1 Introduction

1.1 The early days of quantum mechanics

These lectures are concerned with one of the most important pillars of modern physics namely quantum mechanics.

The first understanding of the quantum aspects of Nature goes back to Planck's black body radiation formula in 1900. Planck based his famous formula on the hypothesis that light - and more generally electromagnetic radiation - could be emitted by matter only by discrete quantities whose energies ϵ are related to their frequencies ν as

$$\epsilon = h\,\nu \quad \text{or} \quad \epsilon = \hbar\,\omega \;, \tag{1.1}$$

where, of course, $\omega = 2\pi\nu$. The constant h is a fundamental constant, called Planck's constant and $\hbar = \frac{h}{2\pi}$ is called the reduced Planck's constant. Their numerical values are

$$h \simeq 6.6261 \; 10^{-34} \; \text{J\,s} \quad , \quad \hbar = \frac{h}{2\pi} \simeq 1.05457 \; 10^{-34} \; \text{J\,s} \;. \tag{1.2}$$

A few years later, in 1905, Einstein published his paper on the photo-electric effect and he advocated the idea that light is not only emitted by discrete energies but that light itself is composed by quanta, later called photons, that have the discrete energies given above in (1.1). The photon also has a momentum $\vec{p}$ related to the wave-vector $\vec{k}$ by a similar relation

$$\vec{p} = \hbar\,\vec{k} \quad , \quad |\vec{k}| = \frac{2\pi}{\lambda} \;, \tag{1.3}$$

where λ is the wave-length. This relation was very directly confirmed by the scattering of light (X-rays) with electrons in a thin aluminium foil in 1923 by Compton.

Another important step was taken by Bohr around 1912-1914 when he tried to understand the structure of atoms and explain their emission and absorption spectra. Bohr postulated that the emission and absorption of light by atoms can only occur at frequencies

$$\nu_{fi} = \frac{|E_f - E_i|}{h} \;, \tag{1.4}$$

and that the discreteness of the spectra corresponds to a discreteness of the energies of the atoms. In particular the known empirical Balmer formula for the hydrogen atom $\hbar\,\nu_{nk} = 13.6\,\text{eV} \times \left(\frac{1}{n^2} - \frac{1}{k^2}\right)$ led to the famous formula of the energy levels of the hydrogen atom

$$E_n = \frac{E_1}{n^2} \quad , \quad E_1 = -13.6\,\text{eV} \;. \tag{1.5}$$

Note that, while the Joule (J) is a convenient unit for energies in the macroscopic world, in the realm of microscopic physics one more often encounters the electron Volt (eV). One has $1\,\text{eV} = 1.602 \; 10^{-19}\,\text{J}$, and hence

$$h \simeq 4.136 \; 10^{-15}\,\text{eV\,s} \quad , \quad \hbar \simeq 6.582 \; 10^{-16}\,\text{eV\,s} \;. \tag{1.6}$$

Somewhat later in 1923, de Broglie postulated that, just as light has a dual particle-like behaviour, matter also has a dual wave-like behaviour with the wave-length or wave-vector related to the momentum by the same relation (1.3). This allowed to explain the discrete energy levels of (hydrogen) atoms (1.5) as corresponding to stationary electron waves in these atoms.

In the following years the full formalism of modern quantum mechanics was very rapidly developed between 1925 and 1927 by a few brilliant (and mostly young) physicists and mathematicians : Schrödinger, Heisenberg, Born, Bohr, Dirac, Hilbert and von Neumann, as well as others. Since then, quantum mechanics has been confirmed over and over again by experiment and has opened the way to almost all major discoveries in physics during the last 100 years.

1.1.1 The Franck-Hertz experiment

In 1914, Frank and Hertz were directing a beam of accelerated electrons on a gas of Mercury atoms. As long as the kinetic energy of the electrons was not too large, they just observed that the electrons are elastically scattered by the Mercury atoms that are about $400\,000$ times more massive than the electrons. However, as the electron energy was increased above $4.9\,\mathrm{eV}$ an important fraction of electrons was scattered inelastically, having lost precisely $4.9\,\mathrm{eV}$ of their energy. Moreover, the Mercury atoms then emit an ultraviolet radiation of the well-known emission wave-length of Mercury of $\lambda = 253\,\mathrm{nm}$ which corresponds to a frequency $\nu = \frac{c}{\lambda} = \frac{3\times10^8\,\mathrm{m\,s^{-1}}}{253\times10^{-9}\,\mathrm{m}} = 1.18\ 10^{15}\,\mathrm{s^{-1}}$. Using (1.1) and the value of h as given in (1.6) this corresponds exactly to the $4.9\,\mathrm{eV}$ lost by the electrons !

This experiment was a direct confirmation of Bohr's hypothesis about the atomic energy levels being discrete and that the atomic electron can make a transition between two such levels by absorbing just the right amount of energy, which can then again be released as a photon of corresponding energy and frequency. The same scattering experiment can be repeated by replacing the Mercury atoms by a gaz of diatomic molecules that can vibrate and act as a harmonic oscillator. Then the observed energy loss of the electrons in the beam are all of the form $n \times h\nu$ and by the same argument one infers the energy levels of the corresponding harmonic oscillator as $E_n = n\,h\nu + E_0\,,\ n = 0, 1, 2, \ldots$. We will indeed see later-on that for a harmonic oscillator of frequency $\nu = 2\pi\omega$ one has

$$E_n = \hbar\omega\big(n + \frac{1}{2}\big)\,. \tag{1.7}$$

1.1.2 Young's double slit interference

In 1927, Davisson and Germer, showed that an electron beam was diffracted by a nickel crystal, just as X-rays are diffracted by crystals. This was the first direct evidence of the wave-like behaviour of matter as predicted by de Broglie. Here, let us discuss instead Young's double slit experiment that is conceptually simpler to analyse, but experimentally more difficult to realize. The modern realization[1] is the following. A cloud of a few millions of neon atoms is captured and cooled to about $10^{-3}\,\mathrm{K}$ by a laser trap, a few cm above a screen with 2 slits, each $2\ 10^{-6}\,\mathrm{m}$ wide, and separated by $a = 6\ 10^{-6}\,\mathrm{m}$. The trap is then removed and the atoms "fall" under the effect of gravity. Some $d = 80\,\mathrm{cm}$ below the double-slit screen they hit a detector plate where their

---

[1]F. Shimizu, K. Shimizu and H Takuma, Phys. Rev A46, R17 (1992)

impact is registered. One sees that the impacts distribute according to an interference pattern with parallel lines of numerous impacts, separated by lines with few or no impact.

The same experiment can be done with other particles like electrons, neutrons or even molecules. One always gets the interference pattern. For a plane wave of wave-length λ hitting the double slit, the separation between to adjacent maximum intensity lines would be[2] $\Delta x = \lambda d/a$ and the observed separation with atoms, electrons, neutrons, etc is perfectly in agreement with this formula for a wave-length[3] given by de Broglie's relation (1.3). In some of the variants of this experiment one can dim the particle source enough so that the particles essentially arrive one by one on the screen. Similarly in the free fall of the neon atoms one can also see how the impacts form over time.

The important observations are :

- Each particle makes an impact at one precise point as expected for a point-like particle.

- There is no way to predict where the impact of any given particle will be : it is random.

- It is only when a large number of impacts has been realized by many particles that the interference picture emerges.

- If one closes one or the other slit one gets a Gaussian type distribution of the impacts. The interference pattern observed with two open slits is *not* the sum of the two individual distributions with one open slit.

One can also try to figure out through which slit a given particle has "travelled" by putting a detector behind the corresponding slit. But this means that the particle gets actually absorbed by the detector which has the same effect as closing the corresponding slit and then the interference pattern no longer is there. This is another general lesson : One cannot measure a microscopic system without perturbing it.

We are led to the following conclusions. The particles do behave as individual particles, but there is no way to establish through which slit a given particle has gone. The outcome of each measurement (impact of the particle) is random and can only be described in terms of a probability. Since all particles are independent, *the observed interference pattern for many particles must already be present in the description of the probability where a single particle will be observed.*

Interference patterns for electromagnetic waves (electric field $\vec{E}$) are observed, because we observe the intensity as $I \sim |\vec{E}|^2$ and for a 2-slit experiment we have $\vec{E} = \vec{E}_1 + \vec{E}_2$ and then

$$I_{12} \sim |\vec{E}_1 + \vec{E}_2|^2 = |\vec{E}_1|^2 + |\vec{E}_2|^2 + 2\,\mathrm{Re}\big(\vec{E}_1 \cdot \vec{E}_2^*\big) \quad \Rightarrow \quad I_{12} \neq I_1 + I_2 \ . \tag{1.8}$$

In analogy with this observation we are led to postulate that

---

[2] Let us remind the reader how this result is obtained. The interference is due to the different path lengths ℓ_1 and ℓ_2 from the slits with coordinates $-\frac{a}{2}$ and $+\frac{a}{2}$ to the point with coordinate x on the screen. The screen is at a distance d from the slits. We suppose that $d \gg a,\, x$. By Pythagor's theorem we have $\ell_1^2 = (\frac{a}{2} + x)^2 + d^2$ and $\ell_2^2 = (\frac{a}{2} - x)^2 + d^2$, as one easily sees from a simple drawing. Then, $\ell_1^2 - \ell_2^2 = 2ax$ and $\ell_1 - \ell_2 = (\ell_1^2 - \ell_2^2)/(\ell_1 + \ell_2) \simeq (\ell_1^2 - \ell_2^2)/(2d) = ax/d$. The distance between two maxima or two minima occurs if $\ell_1 - \ell_2 = \lambda$, hence $\lambda = a\Delta x/d$.

[3] Of course, in the free-fall realization of the experiment, one must take into account that the momentum p increases during the free fall.

- for each particle we can define a probability *amplitude* $\mathcal{A}_1(x)$ that it passes through the slit $i = 1$ and makes an impact at a given point x of the detector plate, and similarly a probability amplitude $\mathcal{A}_2(x)$ that is passes through the slit $i = 2$ and makes an impact at x of the detector plate,

- that the total probability amplitude to make an impact at x is the sum of the individual amplitudes $\mathcal{A}(x) = \mathcal{A}_1(x) + \mathcal{A}_1(x)$, and

- that the probability to make an impact at x is $P(x) = |\mathcal{A}(x)|^2$.

Then if each individual probability amplitude reflects the wave-like nature, this postulate perfectly explains the presence or absence of interferences according to whether both slits are open or one is obstructed. It is important to realize that we assign a probability amplitude to a single particle, and hence that the probability for an individual particle already contains the interference picture. But as always with probabilities, in order to be able to observe a probabilistic distribution we need to realize many events.

1.1.3   The Davisson-Germer experiment

As already mentioned, in 1927 Davisson and Germer studied the diffraction of a beam of electrons of well-defined kinetic energy (and hence momentum) by a cristal and obtained the same diffraction pattern as for the diffraction of X-rays provided the wave-length of the electrons given by de Broglie's formula as $\lambda = \frac{2\pi\hbar}{|\vec{p}|}$ is the same as those of the corresponding X-rays. We could again discuss the diffracation and how the pattern gives information about the cristal, but we refer the reader to the corresponding discussion for X-rays. Let us only mention that the diffraction of electrons or neutrons by crystals or more generally by matter samples has become a powerful method to study the structure of these matter samples.

1.1.4   Quantum mechanics and modern life

Of course our modern life and technology could not be separated from quantum mechanics. It is not possible to give even an incomplete list of where quantum mechanics plays a major role. From lasers and atomic clocks, to semi-conductors, nuclear resonance imaging, atomic force microscopes, to quantum computing and quantum cryptography, quantum mechanics is all over the place !

1.2   Bibliography

There are many excellent books on quantum mechanics. These lectures have been profoundly influenced by :

- C. Cohen-Tannoudji, B. Diu and F. Laloë, *Mécanique quantique*, CNRS éditions, edp sciences, vol 1 and 2;

- J.-L. Basdevant and J. Dalibard, *Quantum mechanics*, éditions Ecole polytechnique, Springer.

Both exist in English and in French.

2 The basics of quantum mechanics for 2-state or N-state systems

We will begin our exploration of quantum mechanics in the simplest setting which are systems having only two states.

2.1 Why 2 state systems ?

Already a simple atom like the hydrogen atom is very complicated to describe, even if we consider the usual approximation of an electron in a static Coulomb potential created by the nucleus considered as fixed in space. There are infinitely many energy levels and the corresponding quantum mechanical description requires an infinite-dimensional vector space which has the mathematical structure of a Hilbert space. A somewhat simpler system is the one-dimensional[4] quantum mechanical oscillator which also has infinitely many energy levels, as already mentioned in (1.7), and thus also requires an infinite-dimensional vector space. Actually almost all quantum mechanical systems of a single particle require such an infinite-dimensional vector space.

However, fortunately, for some systems one can consider approximations where the essential physics reduces to only two states and then the quantum mechanical description only requires a two-dimensional vector space. We will begin by studying such 2-state systems first, as they are mathematically much simpler and still allow to exemplify (almost) all important quantum mechanical features. In a certain way, the above description of Young's two-slit experiment corresponds to such a 2-state system.

Actually 2-state systems are very common. For example, the spin of an electron or of any other spin-$\frac{1}{2}$ particle constitutes a 2 state system, if one can ignore the spatial properties of the particle, e.g. because it is somehow trapped, or because they are just not important for the experiment one wants to describe. Similarly, photons have two possible independent polarisations, and if the experiment only concentrates on the polarisation we have another 2-state system. The ammonia molecule has two equivalent ground state configurations, and if one can neglect all states corresponding to the excited states of higher energy (e.g. because one works at very low temperature) we may also consider it as a 2-state system. This gives rise to the ammonia maser. We may discuss the phenomenon of magnetic nuclear resonance within the framework of a 2-state system. In quantum computing a 2-state system is called a q-bit. We see that studying 2-state systems is not only much nicer mathematically, but they are also physically quite relevant.

2.2 The polarisation of photons

Let us begin by discussing the familiar phenomenon of polarisation of light. In the context of electromagnetic waves, the polarisation is just the direction of the electric field and we know

---

[4]Of course, one should not confuse the dimension of the physical space (3 for the hydrogen atom, 1 for an oscillator vibrating along a single direction only) and the dimension of the vector space used in the quantum mechanical description.

from Maxwell's equations that this direction lies in the plane perpendicular to the direction of propagation. Let us assume that the wave propagates in the z-direction so that the electric field is $\vec{E}(z,t) = \vec{E}_0 e^{i(kz-\omega t)}$ (with $k = \omega/c$) and $\vec{E}_0$ in the $x-y$ plane. Let us send this electromagnetic wave through a polariser that lets pass only polarisations in a direction making an angle θ with the x-axis. We get a wave with

$$\vec{E}_0 = E_0 \, \vec{e}_p \quad , \quad \vec{e}_p = \cos\theta \, \vec{e}_x + \sin\theta \, \vec{e}_y \; . \tag{2.1}$$

We then put another polariser (analyser) either aligned with $\vec{e}_x$ or with $\vec{e}_y$. In the first case we will observe an intensity $|E_0|^2 \cos^2\theta$ and in the second case an intensity $|E_0|^2 \sin^2\theta$.

Let us repeat this experiment with a very low luminosity of light so that we can consider dealing with individual photons. We can consider that each photon present after the first polariser is polarised in the direction $\vec{e}_p$. The effect of the analyser on a single photon is that it either passes or is absorbed. From our discussions of the intensity of the electric field we know that for the analyser in the x-direction it should have a probability $\cos^2\theta$ to pass and probability $\sin^2\theta$ to be absorbed, and for the analyser in the y-direction it should have a probability $\sin^2\theta$ to pass and probability $\cos^2\theta$ to be absorbed. If the photon passes the analyser in the x-direction we know that afterwards it is polarised in this x-direction, and similarly if it passes the analyser in the y-direction it will be polarised in this y-direction.

Let us rephrase these statements in a quantum mechanical language describing the process of preparing a state and the process of measuring. Our first polariser has prepared each photon in a state that is polarised in the $\vec{e}_p$ direction. We consider our analyser as the measuring apparatus. For a given measuring configuration, say the analyser in the x-direction, there are only two possible outcomes : either the photon passes the analyser or it gets absorbed. The probability to pass is $|\vec{e}_p \cdot \vec{e}_x|^2 = \cos^2\theta$ and the probability to be absorbed is $|\vec{e}_p \cdot \vec{e}_y|^2 = \sin^2\theta$. We could then call $\vec{e}_x = \vec{e}_{\text{pass}}$ and $\vec{e}_y = \vec{e}_{\text{abs}}$, so that the probabilities are

$$P(\text{pass}) = |\vec{e}_p \cdot \vec{e}_{\text{pass}}|^2 \quad , \quad P(\text{abs}) = |\vec{e}_p \cdot \vec{e}_{\text{abs}}|^2 \; . \tag{2.2}$$

We can say that our quantum system (the photon) is initially prepared in some superposition $\vec{e}_p$ of the "eigenstates" $\vec{e}_{\text{pass}}$ and $\vec{e}_{\text{abs}}$ of the analyser, and that the probability amplitudes are given by the scalar product of $\vec{e}_p$ with the corresponding eigenstate, the probabilities being the modulus squared of the amplitudes. The measurement changes the state of the photon : if it passes the analyser its state will be $\vec{e}_{\text{pass}}$ afterwards.

The name "eigenstates" is used since more generally, as we will see next, these vectors are associated with eigenvectors of some linear operators (matrices) that represent the physical quantity we want to measure. At present, these two vectors $\vec{e}_{\text{pass}} = \vec{e}_x$ and $\vec{e}_{\text{abs}} = \vec{e}_y$ span a real 2-dimensional vector space isomorphic to $\mathbf{R}^2$.

Let us note that the electric field is real, but it is often *convenient* to work with complex fields instead. In this case one uses complex polarisation vectors, i.e. $\vec{e}_p = a\vec{e}_x + b\vec{e}_y$, with $a, b \in \mathbf{C}$ (and $|a|^2 + |b|^2 = 1$). Then our two-dimensional vector space describing the polarisations is a complex vector space isomorphic to $\mathbf{C}^2$.

2.3 The Hilbert space for 2-state systems (single q-bit)

Drawing on the previous experimental examples and their discussion, we will now embark to formulate the postulates of quantum mechanics for 2-state systems. As we will see later-on, these postulates will carry over to general systems, except that parts of them have to be formulated more carefully mathematically to take into account the subtleties of infinite-dimensional vector spaces. For now these subtleties do not occur. We begin by carefully describing the structure of the relevant complex two-dimensional vector space which is the simplest example of a Hilbert space. While using complex polarisations was a mere convenience in the example of an electric field, it will turn out that quantum mechanics intrinsically requires complex vector spaces.

We denote this Hilbert space as $\mathcal{H}$. Being two-dimensional, this Hilbert space is isomorphic to $\mathbf{C}^2$. Of course, $\mathbf{C}^2$ is just the set of pairs of complex numbers $a, b \in \mathbf{C}$ which we can write as $\begin{pmatrix} a \\ b \end{pmatrix}$. What makes this a Hilbert space is the inner product between two elements $\psi = \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbf{C}^2$ and $\phi = \begin{pmatrix} c \\ d \end{pmatrix} \in \mathbf{C}^2$, defined as

$$(\phi, \psi) = \phi^\dagger \psi = (\phi^*)^{\mathrm{T}} \psi = (c^*, d^*) \begin{pmatrix} a \\ b \end{pmatrix} = c^* a + d^* b \; , \tag{2.3}$$

where $c^* \equiv \bar{c}$ denotes the complex conjugate of c. Note the definitions of the adjoint $\phi^\dagger$ and transposed ϕ^{T}. Note also the important property of the inner product under complex conjugation

$$(\phi, \psi)^* = (\psi, \phi) \; , \tag{2.4}$$

and that this inner product is linear in the second factor and anti-linear in the first, i.e. for any complex numbers λ, μ we have

$$(\phi, \lambda\psi + \mu\chi) = \lambda(\phi, \psi) + \mu(\psi, \chi) \quad , \quad (\lambda\phi + \mu\chi, \psi) = \lambda^*(\phi, \psi) + \mu^*(\chi, \psi) \; . \tag{2.5}$$

It is often useful to introduce a basis in a given vector space. Recall that in an n-dimensional vector space, a basis is a set of n linearly independent vectors such that any element of the vector space can be written as a unique linear combination of these basis vectors. Obviously, a basis in $\mathbf{C}^2$ consists of $\{e_0, e_1\}$ with

$$e_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad , \quad e_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \; , \quad \Rightarrow \quad \psi = a\,e_0 + b\,e_1 \; . \tag{2.6}$$

In the language of (quantum) computing, a classical bit is simply the binary information 0 or 1, while a quantum bit (q-bit) is a state that can be expressed as a linear combination of the basis vectors e_0 and e_1, like ψ, i.e. an arbitrary element of $\mathcal{H} \simeq \mathbf{C}^2$. The basis $\{e_0, e_1\}$ then is called the computational basis or the z-basis. This basis is orthonormal, which means

$$(e_i, e_j) = \delta_{ij} \; , \tag{2.7}$$

where here the indices i, j are just 0 or 1 and δ_{ij} is the Kronecker-delta[5]. Of course, there are infinitely many different orthonormal basis' in $\mathbf{C}^2$. Two other orthonormal basis' are $\{e_+, e_-\}$ and $\{e_R, e_L\}$ with

$$ e_+ = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad , \quad e_- = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad , \quad (e_\alpha, e_\beta) = \delta_{\alpha\beta} \quad \text{where here } \alpha, \beta \in \{+, -\} \ , \quad (2.8) $$

$$ e_R = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix} \quad , \quad e_L = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix} \quad , \quad (e_A, e_B) = \delta_{AB} \quad \text{where here } A, B \in \{R, L\} \quad (2.9) $$

For reasons that will become clear later-on, $\{e_+, e_-\}$ is referred to as the x-basis and $\{e_L, e_R\}$ as the y-basis. The change between the z-basis (2.6) and the x-basis (2.8) or the y-basis (2.9) is given by

$$ e_+ = \frac{1}{\sqrt{2}}(e_0 + e_1) \ , \ e_- = \frac{1}{\sqrt{2}}(e_0 - e_1) \quad , \quad e_0 = \frac{1}{\sqrt{2}}(e_+ + e_-) \ , \ e_1 = \frac{1}{\sqrt{2}}(e_+ - e_-) \ , \ (2.10) $$

$$ e_R = \frac{1}{\sqrt{2}}(e_0 + ie_1) \ , \ e_L = \frac{1}{\sqrt{2}}(e_0 - ie_1) \quad , \quad e_0 = \frac{1}{\sqrt{2}}(e_R + e_L) \ , \ e_1 = \frac{i}{\sqrt{2}}(e_L - e_R) \ . \ (2.11) $$

Doing these changes of basis, the *same* vector ψ will be expressed with different components as

$$ \psi = a\, e_0 + b\, e_1 = \frac{a+b}{\sqrt{2}}\, e_+ + \frac{a-b}{\sqrt{2}}\, e_- = \frac{a-ib}{\sqrt{2}}\, e_R + \frac{a+ib}{\sqrt{2}}\, e_L \ . \tag{2.12} $$

2.4 Dirac's notation, bras and kets

A particularly convenient notation is Dirac's bra and ket notation, where one denotes these same basis vectors as

$$ e_0 \to |0\rangle \ , \ e_1 \to |1\rangle \quad , \quad e_+ \to |+\rangle \ , \ e_- \to |-\rangle \quad , \quad e_R \to |R\rangle \ , \ e_L \to |L\rangle \ . \tag{2.13} $$

The $|...\rangle$ are called kets. Then equation (2.12) for ψ translates into

$$ |\psi\rangle = a\,|0\rangle + b\,|1\rangle = \frac{a+b}{\sqrt{2}}\,|+\rangle + \frac{a-b}{\sqrt{2}}\,|-\rangle = \frac{a-ib}{\sqrt{2}}\,|R\rangle + \frac{a+ib}{\sqrt{2}}\,|L\rangle \ . \tag{2.14} $$

We denote the adjoint vectors as bras $\langle ...|$, so that $\psi^\dagger$ corresponds to

$$ \langle\psi| = a^*\,\langle 0| + b^*\,\langle 1| = \frac{a^*+b^*}{\sqrt{2}}\,\langle +| + \frac{a^*-b^*}{\sqrt{2}}\,\langle -| = \frac{a^*+ib^*}{\sqrt{2}}\,\langle R| + \frac{a^*-ib^*}{\sqrt{2}}\,\langle L| \ . \tag{2.15} $$

The inner product (ϕ, ψ) is just written as the bra-ket $\langle\phi\,|\psi\rangle \equiv \langle\phi|\,|\psi\rangle$. The relation (2.4) then translates into

$$ \langle\phi\,|\psi\rangle = \langle\psi\,|\phi\rangle^* \ . \tag{2.16} $$

The orthonormality relation (2.7) becomes $\langle i\,|j\rangle = \delta_{ij}$, and similarly for the other basis'. Very explicitly we have

$$ \langle 0\,|0\rangle = \langle 1\,|1\rangle = 1 \quad , \quad \langle 0\,|1\rangle = \langle 1\,|0\rangle = 0 \ , $$
$$ \langle +\,|+\rangle = \langle -\,|-\rangle = 1 \quad , \quad \langle +\,|-\rangle = \langle -\,|+\rangle = 0 \ , $$
$$ \langle R\,|R\rangle = \langle L\,|L\rangle = 1 \quad , \quad \langle R\,|L\rangle = \langle L\,|R\rangle = 0 \ . \tag{2.17} $$

---

[5]δ_{ij} equals 1 if the indices are equal and is 0 otherwise. Here $\delta_{00} = \delta_{11} = 1$ and $\delta_{01} = \delta_{10} = 0$.

The change of basis (2.10) now reads

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

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

$$(2.18)$$

One can read off the inner products between the kets of one basis and the bras of the other basis, e.g.

$$\langle + \,|0\rangle = \langle + \,|1\rangle = \langle - \,|0\rangle = \frac{1}{\sqrt{2}} \ , \quad \langle - \,|1\rangle = -\frac{1}{\sqrt{2}} \ , \tag{2.19}$$

and since all these inner products are real, by (2.16) we obviously also have

$$\langle 0 \,|+\rangle = \langle 1 \,|+\rangle = \langle 0 \,|-\rangle = \frac{1}{\sqrt{2}} \ , \quad \langle 1 \,|-\rangle = -\frac{1}{\sqrt{2}} \ . \tag{2.20}$$

It is useful to note that we may write the identity operator using any orthonormal basis as

$$\mathbf{1} = |0\rangle \langle 0| + |1\rangle \langle 1| = |+\rangle \langle +| + |-\rangle \langle -| = |R\rangle \langle R| + |L\rangle \langle L| \ . \tag{2.21}$$

This is called the closure relation. It is easy to check. We may apply this to $|\psi\rangle = a\,|0\rangle + b\,|1\rangle$:

$$|\psi\rangle = \mathbf{1}\,|\psi\rangle = \big(|0\rangle \langle 0| + |1\rangle \langle 1| \big)\big(a\,|0\rangle + b\,|1\rangle \big) = |0\rangle\,a + |1\rangle\,b = a\,|0\rangle + b\,|1\rangle \ , \tag{2.22}$$

since, of course it does not matter whether the complex numbers a and b are written to the left or the right of the bras or kets. We may also define the projectors on the basis states as $P_0 = |0\rangle \langle 0|$ and $P_1 = |1\rangle \langle 1|$, so that $\mathbf{1} = P_0 + P_1$, and similarly for P_+, P_- and P_R, P_L.

2.5 Generalisation to a finite-dimensional Hilbert space

The previous discussion of the two-dimensional Hilbert space immediately carries over to finite-dimensional Hilbert spaces. Any N-dimensional Hilbert $\mathcal{H}$ space is isomorphic to $\mathbf{C}^N$. It is equipped with the obvious generalisation of the inner product (2.3) satisfying (2.16). Any basis of $\mathcal{H}$ contains N linearly independent vectors $|e_j\rangle \equiv |j\rangle$ and we may suppose that they are chosen to be orthonormal :

$$\langle i \,|j\rangle = \delta_{ij} \ , \quad i,j = 1,\ldots N \ . \tag{2.23}$$

Again, there are (infinitely) many different possible basis', but let's suppose that we have chosen one of them. Then any $|\psi\rangle \in \mathcal{H}$ can be decomposed on this basis as

$$|\psi\rangle = \sum_{j=1}^{N} a_j\,|j\rangle \quad , \quad a_j = \langle j \,|\psi\rangle \ . \tag{2.24}$$

Inserting the expression for the a_j into the expansion of $|\psi\rangle$ one gets $|\psi\rangle = \sum_{j=1}^{N} |j\rangle \langle j \,|\psi\rangle$, valid for any $|\psi\rangle$ from which one deduces the general closure relation

$$\mathbf{1} = \sum_{j=1}^{N} |j\rangle \langle j| = \sum_{j=1}^{N} P_j \quad , \quad P_j = |j\rangle \langle j| \ , \tag{2.25}$$

where P_j projects on the basis state $|j\rangle$.

2.6 Linear operators on finite-dimensional Hilbert spaces

We will now discuss linear operators. Since the definitions and properties of these operators discussed in this subsection are equally valid for any finite-dimensional Hilbert space, we will formulate them in this slightly more general setting. (As we will see later-on, with appropriate qualifications most of them also remain valid on infinite-dimensional Hilbert spaces.) So let $\mathcal{H}$ denote the N-dimensional Hilbert space introduced in the previous subsection.

A linear operator A on $\mathcal{H}$ is a map from $\mathcal{H}$ to itself such that

$$A\big(\lambda\,|\psi\rangle + \mu\,|\chi\rangle\big) = \lambda A\,|\psi\rangle + \mu A\,|\chi\rangle\ , \quad |\psi\rangle,\,|\chi\rangle \in \mathcal{H}\ , \quad \lambda,\,\mu \in \mathbf{C}\ . \tag{2.26}$$

The simplest linear operators are the identity operator as given by the closure relation (2.25), or the projectors. Linearity is very important in quantum mechanics and it is at the origin of the superposition principle : it does not matter whether we first add two vectors and then act with the linear operator, or do it the other way round. Linearity also means that if we know how a (linear) operator acts on the basis vectors we know its action in general. Indeed,

$$A\,|\psi\rangle = \mathbf{1}\,A\,\mathbf{1}\,|\psi\rangle = \sum_{i=1}^{N} |i\rangle\,\langle i|\ A\ \sum_{j=1}^{N} |j\rangle\,\langle j\,|\psi\rangle = \sum_{i,j=1}^{N} |i\rangle\,\langle i|\,A\,|j\rangle\,\langle j\,|\psi\rangle\ , \tag{2.27}$$

and the action of A is entirely coded in the complex numbers

$$A_{ij} = \langle i|\,A\,|j\rangle\ , \tag{2.28}$$

called the matrix elements of the linear operator A with respect to the orthonormal basis we have chosen. Obviously one can then rewrite (2.27) as

$$A\ =\ |i\rangle\,A_{ij}\,\langle j| \tag{2.29}$$

$$|\phi\rangle = A\,|\psi\rangle \quad \Leftrightarrow \quad \langle i\,|\phi\rangle = \sum_{j=1}^{N} A_{ij}\,\langle j\,|\psi\rangle\ , \tag{2.30}$$

and one sees that one obtains the column vector of the components of $|\phi\rangle = A\,|\psi\rangle$ by the usual action of the matrix with elements A_{ij} on the column vector of the component of $|\psi\rangle$. One sometimes writes $A\,|\psi\rangle \equiv |A\psi\rangle$. The corresponding bra is simply denoted $\langle A\psi|$.

Just as A is a linear map from $\mathcal{H}$ to $\mathcal{H}$, one may consider, for fixed ψ, the inner product $\langle \psi\,|\phi\rangle$ as a linear map from $\mathcal{H}$ to $\mathbf{C}$ mapping each $|\phi\rangle$ to the complex number $\langle \psi\,|\phi\rangle$. The set of all such linear maps from $\mathcal{H}$ to $\mathbf{C}$ constitutes the set of linear functionals on $\mathcal{H}$ and is referred to as the dual of $\mathcal{H}$, denoted as $\mathcal{H}^{\dagger}$. One can easily show that all such functionals are of the form $|\phi\rangle \to \langle \psi\,|\phi\rangle$ for some ψ, and hence are uniquely determined[6] by a $\psi \in \mathcal{H}$. We may thus identify $\mathcal{H}^{\dagger} \simeq \mathcal{H}$.

---

[6]Let F be a linear functional. Being linear, its action on an arbitrary vector in $\mathcal{H}$ is completely determined by its action on the basis vectors $|j\rangle$. Define $f_j = F(|j\rangle)$. Let $|\psi\rangle = \sum_i f_i^*\,|i\rangle$ so that $\langle \psi| = \sum_i f_i\,\langle i|$ and then we have indeed $\langle \psi\,|j\rangle = f_j = F(|j\rangle)$.

Consider now $\langle \psi \,| A\phi \rangle$. Thus is again a linear map from $\mathcal{H}$ to $\mathbf{C}$, namely $|\phi\rangle \to \langle \psi \,| A\phi \rangle$, and thus of the form $\langle \chi \,| \phi \rangle$. Moreover, it is easy to see that χ depends linearly on ψ, so that there exists a linear operator, called $A^\dagger$ such that $|\chi\rangle = \big| A^\dagger \psi \big\rangle$ and then also $\langle \chi | = \big\langle A^\dagger \psi \big|$. In summary, the linear operator $A^\dagger$ is determined by

$$\langle A^\dagger \psi \,| \phi \rangle = \langle \psi \,| A\phi \rangle \ , \quad \forall \,|\psi\rangle \,, |\phi\rangle \in \mathcal{H} \ . \tag{2.31}$$

$A^\dagger$ is called the adjoint of A.

Exercise 2.1 : *Show that in the above argument, χ indeed depends linearly on ψ. Using the form (2.29) of the operator A show that the adjoint $A^\dagger$ is given by*

$$A^\dagger = \sum_{i,j=0,1} |j\rangle \, A^*_{ij} \, \langle i| \quad \Rightarrow \quad (A^\dagger)_{ij} = A^*_{ji} \ , \tag{2.32}$$

so that the matrix corresponding to $A^\dagger$ equals the hermitian conjugate (transposed and complex conjugate) of the matrix corresponding to A.

Exercise 2.2 : *Show that taking the adjoint twice gives back the original linear operator :*

$$(A^\dagger)^\dagger = A \quad \Leftrightarrow \quad \langle A\psi \,| \phi \rangle = \big\langle \psi \,\big| A^\dagger \phi \big\rangle \ . \tag{2.33}$$

A linear operator is called hermitian or self-adjoint if

$$A^\dagger = A \quad : \quad \text{hermitian or self-adjoint} \ . \tag{2.34}$$

Obviously then the corresponding matrix is a hermitian matrix.

A linear operator U is called unitary if its adjoint equals its inverse, i.e.

$$U^\dagger = U^{-1} \quad : \quad \text{unitary} \ . \tag{2.35}$$

Unitary operators preserve the scalar product in the sense that

$$\langle U\psi \,| U\phi \rangle = \big\langle \psi \,\big| U^\dagger U\phi \big\rangle = \big\langle \psi \,\big| U^{-1} U\phi \big\rangle = \langle \psi \,| \phi \rangle \ . \tag{2.36}$$

Since one defines the norm in terms of the scalar product,

$$|| \,|\psi\rangle \,||^2 = \langle \psi \,| \psi \rangle \ , \tag{2.37}$$

unitary operators also preserve the norm of a vector.

One defines the sum and differences of two linear operators A and B in the obvious way:

$$(A \pm B) \,|\psi\rangle = A \,|\psi\rangle \pm B \,|\psi\rangle \ . \tag{2.38}$$

One defines the "product" of two linear operators A and B as the composite operator, AB means that we act with B first and then with A on the resulting vector :

$$\big(A \, B\big) \,|\psi\rangle = A\big(B \,|\psi\rangle\big) \ . \tag{2.39}$$

Note that AB is in general different from BA. The kth power A^k of a linear operator A is defined by repeating its action k times. One can then define functions of operators by the corresponding power series. In particular one defines the exponential of an operator by the usual power series

$$\exp(A) \equiv e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k \ . \tag{2.40}$$

Of course, one needs to show that this infinite series converges in some appropriate norm defined on the space of linear operators. This can be done without any problem as long as one deals with linear operators on finite-dimensional Hilbert spaces.[7]

Exercise 2.3 : *Show that*

$$\left(A\,B \right)^\dagger = B^\dagger\,A^\dagger \quad , \quad \left(A^k \right)^\dagger = \left(A^\dagger \right)^k \quad \Rightarrow \quad \left(e^A \right)^\dagger = e^{A^\dagger} \ . \tag{2.41}$$

Use this to show that

$$A^\dagger = A \quad \Rightarrow \quad \left(e^{iA} \right)^\dagger = \left(e^{iA} \right)^{-1} \ , \tag{2.42}$$

i.e. if A is hermitian, e^{iA} is unitary.

If the action of a linear operator A on a given vector $|\phi\rangle$ gives a multiple a_ϕ of this same vector, the latter is called an <u>eigen-vector</u> and the multiple a_ϕ the <u>eigenvalue</u> :

$$A\,|\phi\rangle = a_\phi\,|\phi\rangle \ . \tag{2.43}$$

For hermitian linear operators A, a well-known result from linear algebra states that the different eigenvectors of A form a basis of the vector space. This means that on $\mathcal{H}$ there are N linearly independent eigenvectors $|\phi_n\rangle$:

$$A\,|\phi_n\rangle = a_n\,|\phi_n\rangle \ , \quad n = 1, \ldots N \ . \tag{2.44}$$

However, this does not necessarily mean that all a_n are also different. If the same a_n corresponds to $d_n > 1$ different (linearly independent) eigenvectors, we say that the eigenvalue a_n is d_n times degenerate. (If $d_n = 1$ it is non-degenerate.) To make this more clear, one usually writes instead of (2.44)

$$A\,\left|\phi_n^i\right\rangle = a_n\,\left|\phi_n^i\right\rangle \ , \quad i = 1, \ldots d_n, \quad \sum_n d_n = N \ , \tag{2.45}$$

with the understanding that $a_n \neq a_m$ if $n \neq m$. Note the obvious fact that any linear combination of eigenvectors corresponding to the same a_n is still an eigenvector of A with the same eigenvalue a_n. Hence it makes sense to talk about the eigenspace $\mathcal{E}_n$ corresponding to a given eigenvalue a_n. It is the linear subspace of $\mathcal{H}$ of all linear combinations of the eigenvectors corresponding to a_n One has the following

---

[7]With respect to some fixed orthonormal basis, A is determined by its N^2 matrix elements A_{ij}, so the space of linear operators is isomorphic to $\mathbf{C}^{N^2}$ which becomes a Hilbert space itself by introducing the inner product $(A, B) = \text{Tr}\, A^\dagger B = \sum_{ij} A_{ij}^* B_{ij}$. Note that the trace does not depend on the orthonormal basis we have chosen. The corresponding norm then is $||A||^2 = (A, A) = \sum_{ij} |A_{ij}|^2$. In this norm the series (2.40) converges for all A.

Lemma : For hermitian operators (i) the eigenvalues are real and (ii) eigenvectors corresponding to different eigenvalues must be orthogonal.

The proof is easy. Indeed, we have

$$0 = \langle A\phi_n \,|\, \phi_m \rangle - \langle \phi_n \,|\, A\phi_m \rangle = \langle a_n\phi_n \,|\, \phi_m \rangle - \langle \phi_n \,|\, a_m\phi_m \rangle = (a_n^* - a_m)\langle \phi_n \,|\, \phi_m \rangle \ . \tag{2.46}$$

If we take $n = m$ then $\langle \phi_n \,|\, \phi_n \rangle = ||\,|\phi_n\rangle\,||^2 > 0$ and $a_n^* - a_n$ must vanish, i.e. the a_n are real. Let now $n \neq m$ and moreover $a_n \neq a_m$. Then the previous equation implies $\langle \phi_n \,|\, \phi_m \rangle = 0$, i.e. the two eigenvectors are orthogonal.

Exercise 2.4 : *On our two-dimensional Hilbert space $\mathcal{H}$ let $A = |0\rangle \langle 1| + |1\rangle \langle 0|$. Show that $A^\dagger = A$ and determine its eigenvalues and eigenvectors and verify the orthogonality of the eigenvectors.*

The set of eigenvectors forms a basis and by the previous lemma the eigenvectors corresponding to some eigenvalue are orthogonal to all other eigenvectors corresponding to different eigenvalues. Then within any set of eigenvectors corresponding to some eigenvalue a_n (i.e. within any eigenspace $\mathcal{E}_n$) one may always take linear combinations of the eigenvectors so that they form an orthonormal set. Obviously this new set still consists of eigenvectors of A with eigenvalue a_n. By doing so within any set for all the a_n one obtains an orthonormal basis in $\mathcal{H}$ consisting of eigenvectors of A. Hence:

Corollary : Any hermitian operator admits an orthonormal basis of eigenvectors.

Exercise 2.5 : *Let $\{|j\rangle\}_{j=1,\ldots N}$ by an orthonormal basis of $\mathcal{H}$. Consider some subset $\mathcal{S} \subset \{1,\ldots N\}$ and let $P_\mathcal{S} = \sum_{j \in \mathcal{S}} |j\rangle \langle j|$. Show that $P_\mathcal{S}$ is a projector, i.e. satisfies $P_\mathcal{S}^2 = P_\mathcal{S}$. Determine its eigenvalues and eigenvectors. Conversely, show that if the only eigenvalues of A are 0 and 1, then necessarily $A^2 = A$ and A is a projector.*

In general, if $\mathcal{E}$ is some (linear) subspace of dimension k of $\mathcal{H}$ then any vector in $\mathcal{E}$ can be expressed as a linear combination of k basis vectors in $\mathcal{E}$. By taking appropriate linear combinations of these basis vectors one obtains an orthonormal basis $\{|a_j\rangle, j = 1, \ldots k\}$ in $\mathcal{E}$. It is then not difficult to verify that $P_\mathcal{E} = \sum_{j=1}^k |a_j\rangle \langle a_j|$ is the projector on this subspace and satisfies, as any projector, $P_\mathcal{E} P_\mathcal{E} = P_\mathcal{E}$.

For two linear operators A and B one defines the <u>commutator</u> as

$$[A, B] = AB - BA \ . \tag{2.47}$$

Lemma : If A and B are hermitian and $[A, B] = 0$ one can construct a basis of common eigenvectors of A and B. Conversely if such a basis exists, then necessarily $[A, B] = 0$.

The proof is relatively simple. Let $|\phi_n\rangle$ be an eigenvector of A with eigenvalue a_n. Then

$$A(B\,|\phi_n\rangle) = AB\,|\phi_n\rangle = BA\,|\phi_n\rangle = Ba_n\,|\phi_n\rangle = a_n(B\,|\phi_n\rangle) \ , \tag{2.48}$$

and one sees that $B\,|\phi_n\rangle$ (if non-vanishing) is also an eigenvector of A with the same eigenvalue a_n. First suppose that a_n is not degenerate. Then necessarily $B\,|\phi_n\rangle$ must be proportional to $|\phi_n\rangle$,

i.e. $B |\phi_n\rangle = b_n |\phi_n\rangle$ for some complex number b_n. But this means that b_n is an eigenvalue of B (and hence it is real) and $|\phi_n\rangle$ a corresponding eigenvector, also of B. In case $B |\phi_n\rangle = 0$, it is still true that $|\phi_n\rangle$ is an eigenvector also of B (with zero eigenvalue). Next, if the eigenvalue a_n is degenerate, the above equation can be rewritten as $A(B |\phi_n^i\rangle) = a_n(B |\phi_n^i\rangle)$. This means that $B |\phi_n^i\rangle$ is in the eigenspace $\mathcal{E}_n$ of A, i.e. $B |\phi_n^i\rangle = \sum_{j=1}^{d_n} b_{ji} |\phi_n^j\rangle$. Obviously $b_{ji} = \langle \phi_n^j| B |\phi_n^i\rangle$, and since B is hermitian, the b_{ji} form a hermitian $d_n \times d_n$ matrix. This shows that the restriction of B on the subspace $\mathcal{E}_n$ is hermitian and from our previous discussion we know that we can find an orthonormal basis, within this subspace, of eigenvectors $\left|\tilde{\phi}_n^m\right\rangle$ of B with eigenvalues b_m. But the $\left|\tilde{\phi}_n^m\right\rangle$ are all in the eigenspace $\mathcal{E}_n$ of A and hence also eigenvectors of A with eigenvalue a_n. This shows that the set of $\left|\tilde{\phi}_n^m\right\rangle$ is an orthonormalised basis of $\mathcal{H}$, consisting of simultaneous eigenvectors of A and B. Conversely, it is easy to see that if a common basis of eigenvectors exists, then necessarily $AB |\phi_n^m\rangle = a_n b_m |\phi_n^m\rangle = b_m a_n |\phi_n^m\rangle = BA |\phi_n^m\rangle$, so that $AB = BA$ on all states of the basis, i.e. $AB = BA$ and hence $[A, B] = 0$.

The following result is often useful :

Lemma : The eigenvalues a_n of a linear operator A are the same as the eigenvalues of the associated matrix with matrix elements $A_{ij} = \langle i| A |j\rangle$. If $|\phi_n\rangle$ is an eigenvector of A then the $\langle j |\phi_n\rangle$ are the components of the corresponding eigenvector of the matrix.
The proof is again simple. Let $A |\phi_n\rangle = a_n |\phi_n\rangle$ then $a_n\langle i |\phi_n\rangle = \langle i| a_n |\phi_n\rangle = \langle i| A |\phi_n\rangle = \sum_j \langle i| A |j\rangle \langle j |\phi_n\rangle = \sum_j A_{ij}\langle j |\phi_n\rangle$.

2.7   The postulates of quantum mechanics for N-state systems

Enumerating the postulates of quantum mechanics for a simple 2-state system or for an N-state system is exactly the same, so we will give them here for any N-state system for which the Hilbert space $\mathcal{H}$ has a finite dimension N.

- We have a complex N-dimensional Hilbert space $\mathcal{H}$, containing (state) vectors $\psi \in \mathcal{H}$, also denoted as $|\psi\rangle$.

- The Hilbert space is equipped with an inner product (also called hermitian product) that is anti-linear in its first element and linear in its second:

$$(\psi, \phi) \equiv \langle \psi |\phi\rangle = \langle \phi |\psi\rangle^* \equiv (\phi, \psi)^* . \tag{2.49}$$

- The norm of a vector is induced by the inner product as $||\psi|| = \sqrt{\langle \psi |\psi\rangle} \geq 0$.

- A physical state is represented by a <u>ray</u> $\mathcal{R}_\psi$ in $\mathcal{H}$, i.e. a normed state vector $|\psi\rangle$, up to a phase: $|\psi\rangle \simeq e^{i\alpha} |\psi\rangle$ and $\langle \psi |\psi\rangle = 1$.

- Physical observables $\mathcal{A}$ are represented by linear hermitian operators A on $\mathcal{H}$:

$$A^\dagger = A \quad \text{where} \quad \langle \psi \left| A^\dagger \phi \right\rangle \equiv \langle A\psi |\phi\rangle . \tag{2.50}$$

One usually calls such A an observable.

- Any measurement of the physical observable $\mathcal{A}$ can only give values a_n that are eigenvalues of A, i.e. $A\,|\phi_n^i\rangle = a_n\,|\phi_n^i\rangle$, where we may assume that the eigenvectors $|\phi_n^i\rangle$ form an orthonormal basis. The additional index $i = 1, \ldots d_n$ takes into account a possible d_n-fold degeneracy of the eigenvalue a_n. If the state of the system is given by $\mathcal{R}_\psi$ then the probability to measure a_n is

$$\mathcal{P}_n(\psi) \equiv \mathcal{P}(a_n, \psi) = \sum_{i=1}^{d_n} |\langle \phi_n^i \,|\psi\rangle\,|^2 \ , \tag{2.51}$$

  independently of the $|\psi\rangle$ chosen to represent $\mathcal{R}_\psi$.

- If a measurement of a physical observable $\mathcal{A}$ on a state initially in ray $\mathcal{R}_\psi$ gives the result a_n, then immediately *after* the measurement the state is an eigenstate of A corresponding to the measured eigenvalue a_n. This goes under the name of "reduction of the wave-packet". If one measures $\mathcal{A}$ immediately again, one obtains a_n with probability one.

The fact that a physical state is actually a ray has some important physical consequences as we will see later-on. For now we only mention that mathematically this means that $\psi \in \mathcal{H} \simeq \mathbf{C}^N$ and $\lambda\psi$, for $\lambda \in \mathbf{C}$ and $\lambda \neq 0$, get "identified". The space $\mathbf{C}^N$ with the equivalence relation $z^i \simeq \lambda z^i$ defines the projective space $\mathbf{CP}^{N-1}$, so that here $\mathcal{R}_\psi \in \mathbf{CP}^{N-1}$.

In general a projector P on some subspace satisfies $P^2 \equiv PP = P$. Then (provided the $|\phi_n^i\rangle$ are orthonormalised as we supposed)

$$P(a_n) = \sum_{i=1}^{d_n} |\phi_n^i\rangle \langle \phi_n^i| \tag{2.52}$$

is the projector on the eigenspace of a_n, i.e. on the space of eigenvectors corresponding to the eigenvalue a_n. Using this projector, the probability $\mathcal{P}(a_n, \psi)$ to measure a_n as given in (6.71) can be re-written as

$$\mathcal{P}(a_n, \psi) = \langle \psi|\,P(a_n)\,|\psi\rangle \ . \tag{2.53}$$

We can then also reformulate the reduction of the wave packet more precisely : immediately after the measurement giving a_n the state is $P(a_n)\,|\psi\rangle \,/\, ||P(a_n)\,|\psi\rangle\,||$.

The previous set of postulates are rather general. They will carry over to the case of infinite-dimensional Hilbert spaces as appropriate for the harmonic oscillator or the hydrogen atom or even a system of many particles, except for a few mathematical subtleties that need to be made more precise.

Note that we have said noting about the evolution in time of the quantum states between any two measurements. We have just said how to describe a quantum system, at a given instant t_0 in time, including how to measure its properties. It is also clear from the "reduction of the wave packet" that if we successively measure a set of commuting observables, this amounts

to "preparing" a quantum state that is simultaneous eigenstate of all these observables for the corresponding eigenvalues. The analogue in classical mechanics would be e.g. the description of a system in terms of the position and velocities of all its particles, also called a point in the corresponding phase space. The dynamics, i.e. the question how this system evolves in time is a different story. This time evolution can be relativistic or non-relativistic. The point is that our postulates (when adapted to infinite-dimensional Hilbert spaces) are general enough to be valid, independently of the dynamics. They apply in the same way to non-relativistic quantum mechanics, as well as to relativistic quantum mechanics (quantum field theory).[8] Here, however, we will only be concerned with non-relativistic quantum mechanics.

Quite generally, the time evolution can be carried either by the state vectors or by the operators. In the first case, this is called the Schrödinger picture, and in the second case, this is the Heisenberg picture. Both pictures are equivalent. Here we will use the Schrödinger picture. This means that the state vectors depend on time and we write $|\psi(t)\rangle$. Then we have the additional postulate

- The time evolution of the state vector $|\psi(t)\rangle$ is governed by the Schrödinger equation

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t} |\psi(t)\rangle = H(t) |\psi(t)\rangle \ , \tag{2.54}$$

where $H(t)$ is called the Hamiltonian and is the observable corresponding to the (total) energy of the system.

This is very different from the equations governing the dynamics of classical mechanics or electrodynamics. The equations of motion of classical mechanics are in general non-linear differential equations, like $\ddot{\vec{x}}(t) = \vec{F}(\vec{x}(t))/m$. Using phase space variables $\vec{x}(t)$ and $\vec{p}(t)$ this can be recast into first-order equations, but they still remain non-linear in general. On the other hand, Maxwell's equations (in the vacuum) are linear in the electric field $\vec{E}$ and the magnetic field $\vec{B}$ and can be written as first-order equations, however with constraints on these fields. These constraints can be solved by introducing the scalar and vector potentials A^0 and $\vec{A}$, but the latter obey second-order equations. Moreover, these differential equations are *real*, i.e. all coefficients are real, while the Schrödinger equation (2.54) is intrinsically complex due to the presence of i on the left-hand side.

Suppose that we had been looking for some general linear, first-order (in time) differential equation for $|\psi(t)\rangle$. Then (2.54) is indeed the most general equation one can write, with H being some linear operator. We will see that conservation of the norm of $|\psi(t)\rangle$ implies the presence of the i and the fact that the linear operator H must be hermitian. The non-trivial point is the interpretation of this linear operator. Let us argue why this observable H should correspond to the energy. Suppose this hermitian linear operator H does not depend on time (as should be the case for an isolated system) and let $|\varphi_n\rangle$ be its eigenvectors : $H |\varphi_n\rangle = E_n |\varphi_n\rangle$. Then by our postulates, measuring the observable H can result in any of the E_n (which we suppose non-degenerate to simplify the discussion). Suppose at $t = t_0$ we have measured one such E_n. Then immediately

---

[8]Quantum field theory requires to abandon the notion of a Hilbert space of a fixed number of particles, resulting in an even bigger Hilbert space which can describe arbitrary numbers of particles (and anti-particles).

after the measurement, the state is given by $|\psi(t_0)\rangle = |\varphi_n\rangle$. Hence we must solve the first-order differential equation (2.54) with this initial condition. The result is $|\psi(t)\rangle = e^{-iE_n(t-t_0)/\hbar}|\varphi_n\rangle$. This state displays an oscillation with frequency $\omega = \frac{E_n}{\hbar}$ and by Planck's relation (1.1) we then indeed identify E_n with an energy.

However, one needs to go a bit further. Indeed, as already emphasised, the overall phase $e^{-iE_n(t-t_0)/\hbar}$ is irrelevant and unobservable. If instead we suppose that the initial state is given by a superposition of two eigenstates with different eigenvalues, i.e. $|\psi(t_0)\rangle = \frac{1}{\sqrt{2}}(|\varphi_n\rangle + |\varphi_m\rangle)$, then one finds

$$\begin{aligned} |\psi(t)\rangle &= \frac{1}{\sqrt{2}}\big(e^{-iE_n(t-t_0)/\hbar}|\varphi_n\rangle + e^{-iE_m(t-t_0)/\hbar}|\varphi_m\rangle\big) \\ &= e^{-iE_n(t-t_0)/\hbar}\frac{1}{\sqrt{2}}\big(|\varphi_n\rangle + e^{-i(E_m-E_n)(t-t_0)/\hbar}|\varphi_m\rangle\big) , \end{aligned} \qquad (2.55)$$

and then the relative frequency $\omega_{mn} = \frac{E_m-E_n}{\hbar}$ is observable and corresponds to an energy difference according to Planck's relation.

The question how to construct the appropriate Hamiltonian operator for a given system, or more generally how to construct the linear operators associated with the physical observables often is referred to as "quantisation rules", and we will discuss this later-on. However, at a fundamental level, Nature is quantum mechanical and the question is not so much how to quantise the classical physics, but how and why the classical description of physics emerges from the underlying quantum mechanics. There are many complicated and subtle aspects to this question, but a partial answer will be provided by Ehrenfest's theorem we will prove below.

2.8 Some direct implications of the postulates

<u>Conservation of the norm</u>: Taking the hermitian conjugate of the Schrödinger equation gives

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}\langle\psi(t)| = -\langle\psi(t)|H(t) . \qquad (2.56)$$

where we used that $H(t)$ is hermitian. (The minus sign of course arises from $(i)^\dagger = i^* = -i$.) It follows that

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}\langle\psi(t)\,|\,\psi(t)\rangle = \big(i\hbar\frac{\mathrm{d}}{\mathrm{d}t}\langle\psi(t)|\big)\,|\psi(t)\rangle + \langle\psi(t)|\,i\hbar\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = \langle\psi(t)|\big(-H(t)+H(t)\big)|\psi(t)\rangle = 0 . \qquad (2.57)$$

Thus the hermiticity of the Hamiltonian in the Schrödinger equation ensures the conservation (in time) of the norm of any state.

<u>Mean values</u>: Using the projectors $P(a_n)$ on the eigenspaces of an observable A, one can re-write A as

$$A = \sum_n a_n P(a_n) . \qquad (2.58)$$

Indeed, the eigenstates form a basis so it is enough to show that this identity holds an any eigenvector : one has $\sum_n a_n P(a_n)|\phi_m\rangle = \sum_n a_n\delta_{nm}|\phi_m\rangle = a_m|\phi_m\rangle = A|\phi_m\rangle$. Next, the mean

value of an observable A in a given state is

$$\langle A \rangle_\psi = \sum_n \mathcal{P}(a_n, \psi) a_n = \sum_n \langle \psi | P(a_n) | \psi \rangle \, a_n = \langle \psi | \sum_n a_n P(a_n) | \psi \rangle = \langle \psi | A | \psi \rangle \ . \qquad (2.59)$$

<u>Variance</u>: In total analogy one obtains the mean value of A^2 as $\langle A^2 \rangle_\psi = \langle \psi | A^2 | \psi \rangle$. Finally, the mean square deviation is given by

$$(\Delta A)_\psi^2 \equiv \langle (A - \langle A \rangle_\psi)^2 \rangle_\psi = \langle A^2 \rangle_\psi - 2 \langle A \rangle_\psi \langle A \rangle_\psi + \langle A \rangle_\psi^2 \langle \mathbf{1} \rangle_\psi = \langle A^2 \rangle_\psi - \langle A \rangle_\psi^2 \ . \qquad (2.60)$$

Incidentally, since the mean value of the square $(A - \langle A \rangle_\psi)^2$ must be positive, this shows that

$$\langle A^2 \rangle_\psi \geq \langle A \rangle_\psi^2 \ . \qquad (2.61)$$

Lemma : The mean square deviation vanishes if and only if ψ is an eigenstate of A.
Indeed, if $A | \psi \rangle = a_n | \psi \rangle$ then also $A^2 | \psi \rangle = a_n A | \psi \rangle = a_n^2 | \psi \rangle$ and $\langle \psi | A^2 | \psi \rangle = a_n^2 = (\langle \psi | A | \psi \rangle)^2$, so that $(\Delta A)_\psi^2 = 0$. On the other hand, if the mean square deviation vanishes one has

$$0 = \langle \psi | (A - \langle A \rangle_\psi)^2 | \psi \rangle = \langle (A - \langle A \rangle_\psi) \psi | (A - \langle A \rangle_\psi) \psi \rangle = || (A - \langle A \rangle_\psi) \psi ||^2 \ . \qquad (2.62)$$

But this says that the norm squared of $(A - \langle A \rangle_\psi) | \psi \rangle$ vanishes, so $(A - \langle A \rangle_\psi) | \psi \rangle = 0$, i.e. $A | \psi \rangle = \langle A \rangle_\psi | \psi \rangle$ and we see that $| \psi \rangle$ is eigenvector of A with eigenvalue $\langle A \rangle_\psi$.

Consider now two observables A and B. Suppose first that they commute. We have seen that in this case there exists a basis of simultaneous eigenvectors of A and B. Then if $| \psi \rangle$ is any of these common eigenvectors $\left| \tilde{\phi}_n^m \right\rangle$ we have $(\Delta A)_\psi^2 = (\Delta B)_\psi^2 = 0$. This means that if the system is prepared in this eigenstate $\left| \tilde{\phi}_n^m \right\rangle$ we can know the values of A and B without any uncertainty. Indeed, for the eigenstates there is no "reduction of the wave packet" and measuring A gives the eigenvalue a_n and the state remains $\left| \tilde{\phi}_n^m \right\rangle$, so that measuring B just gives b_m and the state still remains $\left| \tilde{\phi}_n^m \right\rangle$. On the other hand, if A and B do not commute there is no basis of common eigenvectors. Of course, this does not exclude the existence of some common eigenvectors. Let us consider the general case where $[A, B] = i\,C$. The factor i is present, so that if A and B are hermitian, C is also hermitian. Indeed, then $[A, B]^\dagger = (AB - BA)^\dagger = BA - AB = -[A, B]$, consistent with $(iC)^\dagger = -iC$. Consider now any state $| \psi \rangle$ and compute the norm squared of $(A + i\lambda B) | \psi \rangle$ for some real λ :

$$\begin{aligned} 0 \ &\leq \ || (A + i\lambda B) | \psi \rangle ||^2 = \langle \psi | (A - i\lambda B)(A + i\lambda B) | \psi \rangle = \langle \psi | (A^2 + i\lambda [A, B] + \lambda^2 B^2) | \psi \rangle \\ &= \ \langle A^2 \rangle_\psi - \lambda \langle C \rangle_\psi + \lambda^2 \langle B^2 \rangle_\psi \equiv Q(\lambda) \ . \end{aligned} \qquad (2.63)$$

This quadratic polynomial $Q(\lambda)$ tends to $+\infty$ as $\lambda \to \pm\infty$ and will be non-negative everywhere (as required) if it has no zeros or a single zero. This will be the case if the discriminant is negative or zero :

$$\langle A^2 \rangle_\psi \, \langle B^2 \rangle_\psi \geq \frac{1}{4} \langle C \rangle_\psi^2 = \frac{1}{4} \langle i[A, B] \rangle_\psi^2 \qquad (2.64)$$

We may now consider the observables $\widetilde{A} = A - \langle A \rangle_\psi \mathbf{1}$ and $\widetilde{B} = B - \langle B \rangle_\psi \mathbf{1}$. Then $[\widetilde{A}, \widetilde{B}] = [A, B] = iC$ and repeating the previous argument with $\widetilde{A}$ and $\widetilde{B}$ one gets $\langle \widetilde{A}^2 \rangle_\psi \langle \widetilde{B}^2 \rangle_\psi \geq \frac{1}{4} \langle i[A, B] \rangle_\psi^2$. But $\langle \widetilde{A}^2 \rangle_\psi = (\Delta A)_\psi^2$ and $\langle \widetilde{B}^2 \rangle_\psi = (\Delta B)_\psi^2$, so that we finally arrive at $(\Delta A)_\psi^2 (\Delta B)_\psi^2 \geq \frac{1}{4} \langle i[A, B] \rangle_\psi^2$. Taking the square root one gets a general version of Heisenberg's uncertainty relation :

Uncertainty relation : One has the following relation for the variance of two observables A and B in any state $|\psi\rangle$:

$$(\Delta A)_\psi (\Delta B)_\psi \geq \frac{1}{2} | \langle [A, B] \rangle_\psi | . \tag{2.65}$$

2.9 General solution of the Schrödinger equation

The Schrödinger equation is a first-order differential equation, and it determines uniquely the state $|\psi(t)\rangle$ if we know the initial state $|\psi(t_0)\rangle$. One has to distinguish the situations where H depends or does not depend on time.

• For a time-*independent* Hamiltonian the solution is most easily written in the basis of eigenvectors of this Hamiltonian. Suppose we have solved the corresponding problem of finding the eigenvalues (energies) E_n and corresponding eigenvectors $|\varphi_n\rangle$ of H. Since these eigenvectors form a basis we may decompose $|\psi(t)\rangle$ on these $|\varphi_n\rangle$:

$$H |\varphi_n\rangle = E_n |\varphi_n\rangle \quad , \quad |\psi(t)\rangle = \sum_n c_n(t) |\varphi_n\rangle . \tag{2.66}$$

Inserting this into the Schrödinger equation and using the independence of the basis vectors yields $i\hbar \dot{c}_n(t) = E_n c_n(t)$ with solution $c_n(t) = c_n(t_0) e^{-iE_n(t-t_0)/\hbar}$ where the initial data $c_n(t_0)$ are determined in terms of the decomposition of $|\psi(t_0)\rangle$. Thus

$$|\psi(t)\rangle = \sum_n c_n(t_0) e^{-iE_n(t-t_0)/\hbar} |\varphi_n\rangle . \tag{2.67}$$

Alternatively, consider the linear operator

$$U(t, t_0) = e^{-iH(t-t_0)/\hbar} = \sum_{p=0}^{\infty} \frac{(-i)^p (t - t_0)^p}{p!} H^p . \tag{2.68}$$

As one readily verifies from the explicit sum, $U(t, t_0)$ satisfies the Schrödinger equation and appropriate initial condition

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t} U(t, t_0) = H U(t, t_0) \quad , \quad U(t_0, t_0) = \mathbf{1} . \tag{2.69}$$

It follows that

$$|\psi(t)\rangle = U(t, t_0) |\psi(t_0)\rangle , \tag{2.70}$$

satisfies the Schrödinger equation and initial condition and, hence, is the correct unique solution and must coincide with the solution (2.67). This also follows from the following remark : We may write $H = \sum_n |\varphi_n\rangle E_n \langle \varphi_n|$ and one can then straightforwardly sees that

$$U(t, t_0) = \sum_n |\varphi_n\rangle e^{-iE_n(t-t_0)/\hbar} \langle \varphi_n| . \tag{2.71}$$

Applying this to $|\psi(t_0)\rangle$ shows again that (2.70) equals (2.67).

• For a <u>time-*dependent* Hamiltonian $H(t)$</u> all this does not work. Indeed, if H depends on t then one can still compute the eigenvalues and eigenvectors at each time t, but obviously they would depend on t and then the previous decomposition of $|\psi(t)\rangle$ into a time-dependent coefficient $c_n(t)$ and a time-independent eigenvector $|\varphi_n\rangle$ does not work. Also, one might try the obvious generalisation of $U(t,t_0)$, namely $\exp\left(-\frac{i}{\hbar}\int_{t_0}^t \mathrm{d}t'\, H(t')\right)$. However this fails to satisfy the Schrödinger equation with $H \to H(t)$. Indeed, $H(t)$ and $\int_{t_0}^t \mathrm{d}t'\, H(t')$ do not commute in general, and then $\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{t_0}^t \mathrm{d}t'\, H(t')\right)^p \neq p\, H(t)\left(\int_{t_0}^t \mathrm{d}t'\, H(t')\right)^{p-1}$, so that $i\hbar\frac{\mathrm{d}}{\mathrm{d}t}\exp\left(-\frac{i}{\hbar}\int_{t_0}^t \mathrm{d}t'\, H(t')\right) \neq H(t)\exp\left(-\frac{i}{\hbar}\int_{t_0}^t \mathrm{d}t'\, H(t')\right)$. This problem due to the non-commutativity can be resolved by ordering the various Hamiltonians in the multiple integrals according to the integration times. One defines a time ordering T of a sequence $H(t_1)H(t_2)\ldots H(t_p)$ such that the operators with the larger times always are to the left of those with the smaller times. One can then see that

$$T\left(\left(\int_{t_0}^t \mathrm{d}t'\, H(t')\right)^p\right) = p!\,\int_{t_0}^t \mathrm{d}t_1\, H(t_1)\int_{t_0}^{t_1} \mathrm{d}t_2\, H(t_2)\ldots\int_{t_0}^{t_{p-1}} \mathrm{d}t_p\, H(t_p)\ , \qquad (2.72)$$

so that

$$\frac{\mathrm{d}}{\mathrm{d}t}\,T\left(\left(\int_{t_0}^t \mathrm{d}t'\, H(t')\right)^p\right) = H(t)\,T\left(\left(\int_{t_0}^t \mathrm{d}t'\, H(t')\right)^{p-1}\right)\ , \qquad (2.73)$$

and then also

$$\frac{\mathrm{d}}{\mathrm{d}t}\,T\exp\left(-\frac{i}{\hbar}\int_{t_0}^t \mathrm{d}t'\, H(t')\right) = H(t)\,T\exp\left(-\frac{i}{\hbar}\int_{t_0}^t \mathrm{d}t'\, H(t')\right)\ . \qquad (2.74)$$

This shows that

$$U(t,t_0) = T\exp\left(-\frac{i}{\hbar}\int_{t_0}^t \mathrm{d}t'\, H(t')\right) \qquad (2.75)$$

now does solve the Schrödinger equation and obeys the initial condition. Although this is an explicit solution, it is much less convenient to work with than e.g. (2.67) or (2.71).

2.10   Density operator

Consider some (abitrary) orthonormal basis which we denote by $\{|\phi_n\rangle\}$. This can be some eigenbasis of some observable B or not. In any case, being a basis one has the closure relation $\sum_n |\phi_n\rangle\langle\phi_n| = \mathbf{1}$. The trace of a linear operator A is defined with respect to such a basis as

$$\mathrm{tr}\, A = \sum_n \langle\phi_n|\, A\,|\phi_n\rangle\ . \qquad (2.76)$$

Let $\{|\varphi_k\rangle\}$ be some other orthonormal basis so that in particular $\sum_k |\varphi_k\rangle\langle\varphi_k| = \mathbf{1}$. Then

$$\mathrm{tr}\, A = \sum_n \sum_k \langle\phi_n|\, A\,|\varphi_k\rangle\langle\varphi_k|\phi_n\rangle = \sum_n \sum_k \langle\varphi_k|\phi_n\rangle\langle\phi_n|\, A\,|\varphi_k\rangle = \sum_k \langle\varphi_k|\, A\,|\varphi_k\rangle\ , \quad (2.77)$$

which shows that the definition of the trace does not depend on the orthonormal basis chosen. In terms of the corresponding matrix elements, the trace is the sum of the diagonal elements.

Exercise 2.6 : *Show similarly that the trace of a product of linear operators is invariant under cyclic permutations of the operators* :

$$\text{tr}\,(A_1 A_2 \ldots A_N) = \text{tr}\,(A_2 \ldots A_N A_1)\,. \tag{2.78}$$

Suppose now that our quantum system is described by the (normalised) state vector $|\psi\rangle$. We define the corresponding *density operator* as

$$\rho = |\psi\rangle\langle\psi|\,. \tag{2.79}$$

Of course, the state $|\psi\rangle$ evolves in time according to the Schrödinger equation $i\hbar\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = H(t)|\psi(t)\rangle$ and, correspondingly, $\rho = \rho(t)$ also depends on time. Since one also has the hermitian conjugate equation (2.56), it follows that

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}\rho(t) = H(t)\rho(t) - \rho(t)H(t) = [H(t), \rho(t)]\,. \tag{2.80}$$

The information contained in ρ is the same as in $|\psi\rangle$. First note that if we multiply $|\psi\rangle$ by a phase factor $e^{i\alpha}$ this will not change ρ. But $|\psi\rangle$ and $e^{i\alpha}|\psi\rangle$ are equivalent and contain the same information. If A is some physical observable one has

$$\text{tr}\,\rho A = \sum_n \langle\phi_n|\rho A|\phi_n\rangle = \sum_n \langle\phi_n|\psi\rangle\langle\psi|A|\phi_n\rangle = \sum_n \langle\psi|A|\phi_n\rangle\langle\phi_n|\psi\rangle = \langle\psi|A|\psi\rangle \equiv \langle A\rangle_\psi\,. \tag{2.81}$$

We see that $\text{tr}\,\rho A$ computes the mean value of A in the state $|\psi\rangle$. Furthermore, by (2.58) one has $A = \sum_k a_k P(a_k)$ with a_k the eigenvalues and $P(a_k)$ the projectors on the corresponding eigenspaces, and then

$$\text{tr}\,\rho\,P(a_k) = \langle\psi|P(a_k)|\psi\rangle = \mathcal{P}(a_k, \psi) \tag{2.82}$$

computes the probability to find a_k if we measure A in the state $|\psi\rangle$. Finally, if one finds a_k the state after the measurement is $P(a_k)|\psi\rangle/||P(a_k)|\psi\rangle||$. Then the corresponding new density operator is

$$\rho_{\text{after}\,a_k} = \frac{1}{||P(a_k)|\psi\rangle||^2}P(a_k)|\psi\rangle\langle\psi|P(a_k) = \frac{P(a_k)\rho P(a_k)}{\mathcal{P}(a_k)}\,, \tag{2.83}$$

where, of course, $\mathcal{P}(a_k) = \langle\psi|P(a_k)|\psi\rangle = \text{tr}\,\rho P(a_k)$. We can the rewrite the previous formula as

$$\rho_{\text{after}\,a_k} = \frac{P(a_k)\rho P(a_k)}{\text{tr}\,\rho\,P(a_k)}\,. \tag{2.84}$$

So far, describing the quantum state by the density matrix, rather than by the state vector, does not seem to present any advantages. However, the same density matrix formalism also allows to describe the more general situation where one deals not just with *one* system, like an atom or a single spin-$\frac{1}{2}$, but with a *statistical mixture* of such systems, like 10^{23} atoms in thermal equilibrium at some temperature T or a statistical mixture of 1000 spin-$\frac{1}{2}$ in the state $|+\rangle$ and 1000 spin-$\frac{1}{2}$ in the state $|-\rangle$. Such a statistical mixture *cannot* be described just in terms of state vectors,

because $\frac{1}{2}|\psi_1\rangle + \frac{1}{2}|\psi_2\rangle$ does *not* correspond to a mixture of half of the particles being in $|\psi_1\rangle$ and the other half being in $|\psi_2\rangle$, but to a *superposition* of these two states. The deep reason is that the probabilities correspond roughly speaking to the (modulus) squares so that if we want to describe such a mixture it should more look like $\frac{1}{2}\big(|\psi_1\rangle\big)^2 + \frac{1}{2}\big(|\psi_2\rangle\big)^2$. Of course, $\big(|\psi_1\rangle\big)^2$ does not make sense, but $|\psi_1\rangle\langle\psi_1|$ does. So we see that such a statistical mixture should be represented by $\frac{1}{2}|\psi_1\rangle\langle\psi_1| + \frac{1}{2}|\psi_2\rangle\langle\psi_2|$. But this is just $\frac{1}{2}\rho_1 + \frac{1}{2}\rho_2$. More generally, if we have a statistical mixture of K different "pure" states, each described by some state vector $|\psi_i\rangle$, and occurring with a statistical probability p_i, then this is described by the density operator

$$\rho = \sum_{i=1}^{K} p_i |\psi_i\rangle\langle\psi_i| = \sum_{i=1}^{K} p_i \rho_i \ . \tag{2.85}$$

This is a "good" definition, since when computing again $\mathrm{tr}\,\rho\,A$ this gives the mean value of A, but now with the mean value being the statistical mean value of the quantum mean values. Indeed, we have

$$\mathrm{tr}\,\rho\,A = \sum_{i=1}^{K} p_i \,\mathrm{tr}\,\rho_i A = \sum_{i=1}^{K} p_i \langle A\rangle_{\psi_i} = \langle\langle A\rangle\rangle \ , \tag{2.86}$$

where we used the notation $\langle\langle A\rangle\rangle$ to design the double average, first the quantum mechanical average in the state $|\psi_i\rangle$, and then the statistical average. Similarly also, we have

$$\mathrm{tr}\,\rho P(a_k) = \sum_{i=1}^{K} p_i \,\mathrm{tr}\,\rho_i P(a_k) = \sum_{i=1}^{K} p_i \mathcal{P}(a_k, \psi_i) = \widetilde{\mathcal{P}}(a_k) \ . \tag{2.87}$$

where we denote by $\widetilde{\mathcal{P}}(a_k)$ the probability to measure a_k when this probability also takes into account the statistical probabilities. Finally, the time evolution is still given by (2.80) provided the statistical probabilities p_i do not themselves depend on time.

What are the general properties of this density operator ? By taking $A = \mathbf{1}$ in (2.86) one sees that $\mathrm{tr}\,\rho = 1$. It also follows from the definition (2.85) that ρ is hermitian :

$$\rho^\dagger = \rho \quad , \quad \mathrm{tr}\,\rho = 1 \ . \tag{2.88}$$

Exercise 2.7 : *Conversely, show that any hermitian operator (on a finite-dimensional Hilbert space) whose trace is 1 corresponds a density operator ρ which can be written in the form* (2.85). *Moreover, show that if ρ has a single eigenvalue 1 and all other eigenvalues are 0, then it corresponds to a pure state $\rho = |\psi\rangle\langle\psi|$. If ρ has more then one non-vanishing eigenvalue it describes a statistical mixture of pure states.*