$

Selecting the ground case ($n_3 = 1$) for the $x_3$ direction, one may construct a wave packet using the range $\kappa \le \kappa_{\max}$ of the wave vectors for the $x_1$ direction; using at $t = 0$ a Gaussian wave packet normalized to unity and centered at $x_1 = x_{10}$, and assuming that the wave packet propagates in the positive direction, one obtains for the initial condition

$$\psi(x_1,x_3,t=0) = \frac{w_3(x_3)}{\sqrt{\Delta x_1 \sqrt{2\pi}}} \exp\left[\mathrm{i}\kappa_0 x_1 - \frac{(x_1 - x_{10})^2}{(2\Delta x_1)^2}\right], \tag{26.80}$$

with $\Delta x_1$ the standard deviation of the wave packet and $\kappa_0$ its center in the $\kappa$ space.

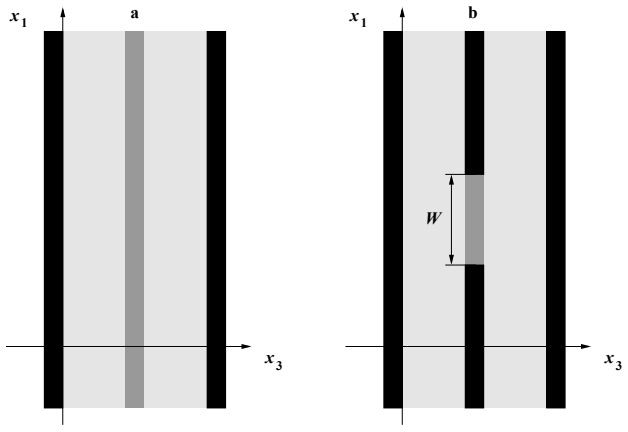



**Fig. 26.14** Schematic picture of two parallel channels. (**a**) The vertical, grey region in the center indicates a barrier, such that the electron tunneling between the two channels is possible. (**b**) The figure shows a modified structure, where the extension of the barrier through which tunneling is possible has been limited to a finite "window" of length $W$, whereas tunneling is impossible through the rest of the barrier (dark region)

Now, assume that a second channel, parallel to the first one, is present (Fig. 26.14a); a section of the potential energy in the $x_3$ direction provides the profile analyzed in Sect. 22.17. If the height and/or width of the barrier between the two channels is infinite, the wave function describing the particle will propagate along the left channel starting from the initial condition (26.80); if, instead, the height and width

of the barrier are finite, the wave function will still propagate, but will also tunnel from the left to the right channel and viceversa. The peak of the wave packet will therefore oscillate along the $x_3$ direction while propagating in the $x_1$ direction. A simplified analysis of the oscillation (with no propagation) is carried out in Sect. 22.17 considering a wave packet made of the superposition of the monochromatic waves corresponding to the two lowest eigenvalues; a full solution of the problem, starting from the initial condition (26.80), requires a numerical approach; Fig. 26.15

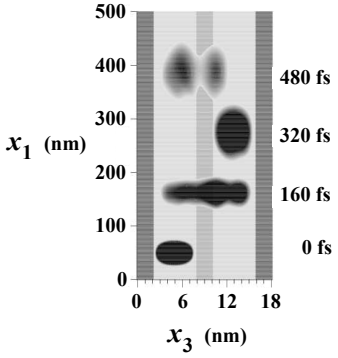



**Fig. 26.15** Numerical solution of the time-dependent Schrödinger equation in the structure of Fig. 26.14a. The extension of each channel in the $x_3$ direction is 6 nm. The scales of the $x_1$ and $x_3$ axes are different (after [8])

(from [8]) shows the oscillation of the square modulus of the wave function between two coupled channels at different times; the barrier height between the two channels is 0.1 eV. The electron is launched in the left channel with the initial condition (26.80), where $\kappa_0$ corresponds to an energy of 0.1 eV.

### *26.9.3 Coupled Channels Implementing Qubits*

Consider now a modified structure, in which the barrier height between the channel is large enough to prevent tunnelling, with the exception of a region of length $W$ ("window") in the $x_1$ direction, where the height of the central barrier is lowered to a value that makes tunnelling possible (Fig. 26.14b). In this way, the two channels are coupled only in the window region; by tuning the parameters of the window and the velocity of the electron, it is possible to obtain an assigned transfer of the wave function from one channel to the other while the electron crosses the region of the window. By way of example, in the case of Fig. 26.16a the parameters are such that the window realizes a complete transfer of the electron from the left to the right wire; clearly, when the electron is launched in the right channel, the window transfers it to the left channel: it is then sensible to associate this action of the two

channels to the logical NOT operation. A different tuning of the parameters, like that corresponding to the simulation shown in Fig. 26.16b, realizes instead an equal splitting of the wave function between the two wires.

One also observes that the wave function of the electron is represented by a linear combination of two linearly-independent functions; also, a measuring apparatus at the end of the channels will detect the electron in one or the other channel, so that it is sensible to associate the logic values 0 and 1 to the two possible states. In conclusion, the states of an electron within the structure described here fulfill the definition of qubit given in Sect. 26.5; in order to distinguish this qubit from another to be introduced later, the present one is indicated with *data qubit*.

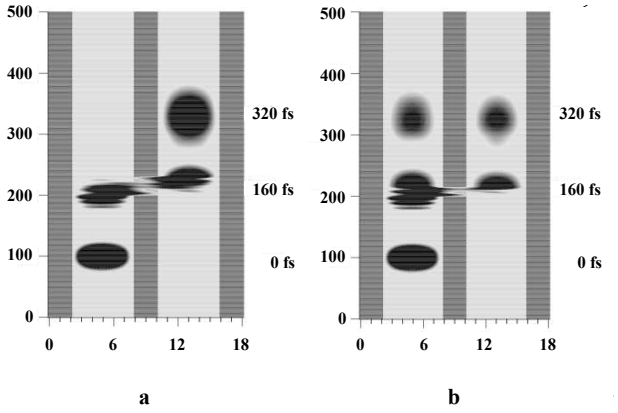



**Fig. 26.16** By realizing a window in the central barrier between the channels (see text) one realizes (**a**) the NOT operation or (**b**) an equal splitting of the wave function betweeen the two channels. The scales and the initial condition are the same as in Fig. 26.15 (after [8])

With reference to the two channels of Fig. 26.14b, let $\tau$ be the interval of time during which the peak of the wave packet, in the $x_1$ direction, faces the window; observing that the group velocity of the packet is $\hbar\kappa_0/m$, one finds $\tau = mW/(\hbar\kappa_0)$. Letting $T = 2\pi/\omega = 1/\nu$ be the period associated to the oscillation of the wave packet between the two channels, it follows that a complete transfer from one channel to the other, across the window, occurs if $\tau = (n+1/2)\,T$, with $n$ a positive integer; in contrast, no transfer occurs if $\tau = n\,T$. Period $T$ can be estimated basing on the analysis of Sect. 22.17, which considers a wave packet made of the superposition of the monochromatic waves corresponding to the two lowest eigenvalues: within this approximation, the complete-transfer and no-transfer conditions read, respectively,

$$W = \left(n+\frac{1}{2}\right) W_s\,, \qquad W = n\,W_s\,, \qquad W_s = \frac{\hbar\kappa_0}{m\nu}\,, \qquad n = 1,2,\ldots \tag{26.81}$$

Using $\nu = \nu_{\max} \simeq 6.8\,10^{11}$ Hz taken from Prob. 22.1, the data of Tab. D.2, and a value of $\kappa_0$ corresponding to an energy of 0.1 eV like in [8], one finds $W_s \simeq 275$ nm. Letting for instance $n = 1$, the complete-transfer condition is achieved with $W = (3/2)\,W_s$. Now, assume that the structure of the double channel has been fixed, and that the group velocity $\hbar\,\kappa_0/m$ of the electron (to be launched in the left or right channel) has been identified such that the first of (26.81) applies; one wants to devise an external action which, before the electron reaches the region of the window, changes the relation to $W = (n+1)\,W_s$: in other terms, one wants to transform the complete-transfer condition into the no-transfer condition. Since $W$ and $T$ are prescribed, one may conceive an external action that decreases the group velocity of the wave packet, such that $\kappa_0$ becomes $\kappa_0' = \kappa_0\,(n+1/2)/(n+1)$. The previous example with $n = 1$ renders $\kappa_0' = (3/4)\,\kappa_0$. This result can be achieved, e.g., with the structure shown in Fig. 26.17, showing the cross-section of the data qubit (channels 1 and 2) onto which another pair of channels (channels 3 and 4) is superimposed; the qubit present in the 3-4 structure is indicated with *control qubit.*

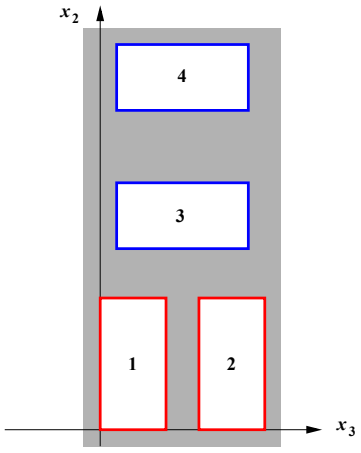



**Fig. 26.17** Rectangles 1 and 2 show the cross section of the channels where the data qubit is present. Here the channels are normal to the page; the orientation of the axes is consistent with that of Fig. 26.13. Rectangles 3 and 4 show the channels where the control qubit is present

The functioning of the structure of Fig. 26.17 is as follows: one electron (*data electron*) is launched in one of the two channels of the data qubit, and another electron (*control electron*) is launched in one of the two channels of the control qubit. Channel 4 is placed at a large distance from the other channels; due to this, if the control electron is injected into channel 4, it does not influence the motion of the data electron. Also, the barrier between channel 3 and 4 is such that tunneling of the control electron to channel 3 is impossible; in summary, injecting the control electron into channel 4 has no effect and, therefore, the complete-transfer condition of the data electron ensues (from channel 1 to channel 2, or vice versa).

Consider, instead, the case where the control electron is injected into channel 3: also in this case the distance between channel 3 and channels 1 or 2 is large enough

to prevent tunneling; however, it is not so large as to prevent the Coulomb interaction between the control electron and the data electron. Also, considering that channel 3 is placed symmetrically with respect to channels 1 and 2, this interaction is the same irrespective of whether the data electron is injected in channel 1 or 2. The slowing down of the data electron necessary to transform the complete-transfer condition into the no-transfer condition is achieved by injecting the control electron into channel 3 slightly before injecting the data electron into channel 1 or 2: the control electron will be ahead of the data electron and will slow it down by Coulomb repulsion; a suitable tuning of the parameters achieves the no-transfer condition [8].

In conclusion, the operation of the structure of Fig. 26.17 is such that the control qubit is never changed; the structure performs the logic NOT operation when the control electron is injected into channel 4, whereas it leaves the logic variable as is when the control electron is injected into channel 3. A comparison with the truth table of Fig. 26.4 shows in fact that the structure realizes the CNOT operator.

### *26.9.4 The Shor Algorithm*

The Shor algorithm [65] succeeds in factoring efficiently a composite integer $m$ by exploiting the parallelism made possible by quantum computation. The goal is finding the period of function $f_m(w;a) \equiv a^w \pmod{m}$ introduced with (4.62).

To proceed, it is necessary to add a prescription to the analysis carried out in Sect. 4.10, where it was shown that, by ascribing successive values $w = 0, 1, 2, \ldots$ to the exponent, one obtains a set of values of $f_m(w;a)$ that exhibit a periodic pattern with some period $r$. The prescription consists in imposing that the left hand side of (4.62) assumes one of the possible values of the pattern, say, $k$; this transforms (4.62) into

$$k \equiv a^w \pmod{m}\,. \tag{26.82}$$

Since both $a$ and $k$ are prescribed, (26.82) can be fulfilled only with some values of the exponent $w$; observing that $k$ is repeatedly found in the pattern, and that the distance between two successive occurrences of $k$ is the period $r$, it follows that, if some value of $w$ provides $k$ in (26.82), then the same $k$ will also be provided by $w+r$, $w+2r$, and so on.

Consider again the example with $m = 15$, $a = 8$ of Sect. 4.10 (Tab. 4.2), that gives rise to the pattern $1, 8, 4, 2, 1, 8, 4, 2, \ldots$ with $r = 4$; then, let $k = 2$. The first instance of the outcome 2 in the pattern occurs when $w = 3$, the second occurs when $w = 7 = 3 + r$, and so on.

Given these premises, the procedure to find the period is made of the following steps:

1. Create a quantum-memory register $R$ and partition it into two parts $R_1$ and $R_2$.

2. Select the set of numbers $w = 0, 1, 2, \ldots, z-1$ to be used as exponents in (4.62), where $2m^2 \leq z \leq 3m^2$.
3. Create a superposition $\mathbf{w}$ of the above numbers, representing all possible inputs to $f_m(w)$, and store it in $R_1$.
4. Calculate all values of $f_m(w)$ in parallel, by applying an operator $\mathcal{Q}$ to the superposition $\mathbf{w}$; in matrix form, $\mathbf{f} = \mathbf{Q}\,\mathbf{w}$. The cost of this operation is equivalent to a single classical calculation instead of many, and provides all the evaluations of $f_m(w)$; the latter are stored in $R_2$.
5. Measure the state of $R_2$; this makes the superposition $\mathbf{f}$ to collapse. Since $\mathbf{f}$ is a superposition of integers, the result of the measurement is some integer $k$ such that $k \equiv a^w \pmod{m}$.
6. Since $R_1$ and $R_2$ are parts of the same quantum register, they are entangled; whence, by making the state of $R_2$ to collapse, one makes the state of $R_1$ to collapse as well: specifically, the superposition in $R_1$ collapses to a superposition of the values of $w$ that correspond to $k$. As shown above, these values are $w$, $w+r$, $w+2r$, …: at this point, the sought-after period is present in $R_1$.
7. As the function stored in $R_1$ is periodic, it makes sense to extract its period by means of the Fourier transform; thus, one computes the Fourier transform of the contents of $R_1$ and puts the result back into $R_1$. The frequency corresponding to $r$ is $1/r$, and one expects to find that the Fourier transform exhibits peaks corresponding to multiples of $1/r$.
8. Computing the Fourier transform a single time provides a multiple of $1/r$; the procedure must then be repeated to extract a series of samples $\lambda_1/r$, $\lambda_2/r$, …, where $\lambda$s are integers; it turns out that, after a few repetitions of the algorithm, it becomes possible to guess $r$.

Another observation is that two different members $r_i \neq r_j$ of sequence (4.58) cannot produce the same remainder when used in (4.59). In fact, assume they do, namely, that the two equalities $r_i\,a = q_i\,m + r'_i$ and $r_j\,a = q_j\,m + r'_i$ hold for a pair of indices $i \neq j$; if one recasts the above as congruences, $r_i\,a \equiv r'_i \pmod{m}$ and $r_j\,a \equiv r'_i \pmod{m}$, the transitivity property (4.17, 4.18) yields $r_i\,a \equiv r_j\,a \pmod{m}$, where $a$ and $m$ are coprime. By the corollary leading to (4.31), the latter congruence can be divided by $a$ to yield $r_i \equiv r_j \pmod{m}$ or, equivalently, $r_i = r_j + k\,m$ with $k$ an integer. On the other hand, since both $r_i$ and $r_j$ are smaller than $m$, the equality above can be fulfilled only with $k = 0$, which yields $r_i = r_j$, contrary to the hypothesis.

The conclusion is that all the $\varphi(m)$ remainders $r'_i$ are different from each other and belong to the sequence (4.58); in other terms, the remainders coincide with the elements of (4.58), albeit in different order. Recasting again (4.59) as $r_i\,a \equiv r'_i \pmod{m}$, taking the product of this congruence from $i = 1$ to $i = \varphi(m)$, and applying the multiplication rule (4.22) yields

$$a^{\varphi(m)}\,r_1\,r_2 \cdots r_{\varphi(m)} \equiv r'_1\,r'_2 \cdots r'_{\varphi(m)} \pmod{m}, \tag{4.60}$$

where the product $r_1\,r_2 \cdots r_{\varphi(m)}$ is equal to the product $r'_1\,r'_2 \cdots r'_{\varphi(m)}$. Letting $M$ be the common value of the two products, congruence (4.60) becomes $a^{\varphi(m)}\,M \equiv M \pmod{m}$ where, by construction, $M$ is coprime with $m$; then, dividing the latter congruence by $M$ yields

$$a^{\varphi(m)} \equiv 1 \pmod{m} \tag{4.61}$$

where, again, the division by the common factor is made possible by the corollary leading to (4.31).

> Applying for instance (4.61) to the first example of Sect. 4.8, namely, $m = 42 = 2 \times 3 \times 7$, $\varphi(42) = 12$, and letting $a = 5$ (coprime with 42), provides $5^{12} - 1 = 244{,}140{,}624 = 5{,}812{,}872 \times 42$.

Congruence (4.61) expresses Euler's theorem; it generalizes the second form (4.46) of Fermat's little theorem; in fact, as mentioned in Sect. 4.8, if $m$ is prime, then $\varphi(m) = m - 1$.

## 4.10 A Special Form of Period

The findings of Sect. 4.9 can further be elaborated upon by considering the congruence

$$a^w \equiv f \pmod{m}, \tag{4.62}$$

where $a$ and $m$ are coprime and $w = 1, 2, \ldots$; the above, written in the form $f = a^w - k\,m$, may also be thought of as the definition of a discrete function $f_m(w; a) \pmod{m}$. On the other hand, Euler's theorem (4.61) holds; applying the multiplication rule (4.22) to (4.61) and (4.62) yields

$$a^{w+\varphi(m)} \equiv f \pmod{m}; \tag{4.63}$$

iterating the procedure provides $a^{w+2\varphi(m)} \equiv f \pmod{m}$, a further iteration yields $a^{w+3\varphi(m)} \equiv f \pmod{m}$, and so on. This shows that $f_m(w;a)$ is periodic $\pmod{m}$, with a period equal to $\varphi(m)$.

The procedure leading to the above result applies to any pair of numbers $a$ and $m$ that are coprime. However, it may happen that $\varphi(m)$ is not the minimum period associated to a given pair $a$ and $m$; in other terms, a divisor $r$ of $\varphi(m)$ may exist such that

$$a^{w+r} \equiv f \pmod{m}, \qquad a^{w+2r} \equiv f \pmod{m}, \qquad \cdots \tag{4.64}$$

A few examples are given in Tab. 4.2: all of them assume $m = 15$, so that $\varphi(15) = 8$. The left part of the table shows the case $a = 8$, corresponding to $8^w = 15k+f$; for $f$ it produces the pattern 1, 8, 4, 2, 1, 8, 4, 2, ... whose period is $r = 4$. The central part shows the case $a = 7$, corresponding to $7^w = 15k+f$; for $f$ it produces the pattern 1, 7, 4, 13, 1, 7, 4, 13, ... whose period turns out to be again $r = 4$. The right part shows the case $a = 14$, corresponding to $14^w = 15k+f$; for $f$ it produces the pattern 1, 14, 1, 14, 1, 14, ... whose period turns out to be $r = 2$. In these three examples it is $\varphi(m) = 8$, whereas the periods are respectively $r = 4, 4, 2$, namely, divisors of $\varphi(m)$.

**Table 4.2** Examples of calculation of $a^w \equiv f \pmod{m}$. The values of $a$ and $m$ are given in the text

| $w$ | $k$ | $f$ | $w$ | $k$ | $f$ | $w$ | $k$ | $f$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 |
| 1 | 0 | 8 | 1 | 0 | 7 | 1 | 0 | 14 |
| 2 | 4 | 4 | 2 | 3 | 4 | 2 | 13 | 1 |
| 3 | 34 | 2 | 3 | 22 | 13 | 3 | 182 | 14 |
| 4 | 273 | 1 | 4 | 160 | 1 | 4 | 2,561 | 1 |
| 5 | 2,184 | 8 | 5 | 1,120 | 7 | 5 | 35,854 | 14 |
| 6 | 17,476 | 4 | 6 | 7,843 | 4 | 6 | 501,969 | 1 |
| 7 | 139,810 | 2 | 7 | 54,902 | 13 | 7 | 7,027,566 | 14 |

Another example of $a^w \equiv f \pmod{m}$, with $a = 2$ and $m = 21$, so that $\varphi(m) = 12$, is given in graphic form in Fig. 4.1; the horizontal red line marks one of the values of the function ($f = 8$ in this case), which is used to identify the period graphically. The pattern of $f$ is 1, 2, 4, 8, 16, 11, so that $r = 6$. Again, the period is a divisor of $\varphi(21) = 12$. If the base is changed from $a = 2$ to $a = -2$, the values of $f$ corresponding to the even powers are left unmodified, whereas those corresponding to the odd powers change sign; in fact, congruence $(-2)^w \equiv f \pmod{21}$ with $w$ odd is equivalent to $2^w = 21(-k) - f$. The pattern of $f$ becomes in this case 1, $-2$, 4, $-8$, 16, $-11$, still with $r = 6$.

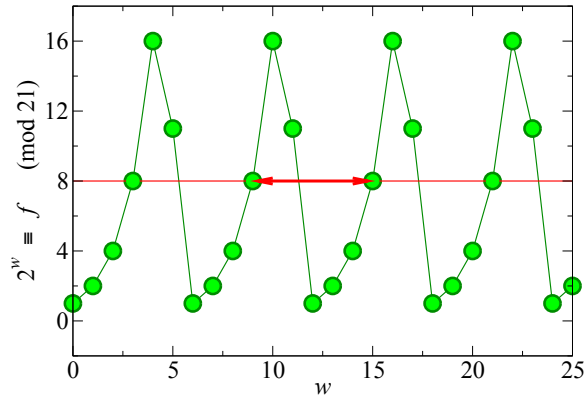



**Fig. 4.1** The discrete, periodic function defined by the congruence $2^w \equiv f \pmod{21}$; the red arrow marks the period $r = 6$

A final example, still in graphic form, is $a = 13$, $m = 15$, $\varphi(m) = 8$, yielding the pattern 1, 13, 4, 7 corresponding to $r = 4$. For reasons that will become apparent later it is convenient to consider a large set of values of $w$; following [80] one takes $2m^2 \leq \max(w) \leq 3m^2$ where, in this case, $2m^2 = 450$ and $3m^2 = 675$. It is also convenient to take the total number of values of $w = 0, 1, 2, \ldots$ equal to a power of 2 (compare with Sect. 4.15); these requirements are fulfilled by taking $\max(w) = 511 = 2^9 - 1$. The graph of $13^w \equiv f \pmod{15}$ is shown in Fig. 4.2 where, for the sake of clarity, the extension of the horizontal axis has been limited to $w = 50$.

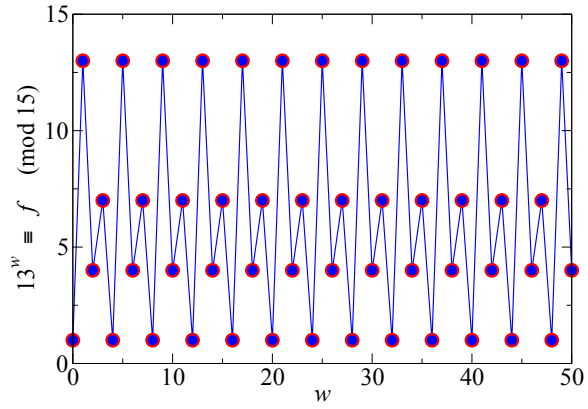



**Fig. 4.2** The discrete, periodic function defined by the congruence $13^w \equiv f \pmod{15}$

Coming back to the general case, it is possible to demonstrate that the period $r$ must necessarily be a divisor of $\varphi(m)$. To better specify the issue, one takes two coprime

numbers $a$ and $m$, and indicates with $r$ the smallest positive integer such that

$$a^r \equiv 1 \pmod{m}; \tag{4.65}$$

some integer $k$ exists such that

$$\varphi(m) = kr + s, \qquad 0 \leq s < r, \tag{4.66}$$

whence

$$b = (a^r)^k \equiv 1 \pmod{m}, \qquad a^{kr+s} = b\,a^s \equiv 1 \pmod{m}. \tag{4.67}$$

The first of (4.67) is obtained by raising (4.65) to the $k$th power, whereas the second one is obtained from (4.61); by construction, $b$ and $m$ are coprime. Applying to (4.67) the transitive property (4.17) yields $b\,a^s \equiv b \pmod{m}$, which has the same form as (4.28). Using the procedure leading to (4.30) yields $a^s \equiv 1 \pmod{m/d}$, where $d = \gcd(b,m)$; on the other hand, $b$ and $m$ are coprime, whence

$$a^s \equiv 1 \pmod{m}. \tag{4.68}$$

This result contradicts the assumption that $r$ is the smallest positive integer such that (4.65) holds, unless one takes $s = 0$; in conclusion, $r$ is necessarily a divisor of $\varphi(m)$.

In the corollary at the end of Sect. 4.8 it has been shown that $\varphi(m)$ is even apart from the trivial cases $\varphi(1) = \varphi(2) = 1$; this property does not apply to $r$: the evenness of $r$ (which, as shown in Sect. 4.11, is necessary in some calculations) must be checked on a case-by-case basis.

## 4.11 Factorization of the Modulus

The findings of Sect. 4.10 lend themselves to a method for factoring the modulus $m$ of (4.61); this is useful in some applications (see, e.g., Sect. 26.9.4). As in Sect. 4.10, let $r$ be the smallest positive integer such that (4.65) holds, so that the latter is recast as $a^r - 1 = k\,m$, with $k$ some integer. Then, assume that $r$ is even;[6] it follows that $r/2$ is an integer, and

$$\frac{(a^{r/2}-1)\,(a^{r/2}+1)}{m} = k, \tag{4.69}$$

where the two factors in the numerator are also integer. Here the general case where $m$ is a composite number is considered; as a consequence, the terms $a^{r/2}-1$ and/or $a^{r/2}+1$ must contain the factors of $m$. Assume that $a^{r/2}-1$ contains some factors of

[6] If $r$ happens to be odd, one must select another value for $a$ and restart the procedure.

$m$: the largest of them is $\gcd(a^{r/2}-1, m)$, to be found by Euclid's algorithm (Sect. 4.1); the same reasoning applies to the case of $\gcd(a^{r/2}+1, m)$.

Note that, depending on the value of $a$, it may happen that $a^{r/2}-1$ or $a^{r/2}+1$ is a multiple of $m$; in this case, only a trivial divisor $m$ is found; another possibility is that the procedure leads to unity, the other trivial divisor of $m$. If, instead, the use of (4.69) yields a non-trivial factor, one concludes that the procedure outlined in this section is an effective method for factoring $m$.

Considering for instance the first example given in Sect. 4.10, $a = 8$, $m = 15$, $r = 4$, it is $a^{r/2} - 1 = 8^2 - 1 = 63$, and the greatest common divisor of 63 and 15 is $p = 3$, which is indeed a non-trivial factor of $m = 15$. Using the other term in the numerator of (4.69) yields $a^{r/2} + 1 = 8^2 + 1 = 65$, and the greatest common divisor of 65 and 15 is $q = 5$, again a non-trivial factor of $m = 15$.

Considering instead the second example of Sect. 4.10, $a = 7$, $m = 15$, $r = 4$, it is $a^{r/2} - 1 = 7^2 - 1 = 48$, and the greatest common divisor of 48 and 15 is again $p = 3$. Using the other term in the numerator yields $a^{r/2} + 1 = 7^2 + 1 = 50$, and the greatest common divisor of 50 and 15 is $q = 5$, the other non-trivial factor of $m = 15$.

Finally, considering the third example of Sect. 4.10, it is $a^{r/2} - 1 = 14^1 - 1 = 13$, and the greatest common divisor of 13 and 15 is $p = 1$, which is a trivial factor of $m = 15$. Using the other term in the numerator yields $a^{r/2} + 1 = 14^1 + 1 = 15$, and the greatest common divisor of 15 and 15 is $q = 15$, which is the other trivial factor of $n = 15$.

The important conclusion is that the factors of a composite integer $m$ can be determined by calculating the period $r$ of a function of the form (4.62). A useful method for determining the period is the Fourier analysis; for this reason, the next sections are devoted to illustrating some properties of it.

## 4.12 Some Properties of the Fourier Series and Transform

In this paragraph a few examples of expansion into a Fourier series are considered, to the purpose of identifying some properties of the spectrum. Given a periodic function $g(t)$ of period $T$, the general expression of its Fourier expansion is

$$g(t) = \frac{1}{2} a_0 + \sum_{n=1}^{\infty} \left[ a_n \cos(n\omega t) + b_n \sin(n\omega t) \right], \qquad \omega = 2\pi/T, \tag{4.70}$$

with

$$a_n = \frac{2}{T} \int_{-T/2}^{T/2} g(t) \cos(n\omega t)\, \mathrm{d}t, \qquad b_n = \frac{2}{T} \int_{-T/2}^{T/2} g(t) \sin(n\omega t)\, \mathrm{d}t. \tag{4.71}$$

One notes that $a_{-n} = a_n$ and $b_{-n} = -b_n$; in particular, $b_0 = 0$. The complex form of (4.70) is obtained from the identities $2\cos(\varphi) = \exp(\mathrm{i}\varphi) + \exp(-\mathrm{i}\varphi)$ and $2\,\mathrm{i}\sin(\varphi) = \exp(\mathrm{i}\varphi) - \exp(-\mathrm{i}\varphi)$ whence, letting

$$c_n = \frac{1}{2}\,(a_n - \mathrm{i}\,b_n) = \frac{1}{T}\int_{-T/2}^{T/2} g(t)\,\exp(-n\,\mathrm{i}\,\omega\,t)\,\mathrm{d}t\,, \tag{4.72}$$

and observing that $c_n^* = c_{-n}$, yields $a_n\cos(n\,\omega\,t) + b_n\sin(n\,\omega\,t) = c_n\exp(n\,\mathrm{i}\,\omega\,t) + c_{-n}\exp(-n\,\mathrm{i}\,\omega\,t)$ for $n = 1,2,\ldots$; also, it is $c_0 = a_0/2$, which equals the average value of $g(t)$ over the period $T$. In conclusion, (4.70) becomes

$$g(t) = \sum_{n=-\infty}^{+\infty} c_n\,\exp(n\,\mathrm{i}\,\omega\,t)\,, \qquad \omega = 2\,\pi/T\,. \tag{4.73}$$

Consider by way of example the function shown in Fig. 4.3; by inspecting the second of (4.71) one finds that $b_n = 0$ because $g(t)$ is even. As for $a_n$, it is found

$$a_0 = \frac{2\,A\,\tau}{T}\,, \qquad a_n = \frac{2\,A}{n\,\pi}\,\sin(n\,\pi\,\tau/T)\,, \quad n = 1,2,\ldots \tag{4.74}$$

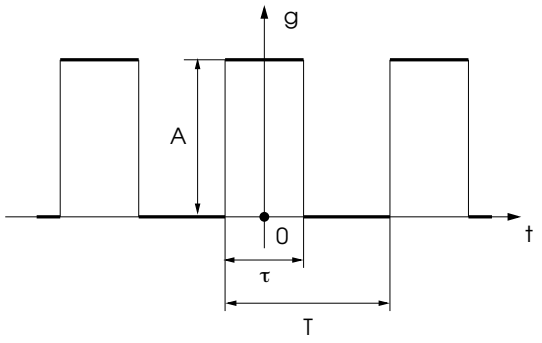



**Fig. 4.3** Square-wave function. The coefficients of the Fourier expansion are given by 4.74

Similarly, consider the function shown in Fig. 4.4; by inspecting the first of (4.71) one finds that $a_n = 0$ because $g(t)$ is odd. As for $b_n$, it is found

$$b_n = -\frac{2\,A}{n\,\pi}\,\cos(n\,\pi)\,. \tag{4.75}$$

The blue dots and the red squares of Fig. 4.5 show the Fourier coefficients $a_n$, given by (4.74), of the square-wave function of Fig. 4.3; the two sets of coefficients correspond to different values of ratio $\tau/T$. Still in Fig. 4.5, the green triangles show the Fourier coefficients $b_n$, given by (4.75), of the saw-like function of Fig. 4.4. As

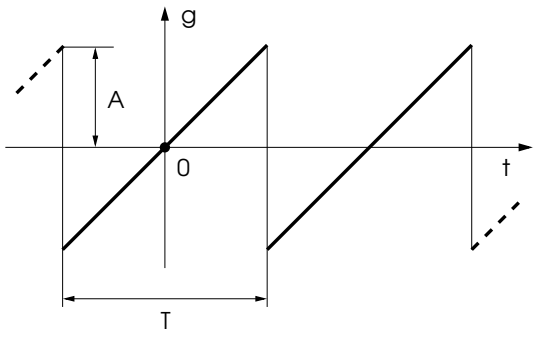



**Fig. 4.4** Saw-like function. The coefficients of the Fourier expansion are given by (4.75)

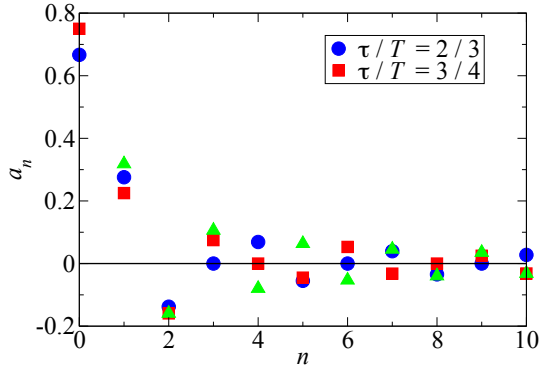



**Fig. 4.5** The blue dots and the red squares show the coefficients (4.74) of the Fourier expansion of the function shown in Fig. 4.3, with $A = 1/2$, for two different values of ratio $\tau/T$. The green triangles show the coefficients (4.75) of the saw-like function of Fig. 4.4, still with $A = 1/2$

mentioned above, $a_0/2$ provides the average value of $g(t)$ over the period $T$. All the other triads of symbols correspond to the index $n = 1, 2, \ldots$ of one of the harmonics, where the coefficient with the largest modulus is that of the fundamental harmonic ($n = 1$).

## 4.13 Discrete Fourier Transform

For the considerations that follow, it is convenient to start by reminding the definition of the Fourier and inverse-Fourier transforms of a function $h(t)$ that depends on time [56, Ch. 12], [81, Ch. 4]:

$$H(f) = \int_{-\infty}^{+\infty} h(t)\, \exp(-2\pi \mathrm{i} f t)\, \mathrm{d}t\,, \quad h(t) = \int_{-\infty}^{+\infty} \frac{H(f)}{2\pi}\, \exp(2\pi \mathrm{i} f t)\, \mathrm{d}f\,. \quad (4.76)$$

In many practical cases, $h$ is sampled at evenly-spaced intervals $\Delta t$, so that the continuous function is transformed into a sequence of $N$ samples $h_k = h(t_k) = h(k\,\Delta t)$, with $k = 0, 1, 2, \ldots, N-1$. It is assumed here that the bandwith $f_M$ of $h$ is limited, $f_M < f_c$, so that, taking $\Delta t = 1/(2\,f_c)$, the number of samples is finite and one can reconstruct $h$ using the Shannon sampling theorem [63, 64]. The first integral in (4.76) then becomes

$$H(f) \simeq \Delta t \sum_{k=0}^{N-1} h_k \exp(-2\,\pi\,\mathrm{i}\,f\,k\,\Delta t)\,, \tag{4.77}$$

where the time variable has been discretized whereas the frequency $f$ is still a continuous variable. Now, assume for simplicity that the total number of samples, $N$, is even; the Discrete Fourier Transform (DFT) of $h$ is calculated from (4.77) by selecting $N$ discrete values of the frequency,

$$f\,\Delta t \leftarrow f_n\,\Delta t = \frac{n}{N}\,, \qquad n = -\frac{N}{2}, \ldots, +\frac{N}{2}\,, \tag{4.78}$$

so that (4.77) becomes

$$H(f) \;\rightarrow\; H(f_n) = \Delta t\,H_n\,, \qquad H_n = \sum_{k=0}^{N-1} h_k \exp(-2\,\pi\,\mathrm{i}\,k\,n/N)\,. \tag{4.79}$$

The above shows that the discrete Fourier transform maps the $N$ numbers $h_k$ into the $N$ complex numbers $H_n$; the memory of the time domain in which $h$ was originally defined is embedded only in the coefficient $\Delta t$. The $h_k \rightarrow H_n$ mapping can be viewed as the multiplication of a vector of length $N$ by an $N \times N$ matrix:

$$H_n = \sum_{k=0}^{N-1} W^{nk} h_k\,, \qquad W = \exp(-2\,\pi\,\mathrm{i}/N)\,, \tag{4.80}$$

so that its computational cost is essentially $\mathrm{O}(N^2)$. By a clever algorithm, called *Fast Fourier Transform* (FFT) and described in Sect. 4.14, the cost is reduced to $\mathrm{O}(N \log_2 N)$.

## 4.14 Fast Fourier Transform

Remembering that $N$ is even, one separates the summands of even index in (4.80) from those of odd index, to obtain

$$H_n = F_n^{\mathrm{e}} + W^n F_n^{\mathrm{o}}\,, \quad F_n^{\mathrm{e}} = \sum_{j=0}^{N/2-1} W^{n\,2\,j} h_{2\,j}\,, \quad F_n^{\mathrm{o}} = \sum_{j=0}^{N/2-1} W^{n\,2\,j} h_{2\,j+1}\,. \tag{4.81}$$

The computational cost of evaluating the right hand side of (4.81) is essentially $2\,\mathrm{O}(N^2/4)$. On the other hand, it is always possible to set the number of samples such that $N = 2^L$; it follows that $N/2$ is also even, and the procedure leading to (4.81) can be repeated, to reduce the cost to $4\,\mathrm{O}(N^2/16)$. By continuing in this manner, after $L$ iterations one eventually reduces the problem to the evaluation of elementary blocks. By way of example, let $N = 8 = 2^3$; one finds $H_n = F_n^{\mathrm{e}} + W^n F_n^{\mathrm{o}}$ with

$$F_n^{\mathrm{e}} = h_0 + h_2 W^{2n} + h_4 W^{4n} + h_6 W^{6n}\,, \quad F_n^{\mathrm{o}} = h_1 + h_3 W^{2n} + h_5 W^{4n} + h_7 W^{6n}\,. \tag{4.82}$$

Then, $F_n^{\mathrm{e}} = F_n^{\mathrm{ee}} + W^{2n} F_n^{\mathrm{eo}}$ and $F_n^{\mathrm{o}} = F_n^{\mathrm{oe}} + W^{2n} F_n^{\mathrm{oo}}$, with

$$F_n^{\mathrm{ee}} = h_0 + h_4 W^{4n}\,, \qquad F_n^{\mathrm{eo}} = h_2 + h_6 W^{4n}\,, \tag{4.83}$$

$$F_n^{\mathrm{oe}} = h_1 + h_5 W^{4n}\,, \qquad F_n^{\mathrm{oo}} = h_3 + h_7 W^{4n}\,, \tag{4.84}$$

so that the last step is

$$F^{\mathrm{eee}} = h_0\,, \quad F^{\mathrm{eeo}} = h_4\,, \quad F^{\mathrm{eoe}} = h_2\,, \quad F^{\mathrm{eoo}} = h_6\,, \tag{4.85}$$

$$F^{\mathrm{oee}} = h_1\,, \quad F^{\mathrm{oeo}} = h_3\,, \quad F^{\mathrm{ooe}} = h_5\,, \quad F^{\mathrm{ooo}} = h_7\,. \tag{4.86}$$

One notes that index $n$ is missing in (4.85) and (4.86) because the right hand sides do not depend on $n$; also, still in (4.85) and (4.86), the number of apices of terms $F$ is $3 = \log_2 N$. As anticipated in Sec. 4.13, the cost of the FFT is $\mathrm{O}(N \log_2 N)$ [56, Sect. 12.2].

The advantage of FFT over DFT is enormous. By way of example, for $N = 10^6$ it is $N/\log_2 N \simeq 5 \times 10^4$, namely, if the FFT takes 10 s on some machine, the DFT on the same machine takes about 6 days.

## 4.15 Discrete Fourier Transform of a Periodic Function

The entries $h_0, h_1, \ldots, h_{N-1}$ of the vector considered in Sects. 4.13 and 4.14 have no special properties; here, it will be assumed that $h_k$ is *periodic*, namely, that some index $r$ exists such that

$$h_{k+gr} = h_k\,, \qquad g = \pm 1, \pm 2, \ldots\,, \tag{4.87}$$

where $k + g\,r \in [0, 1, \ldots, N-1]$. In principle there is no relation between $r$ and $N$; for the sake of simplicity, it is assumed that the period $r$ is a divisor of $N$. Then, letting $m = N/r$, the sum in (4.80) is partitioned into $m$ blocks, each made of $r$ summands:

$$H_n = \sum_{k=0}^{r-1} + \sum_{k=r}^{2r-1} + \cdots + \sum_{k=(m-1)\,r}^{mr-1}\,, \tag{4.88}$$

with $mr = N$. The matrix elements of the first two sums in (4.88) are

$$W^{nk} = \exp(-2\pi \mathrm{i} nk/N)\,, \quad W^{n(k+r)} = W^{nk} q\,, \quad q = \exp(-2\pi \mathrm{i} nr/N)\,. \quad (4.89)$$

By the same token, the matrix elements of the other sums in (4.88) become $W^{nk}q^2, \ldots, W^{nk}q^{m-1}$. On the other hand, due to (4.87) the same set of $h_k$ values is replicated in each sum; in conclusion, observing that $1 + q + \cdots + q^{m-1} = (q^m - 1)/(q-1)$, expression (4.88) of $H_n$ becomes

$$H_n = \frac{1 - \exp(-2\pi \mathrm{i} n)}{1 - \exp(-2\pi \mathrm{i} n/m)} \sum_{k=0}^{r-1} W^{nk} h_k\,. \quad (4.90)$$

One notes that, if $n/m = rn/N$ is not an integer, the denominator of the fraction in (4.90) is different from zero; the numerator, instead, is always equal to zero. It follows that $H_n = 0$ for all indices such that $n/m$ is not an integer. If, instead, $n/m = s$, with $s$ an integer, the fraction in (4.90) takes the $0/0$ form; in this case, one replaces $n$ with $n + \varepsilon$ and calculates the $\varepsilon \to 0$ limit of the fraction, to find

$$\lim_{\varepsilon \to 0} \frac{1 - \cos[2\pi(n+\varepsilon)] + \mathrm{i}\sin[2\pi(n+\varepsilon)]}{1 - \cos[2\pi(n+\varepsilon)/m] + \mathrm{i}\sin[2\pi(n+\varepsilon)/m]} = m = \frac{N}{r}\,. \quad (4.91)$$

Therefore, the non-vanishing elements of the transformed vector read

$$H_n = \frac{N}{r} \sum_{k=0}^{r-1} h_k \exp(-2\pi \mathrm{i} nk/N) = \frac{N}{r} \sum_{k=0}^{r-1} h_k \exp(-2\pi \mathrm{i} sk/r)\,, \quad (4.92)$$

where the admissible values of $n$ are such that $s = nr/N$ is an integer. By way of example, consider the case of the periodic function shown in Fig. 4.2, whose pattern is $h_0 = 1$, $h_1 = 13$, $h_2 = 4$, $h_3 = 7$, and whose period is $r = 4$; assuming $N = 2^9 = 512$, the number of blocks in (4.88) turns out to be $m = N/r = 128$, and the condition by which $s = n/m$ is an integer, when $n = -N/2, \ldots, +N/2$, holds for

$$n = \{-256, -128, 0, 128, 256\}\,, \qquad s = \{-2, -1, 0, 1, 2\}\,. \quad (4.93)$$

The non-vanishing elements of the transformed vector turn out to be

$$H_0 = 128 \sum_{k=0}^{3} h_k\,, \qquad H_{\pm 128} = 128 \sum_{k=0}^{3} h_k \exp(\mp \pi \mathrm{i} k/4)\,,$$

$$H_{\pm 256} = 128 \sum_{k=0}^{3} h_k \exp(\mp \pi \mathrm{i} k/2)\,. \quad (4.94)$$

As the elements $h_k$ are real, it follows that $H_{-128} = H_{128}^*$, $H_{-256} = H_{256}^*$.

As a second example, consider the case where the period of $h_k$ is $r = 8$, and assume that $N = 2^r = 256$; the number of blocks in (4.88) turns out to be $m = N/r = 32$, and the condition by which $s = n/m$ is an integer holds for

$$n = \{0, 32, 64, 96, 128, 160, 192, 224\}\,, \qquad s = \{0, 1, 2, 3, 4, 5, 6, 7\}\,. \tag{4.95}$$

The non-vanishing elements of the transformed vector turn out to be

$$H_0 = 32 \sum_{k=0}^{r-1} h_k\,, \qquad H_{32} = 32 \sum_{k=0}^{r-1} h_k \exp(-2\pi \mathrm{i}\, k/8)\,, \qquad \ldots$$

$$\ldots \qquad H_{224} = 32 \sum_{k=0}^{r-1} h_k \exp(-2\pi \mathrm{i}\, 7k/8)\,. \tag{4.96}$$

Observing that in general these element are complex, it is convenient to consider their square modulus $|H_{sm}|^2$. The set of such square moduli is called the *power spectrum* of vector $h_1$, $h_2$, …, $h_N$.

## 4.16 Connection with Cryptography

With reference to the application of the above algorithms to cryptography, it is worth anticipating some issues related to the RSA encryption algorithm (Sect. 5.6). The most computationally-expensive part of the decryption of a message encrypted with RSA, is the factorization of the modulus. A method for tackling the factorization is based on congruence (4.62) where, as illustrated in Sect. 4.11, function $f_m(w;a)$ is periodic $(\mathrm{mod}\, m)$ with a period equal to $\varphi(m)$; in turn, $a$ is coprime with $m$. It may happen that $\varphi(m)$ is not the minimum period associated to a given pair $a$ and $m$, namely, a positive integer $r < \varphi(m)$ may exist that is also a period; if this happens, $r$ is a divisor of $\varphi(m)$ (compare with (4.64)). Changing the base $a$ gives rise again to a periodic function, with a different pattern.

Instead of factoring the modulus $m$ directly, one may try to find another integer $c$ that has a common factor with $m$; in this case, in fact, the common factor can be calculated with the much cheaper Euclid algorithm (Sect. 4.1), that provides $\gcd(c,m)$. As shown in Sect. 4.11, if the minimum period $r$ of $f_m(w;a)$ is even, then $a^{r/2} - 1$ and/or $a^{r/2} + 1$ must contain the factors of $m$ (compare with (4.69)); the procedure may lead to finding the trivial factors 1 and $m$: if this happens, one must repeat it starting with a different base $a$.

In conclusion, if the minimum period $r$ of (4.62) is known, the factorization of modulus $m$ is computationally affordable on a classical computer; the same applies to the calculation necessary to evaluate $f_m(w;a)$. Unfortunately, there is no known method for calculating the period efficiently using a classical computer; instead, a quantum algorithm provides a more efficient approach (Sect. 26.9.4).