# Quantum Mechanics II

**$2^{nd}$ term 2000**

**Martin Plenio**
**Imperial College**

**Version May 11, 2000**

**Office hours: Tuesdays**
**11am-12noon and 5pm-6pm!**
**Office: Huxley 6M08**
**Available at:**
http://www.lsr.ph.ic.ac.uk/TQO/People/Plenio/lecture.html

# Contents


**I Quantum Mechanics** **5**

**1 Mathematical Foundations** **11**
1.1 The quantum mechanical state space . . . . . 11
1.2 The quantum mechanical state space . . . . . 12
1.2.1 Polarized Light . . . . . 12
1.2.2 Complex vector spaces . . . . . 18
1.2.3 Basis and Dimension . . . . . 21
1.2.4 Scalar products and Norms on Vector Spaces . . . 24
1.2.5 Completeness and Hilbert spaces . . . . . 33
1.2.6 Dirac notation . . . . . 37
1.3 Linear Operators . . . . . 38
1.3.1 Definition in Dirac notation . . . . . 38
1.3.2 Adjoint and Hermitean Operators . . . . . 41
1.3.3 Eigenvectors, Eigenvalues and the Spectral Theorem . . . . . 43
1.3.4 Functions of Operators . . . . . 49
1.4 Operators with continuous spectrum . . . . . 55
1.4.1 The position operator . . . . . 55
1.4.2 The momentum operator . . . . . 59
1.4.3 The position representation of the momentum operator and the commutator between position and momentum . . . . . 60

**2 Quantum Measurements** **63**
2.1 The projection postulate . . . . . 63
2.2 Expectation value and variance. . . . . 68
2.3 Uncertainty Relations . . . . . 69

2.3.1 The trace of an operator . . . . . 72
2.4 The density operator . . . . . 74
2.5 Mixed states, Entanglement and the speed of light . . . . 80
2.5.1 Quantum mechanics for many particles . . . . . 80
2.5.2 How to describe a subsystem of some large system? 84
2.5.3 The speed of light and mixed states. . . . . 88

**3 Dynamics and Symmetries 91**
3.1 The Schrödinger Equation . . . . . 91
3.1.1 The Heisenberg picture . . . . . 94
3.2 Symmetries and Conservation Laws . . . . . 96
3.2.1 The concept of symmetry . . . . . 96
3.2.2 Translation Symmetry and momentum conservation . . . . . 101
3.2.3 Rotation Symmetry and angular momentum conservation . . . . . 103
3.3 General properties of angular momenta . . . . . 106
3.3.1 Rotations . . . . . 106
3.3.2 Group representations and angular momentum commutation relations . . . . . 109
3.3.3 Angular momentum eigenstates . . . . . 111
3.4 Addition of Angular Momenta . . . . . 114
3.4.1 Two angular momenta . . . . . 117

**4 Approximation Methods 123**
4.1 Time-independent Perturbation Theory . . . . . 124
4.1.1 Non-degenerate perturbation theory . . . . . 124
4.1.2 Degenerate perturbation theory . . . . . 127
4.1.3 The van der Waals force . . . . . 130
4.1.4 The Helium atom . . . . . 134
4.2 Variational Principle . . . . . 135
4.2.1 The Rayleigh-Ritz Method . . . . . 135
4.3 Time-dependent Perturbation Theory . . . . . 139
4.3.1 Interaction picture . . . . . 141
4.3.2 Dyson Series . . . . . 143
4.3.3 Transition probabilities . . . . . 144

**II Quantum Information Processing 151**

**5 Quantum Information Theory 153**

5.1 Entanglement and Bell inequalities . . . . . 157

5.2 Quantum State Teleportation . . . . . 163

5.3 A basic description of teleportation . . . . . 164

# Part I

# Quantum Mechanics

# Introduction

This lecture will introduce quantum mechanics from a more abstract point of view than the first quantum mechanics course that you took your second year.

What I would like to achieve with this course is for you to gain a deeper understanding of the structure of quantum mechanics and of some of its key points. As the structure is inevitably mathematical, I will need to talk about mathematics. I will not do this just for the sake of mathematics, but always with a the aim to understand physics. At the end of the course I would like you not only to be able to understand the basic structure of quantum mechanics, but also to be able to solve (calculate) quantum mechanical problems. In fact, I believe that the ability to calculate (finding the quantitative solution to a problem, or the correct proof of a theorem) is absolutely essential for reaching a real understanding of physics (although physical intuition is equally important). I would like to go so far as to state

**If you can't write it down, then you do not understand it!**

With 'writing it down' I mean expressing your statement mathematically or being able to calculate the solution of a scheme that you proposed. This does not sound like a very profound truth but you would be surprised to see how many people actually believe that it is completely sufficient just to be able to talk about physics and that calculations are a menial task that only mediocre physicists undertake. Well, I can assure you that even the greatest physicists don't just sit down and await inspiration. Ideas only come after many wrong tries and whether a try is right or wrong can only be found out by checking it, i.e. by doing some sorts of calculation or a proof. The ability to do calculations is not something that one has or hasn't, but (except for some exceptional cases) has to be acquired by practice. This is one of the reasons why these lectures will be accompanied by problem sheets (Rapid Feedback System) and I really recommend to you that you try to solve them. It is quite clear that solving the problem sheets is one of the best ways to prepare for the exam. Sometimes I will add some extra problems to the problem sheets which are more tricky than usual. They usually intend

to illuminate an advanced or tricky point for which I had no time in the lectures.

The first part of these lectures will not be too unusual. The first chapter will be devoted to the mathematical description of the quantum mechanical state space, the Hilbert space, and of the description of physical observables. The measurement process will be investigated in the next chapter, and some of its implications will be discussed. In this chapter you will also learn the essential tools for studying entanglement, the stuff such weird things as quantum teleportation, quantum cryptography and quantum computation are made of. The third chapter will present the dynamics of quantum mechanical systems and highlight the importance of the concept of symmetry in physics and particularly in quantum mechanics. It will be shown how the momentum and angular momentum operators can be obtained as generators of the symmetry groups of translation and rotation. I will also introduce a different kind of symmetries which are called gauge symmetries. They allow us to 'derive' the existence of classical electrodynamics from a simple invariance principle. This idea has been pushed much further in the 1960's when people applied it to the theories of elementary particles, and were quite successful with it. In fact, t'Hooft and Veltman got a Nobel prize for it in 1999 work in this area. Time dependent problems, even more than time-independent problems, are difficult to solve exactly and therefore perturbation theoretical methods are of great importance. They will be explained in chapter 5 and examples will be given.

Most of the ideas that you are going to learn in the first five chapters of these lectures are known since about 1930, which is quite some time ago. The second part of these lectures, however, I will devote to topics which are currently the object of intense research (they are also my main area of research). In this last chapter I will discuss topics such as entanglement, Bell inequalities, quantum state teleportation, quantum computation and quantum cryptography. How much of these I can cover depends on the amount of time that is left, but I will certainly talk about some of them. While most physicists (hopefully) know the basics of quantum mechanics (the first five chapters of these lectures), many of them will not be familiar with the content of the other chapters. So, after these lectures you can be sure to know about something that quite a few professors do not know themselves! I hope that this motivates

you to stay with me until the end of the lectures.

Before I begin, I would like to thank two students, Vincenzo Vitelli and John Papadimitrou, who took my course last year and spotted errors and suggested improvements in the lecture notes of last years course. These errors are fixed now, but I expect that there are more. If you find errors, please let me know so I can get rid of them.

Last but not least, I would like to encourage you both, to ask questions during the lectures and to make use of my office hours. Questions are essential in the learning process, they help me to improve my lectures and above all it is more fun to lecture when there is some feedback from the audience.

# Chapter 1

# Mathematical Foundations

Before I begin to investigate the mathematical foundations of quantum mechanics, I would like to present a simple example which motivates the first step of this investigation a little bit.

## 1.1 The quantum mechanical state space

When we talk about physics, we usually speak about systems moving from one place to another at a certain speed etc. Such a description requires a number of concepts that are usually taken for granted in classical physics, but are not quite so obvious anymore in quantum physics. Clearly, if we want to describe a physical system we need to be able to specify its 'state' by giving fixed values to a number of variables. The state of a physical system can then be defined as the description of all the known (in fact one should say knowable) properties of that system and it represents your knowledge about the system. The set of all states forms what we usually call the state space. In classical mechanics for example this is the configuration or phase space, which is a **real vector space**. For a classical point-particle moving in one dimension, this space is two dimensional, one dimension for position, one dimension for momentum. We expect, in fact you probably know this from your second year lecture, that the quantum mechanical state space differs from that of classical mechanics. One reason for this can be found in the ability of quantum systems to exist

in coherent superpositions of states with complex amplitudes. This implies, that a good choice for the quantum mechanical state space may be a complex vector space.

Before I begin to investigate the mathematical foundations of quantum mechanics, I would like to present a simple example which motivates the choice of complex vector spaces as state spaces a bit more. It will also remind you of some of the features of quantum mechanics which you have already encountered in your second year course. This will also give you a soft start into lecture before we become more abstract.

## 1.2 The quantum mechanical state space

In the next subsection I will briefly motivate that the quantum mechanical state space should be a complex vector space and also motivate some of the concepts that I will introduce later in a more rigorous fashion. From an example of a system whose properties you will know I will extract some essential features which will then suggest a reasonable choice for the quantum mechanical state space.

### 1.2.1 Polarized Light

Consider plane waves of light propagating along the z-axis. This light is described by the electric field vector $\vec{E}$ orthogonal on the direction of propagation. The electric field vector determines the state of light because in the cgs-system (which I use for convenience in this example) the magnetic field is given by $\vec{B} = \vec{e}_z \times \vec{E}$. Given the electric and magnetic field, the Maxwell equations determine the further time evolution of these fields in the absence of charges. Because $\vec{E}(\vec{r},t)$ cannot have a z-component, we can write

$$\vec{E}(\vec{r},t) = E_x(\vec{r},t)\vec{e}_x + E_y(\vec{r},t)\vec{e}_y = \begin{pmatrix} E_x(\vec{r},t) \\ E_y(\vec{r},t) \end{pmatrix} \quad . \tag{1.1}$$

The electric field is real and the general solution of the free wave equation given by

$$E_x(\vec{r},t) \quad = \quad E_x^0 \cos(kz - \omega t + \alpha_x)$$

$$E_y(\vec{r},t) \;=\; E_y^0\cos(kz-\omega t+\alpha_y)\;\;.$$

Here $k=2\pi/\lambda$ is the wave-number, $\omega=2\pi\nu$ the frequency, $\alpha_x$ and $\alpha_y$ are the real phases and $E_x^0$ and $E_y^0$ the **real valued** amplitudes of the field components. The energy density of the field is given by

$$\begin{aligned}\epsilon(\vec{r},t) &= \frac{1}{8\pi}(\vec{E}(\vec{r},t)^2+\vec{B}(\vec{r},t)^2)\\ &= \frac{1}{4\pi}\left[(E_x^0)^2\cos^2(kz-\omega t+\alpha_x)+(E_y^0)^2\cos^2(kz-\omega t+\alpha_y)\right]\;\;.\end{aligned}$$

For a fixed position $\vec{r}$ we are only really interested in the time-averaged energy density which determines the rate at which energy flows in z-direction. Averaging over one period of the light we obtain the averaged energy density $\bar{\epsilon}(\vec{r})$ with

$$\bar{\epsilon}(\vec{r})=\frac{1}{8\pi}\left[(E_x^0)^2+(E_y^0)^2\right]\;\;. \tag{1.2}$$

For practical purposes it is useful to introduce the **complex** field components

$$E_x(\vec{r},t)=Re(E_xe^{i(kz-\omega t)})\quad E_y(\vec{r},t)=Re(E_ye^{i(kz-\omega t)})\;\;, \tag{1.3}$$

with $E_x=E_x^0e^{i\alpha_x}$ and $E_y=E_y^0e^{i\alpha_y}$. Comparing with Eq. (1.2) we find that the averaged energy density is given by

$$\bar{\epsilon}(\vec{r})=\frac{1}{8\pi}\left[|E_x|^2+|E_y|^2\right]\;\;. \tag{1.4}$$

Usually one works with the complex field

$$\vec{E}(\vec{r},t)=(E_x\vec{e}_x+E_y\vec{e}_y)e^{i(kz-\omega t)}=\begin{pmatrix}E_x\\E_y\end{pmatrix}e^{i(kz-\omega t)}\;\;. \tag{1.5}$$

This means that we are now characterizing the state of light by a vector with **complex** components.

The polarization of light waves are described by $E_x$ and $E_y$. In the general case of complex $E_x$ and $E_y$ we will have elliptically polarized light. There are a number of important special cases (see Figures 1.1 for illustration).

1. $E_y = 0$: linear polarization along the x-axis.
2. $E_x = 0$: linear polarization along the y-axis.
3. $E_x = E_y$: linear polarization along $45^0$-axis.
4. $E_y = iE_x$: Right circularly polarized light.
5. $E_y = -iE_x$: Left circularly polarized light.

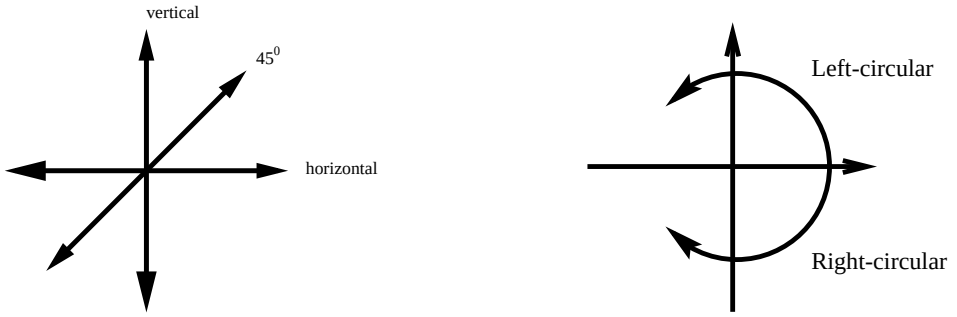


Figure 1.1: Left figure: Some possible linear polarizations of light, horizontally, vertically and 45 degrees. Right figure: Left- and right-circularly polarized light. The light is assumed to propagate away from you.

In the following I would like to consider some simple experiments. Let us first consider a light wave that is falling onto an x-polarizer which allows x-polarized light to pass through (but not y polarized light). This is shown in figure 1.2. After passing the polarizer the light

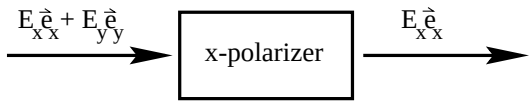


Figure 1.2: Light of arbitrary polarization is hitting a x-polarizer.

is x-polarized and from the expression for the energy density Eq. (1.4) we find that the ratio between incoming intensity $I_0$ (energy density

times speed of light) and outgoing intensity $I$ is given by

$$\frac{I}{I_0} = \frac{|E_x|^2}{|E_x|^2 + |E_y|^2}. \tag{1.6}$$

So far this looks like an experiment in classical electrodynamics or optics. Now I would like to change the way of looking at this problem and thereby turn it into a quantum mechanical experiment. To this end assume, that the light impinging on the polarizer has a very low intensity. In that case we know that we have to consider light as consisting of photons. In fact, a photo detector will register individual 'clicks' and not weaker and weaker 'clicks'. One fundamental property of photons is that they cannot be split – there is no such thing as half a photon going through the polarizer. In this photon picture we have to conclude that sometimes a photon will be absorbed in the polarizer and sometimes it passes it unchanged. The probability $p$ for the photon to pass through the polarizer is obviously the ratio between transmitted and incoming intensities

$$p = \frac{|E_x|^2}{|E_x|^2 + |E_y|^2} \quad . \tag{1.7}$$

If we write the state of the light with normalized intensity

$$\vec{E}_N = \frac{E_x}{\sqrt{|E_x|^2 + |E_y|^2}}\vec{e}_x + \frac{E_y}{\sqrt{|E_x|^2 + |E_y|^2}}\vec{e}_y \quad , \tag{1.8}$$

then in fact we find that the probability for the photon to pass the x-polarizer is just the square of the amplitude in front of the basis vector $\vec{e}_x$! This is just one of the quantum mechanical rules that you have learned in your second year course.

Now we can consider a slightly more elaborate experiment. If after passing the x-polarizer, the light falls onto a y-polarizer (see Fig 1.3), then no light will go through the polarizer. This simple experimental result changes when we place an additional polarizer between the x and the y-polarizer. Assume that we place a x'-polarizer between the two polarizers. Then we will observe light after the y-polarizer (see Fig. 1.4). The light after the last polarizer is described by $\tilde{E}_y\vec{e}_y$. The

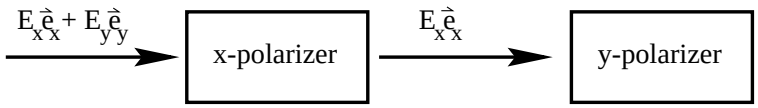


Figure 1.3: Light of arbitrary polarization is hitting a x-polarizer and subsequently a y-polarizer. No light goes through both polarizers.

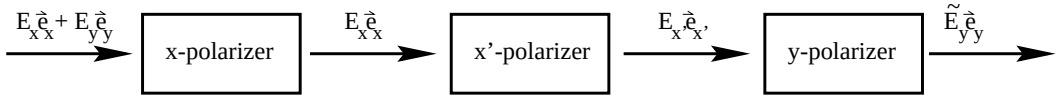


Figure 1.4: An x'-polarizer is placed in between an x-polarizer and a y-polarizer. Now we observe light passing through the y-polarizer.

amplitude $\tilde{E}_y$ is easily calculated by making some basis changes. The new rotated basis $\vec{e}'_x, \vec{e}'_y$ (see Fig. 1.5) can be expressed by the old basis by

$$\vec{e}'_x = \cos\phi\vec{e}_x + \sin\phi\vec{e}_y \qquad \vec{e}'_y = -\sin\phi\vec{e}_x + \cos\phi\vec{e}_y \tag{1.9}$$

and vice versa

$$\vec{e}_x = \cos\phi\vec{e}'_x - \sin\phi\vec{e}'_y \qquad \vec{e}_y = \sin\phi\vec{e}'_x + \cos\phi\vec{e}'_y \ . \tag{1.10}$$

Note that $\cos\phi = \vec{e}'_x \cdot \vec{e}_x$ and $\sin\phi = \vec{e}'_x \cdot \vec{e}_y$ where I have used the real scalar product between vectors.

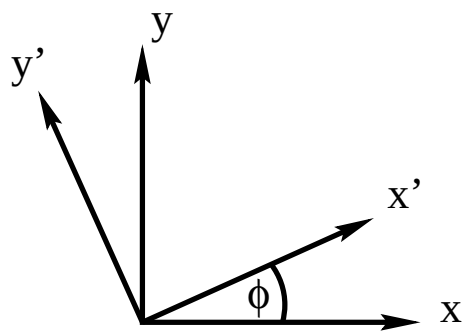


Figure 1.5: An x'-polarizer is placed in between an x-polarizer and a y-polarizer. Now we observe light passing through the y-polarizer.

Now let us describe the (x-x'-y) experiment mathematically. The complex electric field (without the time dependence) is given by

**before the x-polarizer:**

$$\vec{E}_1 = E_x\vec{e}_x + E_y\vec{e}_y.$$

**after the x-polarizer:**

$$\vec{E}_2 = (\vec{E}_1\vec{e}_x)\vec{e}_x = E_x\vec{e}_x = E_x\cos\phi\vec{e}'_x - E_x\sin\phi\vec{e}'_y.$$

**after the x'-polarizer:**

$$\vec{E}_3 = (\vec{E}_2\vec{e}'_x)\vec{e}'_x = E_x\cos\phi\vec{e}'_x = E_x\cos^2\phi\vec{e}_x + E_x\cos\phi\sin\phi\vec{e}_y.$$

**after the y-polarizer:**

$$\vec{E}_4 = (\vec{E}_3\vec{e}_y)\vec{e}_y = E_x\cos\phi\sin\phi\vec{e}_y = \tilde{E}_y\vec{e}_y.$$

Therefore the ratio between the intensity before the x'-polarizer and after the y-polarizer is given by $\frac{I}{I_0} = \cos^2\phi\sin^2\phi = |\vec{e}'_x\cdot\vec{e}_y|^2|\vec{e}'_x\cdot\vec{e}_x|^2$.

Again, if we interpret this result in the photon picture, then we arrive at the conclusion, that the probability for the photon to pass through the whole apparatus is given by $\cos^2\phi\sin^2\phi$. In fact it is easy to see from the behaviour of the light passing through the x'-polarizer that the probability of a photon passing through this polarizer is given by

$$p = \frac{|\vec{E}\vec{e}'_x|^2}{|\vec{E}\vec{e}'_x|^2 + |\vec{E}\vec{e}'_y|^2} \quad . \tag{1.11}$$

This demonstrates that the scalar product between vectors plays an important role in the calculation of the probability of that the photon passes through the beam splitter. Apart from this, the example of polarized light shows that light of one polarization may be interpreted as a superposition of light of other polarizations. This superposition is represented by adding vectors with complex coefficients. If we consider this situation in the photon picture we have to accept that a photon of a particular polarization can also be interpreted as a superposition of different polarization states.

The last thing we should observe is that we can interpret the action of an x-polarizer (or any general polarizer) (e.g. Fig. 1.2) as a linear

transformation between complex vectors. It maps the vector $\begin{pmatrix} E_x \\ E_y \end{pmatrix}$ into the vector $\begin{pmatrix} E_x \\ 0 \end{pmatrix}$.

All these observations suggest that complex vectors, their amplitudes, scalar products and linear transformations between complex vectors are the basic ingredient in the mathematical structure of quantum mechanics as opposed to the real vector space of classical mechanics. Therefore the rest of this chapter will be devoted to a more detailed investigation of the structure of complex vector-spaces and their properties.

**Suggestions for further reading:**
G. Baym *Lectures on Quantum Mechanics*, W.A. Benjamin 1969.
P.A.M. Dirac, *The principles of Quantum Mechanics*, Oxford University Press 1958

———End of $1^{st}$ lecture

### 1.2.2 Complex vector spaces

I will now give you a formal definition of a complex vector space and will then present some of its properties. Before I come to this definition, I introduce a standard notation that I will use in this chapter. Given some set V we define
**Notation:**

1. $\forall x \in V$ means: **For all** x that lie in $V$.
2. $\exists x \in V$ means: **There exists** an element $x$ that lies in $V$.

Now I can state the definition of the complex vector space. It will look a bit abstract at the beginning, but you will soon get used to it, especially when you solve some problems with it.

**Definition 1** *Given a quadruple* $(V, \mathbf{C}, +, \cdot)$ *where* $V$ *is a set of objects (usually called vectors),* $\mathbf{C}$ *denotes the set of complex numbers,* $'+'$ *denotes the group operation of* **addition** *and* $'\cdot'$ *denotes the* **multiplication** *of a vector with a complex number.* $(V, \mathbf{C}, +, \cdot)$ *is called a* **complex vector space** *if the following properties are satisfied:*

1. *(V,+) is an* **Abelian group**, *which means that*

   (a) $\forall \vec{a}, \vec{b} \in V \quad \Rightarrow \vec{a} + \vec{b} \in V.$ *(closure)*

   (b) $\forall \vec{a}, \vec{b}, \vec{c} \in V \Rightarrow \vec{a} + (\vec{b} + \vec{c}) = (\vec{a} + \vec{b}) + \vec{c}.$ *(associative)*

   (c) $\exists \vec{0} \in V$ *so that* $\forall \vec{a} \in V \Rightarrow \vec{a} + \vec{0} = \vec{a}.$ *(zero)*

   (d) $\forall \vec{a} \in V : \exists (-\vec{a}) \in V$ *so that* $\vec{a} + (-\vec{a}) = \vec{0}.$ *(inverse)*

   (e) $\forall \vec{a}, \vec{b} \in V \Rightarrow \vec{a} + \vec{b} = \vec{b} + \vec{a}.$ *(Abelian)*

2. *The Scalar multiplication satisfies*

   (a) $\forall \alpha \in \mathbf{C}, \vec{x} \in V \quad \Rightarrow \alpha \vec{x} \in V$

   (b) $\forall \vec{x} \in V \quad \Rightarrow 1 \cdot \vec{x} = \vec{x}$ *(unit)*

   (c) $\forall c, d \in \mathbf{C}, \vec{x} \in V \quad \Rightarrow (c \cdot d) \cdot \vec{x} = c \cdot (d \cdot \vec{x})$ *(associative)*

   (d) $\forall c, d \in \mathbf{C}, \vec{x}, \vec{y} \in V \Rightarrow c \cdot (\vec{x} + \vec{y}) = c \cdot \vec{x} + c \cdot \vec{y}$
   *and* $(c + d) \cdot \vec{x} = c \cdot \vec{x} + c \cdot \vec{y}.$ *(distributive)*

This definition looks quite abstract but a few examples will make it clearer.

**Example:**

1. *A simple proof*
   I would like to show how to prove the statement $0 \cdot \vec{x} = \vec{0}$. This might look trivial, but nevertheless we need to prove it, as it has not been stated as an axiom. From the axioms given in Def. 1 we conclude.

$$\begin{aligned}
\vec{0} &\overset{(1d)}{=} -\vec{x} + \vec{x} \\
&\overset{(2b)}{=} -\vec{x} + 1 \cdot \vec{x} \\
&= -\vec{x} + (1 + 0) \cdot \vec{x} \\
&\overset{(2d)}{=} -\vec{x} + 1 \cdot \vec{x} + 0 \cdot \vec{x} \\
&\overset{(2b)}{=} -\vec{x} + \vec{x} + 0 \cdot \vec{x} \\
&\overset{(1d)}{=} \vec{0} + 0 \cdot \vec{x} \\
&\overset{(1c)}{=} 0 \cdot \vec{x} \quad .
\end{aligned}$$

2. *The* $\mathbf{C}^2$
   This is the set of two-component vectors of the form

$$\vec{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad , \tag{1.12}$$

   where the $a_i$ are complex numbers. The addition and scalar multiplication are defined as

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} := \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \end{pmatrix} \tag{1.13}$$

$$c \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} := \begin{pmatrix} c \cdot a_1 \\ c \cdot a_2 \end{pmatrix} \tag{1.14}$$

   It is now easy to check that $V = \mathbf{C}^2$ together with the addition and scalar multiplication defined above satisfy the definition of a complex vector space. (You should check this yourself to get some practise with the axioms of vector space.) The vector space $\mathbf{C}^2$ is the one that is used for the description of spin-$\frac{1}{2}$ particles such as electrons.

3. *The set of real functions of one variable* $f : \mathbf{R} \to \mathbf{R}$
   The group operations are defined as

$$\begin{aligned} (f_1 + f_2)(x) &:= f_1(x) + f_2(x) \\ (c \cdot f)(x) &:= c \cdot f(x) \end{aligned}$$

   Again it is easy to check that all the properties of a complex vector space are satisfied.

4. *Complex* $n \times n$ *matrices*
   The elements of the vector space are

$$M = \begin{pmatrix} m_{11} & \dots & m_{1n} \\ \vdots & \ddots & \vdots \\ m_{n1} & \dots & m_{nn} \end{pmatrix} \quad , \tag{1.15}$$

where the $m_{ij}$ are arbitrary complex numbers. The addition and scalar multiplication are defined as

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} + \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \dots & b_{nn} \end{pmatrix} = \begin{pmatrix} a_{11}+b_{11} & \dots & a_{1n}+b_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1}+b_{n1} & \dots & a_{nn}+b_{nn} \end{pmatrix} ,$$

$$c \cdot \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} = \begin{pmatrix} c \cdot a_{11} & \dots & c \cdot a_{1n} \\ \vdots & \ddots & \vdots \\ c \cdot a_{n1} & \dots & c \cdot a_{nn} \end{pmatrix} .$$

Again it is easy to confirm that the set of complex $n \times n$ matrices with the rules that we have defined here forms a vector space. Note that we are used to consider matrices as objects acting **on** vectors, but as we can see here we can also consider them as elements (vectors) of a vector space themselves.

Why did I make such an abstract definition of a vector space? Well, it may seem a bit tedious, but it has a real advantage. Once we have introduced the abstract notion of complex vector space anything we can prove directly from these abstract laws in Definition 1 will *hold true for any vector space* irrespective of how complicated it will look superficially. What we have done, is to isolate the basic structure of vector spaces without referring to any particular representation of the elements of the vector space. This is very useful, because we do not need to go along every time and prove the same property again when we investigate some new objects. What we only need to do is to prove that our new objects have an addition and a scalar multiplication that satisfy the conditions stated in Definition 1.

In the following subsections we will continue our exploration of the idea of complex vector spaces and we will learn a few useful properties that will be helpful for the future.

### 1.2.3 Basis and Dimension

Some of the most basic concepts of vector spaces are those of linear independence, dimension and basis. They will help us to express vec-

tors in terms of other vectors and are useful when we want to define operators on vector spaces which will describe observable quantities.

Quite obviously some vectors can be expressed by linear combinations of others. For example

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 2 \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} \ . \tag{1.16}$$

It is natural to consider a given set of vectors $\{\vec{x}_1, \ldots, \vec{x}_k\}$ and to ask the question, whether a vector in this set can be expressed as a linear combination of the others. Instead of answering this question directly we will first consider a slightly different question. Given a set of vectors $\{\vec{x}_1, \ldots, \vec{x}_k\}$, can the null vector $\vec{0}$ can be expressed as a linear combination of these vectors? This means that we are looking for a linear combination of vectors of the form

$$\lambda_1 \vec{x}_1 + \ldots + \lambda_2 \vec{x}_k = \vec{0} \ . \tag{1.17}$$

Clearly Eq. (1.17) can be satisfied when all the $\lambda_i$ vanish. But this case is trivial and we would like to exclude it. Now there are two possible cases left:
a) There is **no** combination of $\lambda_i$'s, not all of which are zero, that satisfies Eq. (1.17).
b) There are combinations of $\lambda_i$'s,, not all of which are zero, that satisfy Eq. (1.17).

These two situations will get different names and are worth the

**Definition 2** *A set of vectors* $\{\vec{x}_1, \ldots, \vec{x}_k\}$ *is called* **linearly independent** *if the equation*

$$\lambda_1 \vec{x}_1 + \ldots + \lambda_2 \vec{x}_k = \vec{0} \tag{1.18}$$

*has only the trivial solution* $\lambda_1 = \ldots = \lambda_k = 0$.
*If there is a nontrivial solution to Eq. (1.18), i.e. at least one of the* $\lambda_i \neq 0$*, then we call the vectors* $\{\vec{x}_1, \ldots, \vec{x}_k\}$ **linearly dependent**.

Now we are coming back to our original question as to whether there are vectors in $\{\vec{x}_1, \ldots, \vec{x}_k\}$ that can be expressed by all the other vectors in that set. As a result of this definition we can see the following

**Lemma 3** *For a set of* **linearly independent** *vectors* $\{\vec{x}_1, \ldots, \vec{x}_k\}$, **no** $\vec{x}_i$ *can be expressed as a linear combination of the other vectors, i.e. one cannot find* $\lambda_j$ *that satisfy the equation*

$$\lambda_1\vec{x}_1 + \ldots + \lambda_{i-1}\vec{x}_{i-1} + \lambda_{i+1}\vec{x}_{i+1} + \ldots + \lambda_k\vec{x}_k = \vec{x}_i \quad . \tag{1.19}$$

*In a set of* **linearly dependent** *vectors* $\{\vec{x}_1, \ldots, \vec{x}_k\}$ *there is at least one* $\vec{x}_i$ *that can be expressed as a linear combination of all the other* $\vec{x}_j$.

**Proof:** Exercise! □

**Example:** The set $\{\vec{0}\}$ consisting of the null vector only, is linearly dependent.

In a sense that will become clearer when we really talk about quantum mechanics, in a set of linearly independent set of vectors, each vector has some quality that none of the other vectors have.

After we have introduced the notion of linear dependence, we can now proceed to define the dimension of a vector space. I am sure that you have a clear intuitive picture of the notion of dimension. Evidently a plain surface is 2-dimensional and space is 3-dimensional. Why do we say this? Consider a plane, for example. Clearly, every vector in the plane can be expressed as a linear combination of any two linearly independent vectors $\vec{e}_1, \vec{e}_2$. As a result you will not be able to find a set of three linearly independent vectors in a plane, while two linearly independent vectors can be found. This is the reason to call a plane a two-dimensional space. Let's formalize this observation in

**Definition 4** *The* **dimension** *of a vector space* $V$ *is the largest number of linearly independent vectors in* $V$ *that one can find.*

Now we introduce the notion of **basis** of vector spaces.

**Definition 5** *A set of vectors* $\{\vec{x}_1, \ldots, \vec{x}_k\}$ *is called a* **basis** *of a vector space* $V$ *if*
*a)* $\vec{x}_1, \ldots, \vec{x}_k$ *are linearly independent.*
*b)* $\forall \vec{x} \in V : \exists \lambda_i \in \mathbf{C} \Rightarrow x = \sum_{i=1}^{k} \lambda_i \vec{x}_i$.

Condition b) expresses that it is possible to write every vector as a linear combination of the basis vectors. The first condition makes sure that the set $\{\vec{x}_1, \ldots, \vec{x}_k\}$ is the smallest possible set to allow for condition b) to be satisfied. It turns out that any basis of an $N$-dimensional vector space $V$ contains exactly $N$ vectors. Let us illustrate the notion of basis.

**Example:**

1. Consider the space of vectors $\mathbf{C}^2$ with two components. Then the two vectors
$$\vec{x}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \vec{x}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad . \tag{1.20}$$
form a basis of $\mathbf{C}^2$. A basis for the $\mathbf{C}^N$ can easily be constructed in the same way.

2. An example for an infinite dimensional vector space is the space of complex polynomials, i.e. the set
$$V = \{c_0 + c_1 z + \ldots + c_k z^k | \text{arbitrary k and } \forall c_i \in \mathbf{C}\} \quad . \tag{1.21}$$
Two polynomials are equal when they give the same values for all $z \in \mathbf{C}$. Addition and scalar multiplication are defined coefficient wise. It is easy to see that the set $\{1, z, z^2, \ldots\}$ is linearly independent and that it contains infinitely many elements. Together with other examples you will prove (in the problem sheets) that Eq. (1.21) indeed describes a vector space.

End of $2^{nd}$ lecture

### 1.2.4 Scalar products and Norms on Vector Spaces

In the preceding section we have learnt about the concept of a basis. Any set of $N$ linearly independent vectors of an $N$ dimensional vector space $V$ form a basis. But not all such choices are equally convenient. To find useful ways to chose a basis and to find a systematic method to find the linear combinations of basis vectors that give any arbitrary

vector $\vec{x} \in V$ we will now introduce the concept of **scalar product** between two vectors. This is not to be confused with scalar multiplication which deals with a complex number and a vector. The concept of scalar product then allows us to formulate what we mean by **orthogonality**. Subsequently we will define the norm of a vector, which is the abstract formulation of what we normally call a length. This will then allow us to introduce orthonormal bases which are particularly handy.

The scalar product will play an extremely important role in quantum mechanics as it will in a sense quantify how similar two vectors (quantum states) are. In Fig. 1.6 you can easily see qualitatively that the pairs of vectors become more and more different from left to right. The scalar product puts this into a quantitative form. This is the reason why it can then be used in quantum mechanics to quantify how likely it is for two quantum states to exhibit the same behaviour in an experiment.

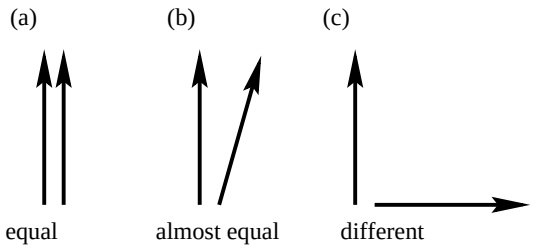


Figure 1.6: The two vectors on the left are equal, the next pair is 'almost' equal and the final pair is quite different. The notion of equal and different will be quantified by the scalar product.

To introduce the scalar product we begin with an abstract formulation of the properties that we would like any scalar product to have. Then we will have a look at examples which are of interest for quantum mechanics.

**Definition 6** *A* **complex scalar product** *on a vector space assigns to any two vectors $\vec{x}, \vec{y} \in V$ a complex number $(\vec{x}, \vec{y}) \in \mathbf{C}$ satisfying the following rules*

1. $\forall \vec{x}, \vec{y}, \vec{z} \in V, \alpha_i \in \mathbf{C}$ :

$$(\vec{x}, \alpha_1 \vec{y} + \alpha_2 \vec{z}) = \alpha_1 (\vec{x}, \vec{y}) + \alpha_2 (\vec{x}, \vec{z}) \qquad \textit{(linearity)}$$

2. $\forall \vec{x}, \vec{y} \in V : (\vec{x}, \vec{y}) = (\vec{y}, \vec{x})^*$ *(symmetry)*

3. $\forall \vec{x} \in V : (\vec{x}, \vec{x}) \geq 0$ *(positivity)*

4. $\forall \vec{x} \in V : (\vec{x}, \vec{x}) = 0 \Leftrightarrow \vec{x} = \vec{0}$

These properties are very much like the ones that you know from the ordinary dot product for real vectors, except for property 2 which we had to introduce in order to deal with the fact that we are now using complex numbers. In fact, you should show as an exercise that the ordinary condition $(\vec{x}, \vec{y}) = (\vec{y}, \vec{x})$ would lead to a contradiction with the other axioms if we have complex vector spaces. Note that we only defined linearity in the second argument. This is in fact all we need to do. As an exercise you may prove that any scalar product is anti-linear in the first component, i.e.

$$\forall \vec{x}, \vec{y}, \vec{z} \in V, \alpha \in \mathbf{C} : (\alpha\vec{x} + \beta\vec{y}, \vec{z}) = \alpha^*(\vec{x}, \vec{z}) + \beta^*(\vec{y}, \vec{z}) \ . \tag{1.22}$$

Note that vector spaces on which we have defined a scalar product are also called **unitary vector spaces**. To make you more comfortable with the scalar product, I will now present some examples that play significant roles in quantum mechanics.

**Examples:**

1. The scalar product in $\mathbf{C}^n$.
   Given two complex vectors $\vec{x}, \vec{y} \in \mathbf{C}^n$ with components $x_i$ and $y_i$ we define the scalar product

   $$(\vec{x}, \vec{y}) = \sum_{i=1}^{n} x_i^* y_i \tag{1.23}$$

   where $^*$ denotes the complex conjugation. It is easy to convince yourself that Eq. (1.23) indeed defines a scalar product. (Do it!).

2. Scalar product on continuous square integrable functions
   A square integrable function $\psi \in \mathcal{L}^2(\mathbf{R})$ is one that satisfies

   $$\int_{-\infty}^{\infty} |\psi(x)|^2 dx < \infty \tag{1.24}$$

Eq. (1.24) already implies how to define the scalar product for these square integrable functions. For any two functions $\psi, \phi \in \mathcal{L}^2(\mathbf{R})$ we define

$$(\psi, \phi) = \int_{-\infty}^{\infty} \psi(x)^* \phi(x) dx \quad . \tag{1.25}$$

Again you can check that definition Eq. (1.25) satisfies all properties of a scalar product. (Do it!) We can even define the scalar product for discontinuous square integrable functions, but then we need to be careful when we are trying to prove property 4 for scalar products. One reason is that there are functions which are nonzero only in isolated points (such functions are discontinuous) and for which Eq. (1.24) vanishes. An example is the function

$$f(x) = \begin{cases} 1 & \text{for } x = 0 \\ 0 & \text{anywhere else} \end{cases}$$

The solution to this problem lies in a redefinition of the elements of our set. If we identify all functions that differ from each other only in countably many points then we have to say that they are in fact the same element of the set. If we use this redefinition then we can see that also condition 4 of a scalar product is satisfied.

An extremely important property of the scalar product is the **Schwarz inequality** which is used in many proofs. In particular I will used it to prove the triangular inequality for the length of a vector and in the proof of the uncertainty principle for arbitrary observables.

**Theorem 7 (The Schwarz inequality)** *For any $\vec{x}, \vec{y} \in V$ we have*

$$|(\vec{x}, \vec{y})|^2 \leq (\vec{x}, \vec{x})(\vec{y}, \vec{y}) \quad . \tag{1.26}$$

**Proof:** For any complex number $\alpha$ we have

$$\begin{aligned} 0 &\leq (\vec{x} + \alpha\vec{y}, \vec{x} + \alpha\vec{y}) \\ &= (\vec{x}, \vec{x}) + \alpha(\vec{x}, \vec{y}) + \alpha^*(\vec{y}, \vec{x}) + |\alpha|^2(\vec{y}, \vec{y}) \\ &= (\vec{x}, \vec{x}) + 2vRe(\vec{x}, \vec{y}) - 2wIm(\vec{x}, \vec{y}) + (v^2 + w^2)(\vec{y}, \vec{y}) \\ &=: f(v, w) \end{aligned} \tag{1.27}$$

In the definition of $f(v,w)$ in the last row we have assumed $\alpha = v + iw$. To obtain the sharpest possible bound in Eq. (1.26), we need to minimize the right hand side of Eq. (1.27). To this end we calculate

$$0 = \frac{\partial f}{\partial v}(v,w) = 2Re(\vec{x},\vec{y}) + 2v(\vec{y},\vec{y}) \tag{1.28}$$

$$0 = \frac{\partial f}{\partial w}(v,w) = -2Im(\vec{x},\vec{y}) + 2w(\vec{y},\vec{y}) \quad . \tag{1.29}$$

Solving these equations, we find

$$\alpha_{min} = v_{min} + iw_{min} = -\frac{Re(\vec{x},\vec{y}) - iIm(\vec{x},\vec{y})}{(\vec{y},\vec{y})} = -\frac{(\vec{y},\vec{x})}{(\vec{y},\vec{y})} \quad . \tag{1.30}$$

Because all the matrix of second derivatives is positive definite, we really have a minimum. If we insert this value into Eq. (1.27) we obtain

$$0 \leq (\vec{x},\vec{x}) - \frac{(\vec{y},\vec{x})(\vec{x},\vec{y})}{(\vec{y},\vec{y})} \tag{1.31}$$

This implies then Eq. (1.26). Note that we have equality exactly if the two vectors are linearly dependent, i.e. if $\vec{x} = \gamma\vec{y}$ $\Box$.

Quite a few proofs in books are a little bit shorter than this one because they just use Eq. (1.30) and do not justify its origin as it was done here.

Having defined the scalar product, we are now in a position to define what we mean by orthogonal vectors.

**Definition 8** *Two vectors $\vec{x},\vec{y} \in V$ are called* **orthogonal** *if*

$$(\vec{x},\vec{y}) = 0 \quad . \tag{1.32}$$

*We denote with $\vec{x}_\perp$ a vector that is orthogonal to $\vec{x}$.*

Now we can define the concept of an orthogonal basis which will be very useful in finding the linear combination of vectors that give $\vec{x}$.

**Definition 9** *An* **orthogonal basis** *of an $N$ dimensional vector space $V$ is a set of $N$ linearly independent vectors such that each pair of vectors are orthogonal to each other.*

**Example:** In $\mathbf{C}$ the three vectors

$$\begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 3 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \tag{1.33}$$

form an orthogonal basis.

———End of $3^{rd}$ lecture

Now let us chose an orthogonal basis $\{\vec{x}_1, \ldots, \vec{x}_N\}$ of an N dimensional vector space. For any arbitrary vector $\vec{x} \in V$ we would like to find the coefficients $\lambda_1, \ldots, \lambda_N$ such that

$$\sum_{i=1}^{N} \lambda_i \vec{x}_i = \vec{x} \quad . \tag{1.34}$$

Of course we can obtain the $\lambda_i$ by trial and error, but we would like to find an efficient way to determine the coefficients $\lambda_i$. To see this, let us consider the scalar product between $\vec{x}$ and one of the basis vectors $\vec{x}_i$. Because of the orthogonality of the basis vectors, we find

$$(\vec{x}_i, \vec{x}) = \lambda_i (\vec{x}_i, \vec{x}_i) \quad . \tag{1.35}$$

Note that this result holds true only because we have used an orthogonal basis. Using Eq. (1.35) in Eq. (1.34), we find that for an orthogonal basis any vector $\vec{x}$ can be represented as

$$\vec{x} = \sum_{i=1}^{N} \frac{(\vec{x}_i, \vec{x})}{(\vec{x}_i, \vec{x}_i)} \vec{x}_i \quad . \tag{1.36}$$

In Eq. (1.36) we have the denominator $(\vec{x}_i, \vec{x}_i)$ which makes the formula a little bit clumsy. This quantity is the square of what we usually call the length of a vector. This idea is illustrated in Fig. 1.7 which shows a vector $\vec{x} = \binom{a}{b}$ in the two-dimensional real vector space $\mathbf{R}^2$. Clearly its length is $\sqrt{a^2 + b^2}$. What other properties does the length in this intuitively clear picture has? If I multiply the vector by a number $\alpha$ then we have the vector $\alpha\vec{x}$ which evidently has the length $\sqrt{\alpha^2 a^2 + \alpha^2 b^2} = |\alpha|\sqrt{a^2 + b^2}$. Finally we know that we have a triangular inequality.

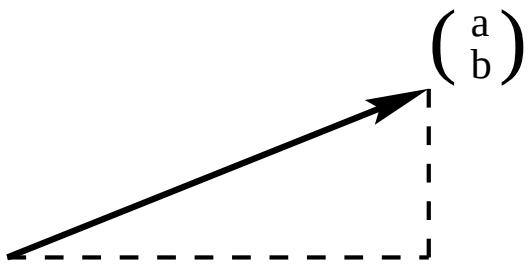


Figure 1.7: A vector $\vec{x} = \binom{a}{b}$ in $\mathbf{R}^2$. From plane geometry we know that its length is $\sqrt{a^2 + b^2}$ which is just the square root of the scalar product of the vector with itself.

This means that given two vectors $\vec{x}_1 = \binom{a_1}{b_1}$ and $\vec{x}_2 = \binom{a_2}{b_2}$ the length of the $\vec{x}_1 + \vec{x}_2$ is smaller than the sum of the lengths of $\vec{x}_1$ and $\vec{x}_2$. This is illustrated in Fig. 1.8. In the following I formalize the concept of a length and we will arrive at the definition of the **norm** of a vector $\vec{x}_i$. The concept of a norm is important if we want to define what we mean by two vectors being close to one another. In particular, norms are necessary for the definition of convergence in vector spaces, a concept that I will introduce in the next subsection. In the following I specify what properties a norm of a vector should satisfy.

**Definition 10** *A* **norm** *on a vector space $V$ associates with every $\vec{x} \in V$ a real number $||\vec{x}||$, with the properties.*

1. $\forall \vec{x} \in V$ : $||\vec{x}|| \geq 0$ *and* $||\vec{x}|| = 0 \Leftrightarrow \vec{x} = \vec{0}$. *(positivity)*
2. $\forall \vec{x} \in V, \alpha \in \mathbf{C}$ : $||\alpha\vec{x}|| = |\alpha| \cdot ||\vec{x}||$. *(linearity)*
3. $\forall \vec{x}, \vec{y} \in V$ : $||\vec{x} + \vec{y}|| \leq ||\vec{x}|| + ||\vec{y}||$. *(triangular inequality)*

A vector space with a norm defined on it is also called a **normed vector space**. The three properties in Definition 10 are those that you would intuitively expect to be satisfied for any decent measure of length. As expected norms and scalar products are closely related. In fact, there is a way of generating norms very easily when you already have a scalar product.

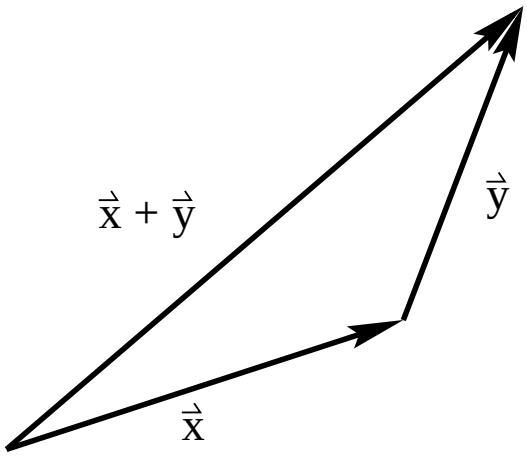


Figure 1.8: The triangular inequality illustrated for two vectors $\vec{x}$ and $\vec{y}$. The length of $\vec{x}+\vec{y}$ is smaller than the some of the lengths of $\vec{x}$ and $\vec{y}$.

**Lemma 11** *Given a scalar product on a complex vector space, we can define the norm of a vector $\vec{x}$ by*

$$||\vec{x}|| = \sqrt{(\vec{x},\vec{x})} \quad . \tag{1.37}$$

**Proof:**

- Properties 1 and 2 of the norm follow almost trivially from the four basic conditions of the scalar product. □

- The proof of the triangular inequality uses the Schwarz inequality.

$$\begin{aligned} ||\vec{x}+\vec{y}||^2 &= |(\vec{x}+\vec{y},\vec{x}+\vec{y})| \\ &= |(\vec{x},\vec{x}+\vec{y})+(\vec{y},\vec{x}+\vec{y})| \\ &\leq |(\vec{x},\vec{x}+\vec{y})|+|(\vec{y},\vec{x}+\vec{y})| \\ &\leq ||\vec{x}||\cdot||\vec{x}+\vec{y}||+||\vec{y}||\cdot||\vec{x}+\vec{y}|| \quad . \end{aligned} \tag{1.38}$$

  Dividing both sides by $||\vec{x}+\vec{y}||$ yields the inequality. This assumes that the sum $\vec{x}+\vec{y} \neq \vec{0}$. If we have $\vec{x}+\vec{y}=\vec{0}$ then the Schwarz inequality is trivially satisfied.□

Lemma 11 shows that any **unitary vector space** can canonically (this means that there is basically one natural choice) turned into a **normed vector space**. The converse is, however, not true. Not every norm gives automatically rise to a scalar product (examples will be given in the exercises).

Using the concept of the norm we can now define an **orthonormal basis** for which Eq. (1.36) can then be simplified.

**Definition 12** *An* **orthonormal basis** *of an $N$ dimensional vector space is a set of $N$ pairwise orthogonal linearly independent vectors $\{\vec{x}_1, \ldots, \vec{x}_N\}$ where each vector satisfies $||\vec{x}_i||^2 = (\vec{x}_i, \vec{x}_i) = 1$, i.e. they are unit vectors. For an orthonormal basis and any vector $\vec{x}$ we have*

$$\vec{x} = \sum_{i=1}^{N} (\vec{x}_i, \vec{x}) \vec{x}_i = \sum_{i=1}^{N} \alpha_i \vec{x}_i \quad , \tag{1.39}$$

*where the* **components** *of $\vec{x}$ with respect to the basis $\{\vec{x}_1, \ldots, \vec{x}_N\}$ are the $\alpha_i = (\vec{x}_i, \vec{x})$.*

**Remark:** Note that in Definition 12 it was not really necessary to demand the linear independence of the vectors $\{\vec{x}_1, \ldots, \vec{x}_N\}$ because this follows from the fact that they are normalized and orthogonal. Try to prove this as an exercise.

Now I have defined what an orthonormal basis is, but you still do not know how to construct it. There are quite a few different methods to do so. I will present the probably most well-known procedure which has the name Gram-Schmidt procedure.

This will not be presented in the lecture. You should study this at home.

There will be an exercise on this topic in the Rapid Feedback class.

The starting point of the Gram-Schmidt orthogonalization procedure is a set of linearly independent vectors $\mathcal{S} = \{\vec{x}_1, \ldots, \vec{x}_n\}$. Now we would like to construct from them an orthonormal set of vectors $\{\vec{e}_1, \ldots, \vec{e}_n\}$. The procedure goes as follows

**First step** We chose $\vec{f}_1 = \vec{x}_1$ and then construct from it the normalized vector $\vec{e}_1 = \vec{x}_1 / ||\vec{x}_1||$.

Comment: We can normalize the vector $\vec{x}_1$ because the set $\mathcal{S}$ is linearly independent and therefore $\vec{x}_1 \neq 0$.

**Second step** We now construct $\vec{f_2} = \vec{x}_2 - (\vec{e}_1, \vec{x}_2)\vec{e}_1$ and from this the normalized vector $\vec{e}_2 = \vec{f_2}/||\vec{f_2}||$.
Comment: 1) $\vec{f_2} \neq \vec{0}$ because $\vec{x}_1$ and $\vec{x}_2$ are linearly independent. 2) By taking the scalar product $(\vec{e}_2, \vec{e}_1)$ we find straight away that the two vectors are orthogonal.

$\vdots$

**k-th step** We construct the vector

$$\vec{f_k} = \vec{x}_k - \sum_{i=1}^{k-1} (\vec{e}_i, \vec{x}_k)\vec{e}_i \quad .$$

Because of linear independence of $\mathcal{S}$ we have $\vec{f_k} \neq \vec{0}$. The normalized vector is then given by

$$\vec{e}_k = \frac{\vec{f_k}}{||\vec{f_k}||} \quad .$$

It is easy to check that the vector $\vec{e}_k$ is orthogonal to all $\rceil_i$ with $i < k$.

**n-th step** With this step the procedure finishes. We end up with a set of vectors $\{\vec{e}_1, \ldots, \vec{e}_n\}$ that are pairwise orthogonal and normalized.

### 1.2.5 Completeness and Hilbert spaces

In the preceding sections we have encountered a number of basic ideas about vector spaces. We have introduced scalar products, norms, bases and the idea of dimension. In order to be able to define a Hilbert space, the state space of quantum mechanics, we require one other concept, that of completeness, which I will introduce in this section.

What do we mean by complete? To see this, let us consider sequences of elements of a vector space (or in fact any set, but we are

only interested in vector spaces). I will write sequences in two different ways

$$\{\vec{x}_i\}_{i=0,\ldots,\infty} \equiv (\vec{x}_0, \vec{x}_1, \vec{x}_2 \ldots) \quad . \tag{1.40}$$

To define what we mean by a convergent sequences, we use norms because we need to be able to specify when two vectors are close to each other.

**Definition 13** *A sequence* $\{\vec{x}_i\}_{i=0,\ldots,\infty}$ *of elements from a normed vector space* $V$ *converges towards a vector* $\vec{x} \in V$ *if* **for all** $\epsilon > 0$ **there is an** $n_0$ **such that for all** $n > n_0$ *we have*

$$||\vec{x} - \vec{x}_n|| \leq \epsilon \quad . \tag{1.41}$$

But sometimes you do not know the limiting element, so you would like to find some other criterion for convergence without referring to the limiting element. This idea led to the following

**Definition 14** *A sequence* $\{\vec{x}_i\}_{i=0,\ldots,\infty}$ *of elements from a normed vector space* $V$ *is called a* **Cauchy sequence** *if for all* $\epsilon > 0$ *there is an* $n_0$ *such that for all* $m, n > n_0$ *we have*

$$||\vec{x}_m - \vec{x}_n|| \leq \epsilon \quad . \tag{1.42}$$

———— End of $4^{th}$ lecture

Now you can wonder whether every Cauchy sequence converges. Well, it sort of does. But unfortunately sometimes the limiting element does not lie in the set from which you draw the elements of your sequence. How can that be? To illustrate this I will present a vector space that is not complete! Consider the set

$$V = \{\vec{x} : \text{only finitely many components of } \vec{x} \text{ are non-zero}\} \quad .$$

An example for an element of $V$ is $\vec{x} = (1, 2, 3, 4, 5, 0, \ldots)$. It is now quite easy to check that $V$ is a vector-space when you define addition of two vectors via

$$\vec{x} + \vec{y} = (x_1 + y_1, x_2 + y_2, \ldots)$$

and the multiplication by a scalar via

$$c\vec{x} = (cx_1, cx_2, \ldots) \quad .$$

Now I define a scalar product from which I will then obtain a norm via the construction of Lemma 11. We define the scalar product as

$$(\vec{x}, \vec{y}) = \sum_{k=1}^{\infty} x_k^* y_k \quad .$$

Now let us consider the series of vectors

$$\begin{aligned}
\vec{x}_1 &= (1, 0, 0, 0, \ldots) \\
\vec{x}_2 &= (1, \frac{1}{2}, 0, \ldots) \\
\vec{x}_3 &= (1, \frac{1}{2}, \frac{1}{4}, 0, \ldots) \\
\vec{x}_4 &= (1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots) \\
&\vdots \\
\vec{x}_k &= (1, \frac{1}{2}, \ldots, \frac{1}{2^{k-1}}, 0, \ldots)
\end{aligned}$$

For any $n_0$ we find that for $m > n > n_0$ we have

$$||\vec{x}_m - \vec{x}_n|| = ||(0, \ldots, 0, \frac{1}{2^n}, \ldots, \frac{1}{2^{m-1}}, 0, \ldots)|| \leq \frac{1}{2^{n-1}} \quad .$$

Therefore it is clear that the sequence $\{\vec{x}_k\}_{k=1,\ldots\infty}$ is a Cauchy sequence. However, the limiting vector is not a vector from the vector space $V$, because the limiting vector contains **infinitely** many nonzero elements.

Considering this example let us define what we mean by a **complete vector space**.

**Definition 15** *A vector space $V$ is called* **complete** *if every Cauchy sequence of elements from the vector space $V$ converges towards an element of $V$.*

Now we come to the definition of Hilbert spaces.

**Definition 16** *A vector space $\mathcal{H}$ is a* **Hilbert space** *if it satisfies the following two conditions*

1. *$\mathcal{H}$ is a unitary vector space.*

2. *$\mathcal{H}$ is complete.*

Following our discussions of the vectors spaces, we are now in the position to formulate the first **postulate** of quantum mechanics.

**Postulate 1** *The state of a quantum system is described by a vector in a Hilbert space $\mathcal{H}$.*

Why did we postulate that the quantum mechanical state space is a Hilbert space? Is there a reason for this choice?

Let us argue physically. We know that we need to be able to represent superpositions, i.e. we need to have a vector space. From the superposition principle we can see that there will be states that are not orthogonal to each other. That means that to some extent one quantum state can be 'present' in another non-orthogonal quantum state – they 'overlap'. The extent to which the states overlap can be quantified by the scalar product between two vectors. In the first section we have also seen, that the scalar product is useful to compute probabilities of measurement outcomes. You know already from your second year course that we need to normalize quantum states. This requires that we have a norm which can be derived from a scalar product. Because of the obvious usefulness of the scalar product, we require that the state space of quantum mechanics is a vector space equipped with a scalar product. The reason why we demand completeness, can be seen from a physical argument which could run as follows. Consider any sequence of physical states that is a Cauchy sequence. Quite obviously we would expect this sequence to converge to a physical state. It would be extremely strange if by means of such a sequence we could arrive at an unphysical state. Imagine for example that we change a state by

smaller and smaller amounts and then suddenly we would arrive at an unphysical state. That makes no sense! Therefore it seems reasonable to demand that the physical state space is complete.

What we have basically done is to distill the essential features of quantum mechanics and to find a mathematical object that represents these essential features without any reference to a special physical system.

In the next sections we will continue this programme to formulate more principles of quantum mechanics.

### 1.2.6 Dirac notation

In the following I will introduce a useful way of writing vectors. This notation, the Dirac notation, applies to any vector space and is very useful, in particular it makes life a lot easier in calculations. As most quantum mechanics books are written in this notation it is quite important that you really learn how to use this way of writing vectors. If it appears a bit weird to you in the first place you should just practise its use until you feel confident with it. A good exercise, for example, is to rewrite in Dirac notation all the results that I have presented so far.

So far we have always written a vector in the form $\vec{x}$. The scalar product between two vectors has then been written as $(\vec{x}, \vec{y})$. Let us now make the following identification

$$\vec{x} \leftrightarrow |x\rangle \quad . \tag{1.43}$$

We call $|x\rangle$ a **ket**. So far this is all fine and well. It is just a new notation for a vector. Now we would like to see how to rewrite the scalar product of two vectors. To understand this best, we need to talk a bit about linear functions of vectors.

**Definition 17** *A function $f : V \to \mathbf{C}$ from a vector space into the complex numbers is called* **linear** *if for any $|\psi\rangle, |\phi\rangle \in V$ and any $\alpha, \beta \in \mathbf{C}$ we have*

$$f(\alpha|\psi\rangle + \beta|\phi\rangle) = \alpha f(|\psi\rangle) + \beta f(|\phi\rangle) \tag{1.44}$$

With two linear function $f_1, f_2$ also the linear combination $\mu f_1 + \nu f_2$ is a linear function. Therefore the linear functions themselves form a

vector space and it is even possible to define a scalar product between linear functions. The space of the linear function on a vector space $V$ is called the **dual space** $V^*$.

Now I would like to show you an example of a linear function which I define by using the scalar product between vectors. I define the function $f_{|\phi\rangle} : V \to \mathbf{C}$, where $|\phi\rangle \in V$ is a fixed vector so that for all $|\psi\rangle \in V$

$$f_{|\phi\rangle}(|\psi\rangle) := (|\phi\rangle, |\psi\rangle) \ . \tag{1.45}$$

Now I would like to introduce a new notation for $f_{|\phi\rangle}$. From now on I will identify

$$f_{|\phi\rangle} \leftrightarrow \langle\phi| \tag{1.46}$$

and use this to rewrite the scalar product between two vectors $|\phi\rangle, |\psi\rangle$ as

$$\langle\phi|\psi\rangle := \langle\phi|(|\psi\rangle) \equiv (|\phi\rangle, |\psi\rangle) \ . \tag{1.47}$$

The object $\langle\phi|$ is called **bra** and the Dirac notation is therefore sometimes called **braket notation**. Note that while the *ket* is a vector in the vector space $V$, the *bra* is an element of the dual space $V^*$.

At the moment you will not really be able to see the usefulness of this notation. But in the next section when I will introduce linear operators, you will realize that the Dirac notation makes quite a few notations and calculations a lot easier.

## 1.3 Linear Operators

So far we have only dealt with the elements (vectors) of vector spaces. Now we need to learn how to transform these vectors, that means how to transform one set of vectors into a different set. Again as quantum mechanics is a linear theory we will concentrate on the description of linear operators.

### 1.3.1 Definition in Dirac notation

**Definition 18** *An* **linear operator** $\hat{A} : \mathcal{H} \to \mathcal{H}$ *associates to every vector* $|\psi\rangle \in \mathcal{H}$ *a vector* $\hat{A}|\psi\rangle \in \mathcal{H}$ *such that*

$$\hat{A}(\lambda|\psi\rangle + \mu|\phi\rangle) = \lambda\hat{A}|\psi\rangle + \mu\hat{A}|\phi\rangle \tag{1.48}$$

*for all* $|\psi\rangle, |\phi\rangle \in \mathcal{H}$ *and* $\lambda, \mu \in \mathbf{C}$.

End of $5^{th}$ lecture

A linear operator $\hat{A} : \mathcal{H} \to \mathcal{H}$ can be specified completely by describing its action on a basis set of $\mathcal{H}$. To see this let us chose an orthonormal basis $\{|e_i\rangle | i = 1, \ldots, N\}$. Then we can calculate the action of $\hat{A}$ on this basis. We find that the basis $\{|e_i\rangle | i = 1, \ldots, N\}$ is mapped into a new set of vectors $\{|f_i\rangle | i = 1, \ldots, N\}$ following

$$|f_i\rangle := \hat{A}|e_i\rangle \quad . \tag{1.49}$$

Of course every vector $|f_i\rangle$ can be represented as a linear combination of the basis vectors $\{|e_i\rangle | i = 1, \ldots, N\}$, i.e.

$$|f_i\rangle = \sum_k A_{ki}|e_k\rangle \quad . \tag{1.50}$$

Combining Eqs. (1.49) and (1.50) and taking the scalar product with $|e_j\rangle$ we find

$$\begin{aligned} A_{ji} &= \langle e_j | \sum_k (\hat{A}_{ki}|e_k\rangle) && (1.51) \\ &= \langle e_j | f_i \rangle \\ &\equiv \langle e_j | \hat{A} | e_i \rangle \quad . && (1.52) \end{aligned}$$

The $A_{ji}$ are called the matrix elements of the linear operator $\hat{A}$ with respect to the orthonormal basis $\{|e_i\rangle | i = 1, \ldots, N\}$.

I will now go ahead and express linear operators in the Dirac notation. First I will present a particularly simple operator, namely the unit operator $\hat{\mathbb{1}}$, which maps every vector into itself. Surprisingly enough this operator, expressed in the Dirac notation will prove to be very useful in calculations. To find its representation in the Dirac notation, we consider an arbitrary vector $|f\rangle$ and express it in an orthonormal basis $\{|e_i\rangle | i = 1, \ldots, N\}$. We find

$$|f\rangle = \sum_{j=1}^{N} f_j |e_j\rangle = \sum_{j=1}^{N} |e_j\rangle\langle e_j | f\rangle \quad . \tag{1.53}$$

To check that this is correct you just form the scalar product between $|f\rangle$ and any of the basis vectors $|e_i\rangle$. Now let us rewrite Eq. (1.53) a little bit, thereby defining the Dirac notation of operators.

$$|f\rangle = \sum_{j=1}^{N} |e_j\rangle\langle e_j|f\rangle =: (\sum_{j=1}^{N} |e_j\rangle\langle e_j|)|f\rangle \tag{1.54}$$

Note that the right hand side is defined in terms of the left hand side. The object in the brackets is quite obviously the identity operator because it maps any vector $|f\rangle$ into the same vector $|f\rangle$. Therefore it is totally justified to say that

$$\mathbb{1} \equiv \sum_{j=1}^{N} |e_j\rangle\langle e_j| \quad . \tag{1.55}$$

This was quite easy. We just moved some brackets around and we found a way to represent the unit operator using the Dirac notation. Now you can already guess how the general operator will look like, but I will carefully derive it using the identity operator. Clearly we have the following identity

$$\hat{A} = \hat{\mathbb{1}}\hat{A}\hat{\mathbb{1}} \quad . \tag{1.56}$$

Now let us use the Dirac notation of the identity operator in Eq. (1.55) and insert it into Eq. (1.56). We then find

$$\begin{aligned} \hat{A} &= (\sum_{j=1}^{N} |e_j\rangle\langle e_j|)\hat{A}(\sum_{k=1}^{N} |e_k\rangle\langle e_k|) \\ &= \sum_{jk}^{N} |e_j\rangle(\langle e_j|\hat{A}|e_k\rangle)\langle e_k| \\ &= \sum_{jk}^{N} (\langle e_j|\hat{A}|e_k\rangle)|e_j\rangle\langle e_k| \\ &= \sum_{jk}^{N} A_{jk}|e_j\rangle\langle e_k| \quad . \end{aligned} \tag{1.57}$$

Therefore you can express any linear operator in the Dirac notation, once you know its matrix elements in an orthonormal basis.

Matrix elements are quite useful when we want to write down linear operator in matrix form. Given an orthonormal basis $\{|e_i\rangle | i = 1, \ldots, N\}$ we can write every vector as a column of numbers

$$|g\rangle = \sum_i g_i |e_i\rangle \doteq \begin{pmatrix} g_1 \\ \vdots \\ g_N \end{pmatrix} \quad . \tag{1.58}$$

Then we can write our linear operator in the same basis as a matrix

$$\hat{A} \doteq \begin{pmatrix} A_{11} & \ldots & A_{1N} \\ \vdots & \ddots & \vdots \\ A_{N1} & \ldots & A_{NN} \end{pmatrix} \quad . \tag{1.59}$$

To convince you that the two notation that I have introduced give the same results in calculations, apply the operator $\hat{A}$ to any of the basis vectors and then repeat the same operation in the matrix formulation.

### 1.3.2 Adjoint and Hermitean Operators

Operators that appear in quantum mechanics are linear. But not all linear operators correspond to quantities that can be measured. Only a special subclass of operators describe physical observables. In this subsection I will describe these operators. In the following subsection I will then discuss some of their properties which then explain why these operators describe measurable quantities.

In the previous section we have considered the Dirac notation and in particular we have seen how to write the scalar product and matrix elements of an operator in this notation. Let us reconsider the matrix elements of an operator $\hat{A}$ in an orthonormal basis $\{|e_i\rangle | i = 1, \ldots, N\}$. We have

$$\langle e_i | (\hat{A} | e_j \rangle) = (\langle e_i | \hat{A}) | e_j \rangle \tag{1.60}$$

where we have written the scalar product in two ways. While the left hand side is clear there is now the question, what the bra $\langle e_i | \hat{A}$ on the right hand side means, or better, to which ket it corresponds to. To see this we need to make the

**Definition 19** *The* **adjoint operator** $\hat{A}^\dagger$ *corresponding to the linear operator* $\hat{A}$ *is the operator such that for all* $|x\rangle, |y\rangle$ *we have*

$$(\hat{A}^\dagger|x\rangle, |y\rangle) := (|x\rangle, \hat{A}|y\rangle) \ , \tag{1.61}$$

*or using the complex conjugation in Eq. (1.61) we have*

$$\langle y|\hat{A}^\dagger|x\rangle := \langle x|\hat{A}|y\rangle^* \ . \tag{1.62}$$

In matrix notation, we obtain the matrix representing $\hat{A}^\dagger$ by transposition and complex conjugation of the matrix representing $\hat{A}$. (Convince yourself of this).
**Example:**

$$\hat{A} = \begin{pmatrix} 1 & 2i \\ i & 2 \end{pmatrix} \tag{1.63}$$

and

$$\hat{A}^\dagger = \begin{pmatrix} 1 & -i \\ -2i & 2 \end{pmatrix} \tag{1.64}$$

The following property of adjoint operators is often used.

**Lemma 20** *For operators* $\hat{A}$ *and* $\hat{B}$ *we find*

$$(\hat{A}\hat{B})^\dagger \ = \ \hat{B}^\dagger\hat{A}^\dagger \ . \tag{1.65}$$

**Proof:** Eq. (1.65) is proven by

$$\begin{aligned} (|x\rangle, (\hat{A}\hat{B})^\dagger|y\rangle &= ((\hat{A}\hat{B})|x\rangle, |y\rangle) \\ &= (\hat{A}(\hat{B}|x\rangle), |y\rangle) \qquad \text{Now use Def. 19.} \\ &= (\hat{B}|x\rangle), \hat{A}^\dagger|y\rangle) \qquad \text{Use Def. 19 again.} \\ &= (|x\rangle), \hat{B}^\dagger\hat{A}^\dagger|y\rangle) \end{aligned}$$

As this is true for any two vectors $|x\rangle$ and $|y\rangle$ the two operators $(\hat{A}\hat{B})^\dagger$ and $\hat{B}^\dagger\hat{A}^\dagger$ are equal $\Box$.

It is quite obvious (see the above example) that in general an operator $\hat{A}$ and its adjoint operator $\hat{A}^\dagger$ are different. However, there are exceptions and these exceptions are very important.

**Definition 21** *An operator $\hat{A}$ is called* **Hermitean** *or* **self-adjoint** *if it is equal to its adjoint operator, i.e. if for all states $|x\rangle, |y\rangle$ we have*

$$\langle y|\hat{A}|x\rangle = \langle x|\hat{A}|y\rangle^* \quad . \tag{1.66}$$

In the finite dimensional case a self-adjoint operator is the same as a Hermitean operator. In the infinite-dimensional case Hermitean and self-adjoint are **not** equivalent.

The difference between self-adjoint and Hermitean is related to the domain of definition of the operators $\hat{A}$ and $\hat{A}^\dagger$ which need not be the same in the infinite dimensional case. In the following I will basically always deal with finite-dimensional systems and I will therefore usually use the term Hermitean for an operator that is self-adjoint.

### 1.3.3 Eigenvectors, Eigenvalues and the Spectral Theorem

Hermitean operators have a lot of nice properties and I will explore some of these properties in the following sections. Mostly these properties are concerned with the eigenvalues and eigenvectors of the operators. We start with the definition of eigenvalues and eigenvectors of a linear operator.

**Definition 22** *A linear operator $\hat{A}$ on an $N$-dimensional Hilbert space is said to have an eigenvector $|\lambda\rangle$ with corresponding eigenvalue $\lambda$ if*

$$\hat{A}|\lambda\rangle = \lambda|\lambda\rangle \quad , \tag{1.67}$$

*or equivalently*

$$(\hat{A} - \lambda\mathbb{1})|\lambda\rangle = 0 \quad . \tag{1.68}$$

This definition of eigenvectors and eigenvalues immediately shows us how to determine the eigenvectors of an operator. Because Eq. (1.68) implies that the $N$ columns of the operator $\hat{A} - \lambda\mathbb{1}$ are linearly dependent we need to have that

$$det(\hat{A} - \lambda\mathbb{1})) = 0 \quad . \tag{1.69}$$

This immediately gives us a complex polynomial of degree $N$. As we know from analysis, every polynomial of degree $N$ has exactly $N$ solutions if one includes the multiplicities of the eigenvalues in the counting. In general eigenvalues and eigenvectors do not have many restriction for an arbitrary $\hat{A}$. However, for Hermitean and unitary (will be defined soon) operators there are a number of nice results concerning the eigenvalues and eigenvectors. (If you do not feel familiar with eigenvectors and eigenvalues anymore, then a good book is for example: K.F. Riley, M.P. Robinson, and S.J. Bence, *Mathematical Methods for Physics and Engineering*.

We begin with an analysis of Hermitean operators. We find

**Lemma 23** *For any Hermitean operator $\hat{A}$ we have*

1. *All eigenvalues of $\hat{A}$ are* **real**.
2. *Eigenvectors to different eigenvalues are* **orthogonal**.

**Proof:** 1.) Given an eigenvalue $\lambda$ and the corresponding eigenvector $|\lambda\rangle$ of a Hermitean operator $\hat{A}$. Then we have using the hermiticity of $\hat{A}$

$$\lambda^* = \langle\lambda|\hat{A}|\lambda\rangle^* = \langle\lambda|\hat{A}^\dagger|\lambda\rangle = \langle\lambda|\hat{A}|\lambda\rangle = \lambda \tag{1.70}$$

which directly implies that $\lambda$ is real $\square$.

2.) Given two eigenvectors $|\lambda\rangle, |\mu\rangle$ for different eigenvalues $\lambda$ and $\mu$. Then we have

$$\lambda\langle\lambda|\mu\rangle = (\lambda\langle\mu|\lambda\rangle)^* = (\langle\mu|\hat{A}|\lambda\rangle)^* = \langle\lambda|\hat{A}|\mu\rangle = \mu\langle\lambda|\mu\rangle \tag{1.71}$$

As $\lambda$ and $\mu$ are different this implies $\langle\lambda|\mu\rangle = 0$. This finishes the proof $\square$.

Lemma 23 allows us to formulate the second postulate of quantum mechanics. To motivate the second postulate a little bit imagine that you try to determine the position of a particle. You would put down a coordinate system and specify the position of the particle as a set of **real** numbers. This is just one example and in fact in any experiment

in which we measure a quantum mechanical system, we will always obtain a real number as a measurement result. Corresponding to each outcome we have a state of the system (the particle sitting in a particular position), and therefore we have a set of real numbers specifying all possible measurement outcomes and a corresponding set of states. While the representation of the states may depend on the chosen basis, the physical position doesn't. Therefore we are looking for an object, that gives real numbers independent of the chosen basis. Well, the eigenvalues of a matrix are independent of the chosen basis. Therefore if we want to describe a physical observable its a good guess to use an operator and identify the eigenvalues with the measurement outcomes. Of course we need to demand that the operator has only real eigenvalues. Thus is guaranteed only by Hermitean operators. Therefore we are led to the following postulate.

**Postulate 2** *Observable quantum mechanical quantities are described by Hermitean operators $\hat{A}$ on the Hilbert space $\mathcal{H}$. The eigenvalues $a_i$ of the Hermitean operator are the possible measurement results.*

End of $6^{th}$ lecture

Now I would like to formulate the spectral theorem for finite dimensional Hermitean operators. Any Hermitean operator on an $N$-dimensional Hilbert space is represented by an $N \times N$ matrix. The characteristic polynomial Eq. (1.69) of such a matrix is of $N$-th order and therefore possesses $N$ eigenvalues, denoted by $\lambda_i$ and corresponding eigenvectors $|\lambda_i\rangle$. For the moment, let us assume that the eigenvalues are all different. Then we know from Lemma 23 that the corresponding eigenvectors are orthogonal. Therefore we have a set of $N$ pairwise orthogonal vectors in an $N$-dimensional Hilbert space. Therefore these vectors form an orthonormal basis. From this we can finally conclude the following important

**Completeness theorem:** For any Hermitean operator $\hat{A}$ on a Hilbert space $\mathcal{H}$ the set of all eigenvectors form an orthonormal basis

of the Hilbert space $\mathcal{H}$, i.e. given the eigenvalues $\lambda_i$ and the eigenvectors $|\lambda_i\rangle$ we find

$$\hat{A} = \sum_i \lambda_i |\lambda_i\rangle\langle\lambda_i| \tag{1.72}$$

and for any vector $|x\rangle \in \mathcal{H}$ we find coefficients $x_i$ such that

$$|x\rangle = \sum_i x_i |\lambda_i\rangle \quad . \tag{1.73}$$

Now let us briefly consider the case for degenerate eigenvalues. This is the case, when the characteristic polynomial has multiple zero's. In other words, an eigenvalue is degenerate if there is more than one eigenvector corresponding to it. An eigenvalue $\lambda$ is said to be $d$-fold degenerate if there is a set of $d$ linearly independent eigenvectors $\{|\lambda_1\rangle, \ldots, |\lambda_d\rangle\}$ all having the same eigenvalue $\lambda$. Quite obviously the space of all linear combinations of the vectors $\{|\lambda_1\rangle, \ldots, |\lambda_d\rangle\}$ is a $d$-dimensional vector space. Therefore we can find an orthonormal basis of this vector space. This implies that any Hermitean operator has eigenvalues $\lambda_1, \ldots, \lambda_k$ with degeneracies $d(\lambda_1), \ldots, d(\lambda_k)$. To each eigenvector$\lambda_i$ I can find an orthonormal basis of $d(\lambda_i)$ vectors. Therefore the above completeness theorem remains true also for Hermitean operators with degenerate eigenvalues.

Now you might wonder whether every linear operator $\hat{A}$ on an $N$ dimensional Hilbert space has $N$ linearly independent eigenvectors? It turns out that this is not true. An example is the $2 \times 2$ matrix

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

which has only one eigenvalue $\lambda = 0$. Therefore any eigenvector to this eigenvalue has to satisfy

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

which implies that $b = 0$. But then the only normalized eigenvector is $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and therefore the set of eigenvectors do not form an orthonormal

basis.

We have seen that any Hermitean operator $\hat{A}$ can be expanded in its eigenvectors and eigenvalues. The procedure of finding this particular representation of an operator is called **diagonalization**. Often we do not want to work in the basis of the eigenvectors but for example in the canonical basis of the vectors

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \ , \dots \ , e_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \ , \dots \ , e_N = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} . \tag{1.74}$$

If we want to rewrite the operator $\hat{A}$ in that basis we need to find a map between the canonical basis and the orthogonal basis of the eigenvectors of $\hat{A}$. If we write the eigenvectors $|\lambda_i\rangle = \sum_{j=1}^{N} \alpha_{ji} |e_j\rangle$ then this map is given by the **unitary operator** (we will define unitary operators shortly)

$$\hat{U} = \sum_{i=1}^{N} |\lambda_i\rangle\langle e_i| \doteq \begin{pmatrix} \alpha_{11} & \dots & \alpha_{1N} \\ \vdots & \ddots & \vdots \\ \alpha_{N1} & \dots & \alpha_{NN} \end{pmatrix} \tag{1.75}$$

which obviously maps a vector $|e_i\rangle$ into the eigenvector corresponding to the eigenvalue $|\lambda_i\rangle$. Using this operator $\hat{U}$ we find

$$\hat{U}^\dagger \hat{A} \hat{U} = \sum_i \lambda_i \hat{U}^\dagger |\lambda_i\rangle\langle\lambda_i| \hat{U} = \sum_i \lambda_i |e_i\rangle\langle e_i| \ . \tag{1.76}$$

The operator in Eq. (1.75) maps orthogonal vectors into orthogonal vectors. In fact, it preserves the scalar product between any two vectors. Let us use this as the defining property of a unitary transformation.

Now as promised the accurate definition of a unitary operator. Avoiding the subtleties of infinite dimensional spaces (which are again problems of the domain of definition of an operator) for the moment we have the following

**Definition 24** *A linear operator $\hat{U}$ on a Hilbert space $\mathcal{H}$ is called* **unitary** *if it is defined for all vectors $|x\rangle, |y\rangle$ in $\mathcal{H}$, maps the whole Hilbert space into the whole Hilbert space and satisfies*

$$\langle x|\hat{U}^\dagger \hat{U}|y\rangle = \langle x|y\rangle \quad . \tag{1.77}$$

In fact we can replace the last condition by demanding that the operator satisfies

$$\hat{U}^\dagger \hat{U} = \mathbb{1} \text{ and } \hat{U}\hat{U}^\dagger = \mathbb{1}. \tag{1.78}$$

Eq. (1.77) implies that a unitary operator preserves the scalar product and therefore in particular the norm of vectors as well as the angle between any two vectors.

Now let us briefly investigate the properties of the eigenvalues and eigenvectors of unitary operators. The eigenvalues of a unitary operators are in general not real, but they not completely arbitrary. We have

**Theorem 25** *Any* **unitary** *operator $\hat{U}$ on an $N$-dimensional Hilbert space $\mathcal{H}$ has a complete basis of eigenvectors and all the eigenvalues are of the form $e^{i\phi}$ with* **real** *$\phi$.*

**Proof:** I will not give a proof that the eigenvectors of $\hat{U}$ form a basis in $\mathcal{H}$. For this you should have a look at a textbook. What I will proof is that the eigenvalues of $\mathcal{H}$ are of the form $e^{i\phi}$ with **real** $\phi$. To see this, we use Eq. (1.78). Be $|\lambda\rangle$ an eigenvector of $\hat{U}$ to the eigenvalue $\lambda$, then

$$\lambda \hat{U}^\dagger |\lambda\rangle = \hat{U}^\dagger \hat{U}|\lambda\rangle = |\lambda\rangle \quad . \tag{1.79}$$

This implies that $\lambda \neq 0$ because otherwise the right-hand side would be the null-vector, which is never an eigenvector. From Eq. (1.79) we find

$$\frac{1}{\lambda} = \langle \lambda|\hat{U}^\dagger|\lambda\rangle = \langle \lambda|\hat{U}|\lambda\rangle^* = \lambda^* \quad . \tag{1.80}$$

This results in

$$|\lambda|^2 = 1 \Leftrightarrow \lambda = e^{i\phi} \quad . \tag{1.81}$$

□.

### 1.3.4 Functions of Operators

In the previous sections I have discussed linear operators and special subclasses of these operators such as Hermitean and unitary operators. When we write down the Hamilton operator of a quantum mechanical system we will often encounter functions of operators, such as the one-dimensional potential $V(\hat{x})$ in which a particle is moving. Therefore it is important to know the definition and some properties of functions of operators. There are two ways of defining functions on an operator, one works particularly well for operators with a complete set of eigenvectors (Definition 26), while the other one works bests for functions that can be expanded into power series (Definition 27).

**Definition 26** *Given an operator $\hat{A}$ with eigenvalues $a_i$ and a complete set of eigenvectors $|a_i\rangle$. Further have a function $f : \mathbf{C} \to \mathbf{C}$ that maps complex numbers into complex numbers then we define*

$$f(\hat{A}) := \sum_{i=1}^{N} f(a_i)|a_i\rangle\langle a_i| \tag{1.82}$$

**Definition 27** *Given a function $f : \mathbf{C} \to \mathbf{C}$ that can be expanded into a power series*

$$f(z) = \sum_{i=0}^{\infty} f_i z^i \tag{1.83}$$

*then we define*

$$f(\hat{A}) = \sum_{i=0}^{\infty} f_i A^i \quad . \tag{1.84}$$

**Definition 28** *The derivative of an operator function $f(\hat{A})$ is defined via $g(z) = \frac{df}{dz}(z)$ as*

$$\frac{df(\hat{A})}{d\hat{A}} = g(\hat{A}) \quad . \tag{1.85}$$

Let us see whether the two definitions Def. 26 and 27 coincide for operators with complete set of eigenvectors and functions that can be

expanded into a power series given in Eq. (1.83).

$$\begin{aligned} f(\hat{A}) &= \sum_{k=1}^{\infty} f_k A^k \\ &= \sum_{k=1}^{\infty} f_k (\sum_{j=1}^{N} a_j |a_j\rangle\langle a_j|)^k \\ &= \sum_{k=1}^{\infty} f_k \sum_{j=1}^{N} a_j^k |a_j\rangle\langle a_j| \\ &= \sum_{j=1}^{N} (\sum_{k=1}^{\infty} f_k a_j^k) |a_j\rangle\langle a_j| \\ &= \sum_{j=1}^{N} f(a_j) |a_j\rangle\langle a_j| \end{aligned} \tag{1.86}$$

For operators that do not have a complete orthonormal basis of eigenvectors of eigenvectors it is not possible to use Definition 26 and we have to go back to Definition 27. In practise this is not really a problem in quantum mechanics because we will always encounter operators that have a complete set of eigenvectors.

As an example consider a Hermitean operator $\hat{A}$ with eigenvalues $a_k$ and eigenvectors $|a_k\rangle$ and compute $\hat{U} = e^{i\hat{A}}$. We find

$$\hat{U} = e^{i\hat{A}} = \sum_{k=1}^{N} e^{ia_k} |a_k\rangle\langle a_k| \quad . \tag{1.87}$$

This is an operator which has eigenvalues of the form $e^{ia_k}$ with real $a_k$. Therefore it is a unitary operator, which you can also check directly from the requirement $\hat{U}\hat{U}^\dagger = \mathbb{1} = \hat{U}^\dagger\hat{U}$. In fact it is possible to show that every unitary operator can be written in the form Eq. (1.87). This is captured in the

**Lemma 29** *To any unitary operator $\hat{U}$ there is a Hermitean operator $\hat{H}$ such that*

$$\hat{U} = e^{i\hat{H}} \quad . \tag{1.88}$$

**Exercise:**
1) Show that for any unitary operator $\hat{U}$ we have $f(\hat{U}^\dagger \hat{A} \hat{U}) = \hat{U}^\dagger f(\hat{A}) \hat{U}$

**Proof:** We use the fact that $\hat{U}\hat{U}^\dagger = \mathbb{1}$ to find

$$\begin{aligned} f(\hat{U}^\dagger \hat{A} \hat{U}) &= \sum_{k=0}^{\infty} f_k (\hat{U}^\dagger \hat{A} \hat{U})^k \\ &= \sum_{k=0}^{\infty} f_k \hat{U}^\dagger \hat{A}^k \hat{U} \\ &= \hat{U}^\dagger (\sum_{k=0}^{\infty} f_k \hat{A}^k) \hat{U} \\ &= \hat{U}^\dagger f(\hat{A}) \hat{U} \quad . \end{aligned}$$

If you have functions, then you will also expect to encounter derivatives of functions. Therefore we have to consider how to take derivatives of matrices. To take a derivative we need to have not only one operator, but a family of operators that is parametrized by a real parameter $s$. An example is the set of operators of the form

$$\hat{A}(s) = \begin{pmatrix} 1+s & i \cdot s \\ -i \cdot s & 1-s \end{pmatrix} \quad . \tag{1.89}$$

Another example which is familiar to you is the time evolution operator $e^{-i\hat{H}s}$.

Now we can define the derivative with respect to $s$ in complete analogy to the derivative of a scalar function by

$$\frac{d\hat{A}}{ds}(s) := \lim_{\Delta s \to 0} \frac{\hat{A}(s+\Delta s) - \hat{A}(s)}{\Delta s} \quad . \tag{1.90}$$

This means that we have defined the derivative of an operator component wise.

Now let us explore the properties of the derivative of operators. First let us see what the derivative of the product of two operators is. We find

**Property:** For any two linear operators $\hat{A}(s)$ and $\hat{B}(s)$ we have

$$\frac{d(\hat{A}\hat{B})}{ds}(s) = \frac{d\hat{A}}{ds}(s)\hat{B}(s) + \hat{A}(s)\frac{d\hat{B}}{ds}(s) \tag{1.91}$$

This looks quite a lot like the product rule for ordinary functions, except that now the order of the operators plays a crucial role.

———End of $7^{th}$ lecture

You can also have functions of operators that depend on more than one variables. A very important example is the commutator of two operators.

**Definition 30** *For two operators $\hat{A}$ and $\hat{B}$ the commutator $[\hat{A}, \hat{B}]$ is defined as*

$$[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A} \quad . \tag{1.92}$$

While the commutator between numbers ($1 \times 1$ matrices) is always zero, this is not the case for general matrices. In fact, you know already a number of examples from your second year lecture in quantum mechanics. For example the operators corresponding to momentum and position do not commute, i.e. their commutator is nonzero. Other examples are the Pauli spin-operators

$$\begin{aligned} \sigma_0 &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ \sigma_2 &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \end{aligned} \tag{1.93}$$

For $i, j = 1, 2, 3$ they have the commutation relations

$$[\sigma_i, \sigma_j] = i\epsilon_{ijk}\sigma_k \quad . \tag{1.94}$$

where $\epsilon_{ijk}$ is the completely antisymmetric tensor. It is defined by $\epsilon_{123} = 1$ and changes sign, when two indices are interchanged, for example $\epsilon_{ijk} = -\epsilon_{jik}$.

There are some commutator relations that are quite useful to know.

**Lemma 31** *For arbitrary linear operators $\hat{A}, \hat{B}, \hat{C}$ on the same Hilbert space we have*

$$[\hat{A}\hat{B}, \hat{C}] = \hat{A}[\hat{B}, \hat{C}] + [\hat{A}, \hat{C}]\hat{B} \tag{1.95}$$

$$0 = [\hat{A}, [\hat{B}, \hat{C}]] + [\hat{B}, [\hat{C}, \hat{A}]] + [\hat{C}, [\hat{A}, \hat{B}]] \tag{1.96}$$

**Proof:** By direct inspection of the equations $\square$.

Commuting observables have many useful and important properties. Of particular significance for quantum physics is the following Lemma 32 because it guarantees that two commuting observables can be simultaneously measured with no uncertainty.

**Lemma 32** *Two commuting observables $\hat{A}$ and $\hat{B}$ have the same eigenvectors, i.e. they can be diagonalized simultaneously.*

**Proof:** For simplicity we assume that both observables have only non-degenerate eigenvalues. Now chose a basis of eigenvectors $\{|a_i\rangle\}$ that diagonalizes $\hat{A}$. Now try to see whether the $|a_i\rangle$ are also eigenvectors of $\hat{B}$. Using $[\hat{A}, \hat{B}]$ we have

$$\hat{A}(\hat{B}|a_i\rangle) = \hat{B}\hat{A}|a_i\rangle = a_i(\hat{B}|a_i\rangle) \quad . \tag{1.97}$$

This implies that $\hat{B}|a_i\rangle$ is an eigenvector of $\hat{A}$ with eigenvalue $a_i$. As the eigenvalue is non-degenerate we must have

$$\hat{B}|a_i\rangle = b_i|a_i\rangle \tag{1.98}$$

for some $b_i$. Therefore $|a_i\rangle$ is an eigenvector to $\hat{B}$ $\square$.

There is a very nice relation between the commutator and the derivative of a function. First I define the derivative of an operator function with respect to the operator.

**Lemma 33** *Given two linear operators $\hat{A}$ and $\hat{B}$ which have the commutator $[\hat{B}, \hat{A}] = \mathbb{1}$. Then for the derivative of an operator function $f(\hat{A})$ we find*

$$[\hat{B}, f(\hat{A})] = \frac{df}{d\hat{A}}(\hat{A}) \quad . \tag{1.99}$$

**Proof:** Remember that a function of an operator is defined via its expansion into a power series, see Eq. (1.84). Therefore we find

$$\begin{aligned} [\hat{B}, f(\hat{A})] &= [\hat{B}, \sum_k f_k \hat{A}^k] \\ &= \sum_k f_k [\hat{B}, \hat{A}^k] \end{aligned}$$

Now we need to evaluate the expression $[\hat{B}, \hat{A}^k]$. We proof by induction that $[\hat{B}, \hat{A}^n] = n\hat{A}^{n-1}$. For $n = 1$ this is true. Assume that the assumption is true for $n$. Now start with $n + 1$ and reduce it to the case for $n$. Using Eq. (1.95) we find

$$[\hat{B}, \hat{A}^{(n+1)}] \equiv [\hat{B}, \hat{A}^n \hat{A}] = [\hat{B}, \hat{A}^n]\hat{A} + \hat{A}^n[\hat{B}, \hat{A}] \quad . \tag{1.100}$$

Now using the $[\hat{B}, \hat{A}] = \mathbb{1}$ and the induction assumption we find

$$[\hat{B}, \hat{A}^{(n+1)}] = n\hat{A}^{n-1}\hat{A} + \hat{A}^n = (n+1)\hat{A}^n \quad . \tag{1.101}$$

Now we can conclude

$$\begin{aligned} [\hat{B}, f(\hat{A})] &= \sum_k f_k [\hat{B}, \hat{A}^k] \\ &= \sum_k f_k k \hat{A}^{k-1} \\ &= \frac{df}{d\hat{A}}(\hat{A}) \end{aligned}$$

This finishes proof.□

A very useful property is

**Lemma 34** (Baker-Campbell-Haussdorff) For general operators $\hat{A}$ and $\hat{B}$ we have

$$e^{\hat{B}} \hat{A} e^{-\hat{B}} = \hat{A} + [\hat{B}, \hat{A}] + \frac{1}{2}[\hat{B}, [\hat{B}, \hat{A}]] + \ldots \quad . \tag{1.102}$$

For operators such that $[\hat{B}, [\hat{B}, \hat{A}]] = 0$ we have the simpler version

$$e^{\hat{B}} \hat{A} e^{-\hat{B}} = \hat{A} + [\hat{B}, \hat{A}] \quad . \tag{1.103}$$

**Proof:** Define the function of one real parameter $\alpha$

$$f(\alpha) = e^{\alpha \hat{B}} \hat{A} e^{-\alpha \hat{B}} \quad . \tag{1.104}$$

We can expand this function around $\alpha = 0$ into a Taylor series $f(\alpha) = \sum_{n=0}^{\infty} \frac{\alpha^n}{n!} \frac{d^n f}{d\alpha^n}(\alpha)|_{\alpha=0}$ and therefore we need to determine the derivatives

of the function $f(\alpha)$. We find

$$\begin{aligned} \frac{df}{d\alpha}(\alpha)|_{\alpha=0} &= [\hat{B}, \hat{A}] \\ \frac{d^2 f}{d\alpha^2}(\alpha)|_{\alpha=0} &= [\hat{B}, [\hat{B}, \hat{A}]] \\ &\vdots \end{aligned}$$

The rest of the proof follows by induction. The proof of Eq. (1.103) follows directly from Eq. (1.102).□

## 1.4 Operators with continuous spectrum

In the preceding sections I have explained the basic ideas of Hilbert spaces and linear operators. All the objects I have been dealing with so far have been finite dimensional an assumption that greatly simplified the analysis and made the new ideas more transparent. However, in quantum mechanics many systems are actually described by an infinite dimensional state space and the corresponding linear operators are infinite dimensional too. Most results from the preceding section hold true also for the infinite dimensional case, so that we do not need to learn too many new things. Nevertheless, there are properties that require some discussion.

### 1.4.1 The position operator

The most natural place where an infinite dimensional state space appears is in the analysis of a particle in free space. Therefore let us briefly reconsider some aspects of wave mechanics as you have learnt them in the second year course. The state of a particle in free space (maybe moving in a potential) is described by the square-integrable wave-function $\psi(x)$. The question is now as to how we connect the wave-function notation with the Dirac notation which we have used to develop our theory of Hilbert spaces and linear operators.

Let us remember what the Dirac notation for finite dimensional systems means mathematically. Given the ket-vector $|\phi\rangle$ for the state

of a finite dimensional system, we can find the components of this ket-vector with respect to a certain basis $\{|e_i\rangle\}$. The i-th component is given by the complex number $\langle e_i|\phi\rangle$. Therefore in a particular basis it makes sense to write the state $|\phi\rangle$ as a column vector

$$|\phi\rangle \leftrightarrow \begin{pmatrix} \langle e_1|\phi\rangle \\ \vdots \\ \langle e_n|\phi\rangle \end{pmatrix} \quad . \tag{1.105}$$

Let us try to transfer this idea to the wave-function of a particle in free space. What we will do is to interpret $\psi(x)$ as the component of a vector with infinitely many components. Informally written this means

$$|\psi\rangle \leftrightarrow \begin{pmatrix} \vdots \\ \psi(x) \\ \vdots \end{pmatrix} \quad , \tag{1.106}$$

where we have given the column vector a name, $|\psi\rangle$. Obviously the set of vectors defined in this way form a vector space as you can easily check. Of course we would like to have a scalar product on this vector space. This is introduced in complete analogy to the finite dimensional case. There we had

$$(|\phi\rangle, |\psi\rangle) = \langle\phi|\psi\rangle = \sum_i \langle\phi|e_i\rangle\langle e_i|\psi\rangle \quad . \tag{1.107}$$

We just replace the summation over products of components of the two vectors by an integration. We have (see also Eq (1.25)

$$(|\phi\rangle, |\psi\rangle) := \int_{-\infty}^{\infty} dx\, \phi^*(x)\psi(x) \quad . \tag{1.108}$$

Now we have a space of vectors (or square integrable wave functions) $\mathcal{H}$ equipped with a scalar product. Indeed it turns out to be a complete space (without proof) so that we have a Hilbert space.

Now that we have introduced ket vectors we also need to define bra-vectors. In the finite dimensional case we obtained the bra vector via the idea of linear functionals on the space of state vectors. Let us

repeat this procedure now for the the case of infinite dimensions. We **define** a linear functional $\langle\phi|$ by

$$\langle\phi|(|\psi\rangle) \equiv \langle\phi|\psi\rangle = (|\phi\rangle, |\psi\rangle) \ . \tag{1.109}$$

Now I would like to investigate a particular ket vector (linear functional) that will allow as to define a position state. We **define** a linear functional $\langle x_0|$ by

$$\langle x_0|(|\psi\rangle) \equiv \langle x_0|\psi\rangle := \psi(x_0) \ . \tag{1.110}$$

We are already writing this functional very suggestively as a bra-vector. Thats perfectly ok, and we just have to check that the so defined functional is indeed linear. Of course we would like to interpret the left hand side of Eq. (1.110) as a scalar product between two ket's, i.e.

$$(|x_0\rangle, |\psi\rangle) := \langle x_0|\psi\rangle = \psi(x_0) \ . \tag{1.111}$$

What does the ket $|x_0\rangle$ corresponding to the bra $\langle x_0|$ mean? Which wave-function $\delta^*_{x_0}(x)$ does it correspond to? Using the scalar product Eq. (1.108), we have

$$\int_{-\infty}^{\infty} dx \delta^*_{x_0}(x)\psi(x) = (|x_0\rangle, |\psi\rangle) = \langle x_0|\psi\rangle = \psi(x_0) \tag{1.112}$$

This means that the function $\delta^*_{x_0}(x)$ has to act like a delta-function! The wave-function corresponding to the bra $\langle x_0|$ is a delta-function. A delta-function however, is **not** square-integrable! Therefore it cannot be an element of the Hilbert space of square integrable functions. However, as we have seen it would be quite convenient to use these wave-functions or states. Therefore we just add them to our Hilbert space, although we will often call them improper states or wave-functions. In fact we can use basically all the rules that we have learned about finite dimensional Hilbert-spaces also for these improper states. All we need to demand is the following rule for the scalar product

$$\langle\psi|x_0\rangle := (\langle x_0|\psi\rangle)^* = \psi^*(x_0) \ . \tag{1.113}$$

Now I can write for arbitrary kets $|\phi\rangle, |\psi\rangle \in \mathcal{H}$

$$\langle\phi|\psi\rangle = \int \phi^*(x)\psi(x)dx = \int \langle\phi|x\rangle\langle x|\psi\rangle dx = \langle\phi|(\int |x\rangle\langle x|dx)|\psi\rangle \ . \tag{1.114}$$

Then we can conclude

$$\int |x\rangle\langle x| dx = \mathbb{1} \ . \tag{1.115}$$

Inserting this identity operator in $\langle x|\psi\rangle$, we obtain the orthogonality relation between position kets

$$\begin{aligned} \int \delta(x-x')\psi(x')dx' = \psi(x) = \langle x|\psi\rangle &= \int \langle x|x'\rangle\langle x'|\psi\rangle dx' \\ &= \int \langle x|x'\rangle\psi(x')dx' \ . \end{aligned}$$

Therefore we have

$$\langle x|x'\rangle = \delta(x-x') \ . \tag{1.116}$$

Now we can derive the form of the position operator from our knowledge of the definition of the position expectation value

$$\begin{aligned} \langle\psi|\hat{x}|\psi\rangle &:= \int x|\psi(x)|^2 dx \\ &= \int \langle\psi|x\rangle x\langle x|\psi\rangle dx \\ &= \langle\psi|(\int x|x\rangle\langle dx)|\psi\rangle \ , \end{aligned} \tag{1.117}$$

where we defined the position operator

$$\hat{x} = \int x|x\rangle\langle x| dx = \hat{x}^\dagger \ . \tag{1.118}$$

Now you see why the improper position kets are so useful. In this basis the position operator is automatically diagonal. The improper position kets $|x_0\rangle$ are eigenvectors of the position operator

$$\hat{x}|x_0\rangle = x_0|x_0\rangle \ . \tag{1.119}$$

This makes sense, as the position kets describe a particle that is perfectly localized at position $x_0$. Therefore a position measurement should always give the result $x_0$. So far we have dealt with one-dimensional systems. All of the above considerations can be generalized to the d-dimensional case by setting

$$\hat{\vec{x}} = \hat{x}_1\vec{e}_1 + \ldots + \hat{x}_d\vec{e}_d \ . \tag{1.120}$$

The different components of the position operator commute are assumed to commute. ———End of $8^{th}$ lecture

### 1.4.2 The momentum operator

Now we are going to introduce the momentum operator and momentum eigenstates using the ideas of linear functionals in a similar fashion to the way in which we introduced the position operator. Let us introduce the linear functional $\langle p|$ defined by

$$\langle p|\psi\rangle := \frac{1}{\sqrt{2\pi\hbar}}\int e^{-ipx/\hbar}\psi(x)dx \quad . \tag{1.121}$$

Now we define the corresponding ket by

$$\langle p|\psi\rangle^* = \tilde{\psi}^*(p) =: \langle\psi|p\rangle \quad . \tag{1.122}$$

Combining Eq. (1.121) with the identity operator as represented in Eq. (1.115) we find

$$\frac{1}{\sqrt{2\pi\hbar}}\int e^{-ipx/\hbar}\psi(x)dx = \langle p|\psi\rangle = \int\langle p|x\rangle\langle x|\psi\rangle dx \tag{1.123}$$

Therefore we find that the state vector $|p\rangle$ represents a plane wave, because

$$\langle x|p\rangle = \frac{1}{\sqrt{2\pi\hbar}}e^{ipx/\hbar} \quad . \tag{1.124}$$

As Eq. (1.124) represents a plane wave with momentum $p$ it makes sense to call $|p\rangle$ a momentum state and expect that it is an eigenvector of momentum operator $\hat{p}$ to the eigenvalue $p$. Before we define the momentum operator, let us find the decomposition of the identity operator using momentum eigenstates. To see this we need to remember from the theory of delta-functions that

$$\frac{1}{2\pi\hbar}\int e^{ip(x-y)/\hbar}dp = \delta(x-y) \quad . \tag{1.125}$$

Then we have for arbitrary $|x\rangle$ and $|y\rangle$

$$\langle x|y\rangle = \delta(x-y) = \frac{1}{2\pi\hbar}\int dp e^{ip(x-y)/\hbar} = \langle x|(\int|p\rangle\langle p|dp)|y\rangle \quad , \tag{1.126}$$

and therefore

$$\int|p\rangle\langle p|dp = \mathbb{1} \quad . \tag{1.127}$$

The orthogonality relation between different momentum kets can be found by using Eq. (1.127) in Eq. (1.121).

$$\langle p|\psi\rangle = \langle p|\mathbb{1}|\psi\rangle = \int \langle p|p'\rangle\langle p'|\psi\rangle dp' \tag{1.128}$$

so that

$$\langle p|p'\rangle = \delta(p-p') \quad . \tag{1.129}$$

The momentum operator $\hat{p}$ is the operator that has as its eigenvectors the momentum eigenstates $|p\rangle$ with the corresponding eigenvalue $p$. This makes sense, because $|p\rangle$ describes a plane wave which has a perfectly defined momentum. Therefore we know the spectral decomposition which is

$$\hat{p} = \int p|p\rangle\langle p|dp \quad . \tag{1.130}$$

Clearly we have

$$\hat{p}|p_0\rangle = p_0|p_0\rangle \quad . \tag{1.131}$$

Analogously to the position operator we can extend the momentum operator to the d-dimensional space by

$$\hat{\vec{p}} = \hat{p}_1\vec{e}_1 + \ldots + \hat{p}_d\vec{e}_d \tag{1.132}$$

The different components of the momentum operator are assumed to commute.

### 1.4.3 The position representation of the momentum operator and the commutator between position and momentum

We have seen how to express the position operator in the basis of the improper position kets and the momentum operator in the basis of the improper momentum kets. Now I would like to see how the momentum operator looks like in the position basis.

To see this, differentiate Eq. (1.124) with respect to x which gives

$$\frac{\hbar}{i}\frac{\partial}{\partial x}\langle x|p\rangle = \frac{\hbar}{i}\frac{\partial}{\partial x}\frac{1}{\sqrt{2\pi\hbar}}e^{ipx/\hbar} = p\langle x|p\rangle \quad . \tag{1.133}$$

Therefore we find

$$\begin{aligned}\langle x|\hat{p}|\psi\rangle &= \int\langle x|p\rangle p\langle p|\psi\rangle dp \\ &= \frac{\hbar}{i}\frac{\partial}{\partial x}\int\langle x|p\rangle\langle p|\psi\rangle dp \\ &= \frac{\hbar}{i}\langle x|(\int|p\rangle\langle p|)|\psi\rangle \\ &= \frac{\hbar}{i}\frac{\partial}{\partial x}\langle x|\psi\rangle \quad . \end{aligned} \tag{1.134}$$

In position representation the momentum operator acts like the differential operator, i.e.

$$\hat{p} \longleftrightarrow \frac{\hbar}{i}\frac{\partial}{\partial x} \quad . \tag{1.135}$$

Knowing this we are now able to derive the commutation relation between momentum and position operator. We find

$$\begin{aligned}\langle x|[\hat{x},\hat{p}]|\psi\rangle &= \langle x|(\hat{x}\hat{p}-\hat{p}\hat{x}|\psi\rangle \\ &= \frac{\hbar}{i}\left[x\frac{\partial}{\partial x}\langle x|\psi\rangle - \frac{\partial}{\partial x}(x\langle x|\psi\rangle)\right] \\ &= i\hbar\langle x|\psi\rangle \\ &= \langle x|i\hbar\mathbb{1}|\psi\rangle\end{aligned}$$

Therefore we have the Heisenberg commutation relations

$$[\hat{x},\hat{p}] = i\hbar\mathbb{1} \quad . \tag{1.136}$$

# Chapter 2

# Quantum Measurements

So far we have formulated two postulates of quantum mechanics. The second of these states that the measurement results of an observable are the eigenvalues of the corresponding Hermitean operator. However, we do not yet know how to determine the probability with which this measurement result is obtained, neither have we discussed the state of the system after the measurement. This is the object of this section.

## 2.1 The projection postulate

We have learned, that the possible outcomes in a measurement of a physical observable are the eigenvalues of the corresponding Hermitean operator. Now I would like to discuss what the state of the system after such a measurement result is. Let us guide by common sense. Certainly we would expect that if we repeat the measurement of the same observable, we should find the same result. This is in contrast to the situation before the first measurement. There we certainly did not expect that a particular measurement outcome would appear with certainty. Therefore we expect that the state of the system after the first measurement has changed as compared to the state before the first measurement. What is the most natural state that corresponds to the eigenvalue $a_i$ of a Hermitean operator $\hat{A}$? Clearly this is the corresponding eigenvector in which the observable has the definite value $a_i$! Therefore it is quite natural to make the following postulate for

observables with non-degenerate eigenvalues.

**Postulate 3 (a)** *The state of a quantum mechanical system after the measurement of observable $\hat{A}$ with the result being the non-degenerated eigenvalue $a_i$ is given by the corresponding eigenvector $|a_i\rangle$.*

For observables with degenerate eigenvalues we have a problem. Which of the eigenvalues should we chose? Does it make sense to chose one particular eigenvector? To see what the right answer is we have to make clear that we are really only measuring observable $\hat{A}$. We do not obtain any other information about the measured quantum mechanical system. Therefore it certainly does **not** make sense to prefer one over another eigenvector. In fact, if I would assume that we have to chose one of the eigenvectors at random, then I would effectively assume that some more information is available. Somehow we must be able to decide which eigenvector has to be chosen. Such information can only come from the measurement of another observable - a measurement we haven't actually performed. This is therefore not an option. On the other hand we could say that we chose one of the eigenvectors at random. Again this is not really an option, as it would amount to saying that someone chooses an eigenvector and but does not reveal to us which one he has chosen. It is quite important to realize, that a) not having some information and b) having information but then choosing to forget it are two quite different situations.

All these problems imply that we have to have all the eigenvectors still present. If we want to solve this problem, we need to introduce a new type of operator - the **projection operator**.

**Definition 35** *An operator $\hat{P}$ is called a* **projection** *operator if it satisfies*

1. $\hat{P} = \hat{P}^\dagger$ ,
2. $\hat{P} = \hat{P}^2$ .

Some examples for projection operators are

1. $\hat{P} = |\psi\rangle\langle\psi|$

2. If $\hat{P}$ is a projection operator, then also $\mathbb{1} - \hat{P}$ is a projection operator.

3. $\hat{P} = \mathbb{1}$

**Exercise:** Prove that the three examples above are projection operators!

**Lemma 36** *The eigenvalues of a projection operator can only have the values* 0 *or* 1.

**Proof:** For any eigenvector $|\lambda\rangle$ of $\lambda$ we have

$$\lambda|\lambda\rangle = \hat{P}|\lambda\rangle = \hat{P}^2|\lambda\rangle = \lambda^2|\lambda\rangle \quad . \tag{2.1}$$

From this we immediately obtain $\lambda = 0$ or $\lambda = 1$ $\square$.

For a set of orthonormal vectors $\{|\psi_i\rangle\}_{i=1,\ldots,N}$ a projection operator $\hat{P} = \sum_{i=1}^{k} |\psi_i\rangle\langle\psi_i|$ projects a state $|\psi\rangle$ onto the subspace spanned by the vectors $\{|\psi_i\rangle\}_{i=1,\ldots,k}$. In mathematics this statement is more clear. If we expand $|\psi\rangle = \sum_{i=1}^{N} \alpha_i|\psi_i\rangle$ then

$$\hat{P}|\psi\rangle = \sum_{i=1}^{k} \alpha_i|\psi_i\rangle \quad . \tag{2.2}$$

This is exactly what we need for a more general formulation of the third postulate of quantum mechanics. We formulate the third postulate again, but now for observables with degenerate eigenvalues.

**Postulate 3 (b)** *The state of a quantum mechanical system after the measurement of general observable $\hat{A}$ with the result being the possibly degenerated eigenvalue $a_i$ is given by*

$$\hat{P}_i|\psi\rangle \quad . \tag{2.3}$$

*where $\hat{P}_i$ is the projection operator on the subspace of $\mathcal{H}$ spanned by all the eigenvectors of $\hat{A}$ with eigenvalue $a_i$, i.e.*

$$\hat{P}_i = \sum_i |\psi_i\rangle\langle\psi_i| \quad . \tag{2.4}$$

Experiments have shown that this postulate describes the results of measurements extremely well and we therefore have to accept it at least as a good working assumption. I say this because there are quite a few people around who do not like this postulate very much. They ask the obvious question as to how and when the reduction of the quantum state is exactly taking place. In fact, we have not answered any of that in this discussion and I will not do so for a simple reason. **Nobody** really knows the answer to these questions and they are a subject of debate since a long time. People have come up with new interpretations of quantum mechanics (the most prominent one being the many-worlds theory) or even changes to the Schrödinger equations of quantum mechanics themselves. But nobody has solved the problem satisfactorily. This is a bit worrying because this means that we do not understand an important aspect of quantum mechanics. However, there is a really good reason for using the projection postulate: **It works superbly well when you want to calculate things**.

———End of $9^{th}$ lecture

Therefore let us now continue with our exploration of the postulates of quantum mechanics. We now know what the measurement outcomes are and what the state of our system after the measurement is, but

we still do not know what the probability for a specific measurement outcome is. To make any sensible guess about this, we need to consider the properties of probabilities to see which of these we would like to be satisfied in quantum mechanics. Once we know what are sensible requirements, then we can go ahead to make a new postulate.

As you know, probabilities $p(A_i)$ to obtain an outcome $A_i$ (which is usually a set of possible results) are positive quantities that are not larger than unity, i.e. $0 \leq p(A_i) \leq 1$. In addition it is quite trivial to demand that the probability for an outcome corresponding to an empty set vanishes, i.e. $p(\emptyset)$, while the probability for the set of all elements is unity, i.e. $p(\mathbb{1})$.

These are almost trivial requirements. Really important is the behaviour of probabilities for joint sets. What we definitively would like to have is that the probabilities for mutually exclusive events add up, i.e. if we are given disjoint sets $A_1$ and $A_2$ and we form the union $A_1 \cup A_2$ of the two sets, then we would like to have

$$p(A_1 \cup A_2) = p(A_1) + p(A_2) \quad . \tag{2.5}$$

In the following postulate I will present a definition that satisfies all of the properties mentioned above.

**Postulate 4** The probability of obtaining the eigenvalue $a_i$ in a measurement of the observable $\hat{A}$ is given by

$$p_i = ||\hat{P}_i|\psi\rangle||^2 = \langle\psi|\hat{P}_i|\psi\rangle \quad . \tag{2.6}$$

For a non-degenerate eigenvalue with eigenvector $|\psi_i\rangle$ this probability reduces to the well known expression

$$p_i = |\langle\psi_i|\psi\rangle|^2 \tag{2.7}$$

It is easy to check that Postulate 4 indeed satisfies all the criteria that we demanded for a probability. The fascinating thing is, that it is essentially the only way of doing so. This very important theorem was

first proved by Gleason in 1957. This came as quite a surprise. Only using the Hilbert space structure of quantum mechanics together with the reasonable properties that we demanded from the probabilities for quantum mechanical measurement outcomes we can't do anything else than using Eq. (2.6)! The proof for this theorem is too complicated to be presented here and I just wanted to justify the Postulate 4 a bit more by telling you that we cannot really postulate anything else.

## 2.2 Expectation value and variance.

Now that we know the measurements outcomes as well as the probabilities for their occurrence, we can use quantum mechanics to calculate the mean value of an observable as well as the variance about this mean value.

What do we do experimentally to determine an mean value of an observable. First, of course we set up our apparatus such that it measures exactly the observable $\hat{A}$ in question. Then we build a machine that creates a quantum system in a given quantum state over and over again. Ideally we would like to have infinitely many of such identically prepared quantum states (this will usually be called an **ensemble**). Now we perform our experiment. Our state preparer produces the first particle in a particular state $|\psi\rangle$ and the particle enters our measurement apparatus. We obtain a particular measurement outcome, i.e. one of the eigenvalues $a_i$ of the observable $\hat{A}$. We repeat the experiment and we find a different eigenvalue. Every individual outcome of the experiment is completely random and after $N$ repetitions of the experiment, we will obtain the eigenvalue $a_i$ as a measurement outcome in $N_i$ of the experiments, where $\sum_i N_i = N$. After finishing our experiment, we can then determine the average value of the measurement outcomes, for which we find

$$\langle \hat{A} \rangle_\psi(N) = \sum_i \frac{N_i}{N} a_i \quad . \tag{2.8}$$

For large numbers of experiments, i.e. large $N$, we know that the ratio $\frac{N_i}{N}$ approaches the probability $p_i$ for the outcome $a_i$. This probability

can be calculated from Postulate 4 and we find

$$\langle \hat{A} \rangle_\psi = \lim_{N\to\infty} \sum_i \frac{N_i}{N} a_i = \sum_i p_i a_i = \sum_i a_i \langle \psi | \hat{P}_i | \psi \rangle = \langle \psi | \hat{A} | \psi \rangle \quad . \tag{2.9}$$

You should realize that the left hand side of Eq. (2.9) is conceptually quite a different thing from the right hand side. The left hand side is an experimentally determined quantity, while the right hand side is a quantum mechanical prediction. This is a testable prediction of Postulate 3 and has been verified in countless experiments. Now we are also able to derive the quantum mechanical expression for the variance of the measurement results around the mean value. Given the probabilities $p_i$ and the eigenvalues $a_i$ of the observable $\hat{A}$ we find

$$(\Delta A)^2 := \sum_i p_i (a_i - \langle A \rangle)^2 = \sum_i p_i a_i^2 - (\sum_i p_i a_i)^2 = \langle \hat{A}^2 \rangle - \langle \hat{A} \rangle^2 \quad . \tag{2.10}$$

After these definitions let us now analyse a very important property of measurement results in quantum mechanics.

## 2.3 Uncertainty Relations

In classical mechanics we are able to measure any physical quantity to an arbitrary precision. In fact, this is also true for quantum mechanics! Nothing prevents us from preparing a quantum system in a very well defined position state and subsequently make a measurement that will tell us with exceeding precision where the particle is. The real difference to classical mechanics arise when we try to determine the values of two different observables simultaneously. Again, in classical mechanics we have no law that prevents us from measuring any two quantities with arbitrary precision. Quantum mechanics, however, is different. Only very special pairs of observables can be measured simultaneously to an arbitrary precision. Such observables are called compatible or commuting observables. In general the uncertainties in the measurements of two arbitrary observables will obey a relation which makes sure that their product has a lower bound which is in general is unequal to zero. This relation is called the **uncertainty relation**.

**Theorem 37** *For any two observables $\hat{A}$ and $\hat{B}$ we find for their uncertainties $\Delta\hat{X} = \sqrt{\langle\hat{X}^2\rangle - \langle\hat{X}\rangle^2}$ the* **uncertainty relation**

$$\Delta A \Delta B \geq \frac{|\langle[\hat{A},\hat{B}]\rangle|}{2} \quad . \tag{2.11}$$

**Proof:** Let us define two ket's

$$|\phi_A\rangle = \left(\hat{A} - \langle\hat{A}\rangle\right)|\psi\rangle \tag{2.12}$$

$$|\phi_B\rangle = \left(\hat{B} - \langle\hat{B}\rangle\right)|\psi\rangle \; . \tag{2.13}$$

Using these vectors we find that

$$\sqrt{|\langle\phi_A|\phi_A\rangle| \cdot |\langle\phi_B|\phi_B\rangle|} = \Delta A \Delta B \; . \tag{2.14}$$

Now we can use the Schwarz inequality to find a lower bound on the product of the two uncertainties. We find that

$$|\langle\phi_A|\phi_B\rangle| \leq \sqrt{|\langle\phi_A|\phi_A\rangle| \cdot |\langle\phi_B|\phi_B\rangle|} = \Delta A \Delta B \; . \tag{2.15}$$

For the real and imaginary parts of $|\langle\phi_A|\phi_B\rangle|$ we find

$$Re\langle\phi_A|\phi_B\rangle = \frac{1}{2}\left(\langle\hat{A}\hat{B} + \hat{B}\hat{A}\rangle - 2\langle\hat{A}\rangle\langle\hat{B}\rangle\right) \tag{2.16}$$

$$Im\langle\phi_A|\phi_B\rangle = \frac{1}{2i}\langle[\hat{A},\hat{B}]\rangle \; . \tag{2.17}$$

We can use Eqs. (2.16-2.17) in Eq. (2.15) to find

$$\Delta A \Delta B \geq \sqrt{(Re\langle\phi_A|\phi_B\rangle)^2 + (Im\langle\phi_A|\phi_B\rangle)^2} \geq \frac{1}{2}|\langle[\hat{A},\hat{B}]\rangle| \tag{2.18}$$

This finishes the proof $\square$.

From Eq. (2.11) we can see that a necessary condition for two observables to be measurable simultaneously with no uncertainty is that the two observables commute. In fact this condition is also a sufficient one. We have

**Theorem 38** *Two observables $\hat{A}$ and $\hat{B}$ can be measured precisely, i.e. $\Delta A = \Delta B = 0$, exactly if they commute.*

**Proof:** From Eq. (2.11) we immediately see that $\Delta A = \Delta B = 0$ implies that $[\hat{A}, \hat{B}] = 0$. For the other direction of the statement we need to remember that two commuting observables have the same eigenvectors. If we assume our quantum mechanical system is in one of these eigenstates $|\psi\rangle$, then we see that in Eqs. (2.12-2.13) that $|\phi_A\rangle$ and $|\phi_B\rangle$ are proportional to $|\psi\rangle$ and therefore proportional to each other. Then we only need to remember that in the case of proportional $|\phi_A\rangle$ and $|\phi_B\rangle$ we have equality in the Schwarz inequality which implies $\Delta A = \Delta B = \frac{1}{2}|\langle[\hat{A}, \hat{B}]\rangle| = 0$ $\square$.

Now let us make clear what the uncertainty relation means, as there is quite some confusion about that. Imagine that we are having an ensemble of infinitely many identically prepared quantum systems, each of which is in state $|\psi\rangle$. We would like to measure two observables $\hat{A}$ and $\hat{B}$. Then we split the ensemble into two halves. On the **first half** we measure exclusively the observable $\hat{\mathbf{A}}$ while on the **second half** we measure the observable $\hat{\mathbf{B}}$. The measurement of the two observables will lead to average values $\langle\hat{A}\rangle$ and $\langle\hat{B}\rangle$ which have uncertainties $\Delta A$ and $\Delta B$. From this consideration it should be clear, that the uncertainties in the measurements of the two observables are **not** caused by a perturbation of the quantum mechanical system by the measurement as we are measuring the two observables on different systems which are in the same state. These uncertainties are an **intrinsic property of any quantum state**. Of course perturbations due to the measurement itself may increase the uncertainty but it is **not** the reason for the existence of the uncertainty relations.

In that context I would also like to point out, that you will sometimes find the uncertainty relation stated in the form

$$\Delta A \Delta B \geq |\langle[\hat{A}, \hat{B}]\rangle| \ , \tag{2.19}$$

that is, where the right-hand side is twice as large as in Eq. (2.11). The reason for this extra factor of 2 is that the uncertainty relation Eq. (2.19) describes a different situation than the one stated in Theorem 37. Eq. (2.19) really applies to the simultaneous measurement of two non-commuting observables on **one** quantum system. This means that we perform a measurement of the observable $\hat{A}$ on every system of the ensemble and subsequently we perform a measurement of the observable

$\hat{B}$ on every system of the ensemble. The measurement of the observable $\hat{A}$ will disturb the system and a subsequent measurement of $\hat{B}$ could therefore be more uncertain. Imagine, for example, that you want to determine the position of an electron and simultaneously the momentum. We could do the position measurement by scattering light from the electron. If we want to determine the position of the electron very precisely, then we need to use light of very short wavelength. However, photons with a short wavelength carry a large momentum. When they collide with the electron, then the momentum of the electron may be changed. Therefore the uncertainty of the momentum of the electron will be larger. This is an example for the problems that arise when you want to measure two non-commuting variables simultaneously.

If you would like to learn more about the uncertainty relation for the joint measurement of non-commuting variables you may have a look at the pedagogical article by M.G. Raymer, Am. J. Phys. **62**, 986 (1994) which you can find in the library.

### 2.3.1 The trace of an operator

In the fourth postulate I have defined the rule for the calculation of the probability that a measurement result is one of the eigenvalues of the corresponding operator. Now I would like to generalize this idea to situations in which we have some form of lack of knowledge about the measurement outcome. An example would be that we are not quite sure which measurement outcome we have (the display of the apparatus may have a defect for example). To give this law a simple formulation, I need to introduce a new mathematical operation.

**Definition 39** *The* **trace** *of an operator $\hat{A}$ on an $N$ dimensional Hilbert space is defined as*

$$tr(\hat{A}) = \sum_{i=1}^{N} \langle \psi_i | \hat{A} | \psi_i \rangle \tag{2.20}$$

*for any orthonormal set of basis vectors $\{|\psi_i\rangle\}$.*

In particular, if the $|\psi_i\rangle$ are chosen as the eigenvectors of $\hat{A}$ we see that the trace is just the sum of all the eigenvalues of $\hat{A}$. The trace will

play a very important role in quantum mechanics which we will see in our first discussion of the measurement process in quantum mechanics.

The trace has quite a few properties that are helpful in calculations. Note that the trace is only well defined, if it is independent of the choice of the orthonormal basis $\{|\psi_i\rangle\}$. If we chose a different basis $\{|\phi_i\rangle\}$, then we know that there is a unitary operator $\hat{U}$ such that $|\psi_i\rangle = \hat{U}|\phi_i\rangle$ and we want the property $\mathrm{tr}\{\hat{A}\} = \mathrm{tr}\{\hat{U}^\dagger \hat{A}\hat{U}\}$. That this is indeed true shows

**Theorem 40** For any two operators $\hat{A}$ and $\hat{B}$ on a Hilbert space $\mathcal{H}$ and unitary operators $\hat{U}$ we have

$$\begin{aligned} tr\{\hat{A}\hat{B}\} &= tr\{\hat{B}\hat{A}\} && (2.21)\\ tr\{\hat{A}\} &= tr\{\hat{U}\hat{A}\hat{U}^\dagger\} \ . && (2.22) \end{aligned}$$

**Proof:** I prove only the first statement, because the second one follows directly from $\hat{U}^\dagger\hat{U} = \mathbb{1}$ and Eq. (2.21). The proof runs as follows

$$\begin{aligned} tr\{\hat{A}\hat{B}\} &= \sum_{n=1}^{N}\langle\psi_n|\hat{A}\hat{B}|\psi_n\rangle \\ &= \sum_{n=1}^{N}\langle\psi_n|\hat{A}\mathbb{1}\hat{B}|\psi_n\rangle \\ &= \sum_{n=1}^{N}\langle\psi_n|\hat{A}\sum_{k=1}^{N}|\psi_k\rangle\langle\psi_k|\hat{B}|\psi_n\rangle \\ &= \sum_{k=1}^{N}\sum_{n=1}^{N}\langle\psi_n|\hat{A}|\psi_k\rangle\langle\psi_k|\hat{B}|\psi_n\rangle \\ &= \sum_{k=1}^{N}\sum_{n=1}^{N}\langle\psi_k|\hat{B}|\psi_n\rangle\langle\psi_n|\hat{A}|\psi_k\rangle \\ &= \sum_{k=1}^{N}\langle\psi_k|\hat{B}\sum_{n=1}^{N}|\psi_n\rangle\langle\psi_n|\hat{A}|\psi_k\rangle \\ &= \sum_{k=1}^{N}\langle\psi_k|\hat{B}\mathbb{1}\hat{A}|\psi_k\rangle \\ &= \sum_{k=1}^{N}\langle\psi_k|\hat{B}\hat{A}|\psi_k\rangle \\ &= tr\{\hat{B}\hat{A}\} \ . \end{aligned}$$

**Comment:** If you write an operator in matrix form then the trace is just the sum of the diagonal elements of the matrix.

I introduced the trace in order to write some of the expressions from the previous chapters in a different form which then allow their generalization to situations which involve classical uncertainty.

In postulate 4 the probability to find eigenvalue $a_i$ in a measurement of the observable $\hat{A}$ has been given for a system in quantum state $|\psi\rangle$. Using the trace it can now be written

$$p_i = \langle\psi|\hat{P}_i|\psi\rangle = tr\{\hat{P}_i|\psi\rangle\langle\psi|\} \quad .$$

Using the trace, the expectation value of an observable $\hat{A}$ measured on a system in a quantum state $|\psi\rangle$ can be written as

$$\langle\hat{A}\rangle_{|\psi\rangle} = \langle\psi|\hat{A}|\psi\rangle = tr\{\hat{A}|\psi\rangle\langle\psi|\} \quad .$$

## 2.4 The density operator

When discussing measurement theory it is quite natural to introduce a new concept for the description of the state of quantum system. So far we have always described a quantum system by a state vector, i.e. by a pure state which may be a **coherent** superposition of many other state vectors, e.g.

$$|\psi\rangle = \alpha_1|\psi_1\rangle + \alpha_2|\psi_2\rangle \quad . \tag{2.23}$$

A quantum mechanical system in such a state is ideal for exhibiting quantum mechanical interference. However, it has to be said that such a state is an idealization as compared to the true state of a quantum system that you can prepare in an experiment. In an experiment you will always have some statistical uncertainty due to imperfections in the preparation procedure of the quantum state or the measurement. For example, due to an occasional error, happening at a random time and unknown to you, the wrong quantum state may be prepared. Then we do not only have the quantum mechanical uncertainty that is described by the state vector, but we also have a **classical** statistical uncertainty. This raises the question as to how to describe such an

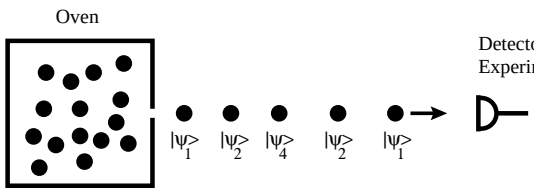


Figure 2.1: An oven emits atomic two-level systems. The internal state of the system is randomly distributed. With probability $p_i$ the system is in the pure state $|\psi_i\rangle$. A person oblivious to this random distribution measures observable $\hat{A}$. What is the mean value that he obtains?

experimental situation in the most elegant way. Certainly it **cannot** be dealt with by just adding the state vectors together, i.e. forming a coherent superposition.

To understand this better and to clarify the definition of the **density operator** I will present an example of an experimental situation where the description using pure quantum states fails, or better, where it is rather clumsy. Consider the situation presented in Fig. 2.1. An oven is filled with atomic two-level systems. We assume that each of these two-level systems is in a random pure state. To be more precise, let us assume that with probability $p_i$ a two-level system is in state $|\psi_i\rangle$. Imagine now that the oven has a little hole in it through which the two-level atoms can escape the oven in the form of an atomic beam. This atomic beam is travelling into a measurement apparatus built by an experimentalist who would like to measure the observable $\hat{A}$. The task of a theoretical physicist is to predict what the experimentalist will measure in his experiment.

We realize that each atom in the atomic beam is in a pure state $|\psi_i\rangle$ with probability $p_i$. For each individual atom it is unknown to the experimentalist in which particular state it is. He only knows the probability distribution of the possible states. If the experimentalist makes measurements on $N$ of the atoms in the beam, then he will perform the measurement $N_i \approx N p_i$ times on an atom in state $|\psi_i\rangle$. For each of these pure states $|\psi_i\rangle$ we know how to calculate the expectation value of the observable $\hat{A}$ that the experimentalist is measuring. It is simply $\langle\psi_i|\hat{A}|\psi_i\rangle = tr\{\hat{A}|\psi_i\rangle\langle\psi_i|\}$. What average value will the experimentalist see in N measurements? For a large $N$ the relative frequencies of oc-

currence of state $|\psi_i\rangle$ is $N_i/N = p_i$. Therefore the mean value observed by the experimentalist is

$$\langle \hat{A} \rangle = \sum_i p_i tr\{\hat{A}|\psi_i\rangle\langle\psi_i|\} \quad . \tag{2.24}$$

This equation is perfectly correct and we are able to calculate the expectation value of any observable $\hat{A}$ for any set of states $\{|\psi_i\rangle\}$ and probabilities $\{p_i\}$. However, when the number of possible states $|\psi_i\rangle$ is really large then we have a lot of calculations to do. For each state $|\psi_i\rangle$ we have to calculate the expectation value $\langle\psi_i|\hat{A}|\psi_i\rangle = tr\{\hat{A}|\psi_i\rangle\langle\psi_i|\}$ and then sum up all the expectation values with their probabilities. If we then want to measure a different observable $\hat{B}$ then we need to do the same lengthy calculation again. That's not efficient at all! As a theoretical physicist I am of course searching for a better way to do this calculation. Therefore let us reconsider Eq. (2.24), transforming it a bit and use it to define the **density operator**.

$$\begin{aligned} \langle \hat{A} \rangle &= \sum_i p_i tr\{\hat{A}|\psi_i\rangle\langle\psi_i|\} \quad . \\ &= tr\{\hat{A} \sum p_i |\psi_i\rangle\langle\psi_i|\} \\ &=: tr\{\hat{A}\hat{\rho}\} \quad . \end{aligned} \tag{2.25}$$

The last equality is really the definition of the **density operator** $\hat{\rho}$. Quite obviously, if I know the density operator $\hat{\rho}$ for the specific situation - here the oven generating the atomic beam - then I can calculate quite easily the expectation value that the experimentalist will measure. If the experimentalist changes his apparatus and now measures the observable $\hat{B}$ then this is not such a big deal anymore. We just need to re-evaluate Eq. (2.25) replacing $\hat{A}$ by $\hat{B}$ and not the potentially huge number of expectation values $tr\{\hat{A}|\psi_i\rangle\langle\psi_i|\}$.

**Exercise: Check that in general we have $\langle f(\hat{A})\rangle = tr\{f(\hat{A})\hat{\rho}\}$.**

Before I present some useful properties of a density operator, let me stress again the difference between a **coherent superposition** of quantum states and a **statistical mixture** of quantum states. In a coherent superposition state of a quantum mechanical system, each

representative of the ensemble is in the same **pure state** of the form

$$|\psi\rangle = \sum_i \alpha_i |\psi_i\rangle \quad , \tag{2.26}$$

which can be written as the density operator

$$\rho = |\psi\rangle\langle\psi| = \sum_{i,j} \alpha_i \alpha_j^* |\psi_i\rangle\langle\psi_j| \; . \tag{2.27}$$

This is completely different to a **statistical mixture** where a representative is with a certain probability $|\alpha_i|^2$ in a pure state $|\psi_i\rangle$! The corresponding density operator is

$$\hat{\rho} = \sum_i |\alpha_i|^2 |\psi_i\rangle\langle\psi_i| \quad . \tag{2.28}$$

and is also called a **mixed state**. You should convince yourself that indeed the two states Eqs. (2.26-2.28) are different, i.e.

$$\hat{\rho} = \sum_i |\alpha_i|^2 |\psi_i\rangle\langle\psi_i| \neq |\psi\rangle\langle\psi| \quad . \tag{2.29}$$

Now let us consider some properties of the density operator which can in fact be used as an alternative definition of the density operator.

**Theorem 41** *Any density operator satisfies*

1. *$\hat{\rho}$ is a Hermitean operator.*
2. *$\hat{\rho}$ is a positive semidefinite operator, i.e. $\forall|\psi\rangle : \langle\psi|\hat{\rho}|\psi\rangle \geq 0$.*
3. *$tr\{\hat{\rho}\} = 1$*

**Proof:** As we have defined the density operator already we need to proof the above theorem. From our definition of the density operator we find

$$\hat{\rho}^\dagger = (\sum_i p_i |\psi_i\rangle\langle\psi_i|)^\dagger = \sum_i p_i |\psi_i\rangle\langle\psi_i| = \hat{\rho} \; . \tag{2.30}$$

This followed because $|\psi_i\rangle\langle\psi_i|$ is a projection operator. This proves part 1). Part 2) follows by

$$\langle\psi|\hat{\rho}|\psi\rangle = \langle\psi|(\sum_i p_i |\psi_i\rangle\langle\psi_i|)|\psi\rangle = \sum_i p_i \langle\psi|\psi_i\rangle\langle\psi_i|\psi\rangle = \sum_i p_i |\langle\psi|\psi_i\rangle|^2 \geq 0 \quad . \tag{2.31}$$

The last property follows from the fact that probabilities are normalized, i.e. $\sum_i p_i = 1$. Then we can see that

$$tr\{\hat{\rho}\} = tr\{\sum_i p_i |\psi_i\rangle\langle\psi_i|\} = \sum_i p_i tr\{|\psi_i\rangle\langle\psi_i|\} = \sum_i p_i = 1 \quad . \tag{2.32}$$

This finishes the proof $\Box$.

Some simple but useful properties that can be derived from Theorem 41 are

1. $\hat{\rho}^2 = \hat{\rho} \Longleftrightarrow \hat{\rho}$ is a pure state
2. All eigenvalues of $\hat{\rho}$ lie in the interval $[0, 1]$.

**Proof:** Exercise!

A natural question that arises when one considers the density operator, is that of its decomposition into pure states, i.e. $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$. Is this decomposition unique, or are there many possible ways to obtain a given density operator as a statistical mixture of pure states? The answer can be found by looking at a particularly simple example. Let us consider the density operator representing the 'completely mixed' state of a two-level system, i.e.

$$\hat{\rho} = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) \quad . \tag{2.33}$$

Looking at Eq. (2.33) we readily conclude that this density operator can be generated by an oven sending out atoms in states $|0\rangle$ or $|1\rangle$ with a probability of 50% each, see part a) of Fig . (2.2). That conclusion is perfectly correct, however, we could also imagine an oven that generates atoms in states $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$ with a probability of 50% each. Let us check that this indeed gives rise to the same density operator.

$$\begin{aligned}
\hat{\rho} &= \frac{1}{2}(|+\rangle\langle +| + |-\rangle\langle -|) \\
&= \frac{1}{4}(|0\rangle\langle 0| + |1\rangle\langle 0| + |0\rangle\langle 1| + |1\rangle\langle 1|) + \frac{1}{4}(|0\rangle\langle 0| - |1\rangle\langle 0| - |0\rangle\langle 1| + |1\rangle\langle 1|) \\
&= \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) \quad .
\end{aligned} \tag{2.34}$$

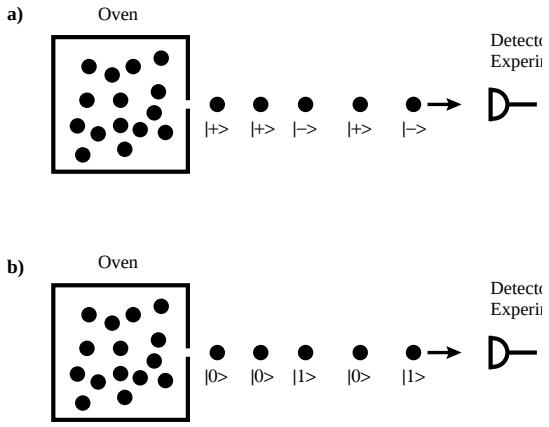


Figure 2.2: a) An oven generates particles in states $|0\rangle$ or $|1\rangle$ with probability 50%. An experimentalist measures observable $\hat{A}$ and finds the mean value $tr\{\hat{A}\rho\}$. b) The oven generates particles in states $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$ with probability 50%. The experimentalist will find the same mean value $tr\{\hat{A}\rho\}$.

Therefore the same density operator can be obtained in different ways. In fact, we can find usually find infinitely many ways of generating the same density operator. What does this mean? Does this make the density operator an ill-defined object, as it cannot distinguish between different experimental realisations? The answer to that is a clear NO! The point is, that if a two realisations give rise to the same density operator, then, as we have seen above, we will also find exactly the same expectation values for **any** observable that we may chose to measure. If a quantum system in a mixed state gives rise to exactly the same predictions for **all** possible measurements, then we are forced to say that the two states are indeed the same! Of course this is something that needs to be confirmed experimentally and it turns out that it is true. Even more interestingly it turns out that if I could distinguish two situations that are described by the same density operator then I would be able to transmit signals faster than light! A clear impossibility as it contradicts special relativity. This is what I will explain to you in the next section using the concept of entanglement which you are going to encounter for the first time.

## 2.5 Mixed states, Entanglement and the speed of light

In the previous section I have introduced the density operator to describe situations in which we have a **lack of knowledge** due to an imperfect preparation of a quantum state. However, this is not the only situation in which we have to use a density operator. In this section I will show you that the density operator description may also become necessary when the system that we are investigating is only the **accessible part of some larger total system**. Even if this larger system is in a pure quantum state, we will see that the smaller system behaves in every respect like a mixed state and has to be described by a density operator. This idea will then lead us to the insight that two different preparations that are described by the same density operator can not be distinguished experimentally, because otherwise we would be able to send messages faster than the speed of light. This would clearly violate the special theory of relativity.

To understand this new way of looking at density operators, I first need to explain how to describe quantum mechanical systems that consist of more than one particle. This is the purpose of the next subsection.

### 2.5.1 Quantum mechanics for many particles

In this section I will show you what you have to do when you want to describe a system of two particles quantum mechanically. The generalization to arbitrary numbers of particles will then be obvious.

If we have two particles, then the state of each particle are elements of a Hilbert space; for particle $A$ we have the Hilbert space $\mathcal{H}_A$ which is spanned by a set of basis states $\{|\phi_i\rangle_A\}_{i=1,...,N}$, while particle $B$ has the Hilbert space $\mathcal{H}_B$ spanned by the set of basis states $\{|\psi_j\rangle_B\}_{j=1,...,M}$. The two Hilbert spaces are not necessarily equal and may describe total different quantities or particles.

Now imagine that system A is in state $|\phi_i\rangle_A$ and system B is in state $|\psi_j\rangle_B$. Then we write the total state of both systems in the tensor

product form

$$|\Psi_{tot}\rangle = |\phi_i\rangle \otimes |\psi_j\rangle \ . \tag{2.35}$$

The symbol $\otimes$ denotes the tensor product, which is **not** to be confused with ordinary products. Clearly there are $N \cdot M$ such combinations of basis vectors. These vectors can be thought of as spanning a larger Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ which describes all possible states that the two particles can be in.

Of course we have to be able to add different states in the new Hilbert space $\mathcal{H}_{AB}$. We have to define this addition such that it is compatible with physics. Imagine that system A is in a superposition state $|\Phi\rangle_A = a_1|\phi_1\rangle + a_2|\phi_2\rangle$ and system B is in the basis state $|\Psi\rangle = |\psi_1\rangle$. Then

$$|\Phi\rangle_A \otimes |\Psi\rangle = (a_1|\phi_1\rangle_A + a_2|\phi_2\rangle_A) \otimes |\psi_1\rangle \tag{2.36}$$

is the joint state of the two particles. Now surely it makes sense to say that

$$|\Phi\rangle_A \otimes |\Psi\rangle = a_1|\phi_1\rangle_A \otimes |\psi_1\rangle + a_2|\phi_2\rangle_A \otimes |\psi_1\rangle \ \ . \tag{2.37}$$

This is so, because the existence of an additional particle (maybe at the end of the universe) cannot change the linearity of quantum mechanics of system A. As the same argument applies for system B we should have in general

$$\left(\sum_i a_i|\phi_i\rangle\right) \otimes \left(\sum_j b_j|\psi_j\rangle\right) = \sum_{ij} a_i b_j |\phi_i\rangle \otimes |\psi_j\rangle \ \ . \tag{2.38}$$

The set of states $\{|\phi_i\rangle_A \otimes |\psi_j\rangle_B\}$ forms a basis of the state space of two particles.

Now let us see how the scalar product between two states $|\psi_1\rangle_A \otimes |\psi_2\rangle_B$ and $|\phi_1\rangle_A \otimes |\phi_2\rangle_B$ has to be defined. Clearly if $|\psi_2\rangle = |\phi_2\rangle$ then we should have

$$(|\psi_1\rangle_A \otimes |\psi_2\rangle_B, |\phi_1\rangle_A \otimes |\phi_2\rangle_B) = (|\psi_1\rangle_A, |\phi_1\rangle_A) \tag{2.39}$$

again because the existence of an unrelated particle somewhere in the world should not change the scalar product of the first state.

Eq. (2.39) can only be true if in general

$$(|\psi_1\rangle_A \otimes |\psi_2\rangle_B, |\phi_1\rangle_A \otimes |\phi_2\rangle_B) = (|\psi_1\rangle_A, |\phi_1\rangle_A)\,(|\psi_2\rangle_B, |\phi_2\rangle_B) \ . \quad (2.40)$$

The tensor product of linear operators is defined by the action of the operator on all possible product states. We define

$$\hat{A} \otimes \hat{B}(|\phi\rangle_A \otimes |\psi\rangle_B) := (\hat{A}|\phi\rangle_A) \otimes (\hat{B}|\psi\rangle_B) \ . \quad (2.41)$$

In all these definitions I have only used states of the form $|\phi\rangle \otimes |\psi\rangle$ which are called **product states**. Are all states of this form? The answer is evidently no, because we can form linear superpositions of different product states. An example of such a linear superposition is the state

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \ . \quad (2.42)$$

Note that I now omit the $\otimes$ and abbreviate $|0\rangle \otimes |0\rangle$ by $|00\rangle$. If you try to write the state Eq. (2.42) as a product

$$\frac{|00\rangle + |11\rangle}{\sqrt{2}} = (\alpha|0\rangle + \beta|1\rangle) \otimes (\gamma|0\rangle + \delta|1\rangle) = \alpha\gamma|00\rangle + \alpha\delta|01\rangle + \beta\gamma|10\rangle + \beta\delta|11\rangle \ .$$

This implies that $\alpha\delta = 0 = \beta\gamma$. If we chose $\alpha = 0$ then also $\alpha\gamma = 0$ and the two sides cannot be equal. Likewise if we chose $\delta = 0$ then $\beta\delta = 0$ and the two sides are different. As either $\alpha = 0$ or $\delta = 0$ we arrive at a contradiction. States which cannot be written in product form, are called **entangled states**. In the last part of these lectures we will learn a bit more about the weird properties of the entangled states. They allow such things as quantum state teleportation, quantum cryptography and quantum computation.

Let us now see how the tensor product looks like when you write them as column vectors? The way you write them is really arbitrary, but of course there are clumsy ways and there are smart ways. In literature you will find only one way. This particular notation for the tensor product of column vectors and matrices ensures that the relation $\hat{A} \otimes \hat{B}|\phi\rangle \otimes |\psi\rangle = \hat{A}|\phi\rangle \otimes \hat{B}|\psi\rangle$ is true when you do all the manipulations

using matrices and column vectors. The tensor product between a $m$-dimensional vector and an $n$-dimensional vector is written as

$$\begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} \otimes \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \\ \vdots \\ a_m \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ \vdots \\ a_m b_n \end{pmatrix} \quad . \tag{2.43}$$

The tensor product of a linear operator $\hat{A}$ on an $m$-dimensional space and linear operator $\hat{B}$ on an $n$-dimensional space in matrix notation is given by

$$\begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mm} \end{pmatrix} \otimes \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \dots & b_{nn} \end{pmatrix} = \begin{pmatrix} a_{11}B & \dots & a_{1m}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \dots & a_{mm}B \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \dots & b_{nn} \end{pmatrix} & \dots & a_{1m} \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \dots & b_{nn} \end{pmatrix} \\ \vdots & \ddots & \vdots \\ a_{m1} \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \dots & b_{nn} \end{pmatrix} & \dots & a_{mm} \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \dots & b_{nn} \end{pmatrix} \end{pmatrix} \quad . \tag{2.44}$$

The simplest explicit examples that I can give are those for two 2-dimensional Hilbert spaces. There we have

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \otimes \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ a_1 b_2 \\ a_2 b_1 \\ a_2 b_2 \end{pmatrix} \quad . \tag{2.45}$$

and

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \otimes \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{pmatrix} . \tag{2.46}$$

### 2.5.2 How to describe a subsystem of some large system?

Let us now imagine the following situation. The total system that we consider consists of two particles each of which are described by an N(M)-dimensional Hilbert space $\mathcal{H}_A$ ($\mathcal{H}_B$) with basis vectors $\{|\phi_i\rangle_A\}_{i=1,...,N}$ ($\{|\psi_i\rangle_B\}_{i=1,...,M}$). Let us also assume that the two particles together are in an arbitrary, possibly mixed, state $\rho_{AB}$. One person, Alice, is holding particle $A$, while particle $B$ is held by a different person (let's call him Bob) which, for whatever reason, refuses Alice access to his particle. This situation is schematically represented in Fig. 2.3. Now imagine that Alice makes a measurement on her system alone (she cannot access Bobs system). This means that she is measuring an operator of the form $\hat{A} \otimes \mathbb{1}$. The operator on Bobs system must be the identity operator because no measurement is performed there. I would now like to calculate the expectation value of this measurement, i.e.

$$\langle \hat{A} \rangle = \sum_{ij} {}_A\langle\phi|_B\langle\psi_j|\hat{A} \otimes \mathbb{1}|\phi\rangle_A|\psi_j\rangle \quad = \quad tr_{AB}\{\hat{A} \otimes \mathbb{1}\rho_{AB}\} \quad ,$$

where $\mathbf{tr_{AB}}$ means the trace over **both** systems, that of Alice and that of Bob. This description is perfectly ok and always yields the correct result. However, if we only ever want to know about outcomes of measurements on Alice's system alone then it should be sufficient to have a quantity that describes the state of her system alone without any reference to Bob's system. I will now derive such a quantity, which is called the **reduced density operator**.

Our task is the definition of a density operator $\rho_A$ for Alices system which satisfies

$$\langle \hat{A} \rangle = tr_{AB}\{\hat{A} \otimes \mathbb{1}\rho_{AB}\} = tr_A\{\hat{A}\rho_A\} \tag{2.47}$$

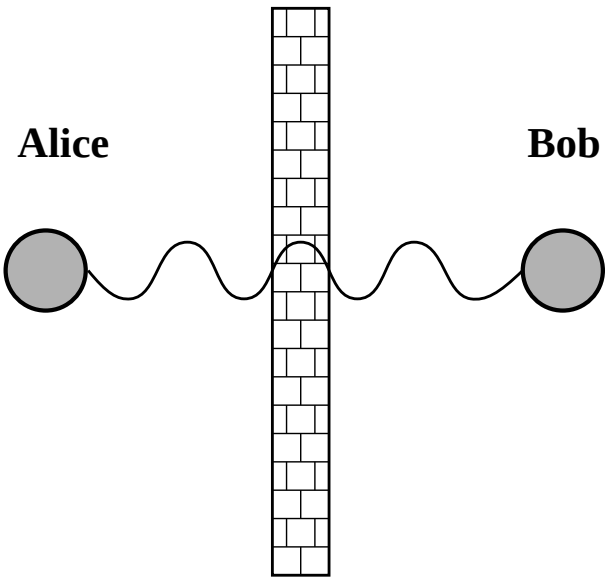


Figure 2.3: Alice and Bob hold a joint system, here composed of two particles. However, Alice does not have access to Bob's particle and vice versa.

for **all** observables $\hat{A}$. How can we construct this operator? Let us rewrite the left hand side of Eq. (2.47)

$$\begin{aligned}
\langle \hat{A} \rangle &= \sum_{i=1}^{N} \sum_{j=1}^{M} {}_A\langle \phi_i |_B \langle \psi_j | \hat{A} \otimes \mathbb{1} \rho_{AB} | \phi_i \rangle_A | \psi_j \rangle_B \\
&= \sum_{i=1}^{N} {}_A\langle \phi_i | \hat{A} \left( \sum_{j=1}^{M} {}_B\langle \psi_j | \rho_{AB} | \psi_j \rangle_B \right) | \phi_i \rangle_A \\
&= tr_A \{ \hat{A} \left( \sum_{j=1}^{M} {}_B\langle \psi_j | \rho_{AB} | \psi_j \rangle_B \right) \} \quad .
\end{aligned}$$

In the last step I have effectively split the trace operation into two parts. This split makes it clear that the state of Alices system is described by

the **reduced density operator**

$$\rho_A := \sum_{j=1}^{M} {}_B\langle\psi_j|\rho_{AB}|\psi_j\rangle_B = tr_B\{\rho_{AB}\} \quad . \tag{2.48}$$

where the last identity describes the operation of taking the **partial trace**. The point is now that the reduced density operator allows Alice to describe the outcome of measurements on her particle without any reference to Bob's system, e.g.

$$\langle\hat{A}\rangle = tr_A\{\hat{A}\rho_A\} \quad .$$

When I am giving you a density operator $\rho_{AB}$ then you will ask how to actually calculate the reduced density operator explicitly. Let me first consider the case where $\rho_{AB}$ is the pure product state, i.e. $\rho_{AB} = |\phi\rangle_A|\psi_1\rangle_{BA}\langle\phi|_B\langle\psi_1|$ Then following Eq. (2.48) the partial trace is

$$\begin{aligned} tr_B\{|\phi_1\psi_1\rangle\langle\phi_1\psi_1|\} &= \sum_{i=1}^{M} {}_B\langle\psi_i|\left(|\phi_1\psi_1\rangle\langle\phi_1\psi_1|\right)|\psi_i\rangle_B \\ &= \sum_{i=1}^{M} |\phi_1\rangle_{AA}\langle\phi_1|_B\langle\psi_i|\psi_1\rangle_{BB}\langle\psi_1|\psi_i\rangle_B \\ &= |\phi_1\rangle_{AA}\langle\phi_1| \quad . \end{aligned} \tag{2.49}$$

Now it is clear how to proceed for an arbitrary density operator of the two systems. Just write down the density operator in a product basis

$$\rho_{AB} = \sum_{i,k=1}^{N}\sum_{j,l=1}^{M} \alpha_{ij,kl}|\phi_i\rangle_{AA}\langle\phi_k| \otimes |\psi_j\rangle_{BB}\langle\psi_l| \quad , \tag{2.50}$$

and then use Eq. (2.49) to find

$$\begin{aligned} tr_B\{\rho_{AB}\} &= \sum_{c=1}^{M} {}_B\langle\psi_c|\left(\sum_{i,k=1}^{N}\sum_{j,l=1}^{M} \alpha_{ij,kl}|\phi_i\rangle_{AA}\langle\phi_k| \otimes |\psi_j\rangle_{BB}\langle\psi_l|\right)|\psi_c\rangle_B \\ &= \sum_{c=1}^{M}\sum_{i,k=1}^{N}\sum_{j,l=1}^{M} \alpha_{ij,kl}|\phi_i\rangle_{AA}\langle\phi_k|({}_B\langle\psi_c|\psi_j\rangle_B)({}_B\langle\psi_l|\psi_c\rangle_B) \end{aligned}$$

$$= \sum_{c=1}^{M}\sum_{i,k=1}^{N}\sum_{j,l=1}^{M}\alpha_{ij,kl}|\phi_i\rangle_{AA}\langle\phi_k|\delta_{cj}\delta_{lc}$$
$$= \sum_{i,k=1}^{N}\left(\sum_{c=1}^{M}\alpha_{ic,kc}\right)|\phi_i\rangle_{AA}\langle\phi_k| \qquad (2.51)$$

Let me illustrate this with a concrete example.

**Example:** Consider two system $A$ and $B$ that have two energy levels each. I denote the basis states in both Hilbert spaces by $|1\rangle$ and $|2\rangle$. The basis states in the tensor product if both spaces are then $\{|11\rangle, |12\rangle, |21\rangle, |22\rangle\}$. Using this basis I write a possible state of the two systems as

$$\begin{aligned}\rho_{AB} &= \frac{1}{3}|11\rangle\langle 11| + \frac{1}{3}|11\rangle\langle 12| + \frac{1}{3}|11\rangle\langle 21| \\ &\quad + \frac{1}{3}|12\rangle\langle 11| + \frac{1}{3}|12\rangle\langle 12| + \frac{1}{3}|12\rangle\langle 21| \\ &\quad + \frac{1}{3}|21\rangle\langle 11| + \frac{1}{3}|21\rangle\langle 12| + \frac{1}{3}|21\rangle\langle 21| \ . \end{aligned} \qquad (2.52)$$

Now let us take the partial trace over the second system B. This means that we collect those terms which contain states of the form $|i\rangle_{BB}\langle i|$. Then we find that

$$\rho_A = \frac{2}{3}|1\rangle\langle 1| + \frac{1}{3}|1\rangle\langle 2| + + \frac{1}{3}|2\rangle\langle 1| + \frac{1}{3}|2\rangle\langle 2| \ . \qquad (2.53)$$

This is **not** a pure state (Check that indeed $det\rho_A \neq 0$). It is not an accident that I have chosen this example, because it shows that the state of Alice's system may be a mixed state (incoherent superposition) although the total system that Alice and Bob are holding is in a pure state. To see this, just check that

$$\rho_{AB} = |\Psi\rangle_{ABAB}\langle\Psi| \ , \qquad (2.54)$$

where $|\Psi\rangle_{AB} = \sqrt{\frac{1}{3}}|11\rangle + \sqrt{\frac{1}{3}}|12\rangle + \sqrt{\frac{1}{3}}|21\rangle$. This is evidently a pure state! Nevertheless, the reduced density operator Eq. (2.53) of Alice's system alone is described by a mixed state!

In fact this is a very general behaviour. Whenever Alice and Bob hold a system that is in a joint pure state but that **cannot** be written

as a product state $|\phi\rangle \otimes |\psi\rangle$ (Check this for Eq. (2.54), then the reduced density operator describing one of the subsystems represents a **mixed** state. Such states are called entangled states and they have quite a lot of weird properties some of which you will encounter in this lecture and the exercises.

Now let us see how these ideas help us to reveal a connection between special relativity and mixed states.

### 2.5.3 The speed of light and mixed states.

In the previous subsection you have seen that a mixed state of your system can also arise because your system is part of some larger inaccessible system. Usually the subsystem that you are holding is then in a mixed state described by a density operator. When I introduced the density operator, I told you that a particular mixed state be realized in different ways. An example was

$$\begin{aligned} \rho &= \frac{1}{2}|1\rangle\langle 1| + \frac{1}{2}|2\rangle\langle 2| \\ &= \frac{1}{2}|+\rangle\langle +| + \frac{1}{2}|-\rangle\langle -| \end{aligned}$$

with $|\pm\rangle = \frac{1}{\sqrt{2}}(|1\rangle \pm |2\rangle)$. However, I pointed out that these two realizations are physically the same because you are unable to distinguish them. I did not prove this statement to you at that point. I will now correct this omission. I will show that if you were able to distinguish two realizations of the same density operator, then you could send **signals faster than the speed of light**, i.e. this would violate special relativity. How would that work?

Assume two persons, Alice and Bob, hold pairs of particles where each particle is described by two-dimensional Hilbert spaces. Let us assume that Alice and Bob are sitting close together, and are able to access each others particle such that they are able to prepare the state

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|11\rangle + |22\rangle\right) \quad . \tag{2.55}$$

Now Alice and Bob walk away from each other until they are separated by, say, one light year. Now Alice discovers the answer to a very tricky

question which is either 'yes' or 'no' and would like to send this answer to Bob. To do this she performs either of the following measurements on her particle.

**'yes'** Alice measures the observable $\hat{A}_{yes} = |1\rangle\langle 1| + 2|2\rangle\langle 2|$.
The probability to find the state $|1\rangle$ is $p_1 = \frac{1}{2}$ and to find $|2\rangle$ is $p_2 = \frac{1}{2}$. This follows from postulate 4 because $p_i = tr\{|i\rangle\langle i|\rho_A\} = \frac{1}{2}$ where $\rho_A = \frac{1}{2}(|1\rangle\langle 1| + |2\rangle\langle 2|)$ is the reduced density operator describing Alices system. After the result $|1\rangle$ the state of the total system is $|1\rangle\langle 1| \otimes \mathbb{1}|\psi\rangle \sim |11\rangle$, after the result $|2\rangle$ the state of the total system is $|2\rangle\langle 2| \otimes \mathbb{1}|\psi\rangle \sim |22\rangle$.

**'no'** Alice measures the observable $\hat{A}_{no} = |+\rangle\langle +| + 2|-\rangle\langle -| \neq \hat{A}_{yes}$.
The probability to find the state $|+\rangle$ is $p_+ = \frac{1}{2}$ and to find $|-\rangle$ is $p_- = \frac{1}{2}$. This follows from postulate 4 because $p_i = tr\{|i\rangle\langle i|\rho_A\} = \frac{1}{2}$ where $\rho_A = \frac{1}{2}(|+\rangle\langle +| + |-\rangle\langle -|)$ is the reduced density operator describing Alices system. After the result $|+\rangle$ the state of the total system is $|+\rangle\langle +| \otimes \mathbb{1}|\psi\rangle \sim |++\rangle$, after the result $|-\rangle$ the state of the total system is $|-\rangle\langle -| \otimes \mathbb{1}|\psi\rangle \sim |--\rangle$.

What does this imply for Bob's system? As the measurement results 1 and 2 occur with probability 50%, Bob holds an equal mixture of the two states corresponding the these outcomes. In fact, when Alice wanted to send 'yes' then the state of Bob's particle after Alices measurement is given by

$$\rho_{yes} = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| \tag{2.56}$$

and if Alice wanted to send 'no' then the state of Bob's particle is given by

$$\rho_{no} = \frac{1}{2}|+\rangle\langle +| + \frac{1}{2}|-\rangle\langle -| \quad . \tag{2.57}$$

You can check quite easily that $\rho_{no} = \rho_{yes}$. If Bob would be able to distinguish the two realizations of the density operators then he could find out which of the two measurements Alice has carried out. In that case he would then be able to infer what Alices answer is. He can do that immediately after Alice has carried out her measurements, which in principle requires negligible time. This would imply that Alice could

send information to Bob over a distance of a light year in virtually no time at all. This clearly violates the theory of special relativity.

As we know that special relativity is an extremely well established and confirmed theory, this shows that Bob is unable to distinguish the two density operators Eqs. (2.56-2.57).

# Chapter 3

# Dynamics and Symmetries

So far we have only been dealing with stationary systems, i.e. we did not have any prescription for the time evolution of quantum mechanical systems. We were only preparing systems in particular quantum mechanical states and then subjected them to measurements. In reality of course any system evolves in time and we need to see what the quantum mechanical rules for time evolution are. This is the subject of this chapter.

## 3.1 The Schrödinger Equation

Most of you will know that the prescription that determines the quantum mechanical time evolution is given by the Schrödinger equation. Before I state this as the next postulate, let us consider what any decent quantum mechanical time evolution operator should satisfy, and then we will convince us that the Schrödinger equation satisfies these criteria. In fact, once we have accepted these properties, we will not have too much freedom of choice for the form of the Schrödinger equation.

**Definition 42** *The time evolution operator* $\hat{U}(t_2, t_1)$ *maps the quantum state* $|\psi(t_1)\rangle$ *to the state* $|\psi(t_2)\rangle$ *obeying the properties*

1. $\hat{U}(t_2, t_1)$ *is unitary.*

2. *We have* $\hat{U}(t_2,t_1)\hat{U}(t_1,t_0) = \hat{U}(t_2,t_0)$. *(semi-group property)*

3. $\hat{U}(t,t_1)$ *is differentiable in* $t$.

What is the reason for these assumptions? First of all, the time evolution operator has to map physical states onto physical states. That means in particular that a normalized state $|\psi(t_0)\rangle$ is mapped into a normalized state $|\psi(t_1)\rangle$, i.e. for any $|\psi(t_0)\rangle$

$$\langle\psi(t_0)|\psi(t_0)\rangle = \langle\psi(t_1)|\psi(t_1)\rangle = \langle\psi(t_0)|\hat{U}^\dagger(t_1,t_0)\hat{U}(t_1,t_0)|\psi(t_0)\rangle \quad . \tag{3.1}$$

Therefore it seems a fair guess that $\hat{U}(t_1,t_0)$ should be a unitary operator.

The second property of the time evolution operator demands that it does not make a difference if we first evolve the system from $t_0$ to $t_1$ and then from $t_1$ to $t_2$ or if we evolve it directly from time $t_0$ to $t_2$. This is a very reasonable assumption. Note however, that this does not imply that we may measure the system at the intermediate time $t_1$. In fact we **must not** interact with the system.

The third condition is one of mathematical convenience and of physical experience. Every system evolves continuously in time. This is an observation that is confirmed in experiments. Indeed dynamics in physics is usually described by differential equations, which already implies that observable quantities and physical state are differentiable in time. You may then wonder what all the fuss about quantum jumps is then about. The point is that we are talking about the time evolution of a **closed** quantum mechanical system. This means that we, e.g. a person from outside the system, do not interact with the system and in particular this implies that we are not measuring the system. We have seen that in a measurement indeed the state of a system can change discontinuously, at least according to our experimentally well tested Postulate 4. In summary, the quantum state of a closed system changes smoothly unless the system is subjected to a measurement, i.e. an interaction with an outside observer.

What further conclusions can we draw from the properties of the time evolution operator? Let us consider the time evolution operator for very short time differences between $t$ and $t_0$ and use that it is

differentiable in time (property 3) of its Definition 42). Then we find

$$\hat{U}(t,t_0) = \mathbb{1} - \frac{i}{\hbar}\hat{H}(t_0)(t-t_0) + \dots \quad . \tag{3.2}$$

Here we have used the fact that we have assumed that the time evolution of a closed quantum mechanical system is smooth. The operator $\hat{H}(t_0)$ that appears on the right hand side of Eq. (3.2) is called the **Hamilton operator** of the system. Let us apply this to an initial state $|\psi(t_0)\rangle$ and take the time derivative with respect to $t$ on both sides. Then we find

$$i\hbar\partial_t|\psi(t)\rangle \equiv i\hbar\partial_t(\hat{U}(t,t_0)|\psi(t_0)\rangle) \approx \hat{H}(t_0)|\psi(t_0)\rangle \quad . \tag{3.3}$$

If we now carry out the limit $t \to t_0$ then we finally find

$$i\hbar\partial_t|\psi(t)\rangle = \hat{H}(t)|\psi(t)\rangle \quad . \tag{3.4}$$

This is the Schrödinger equation in the Dirac notation in a coordinate independent form. To make the connection to the Schrödinger equation as you know it from last years quantum mechanics course let us consider **Example:** Use the Hamilton operator for a particle moving in a one-dimensional potential $V(x,t)$. Write the Schródinger equation in the position representation. This gives

$$\begin{aligned} i\hbar\partial_t\langle x|\psi(t)\rangle &= \langle x|i\hbar\partial_t|\psi(t)\rangle \\ &= \langle x|\hat{H}|\psi(t)\rangle \\ &= \left(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + V(\hat{x},t)\right)\langle x|\psi(t)\rangle \quad . \end{aligned}$$

This is the Schrödinger equation in the position representation as you know it from the second year course.

Therefore, from the assumptions on the properties of the time evolution operator that we made above we were led to a differential equation for the time evolution of a state vector. Again, of course, this result has to be tested in experiments, and a (sometimes) difficult task is to find the correct Hamilton operator $\hat{H}$ that governs the time evolution of the system. Now let us formulate the result of our considerations in

the following

**Postulate 5** *The time evolution of the quantum state of an isolated quantum mechanical system is determined by the Schrödinger equation*

$$i\hbar\partial_t|\psi(t)\rangle = \hat{H}(t)|\psi(t)\rangle \ , \tag{3.5}$$

*where $\hat{H}$ is the* **Hamilton operator** *of the system.*

As you may remember from your second year course, the Hamilton operator is the observable that determines the energy eigenvalues of the quantum mechanical system. In the present course I cannot justify this in more detail, as we want to learn a lot more things about quantum mechanics. The following discussion will however shed some more light on the time evolution of a quantum mechanical system and in particular it will pave our way towards the investigation of quantum mechanical symmetries and conserved quantities.

### 3.1.1 The Heisenberg picture

In the preceding section we have considered the time evolution of the quantum mechanical state vector. In fact, we have put all the time evolution into the state and have left the observables time independent. This way of looking at the time evolution is called the Schrödinger picture. This is not the only way to describe quantum dynamics (unless they had some intrinsic time dependence). We can also go to the other extreme, namely leave the states time invariant and evolve the observables in time. How can we find out what the time evolution of a quantum mechanical observable is? We need some guidance, some property that must be the same in both pictures of quantum mechanics. Such a property must be experimentally observable. What I will be using here is the fact that expectation values of **any** observable has to be the same in both pictures. Let us start by considering the expectation value of an operator $\hat{A}_S$ in the Schrödinger picture at time time

$t$ given an initial state $|\psi(t_0)\rangle$. We find

$$\langle A\rangle = \langle\psi(t)|\hat{A}_S|\psi(t)\rangle = \langle\psi(t_0)|\hat{U}^\dagger(t,t_0)\hat{A}_S\hat{U}(t,t_0)|\psi(t_0)\rangle \ . \tag{3.6}$$

Looking at Eq. (3.6) we see that we can interpret the right hand side also as the expectation value of a time-dependent operator

$$A_H(t) = \hat{U}^\dagger(t,t_0)\hat{A}_S\hat{U}(t,t_0) \tag{3.7}$$

in the initial state $|\psi(t_0)\rangle$. That is

$$\langle\psi(t)|\hat{A}_S|\psi(t)\rangle = \langle A\rangle = \langle\psi(t_0)|A_H(t)|\psi(t_0)\rangle \ . \tag{3.8}$$

Viewing the state of the system as time-independent, and the observables as evolving in time is called the **Heisenberg-picture** of quantum mechanics. As we have seen in Eq. (3.8) both, the Schrödinger picture and the Heisenberg picture, give exactly the same predictions for physical observables. But, depending on the problem, it can be advantageous to chose one or the other picture.

Like the states in the Schrödinger picture the Heisenberg operators obey a differential equation. This can easily be obtained by taking the time derivative of Eq. (3.7). To see this we first need to know the time derivative of the time evolution operator. Using $|\psi(t)\rangle = \hat{U}(t,t_0)|\psi(t_0)\rangle$ in the Schrödinger equation Eq. (3.4) we find for any $|\psi(t_0)\rangle$

$$i\hbar\partial_t\hat{U}(t,t_0)|\psi(t_0)\rangle = \hat{H}(t)\hat{U}(t,t_0)|\psi(t_0)\rangle \tag{3.9}$$

and therefore

$$i\hbar\partial_t\hat{U}(t,t_0) = \hat{H}(t)\hat{U}(t,t_0) \ . \tag{3.10}$$

Assuming that the the operator in the Schrödinger picture has no explicit time dependence, we find

$$\begin{aligned}
\frac{d}{dt}A_H(t) &= \frac{d}{dt}\left(\hat{U}^\dagger(t,t_0)\hat{A}_S\hat{U}(t,t_0)\right) \\
&= \frac{\partial\hat{U}^\dagger(t,t_0)}{\partial t}\hat{A}_S\hat{U}(t,t_0) + \hat{U}^\dagger(t,t_0)\hat{A}_S\frac{\partial\hat{U}(t,t_0)}{\partial t} \\
&= \frac{i}{\hbar}\hat{U}^\dagger(t,t_0)\hat{H}(t)\hat{A}_S\hat{U}(t,t_0) + \hat{U}^\dagger(t,t_0)\hat{A}_S\frac{i}{\hbar}\hat{H}(t)\hat{U}(t,t_0)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{i}{\hbar}\hat{U}^{\dagger}(t,t_0)[\hat{H}(t),A_S]\hat{U}(t,t_0) \\
&= \frac{i}{\hbar}[\hat{H}(t),A_S]_H \qquad (3.11) \\
&= \frac{i}{\hbar}[\hat{H}_H(t),A_H] \ . \qquad (3.12)
\end{aligned}
$$

It is easy to check that for an operator that has an explicit time dependence in the Schrödinger picture, we find the **Heisenberg equation**

$$\frac{d}{dt}A_H(t) = \frac{i}{\hbar}[\hat{H}(t),A_H(t)] + \left(\frac{\partial A_S}{\partial t}\right)_H(t) \ . \qquad (3.13)$$

One of the advantages of the Heisenberg equation is that it has a direct analogue in classical mechanics. In fact this analogy can be viewed as a justification to identify the Hamiltonian of the system $\hat{H}$ with the energy of the system. I will not go into details here and for those of you who would like to know more about that I rather recommend the book: H. Goldstein, *Classical Mechanics*, Addison-Wesley (1980).

End of $14^{th}$ lecture

## 3.2 Symmetries and Conservation Laws

After this introduction to the Heisenberg picture it is now time to discuss the concept of symmetries and their relation to conservation laws.

### 3.2.1 The concept of symmetry

If we are able to look at a system in different ways and it appears the same to us then we say that a system has a symmetry. This simple statement will be put on solid ground in this section. Why are we interested in symmetries? The answer is simple! It makes our life easier. If we have a symmetry in our problem, then the solution will also have such a symmetry. This symmetry of the solution expresses itself in the existence of a quantity which is conserved for any solution

to the problem. Of course we are always interested to know quantities that are conserved and the natural question is how these conserved quantities are related to the symmetry of the problem! There are two ways to formulate symmetries, in an active way and a passive way. If we assume that we transform the system $S$ into a new system $S'$ then we speak of an active transformation, i.e. we actively change the system. If we change the coordinate system in which we consider the same system, then we speak of a passive transformation. In these lectures I will adopt the active point of view because it appears to be the more natural way of speaking. An extreme example is the time reversal symmetry. Certainly we can make a system in which all momenta of particles are reversed (the active transformation) while we will not be able to actually reverse time, which would amount to the corresponding passive transformation.

Now let us formalize what we mean by a symmetry. In the active viewpoint, we mean that the systems $S$ and $S' = TS$ look the same. We have to make sure what we mean by '$S$ and $S'$ look the same'. What does this mean in quantum mechanics? The only quantities that are accessible in quantum mechanics are expectation values or, even more basic, transition probabilities. Two systems are the same in quantum mechanics if the transition probabilities between corresponding states are the same. To be more precise let us adopt the

**Definition 43** *A transformation $\hat{T}$ is called a* **symmetry transformation** *when for all vectors $|\phi\rangle$ and $|\psi\rangle$ in system $S$ and the corresponding vectors $|\phi'\rangle = \hat{T}|\phi\rangle$ and $|\psi'\rangle = \hat{T}|\psi\rangle$ in system $S'$ we find*

$$|\langle\psi|\phi\rangle| = |\langle\psi'|\phi'\rangle| \quad . \tag{3.14}$$

In this definition we have seen that quantum states are transformed as $|\phi'\rangle = \hat{T}|\phi\rangle$. This implies that observables, which are of the form $\hat{A} = \sum_i a_i |a_i\rangle\langle a_i|$ will be transformed according to the prescription $\hat{A}' = \hat{T}\hat{A}\hat{T}^\dagger$.

Which transformations $\hat{T}$ are possible symmetry transformations? As all transition probabilities have to be preserved one would expect that symmetry transformations are automatically unitary transformations. But this conclusion would be premature because of two reasons.

Firstly, we did not demand that the scalar product is preserved, but only the **absolute** value of the scalar product and secondly we did not even demand the linearity of the transformation $T$. Fortunately it turns out that a symmetry transformation $\hat{T}$ cannot be very much more general than a unitary transformation. This was proven in a remarkable theorem by Wigner which states

**Theorem 44** *Any symmetry transformation can either be represented by a unitary transformation or an anti-unitary transformation.*

For a proof (which is rather lengthy) of this theorem you should have a look at books such as: K. Gottfried, *Quantum Mechanics I*, Benjamin (1966).

An anti-unitary transformation has the property

$$\hat{U}(\alpha|\phi\rangle + \beta|\psi\rangle) = \alpha^*\hat{U}|\phi\rangle + \beta^*\hat{U}|\psi\rangle \tag{3.15}$$

and preserves all transition probabilities. An example of an anti-unitary symmetry is time-reversal, but this will be presented later.

Of course a system will usually be symmetric under more than just one particular transformation. In fact it will be useful to consider symmetry groups. Such groups may be discrete (you can number them with whole numbers) or continuous (they will be parametrized by a real number).

**Definition 45** *A continuous symmetry group is a set of symmetry transformation that can be parametrized by a real parameter such that the symmetry transformations can be differentiated with respect to this parameter.*

**Example:** a) Any Hermitean operator $\hat{A}$ gives rise to a continuous symmetry group using the definition

$$\hat{U}(\epsilon) := e^{i\hat{A}\epsilon/\hbar} \quad . \tag{3.16}$$

Obviously the operator $\hat{U}(\epsilon)$ can differentiated with respect to the parameter $\epsilon$.

So far we have learned which transformations can be symmetry transformations. A symmetry transformation of a quantum mechanical system is either a unitary or anti-unitary transformation. However, for a given quantum system not every symmetry transformation is a symmetry of that quantum system. The reason for this is that it is not sufficient to demand that the configuration of a system is symmetric at a given instance in time, but also that this **symmetry is preserved under the dynamics of the system**! Otherwise we would actually be able to distinguish two 'symmetric' states by just waiting and letting the system evolve. The invariance of the configurational symmetry under a time evolution will lead to observables that are conserved in time. As we are now talking about the time evolution of a quantum system we realize that the Hamilton operator plays a major role in the theory of symmetries of quantum mechanical systems. In fact, this is not surprising because the initial state (which might posses some symmetry) and the Hamilton operator determine the future of a system completely.

Assume that we have a symmetry that is preserved in time. At the initial time the symmetry takes a state $|\psi\rangle$ into $|\psi\rangle' = \hat{T}|\psi\rangle$. The time evolution of the original system $S$ is given by the Hamilton operator $\hat{H}$ while that of the transformed system $S'$ is given by $\hat{H}' = \hat{T}\hat{H}\hat{T}^\dagger$. This leads to

$$|\psi(t)\rangle = e^{-i\hat{H}t/\hbar}|\psi(0)\rangle \tag{3.17}$$

for the time evolution in $S$. The time evolution of the transformed system $S'$ is governed by the Hamilton operator $\hat{H}'$ so that we find

$$|\psi(t)\rangle' = e^{-i\hat{H}'t/\hbar}|\psi(0)\rangle' \ . \tag{3.18}$$

However, the state $|\psi\rangle'$ could also be viewed as quantum state of the original system $S$! This means that the time evolution of system $S$ given by the Hamilton operator $\hat{H}$ could be applied and we would find

$$|\psi(t)\rangle'' = e^{-i\hat{H}'t/\hbar}|\psi(0)\rangle' \ . \tag{3.19}$$

If the symmetry of the system is preserved for all times then the two states $|\psi(t)\rangle'$ and $|\psi(t)\rangle''$ cannot differ by more than a phase factor, i.e.

$$|\psi(t)\rangle'' = e^{i\phi}|\psi(t)\rangle' \ . \tag{3.20}$$

because we need to have

$$|\langle\psi'|\psi'\rangle| = |\langle\psi(t)'|\psi(t)''\rangle| \ . \tag{3.21}$$

As Eq. (3.21) has to be true for all state vectors, the Hamilton operators $\hat{H}$ and $\hat{H}'$ can differ by only a constant which is physically unobservable (it gives rise to the same global $e^{i\phi}$ in all quantum states) and therefore the two Hamilton operators are essentially equal. Therefore we can conclude with

**Lemma 46** *A symmetry of a system S that holds for all times needs to be a unitary or anti-unitary transformation $\hat{T}$ that leaves the Hamilton operator $\hat{H}$ invariant, i.e.*

$$\hat{H} = \hat{H}' = \hat{T}\hat{H}\hat{T}^\dagger \ . \tag{3.22}$$

Given a symmetry group which can be parametrized by a parameter such that the symmetry transformations are differentiable with respect to this parameter (see the Example in this section above), we can determine the conserved observable $\hat{G}$ of the system quite easily from the following relation

$$\hat{G} = \lim_{\epsilon\to 0} i\frac{\hat{U}(\epsilon) - \mathbb{1}}{\epsilon} \ . \tag{3.23}$$

The quantity $\hat{G}$ is also called the **generator of the symmetry group**. Why is $\hat{G}$ conserved? From Eq. (3.23) we can see that for small $\epsilon$ we can write

$$\hat{U}(\epsilon) = \mathbb{1} - i\epsilon\hat{G} + O(\epsilon^2) \ , \tag{3.24}$$

where $\hat{G}$ is a Hermitean operator and $O(\epsilon^2)$ means that all other terms contain at least a factor of $\epsilon^2$ and are therefore very small. (see in the first chapter for the proof). Now we know that the Hamilton operator is invariant under the symmetry transformation, i.e.

$$\hat{H} = \hat{H}' = \hat{U}(\epsilon)\hat{H}\hat{U}^\dagger(\epsilon) \ . \tag{3.25}$$

For small values of $\epsilon$ we find

$$\begin{aligned} \hat{H} &= (\mathbb{1} - i\hat{G}\epsilon)\hat{H}(\mathbb{1} + i\hat{G}\epsilon) \\ &= \hat{H} - i[\hat{G}, \hat{H}]\epsilon + O(\epsilon^2) \ . \end{aligned} \tag{3.26}$$

As the equality has to be true for arbitrary but small $\epsilon$, this implies that

$$[\hat{G}, \hat{H}] = 0 \quad . \tag{3.27}$$

Now we remember the Heisenberg equation Eq. (3.13) and realize that the rate of change of the observable $\hat{G}$ in the Heisenberg picture is proportional to the commutator Eq. (3.27),

$$\frac{d\hat{G}_H}{dt} = i\hbar[\hat{G}, \hat{H}]_H = 0 \quad , \tag{3.28}$$

i.e. it vanishes. Therefore the expectation value of the observable $\hat{G}$ is constant, which amounts to say that $\hat{G}$ is a conserved quantity, i.e.

$$\frac{\langle\psi(t)|\hat{G}|\psi(t)\rangle}{dt} = \frac{\langle\psi(0)|\hat{G}_H(t)|\psi(0)\rangle}{dt} = 0 \quad . \tag{3.29}$$

Therefore we have the important

**Theorem 47** *The generator of a continuous symmetry group of a quantum system is a conserved quantity under the time evolution of that quantum system.*

After these abstract consideration let us now consider some examples of symmetry groups.

End of $15^{th}$ lecture

### 3.2.2 Translation Symmetry and momentum conservation

Let us now explore translations of quantum mechanical systems. Again we adopt the active viewpoint of the transformations. First we need to define what we mean by translating the system by a distance $a$ to the right. Such a transformation $\hat{T}_a$ has to map the state of the system $|\psi\rangle$ into $|\psi_a\rangle = \hat{T}_a|\psi\rangle$ such that

$$\langle x|\psi_a\rangle = \langle x - a|\psi\rangle \quad . \tag{3.30}$$

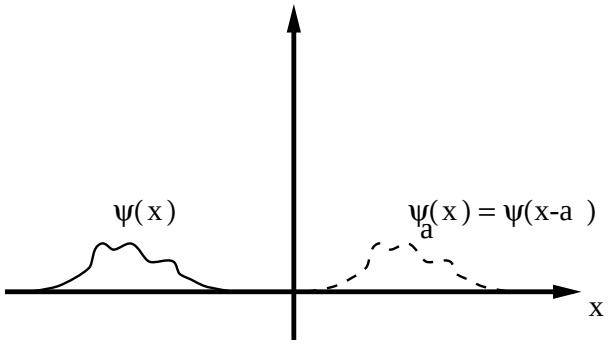


Figure 3.1: The original wave function $\psi(x)$ (solid line) and the shifted wave function $\psi_a(x) = \psi(x-a)$. $|\psi_a\rangle$ represents a system that has been shifted to the right by $a$.

In Fig. 3.1 you can see that this is indeed a shift of the wavefunction (i.e. the system) by $a$ to the right. Which operator represents the translation operator $\hat{T}_a$? To see this, we begin with the definition Eq. (3.30). Then we use the representation of the identity operator $\mathbb{1} = \int dp|p\rangle\langle p|$ (see Eq. (1.127)) and Eq. (1.124). We find

$$\begin{aligned}
\langle x|\hat{T}_a|\psi\rangle &= \langle x-a|\psi\rangle \\
&= \int dp\langle x-a|p\rangle\langle p|\psi\rangle \\
&= \int dp\frac{1}{\sqrt{2\pi\hbar}}e^{ip(x-a)/\hbar}\langle p|\psi\rangle \\
&= \int dp\frac{1}{\sqrt{2\pi\hbar}}e^{-ipa/\hbar}e^{ipx/\hbar}\langle p|\psi\rangle \\
&= \int dpe^{-ipa/\hbar}\langle x|p\rangle\langle p|\psi\rangle \\
&= \langle x|(\int dpe^{-ipa/\hbar}|p\rangle\langle p|)|\psi\rangle \\
&= \langle x|e^{-i\hat{p}a/\hbar}|\psi\rangle \quad . \qquad (3.31)
\end{aligned}$$

As this is true for any $|\psi\rangle$, we that the translation operator has the form

$$\hat{T}_a = e^{-i\hat{p}a/\hbar} \quad . \tag{3.32}$$

This means that the momentum operator is the generator of translations. From Eq. (3.23) it follows then that in a translation invariant system momentum is conserved, i.e.

$$\frac{d\langle\psi(t)|\hat{p}|\psi(t)\rangle}{dt} = 0 \quad .$$

Consider the Hamilton operator of a particle in a potential, which

$$\hat{H} = \frac{\hat{p}^2}{2m} + V(\hat{x}) \quad . \tag{3.33}$$

The momentum operator is obviously translation invariant, but a translation invariant potential has to satisfy $V(x) = V(x+a)$ for all $a$ which means that it is a constant.

### 3.2.3 Rotation Symmetry and angular momentum conservation

Another very important symmetry is rotation symmetry. To consider rotation symmetry it is necessary to go to three-dimensional systems. Let us first consider rotations $\hat{R}_z(\alpha)$ with an angle $\alpha$ around the z-axis. This is defined as

$$\psi_\alpha(r,\phi,\theta) = \psi(r,\phi-\alpha,\theta) \quad . \tag{3.34}$$

Here we have written the wavefunction in polar coordinates

$$x_1 = r\cos\phi\sin\theta \tag{3.35}$$
$$x_2 = r\sin\phi\sin\theta \tag{3.36}$$
$$x_3 = r\cos\theta \quad . \tag{3.37}$$

In Fig. 3.2 we can see that transformation Eq. (3.34) indeed rotates a wave-function by an angle $\alpha$ counter clockwise.

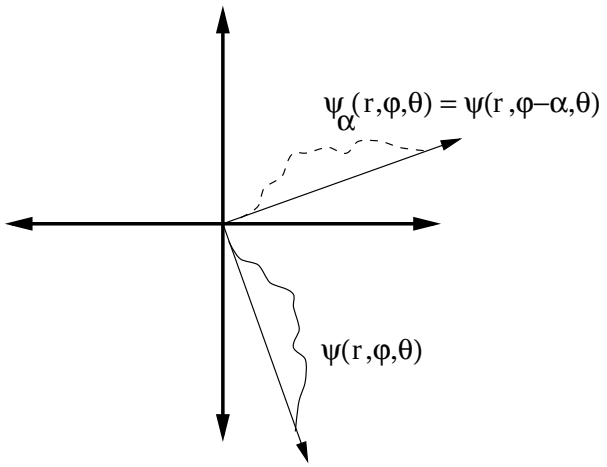


Figure 3.2: The original wave function $\psi(r,\phi,\theta)$ (solid line) and the rotated wave function $\psi_\alpha(r,\phi,\theta) = \psi(r,\phi-\alpha,\theta)$. $|\psi_\alpha\rangle$ represents a system that has been rotated clockwise around the $x_3 - axis$.

As translations are generated by the momentum operator, we expect rotations to be generated by the angular momentum. Let us confirm our suspicion. The angular momentum of a particle is given by

$$\hat{\vec{l}} = \hat{\vec{x}} \times \hat{\vec{p}} \tag{3.38}$$

$$= (\hat{x}_2\hat{p}_3 - \hat{x}_3\hat{p}_2)\vec{e}_1 + (\hat{x}_3\hat{p}_1 - \hat{x}_1\hat{p}_3)\vec{e}_2 + (\hat{x}_1\hat{p}_2 - \hat{x}_2\hat{p}_1)\vec{e}_3 \tag{3.39}$$

where $\times$ is the ordinary vector-product, $\hat{\vec{p}} = \hat{p}_1\vec{e}_1 + \hat{p}_2\vec{e}_2 + \hat{p}_3\vec{e}_3$ and $\hat{\vec{x}} = \hat{x}_1\vec{e}_1 + \hat{x}_2\vec{e}_2 + \hat{x}_3\vec{e}_3$ with $[\hat{x}_i, \hat{p}_j] = i\hbar\delta_{ij}$.

As we are considering rotations around the z-axis, i.e. $\vec{e}_3$, we will primarily be interested in the $\hat{l}_3$ component of the angular momentum operator. In the position representation we find

$$\hat{l}_3 = \frac{\hbar}{i}\left(x_1\frac{\partial}{\partial x_2} - x_2\frac{\partial}{\partial x_1}\right) \quad . \tag{3.40}$$

This can be converted to polar coordinates by using the chain rule

$$\frac{\partial}{\partial\phi} = \frac{\partial x_1}{\partial\phi}\frac{\partial}{\partial x_1} + \frac{\partial x_2}{\partial\phi}\frac{\partial}{\partial x_2}$$

$$
\begin{aligned}
&= \frac{\partial(r\cos\phi\sin\theta)}{\partial\phi}\frac{\partial}{\partial x_1} + \frac{\partial(r\sin\phi\sin\theta)}{\partial\phi}\frac{\partial}{\partial x_2} \\
&= -r\sin\phi\sin\theta\frac{\partial}{\partial x_1} + r\cos\phi\sin\theta\frac{\partial}{\partial x_2} \\
&= -x_2\frac{\partial}{\partial x_1} + x_1\frac{\partial}{\partial x_2} \\
&= \frac{i}{\hbar}\hat{l}_3 \quad . \qquad (3.41)
\end{aligned}
$$

Now we can proceed to find the operator that generates rotations by deriving a differential equation for it.

$$
\begin{aligned}
\langle\vec{x}|\frac{\partial R_z(\alpha)}{\partial\alpha}|\psi\rangle &= \frac{\partial}{\partial\alpha}\langle\vec{x}|R_z(\alpha)|\psi\rangle \\
&= \frac{\partial}{\partial\alpha}\langle r,\phi-\alpha,\theta|\psi\rangle \\
&= -\frac{\partial}{\partial\phi}\langle r,\phi-\alpha,\theta|\psi\rangle \\
&= -\frac{\partial}{\partial\phi}\langle\vec{x}|R_z(\alpha)|\psi\rangle \\
&= -\frac{i}{\hbar}\langle\vec{x}|\hat{l}_3 R_z(\alpha)|\psi\rangle \quad . \qquad (3.42)
\end{aligned}
$$

As this is true for all wave functions $|\psi\rangle$ we have the differential equation for the rotation operator

$$
\frac{\partial\hat{R}_z(\alpha)}{\partial\alpha} = -\frac{i}{\hbar}\hat{l}_3\hat{R}_z(\alpha) \qquad (3.43)
$$

With the initial condition $\hat{R}_z(0) = \mathbb{1}$ the solution to this differential equation is given by

$$
\hat{R}_z(\alpha) = e^{-i\hat{l}_3\alpha/\hbar} = e^{-i\hat{\vec{l}}\vec{e}_3\alpha/\hbar} \quad . \qquad (3.44)
$$

We obtain a rotation around an arbitrary axis $\vec{n}$ by

$$
\hat{R}_z(\alpha) = e^{-i\hat{\vec{l}}\vec{n}\alpha/\hbar} \quad . \qquad (3.45)
$$

Any system that is invariant under arbitrary rotations around an axis $\vec{n}$ preserves the component of the angular momentum in that direction, i.e. it preserves $\hat{\vec{l}}\vec{n}$. As the kinetic energy of a particle is invariant under any rotation (Check!) the angular momentum is preserved if the particle is in a potential that is symmetric under rotation around the axis $\vec{n}$. The electron in a hydrogen atom has the Hamilton operator

$$\hat{H} = \frac{\hat{p}^2}{2m} + V(|\hat{\vec{x}}|) \ . \tag{3.46}$$

Rotations leave the length of a vector invariant and therefore $|\hat{\vec{x}}|$ and with it $V(|\hat{\vec{x}}|)$ are invariant under rotations around any axis. This means that any component of the angular momentum operator is preserved, i.e. the total angular momentum is preserved.

End of $16^{th}$ lecture

## 3.3 General properties of angular momenta

In the preceding section we have investigated some symmetry transformations, the generators of these symmetry transformations and some of their properties. Symmetries are important and perhaps the most important of all is the concept of rotation symmetry and its generator, the angular momentum. In the previous section we have considered the specific example of the orbital angular momentum which we could develop from the classical angular momentum using the correspondence principal. However, there are manifestations of angular momentum in quantum mechanics that have no classical counterpart, the spin of an electron being the most important example. In the following I would like to develop the theory of the quantum mechanical angular momentum in general, introducing the notion of group representations.

### 3.3.1 Rotations

Whenever we aim to generalize a classical quantity to the quantum domain, we first need to investigate it carefully to find out those properties that are most characteristic of it. Then we will use these basic

properties as a definition which will guide us in our quest to find the correct quantum mechanical operator. Correctness, of course, needs to be tested experimentally, because nice mathematics does not necessarily describe nature although the mathematics itself might be consistent.

In the case of rotations I will not look at rotations about arbitrary angles, but rather at rotations about very small, infinitesimal angles. This will be most convenient in establishing some of the basic properties of rotations.

As we have done in all our discussion of symmetries, we will adopt the viewpoint of active rotations, i.e. the system is rotated while the coordinate system remains unchanged. A rotation around an axis $\mathbf{n}$ for a positive angle $\phi$ is one that follows the right hand rule (thumb in direction of the axis of rotation).

In three dimensional space rotations are given by real $3\times3$ matrices. A rotation around the x-axis with an angle $\phi$ is given by

$$\mathcal{R}_x(\phi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{pmatrix} \quad . \tag{3.47}$$

The rotations around the y-axis is given by

$$\mathcal{R}_y(\phi) = \begin{pmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{pmatrix} \tag{3.48}$$

and the rotation about the z-axis is given by

$$\mathcal{R}_z(\phi) = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad . \tag{3.49}$$

Using the expansions for sin and cos for small angles

$$\sin\epsilon = \epsilon + O(\epsilon^3) \qquad \cos\epsilon = 1 - \frac{\epsilon^2}{2} \quad , \tag{3.50}$$

we may now expand the rotation matrices Eqs. (3.47-3.49) up to second order in the small angle $\epsilon$. We find

$$\mathcal{R}_x(\epsilon) \;=\; \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1-\frac{\epsilon^2}{2} & -\epsilon \\ 0 & \epsilon & 1-\frac{\epsilon^2}{2} \end{pmatrix} \quad . \tag{3.51}$$

$$\mathcal{R}_y(\epsilon) = \begin{pmatrix} 1-\frac{\epsilon^2}{2} & 0 & \epsilon \\ 0 & 1 & 0 \\ -\epsilon & 0 & 1-\frac{\epsilon^2}{2} \end{pmatrix} \tag{3.52}$$

$$\mathcal{R}_z(\epsilon) = \begin{pmatrix} 1-\frac{\epsilon^2}{2} & -\epsilon & 0 \\ \epsilon & 1-\frac{\epsilon^2}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} . \tag{3.53}$$

We know already from experience, that rotations around the same axis commute, while rotations around different axis generally do not commute. This simple fact will allow us to derive the canonical commutation relations between the angular momentum operators.

To see the effect of the non-commutativity of rotations for different axis, we compute a special combination of rotations. First rotate the system about an infinitesimal angle $\epsilon$ about the y-axis and then by the same angle about the x-axis

$$\begin{aligned} \mathcal{R}_x(\epsilon)\mathcal{R}_y(\epsilon) &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1-\frac{\epsilon^2}{2} & -\epsilon \\ 0 & \epsilon & 1-\frac{\epsilon^2}{2} \end{pmatrix} \begin{pmatrix} 1-\frac{\epsilon^2}{2} & 0 & \epsilon \\ 0 & 1 & 0 \\ -\epsilon & 0 & 1-\frac{\epsilon^2}{2} \end{pmatrix} \\ &= \begin{pmatrix} 1-\frac{\epsilon^2}{2} & 0 & \epsilon \\ \epsilon^2 & 1-\frac{\epsilon^2}{2} & -\epsilon \\ -\epsilon & \epsilon & 1-\epsilon^2 \end{pmatrix} . \end{aligned} \tag{3.54}$$

Now we calculate

$$\begin{aligned} &\mathcal{R}_x(-\epsilon)\mathcal{R}_y(-\epsilon)\mathcal{R}_x(\epsilon)\mathcal{R}_y(\epsilon) = \\ &= \begin{pmatrix} 1-\frac{\epsilon^2}{2} & 0 & -\epsilon \\ \epsilon^2 & 1-\frac{\epsilon^2}{2} & \epsilon \\ \epsilon & -\epsilon & 1-\epsilon^2 \end{pmatrix} \begin{pmatrix} 1-\frac{\epsilon^2}{2} & 0 & \epsilon \\ \epsilon^2 & 1-\frac{\epsilon^2}{2} & -\epsilon \\ -\epsilon & \epsilon & 1-\epsilon^2 \end{pmatrix} . \\ &= \begin{pmatrix} 1 & -\epsilon^2 & 0 \\ \epsilon^2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} . \\ &= \mathcal{R}_z(\epsilon^2) \end{aligned} \tag{3.55}$$

This will be the relation that we will use to derive the commutation relation between the angular momentum operators.

### 3.3.2 Group representations and angular momentum commutation relations

We already know that there are many different angular momenta in quantum mechanics, orbital angular momentum and spin are just two examples. Certainly they cannot necessarily represented by $3 \times 3$ matrices as in the rotations in the previous section. Nevertheless, all these angular momenta have some common properties that justify their classification as angular momenta. This will lead us to a discussion of groups and their representations.

We have already seen in the first section on vector spaces what a group is. Let us repeat the properties of a group $G$ and its elements $x, y, z$.

1. $\forall x, y \in G : x \cdot y = z \in G$
2. $\exists 1 \in G : \forall x \in G : x \cdot 1 = x$
3. $\forall x \in G : \exists x^{-1} \in G : x^{-1} \cdot x = 1 \,\text{and}\, x \cdot x^{-1} = 1$
4. $\forall x, y, z \in G : x \cdot (y \cdot z) = (x \cdot y) \cdot z$

Note that we did not demand that the group elements commute under the group operation. The reason is that we intend to investigate rotations which are surely not commutative. The fundamental group that we are concerned with is the group of rotations in three-dimensional space, i.e. the group of $3 \times 3$ matrices. It represents all our intuitive knowledge about rotations. Every other group of operators on wavefunctions that we would like to call a group of rotations will have to share the basic properties of the group of three-dimensional rotations. This idea is captured in the notion of a group representation.

**Definition 48** *A representation of a group $G$ of elements $x$ is a group $\mathcal{S}$ of unitary operators $\mathcal{D}(x)$ such that there is a map $\mathcal{T} : G \to \mathcal{S}$ associating with every $x \in G$ an operator $\mathcal{D}(x) \in \mathcal{S}$ such that the group operation is preserved, i.e. for all $x, y \in G$ with $x \cdot y = z$ we have*

$$\mathcal{D}(x)\mathcal{D}(y) = \mathcal{D}(z) \quad . \tag{3.56}$$

This means that both groups $G$ and $\mathcal{S}$ are essentially equal because their elements share the same relations between each other. If we now have a representation of the rotation group then it is natural to say that the operators that make up this representation are rotation operators.

In quantum mechanics a rotation has to be represented by a unitary operator acting on a state vector. For rotations around a given axis $\mathbf{n}$ these unitary operators have to form a set of operators $\hat{U}(\phi)$ that is parametrized by a single parameter, the angle $\phi$. As we know that the operators are unitary we know that they have to be of the form

$$\hat{U}(\phi) = e^{-i\hat{J}_{\mathbf{n}}\phi/\hbar}$$

with a Hermitean operator $\hat{J}_n$ and where we have introduced $\hbar$ for convenience. The operator is the angular momentum operator along direction $\mathbf{n}$. In general we can define the angular momentum operator by $\hat{\vec{J}} = \hat{J}_x\vec{e}_x + \hat{J}_y\vec{e}_y + \hat{J}_z\vec{e}_z$ which yields $\hat{J}_n = \hat{\vec{J}} \cdot \vec{n}$. Now let us find out what the commutation relations between the different components of the angular momentum operator are. We use the fact that we have a representation of the rotation group, i.e. Eq. (3.56) is satisfied. Eq. (3.55) then implies that for an infinitesimal rotation we have the operator equation

$$e^{i\hat{J}_x\epsilon/\hbar}e^{i\hat{J}_y\epsilon/\hbar}e^{-i\hat{J}_x\epsilon/\hbar}e^{-i\hat{J}_y\epsilon/\hbar} = e^{-i\hat{J}_z\epsilon^2/\hbar} \quad .$$

Expanding both sides to second order in $\epsilon$, we find

$$\mathbb{1} + [\hat{J}_y, \hat{J}_x]\epsilon^2/\hbar^2 = \mathbb{1} - i\hat{J}_z\epsilon^2/\hbar$$

which gives rise to the commutation relation

$$[\hat{J}_x, \hat{J}_y] = i\hbar\hat{J}_z \quad . \tag{3.57}$$

In the same way one can obtain the commutation relations between any other component of the angular momentum operator, which results in

$$[\hat{J}_i, \hat{J}_j] = i\hbar\epsilon_{ijk}\hat{J}_k \quad . \tag{3.58}$$

This fundamental commutation relation is perhaps the most important feature of the quantum mechanical angular momentum and can be used as its definition.

———— End of $17^{th}$ lecture

### 3.3.3 Angular momentum eigenstates

Using the basic commutation relation of angular momentum operators Eq. (3.58), we will now derive the structure of the eigenvectors of the angular momentum operators. Eq. (3.58) already shows that we are not able to find simultaneous eigenvectors to all the three components of the angular momentum. However, the square of the magnitude of the angular momentum $\hat{\vec{J}}^2 = \hat{J}_x^2 + \hat{J}_y^2 + \hat{J}_z^2$ commutes with each of its components. In the following we will concentrate on the z-component of the angular momentum, for which we find

$$[\hat{\vec{J}}^2, \hat{J}_z] = 0 \quad . \tag{3.59}$$

As the two operators commute, we can find simultaneous eigenvectors for them, which we denote by $|j, m\rangle$. We know that $\hat{\vec{J}}^2$ is a positive operator and for convenience we chose

$$\hat{\vec{J}}^2 |j, m\rangle = \hbar^2 j(j+1)|j, m\rangle \quad . \tag{3.60}$$

By varying $j$ from zero to infinity the expression $j(j+1)$ can take any positive value from zero to infinity. The slightly peculiar form for the eigenvalues of $\hat{\vec{J}}^2$ has been chosen to make all the following expressions as transparent as possible. $\hat{J}_z$ is not know to be positive, in fact it isn't, and we have the eigenvalue equation

$$\hat{J}_z |j, m\rangle = \hbar m |j, m\rangle \quad . \tag{3.61}$$

Given one eigenvector for the angular momentum we would like to have a way to generate another one. This is done by the ladder operators for the angular momentum. They are defined as

$$\hat{J}_\pm = \hat{J}_x \pm i\hat{J}_y \quad . \tag{3.62}$$

To understand why they are called ladder operators, we first derive their commutation relations with $\hat{J}_z$. They are

$$[\hat{J}_z, \hat{J}_\pm] = \pm\hbar \hat{J}_\pm \quad . \tag{3.63}$$

Using these commutation relations, we find

$$\begin{aligned} \hat{J}_z(\hat{J}_+|j,m\rangle) &= (\hat{J}_+\hat{J}_z + \hbar\hat{J}_+)|j,m\rangle \\ &= (\hat{J}_+\hbar m + \hbar\hat{J}_+)|j,m\rangle \\ &= \hbar(m+1)\hat{J}_+|j,m\rangle \ . \end{aligned} \tag{3.64}$$

This implies that

$$(\hat{J}_+|j,m\rangle) = \alpha_+(j,m)|j,m+1\rangle \ , \tag{3.65}$$

i.e. $\hat{J}_+$ generates a new eigenvector of $\hat{J}_z$ with an eigenvalue increased by one unit and $\alpha_+(j,m)$ is a complex number. Likewise we find that

$$(\hat{J}_-|j,m\rangle) = \alpha_-(j,m)|j,m-1\rangle \ , \tag{3.66}$$

i.e. $\hat{J}_-$ generates a new eigenvector of $\hat{J}_z$ with an eigenvalue decreased by one unit and $\alpha_-(j,m)$ is a complex number. Now we have to establish the limits within which $m$ can range. Can $m$ take any value or will there be bounds given by $j$? To answer this we need consider the positive operator

$$\hat{\vec{J}}^2 - \hat{J}_z^2 = \hat{J}_x^2 + \hat{J}_y^2 \ . \tag{3.67}$$

Applying it to an eigenvector, we obtain

$$\langle j,m|(\hat{\vec{J}}^2 - \hat{J}_z^2)|j,m\rangle = \hbar^2(j(j+1) - m^2) \geq 0 \ . \tag{3.68}$$

We conclude that the value of $m$ is bounded by

$$|m| \leq \sqrt{j(j+1)} \ . \tag{3.69}$$

While from Eqs. (3.65-3.66) we can see that the ladder operators allow us to increase and decrease the value of $m$ by one unit, Eq. (3.69) implies that there has to be a limit to this increase and decrease. In fact, this implies that for the largest value $m_{max}$ we have

$$\hat{J}_+|j,m_{max}\rangle = 0|j,m_{max}+1\rangle = 0 \ , \tag{3.70}$$

while for the smallest value $m_{min}$ we have

$$\hat{J}_-|j,m_{min}\rangle = 0|j,m_{min}-1\rangle = 0 \ . \tag{3.71}$$

This implies that we have

$$\hat{J}_-\hat{J}_+|j,m_{max}\rangle = 0 \ , \tag{3.72}$$
$$\hat{J}_+\hat{J}_-|j,m_{min}\rangle = 0 \ . \tag{3.73}$$
$$\tag{3.74}$$

To determine $m_{max}$ and $m_{min}$ we need to express $\hat{J}_-\hat{J}_+$ and $\hat{J}_+\hat{J}_-$ in terms of the operators $\hat{\vec{J}}^2$ and $\hat{J}_z$. This can easily be done by direct calculation which gives

$$\hat{J}_+\hat{J}_- = (\hat{J}_x+i\hat{J}_y)(\hat{J}_x-i\hat{J}_y) = \hat{J}_x^2+\hat{J}_y^2-i[\hat{J}_x,\hat{J}_y] = \hat{\vec{J}}^2-\hat{J}_z^2+\hbar\hat{J}_z \ , \tag{3.75}$$

and

$$\hat{J}_-\hat{J}_+ = (\hat{J}_x-i\hat{J}_y)(\hat{J}_x+i\hat{J}_y) = \hat{J}_x^2+\hat{J}_y^2+i[\hat{J}_x,\hat{J}_y] = \hat{\vec{J}}^2-\hat{J}_z^2-\hbar\hat{J}_z \ . \tag{3.76}$$

Using these expressions we find for $m_{max}$

$$\begin{aligned} 0 &= \hat{J}_-\hat{J}_+|j,m_{max}\rangle = (\hat{\vec{J}}^2 - \hat{J}_z^2 - \hbar\hat{J}_z)|j,m_{max}\rangle \\ &= \hbar^2(j(j+1) - m_{max}(m_{max}+1))|j,m_{max}\rangle \ . \end{aligned} \tag{3.77}$$

This implies that

$$m_{max} = j \ . \tag{3.78}$$

Likewise we find

$$\begin{aligned} 0 &= \hat{J}_+\hat{J}_-|j,m_{min}\rangle = (\hat{\vec{J}}^2 - \hat{J}_z^2 + \hbar\hat{J}_z)|j,m_{min}\rangle \\ &= \hbar^2(j(j+1) + m_{min}(1-m_{min}))|j,m_{min}\rangle \ . \end{aligned} \tag{3.79}$$

This implies that

$$m_{min} = -j \ . \tag{3.80}$$

We have to be able to go in steps of one unit from the maximal value $m_{max}$ to $m_{min}$. This implies that $2j$ is a whole number and therefore we find that

$$j = \frac{n}{2} \ \text{ with } \ n \in N \ . \tag{3.81}$$

Every real quantum mechanical particle can be placed in one of two classes. Either it has integral angular momentum, $0, 1, \ldots$ (an example

is the orbital angular momentum discussed in the previous section) and is called a boson, or it has half-integral angular momentum $\frac{1}{2}, \frac{3}{2}, \ldots$ (the electron spin is an example for such a particle) and is called a fermion.

Finally we would like to determine the constant $\alpha_\pm(j,m)$ in

$$\hat{J}_\pm|j,m\rangle = \alpha_\pm(j,m)|j,m\pm 1\rangle \quad . \tag{3.82}$$

This is easily done by calculating the norm of both sides of Eq. (3.82).

$$\begin{aligned} |\alpha_+(j,m)|^2 &= ||\hat{J}_+|j,m\rangle||^2 = \langle j,m|(\hat{\vec{J}}^2 - \hat{J}_z^2 - \hbar\hat{J}_z)|j,m\rangle \\ &= \hbar(j(j+1) - m(m+1)) \quad . \end{aligned} \tag{3.83}$$

Analogously we find

$$\begin{aligned} |\alpha_-(j,m)|^2 &= ||\hat{J}_-|j,m\rangle||^2 = \langle j,m|(\hat{\vec{J}}^2 - \hat{J}_z^2 + \hbar\hat{J}_z)|j,m\rangle \\ &= \hbar(j(j+1) - m(m-1)) \quad . \end{aligned} \tag{3.84}$$

We chose the phase of the states $|j,m\rangle$ in such a way that $\alpha_\pm(j,m)$ is always positive, so that we obtain

$$\hat{J}_+|j,m\rangle = \hbar\sqrt{j(j+1) - m(m+1)}|j,m+1\rangle \tag{3.85}$$

$$\hat{J}_-|j,m\rangle = \hbar\sqrt{j(j+1) - m(m-1)}|j,m-1\rangle \quad . \tag{3.86}$$

End of $18^{th}$ lecture

## 3.4 Addition of Angular Momenta

In the previous section I have reviewed the properties of the angular momentum of a single particle. Often, however, you are actually holding a quantum system consisting of more than one particle, e.g. a hydrogen atom, and you may face a situation where more than one of those particles possesses angular momentum. Even a single particle may have more than one angular momentum. An electron in a central potential, for example, has an **orbital** angular momentum as well as an **internal** angular momentum, namely the spin. Why do we need to

add these angular momenta? To see this, consider the example of an electron in a central potential.

The Hamilton operator of a particle in a central potential is given by

$$\hat{H}_0 = \frac{\hat{\vec{p}}^2}{2m} + V(|\hat{\vec{x}}|) \quad , \tag{3.87}$$

where $\hat{\vec{p}}$ is the linear momentum of the particle, $m$ its mass and $V(|\hat{\vec{x}}|)$ is the operator describing the central potential. From subsection 3.2.3 we know that under this Hamilton operator the orbital angular momentum is preserved because the Hamilton operator is invariant under rotations around an arbitrary axis. However, if you do experiments, then you will quickly realize that the Hamilton operator $\hat{H}_0$ is only the first approximation to the correct Hamilton operator of an electron in a central potential (you may have heard about that in the 2nd year atomic physics course). The reason is, that the electron possesses an internal angular momentum, the spin, and therefore there have to be additional terms in the Hamilton operator. These additional terms can be derived from a relativistic theory of the electron (Dirac equation), but here I will just give a heuristic argument for one of them because the full description of the Dirac equation would take far too long.) Intuitively, an electron in a central potential rotates around the origin and therefore creates a circular current. Such a circular current gives rise to a magnetic field. The spin of an electron on the other hand gives rise to a magnetic moment of the electron whose orientation depends on the orientation of the spin - which is either up or down. This means that the electron, has different energies in a magnetic field $\vec{B}$ depending on the orientation of its spin. This energy is given by $-\mu_0 \hat{\vec{S}}\vec{B}$, where $-\mu_0 \vec{S}$ is the magnetic moment of the electron and $\hat{\vec{S}} = \hbar(\sigma_1\vec{e}_1 + \sigma_2\vec{e}_2 + \sigma_3\vec{e}_3)$ is the electron spin operator. The magnetic field created by the rotating electron is proportional to the orbital angular momentum (the higher the angular momentum, the higher the current induced by the rotating electron and therefore the higher the magnetic field) so that we find that the additional part in the Hamilton operator Eq. (3.87) is given by

$$\hat{H}_1 = -\xi \hat{\vec{S}} \otimes \hat{\vec{L}} = \xi \left( \hat{S}_1 \otimes \hat{L}_1 + \hat{S}_2 \otimes \hat{L}_2 + \hat{S}_3 \otimes \hat{L}_3 \right) \quad , \tag{3.88}$$

where $\xi$ is a constant. Note that the operators $\hat{\vec{S}}$ and $\hat{\vec{L}}$ act on two different Hilbert spaces, $\hat{\vec{S}}$ acts on the 2-dimensional Hilbert space of the electron spin and $\hat{\vec{L}}$ on the Hilbert space describing the motion of the electron.

Under the Hamilton operator Eq. (3.87) the spin was a constant of motion as the spin operator $\hat{\vec{S}}$ evidently commutes with the Hamiltonian Eq. (3.87), i.e. $[\hat{H}_0, \hat{S}_i] = 0$. For the total Hamilton operator $\hat{H} = \hat{H}_0 + \hat{H}_1$, however, **neither** the orbital angular momentum $\hat{\vec{L}}$ nor the spin angular momentum $\hat{\vec{S}}$ are constants of motion anymore. This can easily be seen by checking that now the commutators $[\hat{L}_i, \hat{H}]$ and $[\hat{S}_i, \hat{H}]$ are non-vanishing! See for example that

$$\begin{aligned} [\hat{L}_1, \hat{H}] &= [\hat{L}_1, \hat{H}_1] \\ &= [\hat{L}_1, -\xi \hat{\vec{S}} \cdot \hat{\vec{L}}] \\ &= -\xi [\hat{L}_1, \sum_{i=1}^{3} \hat{S}_i \cdot \hat{L}_i] \\ &= -i\hbar\xi \hat{S}_2 \hat{L}_3 + i\hbar\xi \hat{S}_3 \hat{L}_2 \end{aligned} \tag{3.89}$$

and

$$\begin{aligned} [\hat{S}_1, \hat{H}] &= [\hat{S}_1, \hat{H}_1] \\ &= -\xi [\hat{S}_1, \sum_{i=1}^{3} \hat{S}_i \cdot \hat{L}_i] \\ &= -\xi i\hbar \hat{S}_3 \hat{L}_2 + \xi i\hbar \hat{S}_2 \hat{L}_3 \end{aligned}$$

which both are evidently non-zero. However, the sum of the two spin operators $\hat{\vec{J}} = \hat{\vec{L}} + \hat{\vec{S}}$ is a conserved quantity because all its components commute with the Hamiltonian, i.e.

$$[\hat{L}_i + \hat{S}_i, \hat{H}] = 0 \quad . \tag{3.90}$$

You may check this easily for the first component of $\hat{J}_1$ using the two commutators Eqs. (3.89-3.90).

We know a basis of eigenstates to the orbital angular momentum and also one for the spin angular momentum. However, as the angular momenta separately are not conserved anymore such a choice

of eigenstates is quite inconvenient because eigenstates to the angular momenta $\hat{\vec{L}}$ and $\hat{\vec{S}}$ are not eigenstates of the total Hamilton operator $\hat{H} = \hat{H}_0 + \hat{H}_1$ anymore. Of course it would be much more convenient to **find a basis that is composed of simultaneous eigenvectors of the total angular momentum and the Hamilton operator**. This is the aim of the following subsection which describes a general procedure how one can construct the eigenvectors of the total angular momentum from the eigenstates of orbital angular momentum and spin. As different components of the total angular momentum do not commute, we can of course only find joint eigenstates of the total angular momentum operator $\hat{\vec{J}}^2$, its z-component $\hat{J}_z$ and the total Hamilton operator $\hat{H}_0 + \hat{H}_1$.

### 3.4.1 Two angular momenta

First I will present the general idea behind the addition of two angular momenta represented by the operators $\hat{\vec{j}}^{(1)}$ for the first particle and $\hat{\vec{j}}^{(2)}$ for the second particle. Then I will give the simplest possible explicit example. Note that I am writing $\hat{\vec{j}}^{(1)}$ instead of the more precise $\hat{\vec{j}}^{(1)} \otimes \mathbb{1}$ to shorten the equations a little bit.

Now I would like to consider the total angular momentum

$$\hat{\vec{J}} := \hat{\vec{j}}^{(1)} + \hat{\vec{j}}^{(2)} \quad , \tag{3.91}$$

Here the upper bracketed index indicates on which particle the operator is acting.

Like in the previous section $\hat{\vec{J}}^2$ commutes with each of its components. (Check e.g. that $[\hat{\vec{J}}^2, \hat{J}_z] = [\hat{\vec{J}}^2, \hat{j}_z^{(1)} + \hat{j}_z^{(2)}] = 0$). Therefore I would like to find a set of joint eigenvectors of $\hat{\vec{J}}^2$ and the z-component of the total angular momentum $\hat{J}_z$. These eigenvectors which are written as $|J, M\rangle$ have to satisfy

$$\begin{aligned} \hat{\vec{J}}^2 |J, M\rangle &= \hbar J(J+1)|J, M\rangle \\ \hat{J}_z |J, M\rangle &= \hbar M |J, M\rangle \quad . \end{aligned}$$

As the eigenstates $|j^{(1)}, m^{(1)}; j^{(2)}, m^{(2)}\rangle$ of the separate angular momenta $\hat{j}^{(1)}$ and $\hat{j}^{(2)}$ form a basis, we have to be able to write $|J, M\rangle$ as linear combinations of the eigenvectors to the individual angular momenta $|j^{(1)}, m^{(1)}; j^{(2)}, m^{(2)}\rangle$ of which there are $(2j^{(1)}+1)(2j^{(2)}+1)$ states. It might seem difficult to find these linear combinations, but luckily there is a general recipe which I am going to explain in the following.

In the previous section we constructed all the possible angular momentum eigenstates of a single particle by starting with the state with highest m-quantum number and then working our way down by applying the angular momentum ladder operator $\hat{j}_- = \hat{j}_x - i\hat{j}_y$. We will apply an analogous strategy for the total angular momentum $\hat{\vec{J}}$. Firstly we need to identify the ladder operators for the total angular momentum. They are

$$\hat{J}_- = \hat{j}_-^{(1)} + \hat{j}_-^{(2)} \tag{3.92}$$

$$\hat{J}_+ = \hat{j}_+^{(1)} + \hat{j}_+^{(2)} \quad . \tag{3.93}$$

Let us check whether these operators satisfy commutation relations analogous to Eq. (3.63).

$$\begin{aligned} [\hat{J}_z, \hat{J}_\pm] &= [\hat{j}_z^{(1)} + \hat{j}_z^{(2)}, \hat{j}_\pm^{(1)} + \hat{j}_\pm^{(2)}] \\ &= [\hat{j}_z^{(1)}, \hat{j}_\pm^{(1)}] + [\hat{j}_z^{(2)}, \hat{j}_\pm^{(2)}] \\ &= \pm\hbar\hat{j}_\pm^{(1)} \pm \hbar\hat{j}_\pm^{(2)} \\ &= \pm\hbar\hat{J}_\pm \quad . \end{aligned} \tag{3.94}$$

The commutation relation Eq. (3.94) allows us, just as in the case of a single angular momentum, to obtain the state $|J, M-1\rangle$ from the state $|J, M\rangle$ via

$$\hat{J}_-|J, M\rangle = \hbar\sqrt{J(J+1) - M(M-1)}|J, M-1\rangle \quad . \tag{3.95}$$

This relation can be derived in exactly the same way as relation Eq. (3.86). The procedure for generating all the states $|J, M\rangle$ has three steps and starts with

**Step 1:** Identify the state with maximal total angular momentum and maximal value for the $M$-quantum number. If we have two angular

momenta $j^{(1)}$ and $j^{(2)}$ then their sum can not exceed $J_{max} = j^{(1)} + j^{(2)}$. Which state could have total angular momentum $J_{max} = j^{(1)} + j^{(2)}$ and the $M = J_{max}$? Clearly this must be a combination of two angular momenta that are both parallel, i.e. we guess they should be in the state

$$|J = j^{(1)} + j^{(2)}, M = j^{(1)} + j^{(2)}\rangle = |j^{(1)}, j^{(1)}\rangle \otimes |j^{(2)}, j^{(2)}\rangle \ . \qquad (3.96)$$

To check whether this assertion is true we need to verify that

$$\begin{aligned} \hat{\vec{J}}^2 |J, M\rangle &= \hbar^2 J(J+1)|J, M\rangle \\ \hat{J}_z |J, M\rangle &= \hbar M |J, M\rangle \ , \end{aligned}$$

with $J = M = j^{(1)} + j^{(2)}$. I will not present this proof here, but will let you do it in the problem sheets. Now follows

**Step 2:** Apply the ladder operator defined in Eq. (3.92) using Eq. (3.95) to obtain all the states $|J, M\rangle$. Repeat this step until you reached state $|J, M = -J\rangle$.

However, these are only $2J_{max} + 1$ states which is less than the total of $(2j_1 + 1)(2j_2 + 1)$ states of the joint state space of both angular momenta. To obtain the other states we apply

**Step 3:** Identify the state that is of the form $|J - 1, M = J - 1\rangle$ and then go to step 2. The procedure stops when we have arrived at the smallest possible value of $J$, which is obtained when the two angular momenta are oriented in opposite direction so that we need to subtract them. This implies that $J_{min} = |j^{(1)} - j^{(2)}|$. The absolute value needs to be taken, as the angular momentum is by definition a positive quantity.

Following this strategy we obtain a new basis of states consisting of $\sum_{i=J_{min}}^{J_{max}} (2i+1) = \sum_{i=0}^{J_{max}} 2i + 1 - \sum_{i=0}^{J_{min}-1} 2i + 1 = (J_{max}+1)^2 - J_{min}^2 = (j^{(1)} + j^{(2)} + 1)^2 - (j^{(1)} - j^{(2)})^2 = (2j^{(1)} + 1)(2j^{(2)} + 1)$ states.

As this is a rather formal description I will now give the simplest possible example for the addition of two angular momenta.

**Example:**

Now I am going to illustrate this general idea by solving the explicit example of adding two spin-$\frac{1}{2}$.

The angular momentum operators for a spin-$\frac{1}{2}$ particle are

$$\hat{j}_x = \hbar\hat{\sigma}_x \quad \hat{j}_y = \hbar\hat{\sigma}_y \quad \hat{j}_z = \hbar\hat{\sigma}_z \tag{3.97}$$

with the Pauli spin-operators $\hat{\sigma}_i$. It is easy to check that this definition satisfies the commutation relations for an angular momentum. The angular momentum operator of an individual particle is therefore given by $\hat{\vec{j}} = \hbar\left(\hat{\sigma}_x\vec{e}_x + \hat{\sigma}_y\vec{e}_y + \hat{\sigma}_z\vec{e}_z\right)$ and the total angular momentum operator is given by

$$\hat{\vec{J}} = \hat{\vec{j}}^{(1)} + \hat{\vec{j}}^{(2)} \quad . \tag{3.98}$$

The angular momentum ladder operators are

$$\begin{aligned} \hat{J}_- &= \hat{j}_-^{(1)} + \hat{j}_-^{(2)} \\ \hat{J}_+ &= \hat{j}_+^{(1)} + \hat{j}_+^{(2)} \quad . \end{aligned}$$

The maximal value for the total angular momentum is $J = 1$. The state corresponding to this value is given by

$$|J=1, M=1\rangle \;=\; |j^{(1)} = \frac{1}{2}, m^{(1)} = \frac{1}{2}\rangle \otimes |j^{(2)} = \frac{1}{2}, m^{(2)} = \frac{1}{2}\rangle \quad , \tag{3.99}$$

or using the shorthand

$$|\uparrow\rangle \;=\; |j^{(1)} = \frac{1}{2}, m^{(1)} = \frac{1}{2}\rangle \tag{3.100}$$

$$|\downarrow\rangle \;=\; |j^{(1)} = \frac{1}{2}, m^{(1)} = -\frac{1}{2}\rangle \tag{3.101}$$

we have

$$|J=1, M=1\rangle = |\uparrow\rangle \otimes |\uparrow\rangle \quad .$$

Now we want to find the representation of the state $|J=1, M=0\rangle$ using the operator $\hat{J}_-$. We find

$$\begin{aligned} |J=1, M=0\rangle &= \frac{\hat{J}_-|J=1, M=1\rangle}{\hbar\sqrt{2}} \\ &= \frac{(\hat{j}_-^{(1)} + \hat{j}_-^{(2)})|\uparrow\rangle \otimes |\uparrow\rangle}{\hbar\sqrt{2}} \\ &= \frac{|\downarrow\rangle \otimes |\uparrow\rangle}{\sqrt{2}} + \frac{|\uparrow\rangle \otimes |\downarrow\rangle}{\sqrt{2}} \end{aligned} \tag{3.102}$$

To find the state $|J=1, M=-1\rangle$ we apply $\hat{J}_-$ again and we find that

$$\begin{aligned} |J=1, M=-1\rangle &= \frac{\hat{J}_-|J=1, M=0\rangle}{\hbar\sqrt{2}} \\ &= \frac{(\hat{j}_-^{(1)} + \hat{j}_-^{(2)})|J=1, M=0\rangle}{\hbar\sqrt{2}} \\ &= |\downarrow\rangle \otimes |\downarrow\rangle \end{aligned} \tag{3.103}$$

Now we have almost finished our construction of the eigenstates of the total angular momentum operator. What is still missing is the state with total angular momentum $J=0$. Because $J=0$ this state must also have $M=0$.The state $|J=0, M=0\rangle$ must be orthogonal to the three states $|J=1, M=1\rangle, |J=1, M=0\rangle, |J=1, M=-1\rangle$. Therefore the state must have the form

$$|J=0, M=0\rangle = \frac{|\downarrow\rangle \otimes |\uparrow\rangle}{\sqrt{2}} - \frac{|\uparrow\rangle \otimes |\downarrow\rangle}{\sqrt{2}} \quad . \tag{3.104}$$

To check that this is the correct state, just verify that it is orthogonal to $|J=1, M=1\rangle, |J=1, M=0\rangle, |J=1, M=-1\rangle$. This concludes the construction of the eigenvectors of the total angular momentum.

End of $19^{th}$ lecture

# Chapter 4

# Approximation Methods

Most problems in quantum mechanics (like in all areas of physics) cannot be solved analytically and we have to resort to approximation methods. In this chapter I will introduce you to some of these methods. However, there are a large number of approximation methods in physics and I can only present very few of them in this lecture.

Often a particular problem will be in a form that it is exactly solvable would it not be for a small additional term (a perturbation) in the Hamilton operator. This perturbation may be both, time-independent (e.g. a static electric field) or time-dependent (a laser shining on an atom). It is our task to compute the effect that such a small perturbation has. Historically these methods originated from classical mechanics when in the 18th century physicists such as Lagrange tried to calculate the mechanical properties of the solar system. In particular he and his colleagues have been very interested in the stability of the earths orbit around the sun. This is a prime example of a problem where we can solve part of the dynamics exactly, the path of a single planet around the sun, while we are unable to solve this problem exactly once we take into account the small gravitational effects due to the presence of all the other planets. There are quite a few such problems in quantum mechanics. For example we are able to solve the Schrödinger equation for the hydrogen atom, but if we place two hydrogen atoms maybe 100 Angstrom away from each other then the problem has no exact solution anymore. The atoms will start to perturb each other and a weak perturbation to their dynamics will be the effect. While not be-

ing strong enough to form a molecule this effect leads to an attractive force, the van der Waals force. This force, although being quite weak, is responsible for the sometimes remarkable stability of foams.

## 4.1 Time-independent Perturbation Theory

In the second year quantum mechanics course you have seen a method for dealing with time independent perturbation problems. I will rederive these results in a more general, shorter and elegant form. Then I will apply them to derive the van der Waals force between two neutral atoms.

### 4.1.1 Non-degenerate perturbation theory

Imagine that you have a system with a Hamilton operator $\hat{H}_0$, eigenvectors $|\phi_i\rangle$ and corresponding energies $\epsilon_i$. We will now assume that we can solve this problem, i.e. we know the expressions for the eigenstates $|\phi_i\rangle$ and energies $\epsilon_i$ already. Now imagine a small perturbation in the form of the operator $\lambda\hat{V}$ is added to the Hamilton operator. The real parameter $\lambda$ can be used to count the order to which we perform perturbation theory. Now we have the new total Hamilton operator $\hat{H} = \hat{H}_0 + \lambda\hat{V}$ with the new eigenvectors $|\psi_i(\lambda)\rangle$ and energies $E_i(\lambda)$. For $\lambda = 0$ we recover the old unperturbed system. For the new system, the time-independent Schrödinger equation now reads

$$(\hat{H}_0 + \lambda\hat{V})|\psi_i(\lambda)\rangle = E_i(\lambda)|\psi_i(\lambda)\rangle \quad . \tag{4.1}$$

For the following it is useful to introduce the slightly unusual 'normalization '

$$\langle\phi_i|\psi_i(\lambda)\rangle = 1 \tag{4.2}$$

for the perturbed eigenstates $|\psi_i(\lambda)\rangle$, i.e. we **do not** assume that $\langle\psi_i(\lambda)|\psi_i(\lambda)\rangle = 1$!

Now we can multiply the Schrödinger equation (4.1) from the left with $\langle\phi_i|$ and find

$$\langle\phi_i|(\hat{H}_0 + \lambda\hat{V})|\psi_i(\lambda)\rangle = \langle\phi_i|E_i(\lambda)|\psi_i(\lambda)\rangle \quad . \tag{4.3}$$

Using the 'normalization' relation Eq. (4.2) we then obtain

$$E_i(\lambda) = \epsilon_i + \lambda \langle \phi_i | \hat{V} | \psi_i(\lambda) \rangle \quad . \tag{4.4}$$

To obtain the energy eigenvalues of the perturbed Hamilton operator we therefore need to compute the corresponding eigenvector $|\psi_i(\lambda)\rangle$. In the following I will do this for the state $|\phi_0\rangle$ to obtain the perturbed energy $E_0(\lambda)$. (You can easily generalize this to an arbitrary eigenvalue, just by replacing in all expressions $0's$ by the corresponding value.) For the moment I will assume that the eigenvector $|\phi_0\rangle$ of the unperturbed Hamilton operator $\hat{H}_0$ is **non-degenerate**. You will see soon why I had to make this assumption.

In order to do the perturbation theory as systematic as possible I introduce the two projection operators

$$\hat{P}_0 = |\phi_0\rangle\langle\phi_0| \quad \text{and} \quad \hat{Q}_0 = \mathbb{1} - \hat{P}_0 \quad . \tag{4.5}$$

The second projector projects onto the subspace that supports the corrections to the unperturbed wave-function, i.e.

$$\hat{Q}_0 |\psi_0(\lambda)\rangle = |\psi_0(\lambda)\rangle - |\phi_0\rangle \tag{4.6}$$

Now let us rewrite the Schrödinger equation introducing an energy $\epsilon$ which will be specified later. We write

$$(\epsilon - E_0(\lambda) + \lambda \hat{V}) |\psi_0(\lambda)\rangle = (\epsilon - \hat{H}_0) |\psi_0(\lambda)\rangle \quad . \tag{4.7}$$

This can be written as

$$|\psi_0(\lambda)\rangle = (\epsilon - \hat{H}_0)^{-1} (\epsilon - E_0(\lambda) + \lambda \hat{V}) |\psi_0(\lambda)\rangle \quad . \tag{4.8}$$

Multiplying this with the projector $\hat{Q}_0$ using Eq. (4.6) we find

$$|\psi_0(\lambda)\rangle = |\phi_0\rangle + \hat{Q}_0 (\epsilon - \hat{H}_0)^{-1} (\epsilon - E_0(\lambda) + \lambda \hat{V}) |\psi_0(\lambda)\rangle \quad . \tag{4.9}$$

Now we can iterate this equation and we find

$$|\psi_0(\lambda)\rangle = \sum_{n=0}^{\infty} \left[ \hat{Q}_0 (\epsilon - \hat{H}_0)^{-1} (\epsilon - E_0(\lambda) + \lambda \hat{V}) \right]^n |\phi_0\rangle \quad . \tag{4.10}$$

Now we can plug this into Eq. (4.4) and we find the expression for the perturbed energies

$$E_0(\lambda) = \epsilon_0 + \sum_{n=0}^{\infty} \langle \phi_0 | \lambda \hat{V} \left[ \hat{Q}_0 (\epsilon - \hat{H}_0)^{-1} (\epsilon - E_0(\lambda) + \lambda \hat{V}) \right]^n | \phi_0 \rangle \ . \tag{4.11}$$

Eqs. (4.9) and (4.11) give the perturbation expansion of the energies and corresponding eigenvectors to all orders.

**Remark:** The choice of the constant $\epsilon$ is not completely arbitrary. If it is equal to the energy of one of the excited states of the unperturbed Hamilton operator, then the perturbation expansion will not converge as it will contain infinite terms. This still leaves a lot of freedom. However, we also would like to be able to say that by taking into account the first $k$ terms of the sum Eq. (4.11), we have all the terms to order $\lambda^k$. This is ensured most clearly for two choices for the value of $\epsilon$. One such choice is evidently $\epsilon = E_0(\lambda)$. In that case we have

$$E_0(\lambda) = \epsilon_0 + \sum_{n=0}^{\infty} \langle \phi_0 | \lambda \hat{V} \left[ \hat{Q}_0 (E_0(\lambda) - \hat{H}_0)^{-1} (\lambda \hat{V}) \right]^n | \phi_0 \rangle \ . \tag{4.12}$$

This choice is called the Brillouin-Wigner perturbation theory. The other choice, which is often easier to handle, is that where $\epsilon = \epsilon_0$. In that case we obtain

$$E_0(\lambda) = \epsilon_0 + \sum_{n=0}^{\infty} \langle \phi_0 | \lambda \hat{V} \left[ \hat{Q}_0 (\epsilon_0 - \hat{H}_0)^{-1} (\epsilon_0 - E_0(\lambda) + \lambda \hat{V}) \right]^n | \phi_0 \rangle \ . \tag{4.13}$$

This is the Rayleigh perturbation theory. For this choice $(\epsilon_0 - E_0(\lambda) + \lambda \hat{V})$ is proportional to $\lambda$ and therefore we obtain higher order corrections in $\lambda$ by choosing more terms in the sum.

Now let us look at the first and second order expression so that you can compare it with the results that you learned in the second year course. To do this we now specify the value of the arbitrary constant $\epsilon$ to be the unperturbed energy $\epsilon_0$. For the energy of the perturbed Hamilton operator we then obtain

$$E_0(\lambda) = \epsilon_0 + \langle \phi_0 | \lambda \hat{V} | \phi_0 \rangle + \langle \phi_0 | \lambda \hat{V} \hat{Q}_0 (\epsilon - \hat{H}_0)^{-1} (\epsilon_0 - E_0(\lambda) + \lambda \hat{V}) | \phi_0 \rangle + \ldots$$

$$= \epsilon_0 + \lambda\langle\phi_0|\hat{V}|\phi_0\rangle + \langle\phi_0|\lambda\hat{V}\sum_{n\neq 0}|\phi_n\rangle\langle\phi_n|\frac{1}{\epsilon_0-\epsilon_n}(\epsilon_0 - E_0(\lambda) + \lambda\hat{V})|\phi_0\rangle + \dots$$

$$= \epsilon_0 + \lambda\langle\phi_0|\hat{V}|\phi_0\rangle + \lambda^2\sum_{n\neq 0}\frac{|\langle\phi_0|\hat{V}|\phi_n\rangle|^2}{\epsilon_0-\epsilon_n} + \dots \tag{4.14}$$

**Remark:** If $|\phi_0\rangle$ is the ground state of the unperturbed Hamilton operator, then the second order contribution in the perturbation expansion is always negative. This has a good reason. The ground state of the total Hamilton operator $\hat{H} = \hat{H}_0 + \hat{V}$ is $|\psi_0\rangle$ and we find for $|\phi_0\rangle$ that

$$E_0 = \langle\psi_0|\hat{H}_0 + \hat{V}|\psi_0\rangle < \langle\phi_0|\hat{H}_0 + \hat{V}|\phi_0\rangle = E_0 - \sum_{n\neq 0}\frac{|\langle\phi_0|\hat{V}|\phi_i\rangle|^2}{\epsilon_0-\epsilon_n}$$

This implies that the second order perturbation theory term is negative!

The new eigenstate to the eigenvector Eq. (4.14) can be calculated too, and the first contributions are given by

$$\begin{aligned}
|\psi_0\rangle &= |\phi_0\rangle + \hat{Q}_0(\epsilon - \hat{H}_0)^{-1}(\epsilon_0 - E_0 + \hat{V})|\phi_0\rangle + \dots \\
&= |\phi_0\rangle + \sum_{n\neq 0}|\phi_n\rangle\langle\phi_n|\frac{1}{\epsilon_0-\epsilon_n}(\epsilon_0 - E_0 + \hat{V})|\phi_0\rangle + \dots \\
&= |\phi_0\rangle + \sum_{n\neq 0}|\phi_n\rangle\frac{1}{\epsilon_0-\epsilon_n}\langle\phi_n|\hat{V}|\phi_0\rangle + \dots \quad .
\end{aligned}$$

Looking at these two expressions it is clear why I had to demand, that the eigenvalue $\epsilon_0$ of the unperturbed Hamilton operator $\hat{H}_0$ has to be non-degenerate. In order to obtain a meaningful perturbation expansion the individual terms should be finite and therefore the factors $\frac{1}{\epsilon_0-\epsilon_n}$ need to be finite for $n \neq 0$.

End of $20^{th}$ lecture

### 4.1.2 Degenerate perturbation theory

**This part has not been presented in the lecture**

What do we do in the case of degenerate eigenvalues? We can follow

a similar strategy but as you can expect things get a little bit more complicated.

Let us assume that we have a Hamilton operator $\hat{H}_0$, eigenvectors $|\phi_i^\nu\rangle$ and corresponding energies $\epsilon_i$ where the upper index $\nu$ numerates an orthogonal set of eigenvectors to the degenerate eigenvalue $\epsilon_i$. Again let us deal with the eigenvalue $\epsilon_0$.

Now we write down the projectors

$$\hat{P}_0 = \sum_\nu |\phi_0^\nu\rangle\langle\phi_0^\nu| \quad \text{and} \quad \hat{Q}_0 = \mathbb{1} - \hat{P}_0 \quad . \tag{4.15}$$

Then we find

$$\hat{Q}_0|\psi_0^\mu(\lambda)\rangle = |\psi_0^\mu(\lambda)\rangle - \hat{P}_0|\psi_0^\mu(\lambda)\rangle \quad , \tag{4.16}$$

where the eigenvector $|\psi_0^\mu\rangle$ of the perturbed Hamilton operator originates from the unperturbed eigenvector $|\phi_0^\mu\rangle$.

To obtain the analogue of Eq. (4.3) we multiply the Schrödinger equation for the state $|\psi^\mu(\lambda)\rangle$ from the left with $\langle\psi^\nu(\lambda)|\hat{P}_0$ and we obtain

$$\langle\psi_0^\nu(\lambda)|\hat{P}_0(\hat{H}_0 + \lambda\hat{V} - E_0^\mu(\lambda))|\psi^\mu(\lambda)\rangle = 0 \quad . \tag{4.17}$$

Note that I have proceeded slightly different from the non-degenerate case because there I multiplied from the left by $\langle\phi^\nu|$. I could have done that too, but then I would have been forced to introduce a 'normalization' condition of the form $\langle\phi^\nu|\psi^\mu(\lambda)\rangle = \delta_{\mu\nu}$ which is quite hard to compute.

Eq. (4.17) will eventually help us determine the new energy eigenvalues. However, this time this will involve the diagonalization of a matrix. The reason of course being that we have a degeneracy in the original unperturbed eigenvalue.

Now we proceed very much along the lines of the unperturbed perturbation theory. Using Eq. (4.16) we rewrite Eq. (4.9) and obtain

$$|\psi_0^\mu(\lambda)\rangle = \hat{P}_0|\psi_0^\mu(\lambda)\rangle + \hat{Q}_0(\epsilon_0 - \hat{H}_0)^{-1}(\epsilon_0 - E_0^\mu(\lambda) + \lambda\hat{V})|\psi_0^\mu(\lambda)\rangle \quad . \tag{4.18}$$

Iterating this equation yields

$$|\psi_0^\mu(\lambda)\rangle = \sum_{n=0}^{\infty} \left[\hat{Q}_0(\epsilon_0 - \hat{H}_0)^{-1}(\epsilon - E_0^\mu(\lambda) + \lambda\hat{V})\right]^n \hat{P}_0|\psi_0^\mu(\lambda)\rangle \quad . \tag{4.19}$$

This implies that we can calculate the whole eigenvector $|\psi_0^\mu(\lambda)\rangle$ from the eigenvalues $E^\mu(\lambda)$ and the components $\hat{P}_0|\psi_0^\mu(\lambda)\rangle$ of the eigenvector in the eigenspace of the eigenvalue $\epsilon_0$ of the Hamilton operator $\hat{H}_0$.

This may now be inserted in Eq. (4.17) so that we find with the shorthand notation $|\psi_0^{\nu,P}\rangle = \hat{P}_0|\psi_0^\mu\rangle$

$$\langle\psi_0^{\nu,P}|(\hat{H}_0+\lambda\hat{V}-E_0^\mu(\lambda))\sum_{n=0}^{\infty}\left[\hat{Q}_0(\epsilon_0-\hat{H}_0)^{-1}(\epsilon_0-E_0^\mu(\lambda)+\lambda\hat{V})\right]^n|\psi_0^{\mu,P}\rangle = 0 \quad . \tag{4.20}$$

Now we need to determine the energies $E_0^\mu$ from Eq. (4.20). Let us first determine the lowest order approximation, which means that we set $n = 0$. We find for any $\mu$ and $\nu$ that

$$\langle\psi_0^{\nu,P}|\epsilon_0+\lambda\hat{V}-E_0^\mu(\lambda)|\psi_0^{\mu,P}\rangle = 0 \quad . \tag{4.21}$$

Therefore the new energy eigenvalues are obtained by diagonalizing the operator $\hat{V}$ in the eigenspace of the eigenvalue $\epsilon_0$ of the Hamilton operator $\hat{H}_0$ spanned by the $|\phi_0^\nu\rangle$. We find

$$E_0^\mu(\lambda) = \epsilon_0 + \lambda V_{\mu\mu} \tag{4.22}$$

where the $V_{\mu\mu}$ are the eigenvalues of $\hat{V}$.

The second order contribution is then obtained by using the eigenvectors obtained from Eq. (4.21) and use them in the next order expression, e.g. for $n = 1$. The second order expression for the perturbation theory is therefore obtained from taking the expectation values

$$\langle\psi_0^{\nu,P}|(\epsilon_0+\lambda V_{\nu\nu}-E_0^\mu(\lambda))+\lambda\hat{V}\hat{Q}_0(\epsilon_0-\hat{H}_0)^{-1}\lambda\hat{V}|\psi_0^{\nu,P}\rangle = 0 \quad .$$

This gives

$$E_0^\nu = \epsilon_0 + \lambda V_{\nu\nu} + \lambda^2\sum_{m\neq 0}\frac{|\langle\psi_0^{\nu,P}|\hat{V}|\phi_m\rangle|^2}{\epsilon_0-\epsilon_m} \quad . \tag{4.23}$$

In the following section I will work out an example for the use of perturbation theory which gives a non-trivial result.

———— Start of $21^{th}$ lecture

### 4.1.3 The van der Waals force

In this section I am going to derive the force that two neutral atoms are exerting on each other. This force, the van der Waals force, can play a significant role in the physics (also the chemistry and mechanics) of neutral systems. In the derivation of the van der Waals force I will use time-independent perturbation theory. The calculations will be simplified by the fact that the system of two atoms exhibits symmetries which have the effect, that the first order contribution to perturbation theory vanishes.

The basic experimental situation is presented in Fig. 4.1. We discuss the van der Waals force between two hydrogen atoms. The two

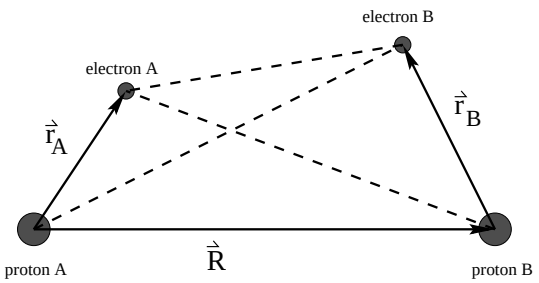


Figure 4.1: Two hydrogen atoms are placed near each other, with a distance $R$ much larger than the Bohr radius $a_0$.

atomic nuclei (proton) are positioned at a fixed distance from each other. We assume that the protons are not moving. The proton of atom A is placed at the origin, while the proton of atom B is located at the position $\vec{R}$. The electron of atom A is displaced from the origin by $\vec{r}_A$, while atom B is displaced from the nucleus B by $\vec{r}_B$. To make sure that we can still speak of separate hydrogen atoms we need to demand that the separation between the two protons is much larger than the radius of each of the atoms, i.e.

$$|\vec{R}| \gg a_0 \quad , \tag{4.24}$$

where $a_0$ is the Bohr radius.

After we have described the physical situation we will now have to determine the Hamilton operator of the total system. The Hamilton

operator has the form

$$\hat{H} = \hat{H}_0 + \hat{V} \quad , \tag{4.25}$$

where $\hat{V}$ describes the small perturbation that gives rise to the van der Waals force. The unperturbed Hamilton-operator $\hat{H}_0$ is that of two hydrogen atoms, i.e.

$$\hat{H}_0 = \frac{\hat{\vec{p}}_A^2}{2m} + \frac{\hat{\vec{p}}_B^2}{2m} - \frac{1}{4\pi\epsilon_0|\hat{\vec{r}}_A|} - \frac{1}{4\pi\epsilon_0|\hat{\vec{r}}_B|} \quad , \tag{4.26}$$

which can be solved exactly. We have energy eigenfunctions of the form $|\phi^A_{n,l,m}\rangle|\phi^B_{n',l',m'}\rangle$ with energies $E_n + E_{n'}$. The perturbation to this Hamilton operator is due to the electrostatic forces between electrons and protons from different atoms. From Fig. 4.1 we can easily see that this gives

$$\hat{V} = \frac{e^2}{4\pi\epsilon_0 R} + \frac{e^2}{4\pi\epsilon_0|\vec{R} + \hat{\vec{r}}_B - \hat{\vec{r}}_A|} - \frac{e^2}{4\pi\epsilon_0|\vec{R} + \hat{\vec{r}}_B|} - \frac{e^2}{4\pi\epsilon_0|\vec{R} + \hat{\vec{r}}_A|} \quad . \tag{4.27}$$

As we have assumed that the Bohr radius is much smaller than the separation between the two hydrogen atoms, we can conclude that $|\vec{r}_A|, |\vec{r}_B| \ll R$. Therefore we are able to expand the perturbation Hamilton operator and we obtain after lengthy but not too difficult calculation

$$\hat{V}_{dd} = \frac{e^2}{4\pi\epsilon_0}\left[\frac{\hat{\vec{r}}_A\hat{\vec{r}}_B}{R^3} - 3\frac{(\hat{\vec{r}}_A\vec{R}_u)(\hat{\vec{r}}_B\vec{R}_u)}{R^3}\right] \quad , \tag{4.28}$$

where $\vec{R}_u$ is the unit vector in direction of $\vec{R}$. Now we are in a position to determine the first and second order contribution of the perturbation theory. Because we are dealing with the ground state, we have no degeneracy and we can apply non-degenerate perturbation theory.

**First order correction** The first order contribution to perturbation theory is given by

$$\Delta E_1 = \langle\phi^A_{0,0,0}|\langle\phi^B_{0,0,0}|\hat{V}_{dd}|\phi^A_{0,0,0}\rangle|\phi^B_{0,0,0}\rangle \quad , \tag{4.29}$$

where $|\phi^{A/B}_{0,0,0}\rangle$ are the unperturbed ground-states of atom A/B. Inserting Eq. (4.27) into Eq. (4.29) gives

$$\begin{aligned}\Delta E_1 &= \frac{e^2}{4\pi\epsilon_0}\left[\frac{\langle\phi^A_{0,0,0}|\hat{\vec{r}}_A|\phi^A_{0,0,0}\rangle\langle\phi^B_{0,0,0}|\hat{\vec{r}}_B|\phi^B_{0,0,0}\rangle}{R^3}\right. \\ &\left. -3\frac{\langle\phi^A_{0,0,0}|\hat{\vec{r}}_A\vec{R}_u|\phi^A_{0,0,0}\rangle\langle\phi^B_{0,0,0}|\hat{\vec{r}}_B\vec{R}_u|\phi^B_{0,0,0}\rangle}{R^3}\right] . \quad (4.30)\end{aligned}$$

Now we can use a symmetry argument to show that this whole, rather lengthy expression must be zero. Why is that so? What appears in Eq. (4.30) are expectation values of components of the position operator in the unperturbed ground state of the atom. However, the unperturbed Hamilton operator of the hydrogen atom possesses some symmetries. The relevant symmetry here is that of the parity, i.e. the Hamilton operator $\hat{H}_0$ commutes with the operator $\hat{P}$ which is defined by $\hat{P}|x\rangle = |-x\rangle$. This implies that both operators, $\hat{H}_0$ and $\hat{P}$, can be diagonalized simultaneously, i.e. they have the same eigenvectors. This is the reason why all eigenvectors of the unperturbed Hamilton operator $\hat{H}_0$ are also eigenvectors of the parity operator. Therefore we have $\hat{P}|\phi^{A/B}_{0,0,0}\rangle = \pm|\phi^{A/B}_{0,0,0}\rangle$. This implies that

$$\begin{aligned}\langle\phi_{0,0,0}|\hat{\vec{x}}|\phi_{0,0,0}\rangle &= \langle\phi_{0,0,0}|\hat{P}\hat{\vec{x}}\hat{P}|\phi_{0,0,0}\rangle \\ &= \langle\phi_{0,0,0}|-\hat{\vec{x}}|\phi_{0,0,0}\rangle\end{aligned}$$

This equality implies that $\langle\phi_{0,0,0}|\hat{\vec{x}}|\phi_{0,0,0}\rangle = 0$ and therefore $\Delta E_1 = 0$! This quick argument illustrates the usefulness of symmetry arguments.

**Second order correction** The second order contribution of the perturbation theory is given by

$$\Delta E_2 = \sum_{(n',l',m'),(n,l,m)}{}' \frac{|\langle\phi^A_{n',l',m'}|\langle\phi^B_{n,l,m}|\hat{V}_{dd}|\phi^A_{0,0,0}\rangle|\phi^B_{0,0,0}\rangle|^2}{2E_0 - E_{n'} - E_n} \quad (4.31)$$

where the $\sum'$ means that the state $|\phi^A_{0,0,0}\rangle|\phi^B_{0,0,0}\rangle$ is excluded from the summation. The first conclusion that we can draw from expression Eq.

(4.31) is that the second order correction $\Delta E_2$ is negative so that we have

$$\Delta E_2 = -\frac{C}{R^6} \tag{4.32}$$

with a positive constant $C$. The force exerted on one of the atoms is therefore given by

$$F = -\frac{dE}{dR} \approx -\frac{6C}{R^7} \tag{4.33}$$

and therefore the force aims to reduce the distance between the atoms and we conclude that the van der Waals force is attractive. This is also the classical result that has been known before the advent of quantum mechanics.

The physical interpretation of the van der Waals force is the following: A hydrogen atom should not be viewed as a stationary charge distribution but rather as a randomly fluctuating electric dipole due to the motion of the electron. Assume that in the first atom a dipole moment appears. This generates an electric field, which scales like $\frac{1}{R^3}$ with distance. This electric field will now induce an electric dipole moment in the other atom. An electric dipole will always tend to move towards areas with increasing field strength and therefore the second atom will be attracted to the first one. The induced dipole moment will be proportional to the electric field. Therefore the net force will scale as $-\frac{d}{dR}\frac{1}{R^6}$. This semi-classical argument gives you an idea for the correct $R$-dependence of the van der Waals force.

However, this semi-classical explanation fails to account for some of the more intricate features of the van der Waals force. One example is that of two hydrogen atoms, one in the ground state and one in an excited state. In that case the van der Waals force scales as $\frac{C}{R^3}$ and the constant $C$ can be either positive or negative, i.e. we can have a repulsive as well as an attractive van der Waals force. The computation of this case is slightly more complicated because excited states in the hydrogen atom are degenerate, which requires the use of degenerate perturbation theory. On the other hand it only involves first order terms.

### 4.1.4 The Helium atom

Now let us try to apply perturbation theory to the problem of the Helium atom. In fact it is not at all clear that we can do that because the two electrons in the Helium atom are very close together and their mutual exerted force is almost as strong as that between the nucleus and the individual electrons. However, we may just have a look how far we can push perturbation theory. Again I am interested in the ground state of the Helium atom. The Hamilton operator is given by

$$\hat{H} = \frac{\hat{\vec{p}}_A^2}{2m} + \frac{\hat{\vec{p}}_B^2}{2m} - \frac{2e^2}{4\pi\epsilon_0|\hat{\vec{r}}_A|} - \frac{2e^2}{4\pi\epsilon_0|\hat{\vec{r}}_B|} + \frac{e^2}{4\pi\epsilon_0|\hat{\vec{r}}_A - \hat{\vec{r}}_B|} \quad . \tag{4.34}$$

The first four terms, $\hat{H}_0$ describe two non-interacting electrons in a central potential and for this part we can find an exact solution, given by the energies and wave-functions of a $He^+$ ion. Now we perform perturbation theory up to first order. The zeroth order wavefunction is given by

$$|\psi_{He}\rangle = |\psi_{He^+}\rangle \otimes |\psi_{He^+}\rangle \tag{4.35}$$

where $|\psi_{He^+}\rangle$ is the ground state of an electron in a Helium ion, i.e. a system consisting of two protons and one electron. In position space this wavefunction is given by

$$\psi_{He}(r) = \frac{Z^3}{\pi a_0^3} e^{-\frac{Z(r_A+r_B)}{a_0}} \quad , \tag{4.36}$$

where $a_0 = 0.511\,10^{-10}m$ is the Bohr radius and $Z = 2$ is the number protons in the nucleus. The energy of this state for the unperturbed system is given by $E = -108.8eV$. The first order correction to the energy of the ground state is now given by

$$\Delta E = \int d^3r_A \int d^3r_B |\psi_{He}(r)|^2 \frac{e^2}{4\pi\epsilon_0|\hat{\vec{r}}_A - \hat{\vec{r}}_B|} = 34eV \quad . \tag{4.37}$$

Therefore our estimate for the ground state energy of the Helium atom is $E_1 = -108.8eV + 34eV = -74.8eV$. This compares quite well with the measured value of $E_1 = -78.9eV$. However, in the next section we will see how we can actually improve this value using a different approximation technique. ———End of $21^{th}$ lecture

## 4.2 Variational Principle

After this section on perturbation methods I am now moving on to a different way of obtaining approximate solutions to quantum mechanical problems. Previously we have investigated problems which were 'almost' exactly solvable, i.e. the exactly solvable Hamilton operator has a small additional term. Now I am going to deal with problems which can not necessarily be cast into such a form. An example would be the Helium atom (which cannot be solved exactly) as compared to a negative Helium ion (which is basically like a hydrogen atom and therefore exactly solvable). Because the two electrons in the Helium atom are very close, it is not obvious that a perturbation expansion gives a good result, simply because the electrostatic forces between the electrons are almost as large as those between electrons and nucleus. Here variational principles can be very useful. The example of the Helium atom already indicates that variational methods are paramount importance e.g. in atomic physics, but also in quantum chemistry or solid state physics.

### 4.2.1 The Rayleigh-Ritz Method

Here I will explain a simple variational method, which nevertheless exhibits the general idea of the method. Variational methods are particularly well suited to determine the ground state energy of a quantum mechanical system and this is what I will present first. The whole method is based on the following theorem concerning the expectation value of the Hamilton operator.

**Theorem 49** *Given a Hamilton operator $\hat{H}$ with ground state $|\phi_0\rangle$ and ground state energy $E_0$. For any $|\psi\rangle$ we have*

$$\langle\phi_0|\hat{H}|\phi_0\rangle \leq \langle\psi|\hat{H}|\psi\rangle \quad . \tag{4.38}$$

**Proof:** For the proof we need to remember that any state vector $|\psi\rangle$ can be expanded in terms of the eigenvectors $|\phi_i\rangle$ of the Hermitean operator $\hat{H}$. This means that we can find coefficients such that

$$|\psi\rangle = \sum_i \alpha_i |\phi_i\rangle \quad . \tag{4.39}$$

I will use this in the proof of theorem 49.

$$\begin{aligned}\langle\psi|\hat{H}|\psi\rangle &= \sum_{i,j}\alpha_i^*\alpha_j\langle\phi_i|\hat{H}|\phi_j\rangle \\ &= \sum_{i,j}\alpha_i^*\alpha_j E_i\langle\phi_i|\phi_j\rangle \\ &= \sum_i |\alpha_i|^2 E_i \quad .\end{aligned}$$

Now we can see that

$$\begin{aligned}\langle\psi|\hat{H}|\psi\rangle - E_0 &= \sum_i |\alpha_i|^2 E_i - E_0 \\ &= \sum_i |\alpha_i|^2 (E_i - E_0) \\ &\geq 0 \quad ,\end{aligned}$$

because $E_i \geq E_0$. If the lowest energy eigenvalue $E_0$ is not degenerate then we have equality exactly if $|\psi\rangle = |\phi_0\rangle$ $\Box$.

How does this help us in finding the energy of the ground state of a quantum mechanical system that is described by the Hamilton operator $\hat{H}$? The general recipe proceeds in two steps.

**Step 1:** Chose a 'trial' wave function $|\psi(\alpha_1, \ldots, \alpha_N)\rangle$ that is parametrized by parameters $\alpha_1, \ldots, \alpha_N$. This choice of the trial wave function will often be governed by the symmetries of the problem.
**Step 2:** Minimize the expression

$$\bar{E} = \langle\psi(\alpha_1, \ldots, \alpha_N)|\hat{H}|\psi(\alpha_1, \ldots, \alpha_N)\rangle \quad , \tag{4.40}$$

with respect to the $\alpha_i$. The value $E_{min}$ that you obtain this way is your estimate of the ground state wave function.

**The Helium atom** Now let us come back to the Helium atom to see whether we can make use of the variational principle to obtain a better estimate of the ground state energy.

In our perturbation theoretical calculation we have used the wave-function Eq. (4.36) with $Z = 2$. Now let us introduce the adjustable

parameter $\sigma$ and use the trial wavefunction

$$\psi_{He}(r) = \frac{Z_{eff}^3}{\pi a^3} e^{-\frac{Z_{eff}(r_A + r_B)}{a_0}} \quad , \tag{4.41}$$

where $Z_{eff} = Z - \sigma$. This is a physically reasonable assumption because each electron sees effectively a reduced nuclear charge due to the shielding effect of the other electron. Now we can use this new trial wavefunction to calculate the energy expectation value of the total Hamilton operator with this wavefunction. We find after some computations

$$E(\sigma) = -2R_H \left( Z^2 - \frac{5}{8} Z + \frac{5}{8}\sigma - \sigma^2 \right) \quad , \tag{4.42}$$

where $R_H = \frac{m_0 e^4}{64\pi^3 \epsilon_0^2 \hbar^3 c}$ is the Rydberg constant. Now we need to compute that value of $\sigma$ for which Eq. (4.42) assumes its minimal value. The value that we obtain is $\sigma = \frac{5}{16}$, independently of the value of $Z$. Inserting this into Eq. (4.42) we find

$$E_{min} = -2R_H (Z - \frac{5}{16})^2 \quad . \tag{4.43}$$

For the Helium atom (Z=2) we therefore obtain

$$E_{min} = -77.4eV \quad , \tag{4.44}$$

for the ground state energy. This is quite close to the true value of $-78.9eV$ and represents a substantial improvement compared to the value obtained via perturbation theory. Making a slightly more detailed choice for the trial wavefunction (basically using one more parameter) would lead to substantially improved values. This shows that the variational method can indeed be quite useful.

However, the next example will illustrate the shortcomings of the variational method and proves that this method has to be applied with some care.

**The harmonic oscillator** The Hamilton operator for the harmonic oscillator is given $\hat{H} = \frac{\hat{p}^2}{2m} + \frac{1}{2} m\omega^2 \hat{x}^2$. Now, assume that we do not

know the ground state energy and the ground state wavefunction of the harmonic oscillator. This Hamilton operator commutes with the parity operator and therefore the eigenstates should also be chosen as eigenfunctions of the parity operator. Therefore let us chose the normalized symmetric wave-function

$$\psi_a(x) = \sqrt{\frac{2a^{3/2}}{\pi}} \frac{1}{x^2 + a} \quad . \tag{4.45}$$

Of course I have chosen this slightly peculiar function in order to make the calculations as simple as possible. Omitting the detailed calculations we find that the mean value of the energy of the harmonic oscillator in this state is given by

$$\begin{aligned} \langle \psi_a | \hat{H} | \psi_a \rangle &= \int \psi_a^*(x) \left( -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + \frac{1}{2} m\omega^2 x^2 \right) \psi_a(x) dx & (4.46) \\ &= \frac{\hbar^2}{4m} \frac{1}{a} + \frac{1}{2} m\omega^2 a \quad . & (4.47) \end{aligned}$$

Now we need to find the value for which this expression becomes minimal. We obtain

$$a_{min} = \frac{1}{\sqrt{2}} \frac{\hbar}{m\omega} \quad , \tag{4.48}$$

and for the energy

$$E_{min} = \frac{1}{\sqrt{2}} \hbar\omega \quad . \tag{4.49}$$

Therefore our estimate gives an error

$$\frac{E_{min} - E_0}{E_0} = \frac{\sqrt{2} - 1}{2} \approx 0.2 \quad . \tag{4.50}$$

This is not too good, but would have been a lot better for other trial functions.

This example shows the limitations of the variational principle. It very much depends on a good choice for the trial wave function. It should also be noted, that a good approximation to the ground state energy does not imply that the chosen trial wave function will give good results for other physical observables. This can be seen from the above

example of the harmonic oscillator. If we compute the expectation value of the operator $\hat{x}^2$ then we find

$$\langle \psi_{a_{min}} | \hat{x}^2 | \psi_{a_{min}} \rangle = \frac{1}{\sqrt{2}} \frac{\hbar}{m\omega}$$

which is quite close to the true value of $\frac{1}{2}\hbar\omega$. On the other hand, the expectation value of the operator $\hat{x}^4$ **diverges** if we calculate it with the trial wave function, while we obtain a finite result for the true ground-state wave function.

**The variational principle for excited states.** The variational method shown here can be extended to the calculation of the energies of excited states. The basic idea that we will be using is that eigenvectors to different eigenvalues of the Hamilton operator are necessarily orthogonal. If we know the ground state of a system, or at least a good approximation to it, then we can use the variational method. The only thing we need to make sure is that our trial wavefunction is always orthogonal to the ground state wave function or the best approximation we have for it. If we have ensured this with the choice of our trial function, then we can proceed analogously to the variational principle for the ground state. ———End of $22^{th}$ lecture

## 4.3 Time-dependent Perturbation Theory

In the previous sections I have explained some of the possible methods of stationary perturbation theory. Using these methods we are able, in principle, to approximate the energy eigenvalues as well as the eigenvectors of the Hamilton operator of a quantum mechanical system that has a small perturbation. We are then able to approximate the spectral decomposition of the Hamilton operator, which is given by

$$\hat{H} = \sum_i E_i |\psi_i\rangle\langle\psi_i| \quad , \tag{4.51}$$

where the $E_i$ are the energy eigenvalues of the Hamilton operator and the $|\psi_i\rangle$ the corresponding eigenvectors. Equation (4.51) is sufficient to

compute the time evolution of any possible initial state $|\phi(t_0)\rangle$ using the solution of the Schrödinger equation

$$e^{-i\hat{H}(t-t_0)}|\phi\rangle = \sum_i e^{-iE_i(t-t_0)/\hbar}|\psi_i\rangle\langle\psi_i|\phi(t_0)\rangle \quad . \tag{4.52}$$

Using stationary perturbation theory we are then able to obtain the approximate time-evolution of the system.

However, there are reasons why this approach is not necessarily the best. First of all, we might have a situation where the Hamilton operator of the system is time-dependent. In that case the solution of the Schrödinger equation is generally not of the form $e^{-i\int \hat{H}(t')dt'/\hbar}$ anymore, simply because Hamilton operators for different times do not necessarily commute with each other. For time-dependent Hamilton operators we have to proceed in a different way in order to obtain approximations to the time evolution.

There are two situations in which we may encounter time-dependent Hamilton operators. While the full Hamilton operator of a closed system is always time-independent, this is not the case anymore if we have a system that is interacting with its environment, i.e. an open system. An example is an atom that is being irradiated by a laser beam. Clearly the electro-magnetic field of the laser is time-dependent and therefore we find a time-dependent Hamilton operator. A time-dependent Hamilton operator may also appear in another situation. If we have given the Hamilton operator of a quantum system which is composed of two parts, the solvable part $\hat{H}_0$ and the perturbation $\hat{V}$, both of which may be time independent, then it can be of advantage to go to an 'interaction' picture that is 'rotating ' with frequencies determined by the unperturbed Hamilton operator $\hat{H}_0$. In this case the remaining Hamilton operator will be a time-dependent version of $\hat{V}$.

The rest of this section will be devoted to the explanation of a method that allows us to approximate the time evolution operator of a time-dependent Hamilton operator. The convergence of this method is very much enhanced if we go into an interaction picture which eliminates the dynamics of the system due to the unperturbed Hamilton operator. The definition of the 'interaction' picture in a precise manner will be the subject of the first part of this section.

### 4.3.1 Interaction picture

Often the total Hamilton operator can be split into two parts $\hat{H} = \hat{H}_0 + \hat{V}$, the exactly solvable Hamilton operator $\hat{H}_0$ and the perturbation $\hat{V}$. If we have a Hamilton operator, time-dependent or not, and we want to perform perturbation theory it will always be useful to make use of the exactly solvable part $\hat{H}_0$ of the Hamilton operator. In the case of time-independent perturbation theory we have used the eigenvalues and eigenvectors of the unperturbed Hamilton operator and calculated corrections to them in terms of the unperturbed eigenvectors and eigenvalues. Now we are going to do the analogous step for the time evolution operator. We use the known dynamics due to the unperturbed Hamilton operator $\hat{H}_0$ and then determine corrections to this time evolution. In fact, the analogy goes so far that it is possible in principle to rederive time-independent perturbation theory as a consequence of time-dependent perturbation theory. I leave it to the slightly more ambitious student to work this out in detail after I have explained the ideas of time-dependent perturbation theory.

The Schrödinger equation reads

$$i\hbar \frac{d}{dt}|\psi(t)\rangle = (\hat{H}_0 + \hat{V})(t)|\psi(t)\rangle \quad , \tag{4.53}$$

with a potentially time dependent total Hamilton operator $\hat{H}(t)$. The solution of the Schrödinger equation can be written formally as

$$|\psi(t)\rangle = \hat{U}(t,t')|\psi(t')\rangle \quad , \tag{4.54}$$

where the time evolution operator $\hat{U}(t,t')$ obeys the differential equation

$$i\hbar \frac{d}{dt}\hat{U}(t,t') = \hat{H}(t)\hat{U}(t,t') \quad . \tag{4.55}$$

This can easily be checked by inserting Eq. (4.54) into the Schrödinger equation. Now let us assume that we can solve the time-evolution that is generated by the unperturbed Hamilton operator $\hat{H}_0(t)$. The corresponding time evolution operator is given by $\hat{U}_0(t,t')$ which gives the solution $|\psi_0(t)\rangle = \hat{U}_0(t,t')|\psi_0(t')\rangle$ of the Schrödinger equation

$$i\hbar \frac{d}{dt}|\psi_0(t)\rangle = \hat{H}_0(t)|\psi_0(t)\rangle \quad . \tag{4.56}$$

What we are really interested in is the time evolution according to the full Hamilton operator $\hat{H}$. As we already know the time evolution operator $\hat{U}_0(t,t')$, the aim is now the calculation of the deviation from this unperturbed time evolution. Let us therefore derive a Schrödinger equation that describes this deviation. We obtain this Schrödinger equation by going over to an **interaction picture** with respect to the time $t_0$ which is defined by choosing the state vector

$$|\psi_I(t)\rangle = \hat{U}_0^\dagger(t,t_0)|\psi(t)\rangle \quad . \tag{4.57}$$

Clearly the state in the interaction picture has been obtained by 'undoing' the part of the time evolution that is due to the unperturbed Hamilton operator in the state $|\psi(t)\rangle$ and therefore describes that part of the time evolution that is due to the perturbation $\hat{V}(t)$. Now we have to derive the Schrödinger equation for the interaction picture wavefunction and in particular we have to derive the Hamilton-operator in the interaction picture. This is achieved by inserting Eq. (4.57) into the Schrödinger equation Eq. (4.53). We find

$$i\hbar\frac{d}{dt}\left(\hat{U}_0(t,t_0)|\psi_I(t)\rangle\right) = (\hat{H}_0+\hat{V})(t)\hat{U}_0(t,t_0)|\psi_I(t)\rangle \tag{4.58}$$

and then

$$i\hbar\left(\frac{d}{dt}\hat{U}_0(t,t_0)\right)|\psi_I(t)\rangle + i\hbar\hat{U}_0(t,t_0)\frac{d}{dt}|\psi_I(t)\rangle = (\hat{H}_0+\hat{V})(t)\hat{U}_0(t,t_0)|\psi_I(t)\rangle \quad . \tag{4.59}$$

Now we use the differential equation

$$i\hbar\frac{d}{dt}\hat{U}_0(t,t_0) = \hat{H}_0(t)\hat{U}_0(t,t_0) \quad . \tag{4.60}$$

Inserting this into Eq. (4.59) gives

$$\hat{H}_0\hat{U}_0(t,t_0)|\psi_I(t)\rangle + i\hbar\hat{U}_0(t,t_0)\frac{d}{dt}|\psi_I(t)\rangle \quad = \quad (\hat{H}_0+\hat{V})(t)\hat{U}_0(t,t_0)|\psi_I(t)\rangle$$

and finally

$$i\hbar\frac{d}{dt}|\psi_I(t)\rangle \quad = \quad \hat{U}_0^\dagger(t,t_0)\hat{V}(t)\hat{U}_0(t,t_0)|\psi_I(t)\rangle \quad .$$

Using the definition for the Hamilton operator in the interaction picture

$$\hat{H}_I(t) = \hat{U}_0^\dagger(t,t_0)(\hat{H} - \hat{H}_0)\hat{U}_0(t,t_0) \tag{4.61}$$

we find the interaction picture Schrödinger equation

$$i\hbar\frac{d}{dt}|\psi_I(t)\rangle = \hat{H}_I(t)|\psi_I(t)\rangle \quad . \tag{4.62}$$

The formal solution to the Schrödinger equation in the interaction picture can be written as

$$|\psi_I(t)\rangle = \hat{U}_I(t,t')|\psi_I(t')\rangle \quad . \tag{4.63}$$

From Eq. (4.63) we can then obtain the solution of the Schrödinger equation Eq. (4.53) as

$$\begin{aligned} |\psi(t)\rangle &= \hat{U}_0(t,t_0)|\psi_I(t)\rangle \\ &= \hat{U}_0(t,t_0)\hat{U}_I(t,t')|\psi_I(t')\rangle \\ &= \hat{U}_0(t,t_0)\hat{U}_I(t,t')\hat{U}_0^\dagger(t',t_0)|\psi(t')\rangle \quad . \end{aligned} \tag{4.64}$$

This result shows that even in a system with a time independent Hamilton operator $\hat{H}$ we may obtain a time dependent Hamilton operator by going over to an interaction picture. As time dependent Hamilton operators are actually quite common place it is important that we find out how the time evolution for a time-dependent Hamilton operator can be approximated.

### 4.3.2 Dyson Series

Given the Schrödinger equation for a time dependent Hamilton operator we need to find a systematic way of obtaining approximations to the time evolution operator. This is achieved by time-dependent perturbation theory which is based on the Dyson series. Here I will deal with the interaction picture Schrödinger equation Eq. (4.62). To find the Dyson series, we first integrate the Schrödinger equation Eq. (4.62) formally. This gives

$$|\psi_I(t)\rangle = |\psi_I(t')\rangle - \frac{i}{\hbar}\int_{t'}^{t} dt_1 \hat{H}_I(t_1)|\psi_I(t_1)\rangle \quad . \tag{4.65}$$

This equation can now be iterated. We then find the Dyson series

$$|\psi_I(t)\rangle = |\psi_I(t')\rangle - \frac{i}{\hbar}\int_{t'}^{t} dt_1 \hat{H}_I(t_1)|\psi_I(t')\rangle - \frac{1}{\hbar^2}\int_{t'}^{t} dt_1 \int_{t'}^{t_1} dt_2 \hat{H}_I(t_1)\hat{H}_I(t_2)|\psi_I(t')\rangle + \dots \quad . \tag{4.66}$$

From this we immediately observe that the time evolution operator in the interaction picture is given by

$$\hat{U}_I(t,t') = \mathbb{1} - \frac{i}{\hbar}\int_{t'}^{t} dt_1 \hat{V}_I(t_1) - \frac{1}{\hbar^2}\int_{t'}^{t} dt_1 \int_{t'}^{t_1} dt_2 \hat{V}_I(t_1)\hat{V}_I(t_2) + \dots \quad . \tag{4.67}$$

Equations (4.66) and (4.67) are the basis of time-dependent perturbation theory.

**Remark:** It should be pointed out that these expressions have some limitations. Quite obviously every integral in the series will generally have the tendency to grow with increasing time differences $|t - t'|$. For sufficiently large $|t - t'|$ the Dyson series will then fail to converge. This problem can be circumvented by first splitting the time evolution operator into sufficiently small pieces, i.e.

$$\hat{U}_I(t,0) = \hat{U}_I(t,t_n)\hat{U}_I(t_n,t_{n-1})\dots\hat{U}_I(t_1,0) \quad , \tag{4.68}$$

such that each time interval $t_i - t_{i-1}$ is very small. Then each of the time evolution operators is calculated using perturbation theory and finally they are multiplied together.

### 4.3.3 Transition probabilities

For the following let us assume that the unperturbed Hamilton operator is time independent and that we take an interaction picture with respect to the time $t_0 = 0$. In this case we can write

$$\hat{U}_0(t,0) = e^{-i\hat{H}_0 t/\hbar} \quad . \tag{4.69}$$

Under the time evolution due to the unperturbed Hamilton operator $\hat{H}_0$ the eigenstates $|\phi_n\rangle$ to the energy $E_n$ of that Hamilton operator only obtain phase factors in the course of the time evolution, i.e.

$$e^{-i\hat{H}_0 t/\hbar}|\phi_n\rangle = e^{-iE_n t/\hbar}|\phi_n\rangle = e^{-i\omega_n t}|\phi_n\rangle \quad , \tag{4.70}$$

where $\omega_n = E_n/\hbar$. Under the time evolution due to the total Hamilton operator $\hat{H}$ this is not the case anymore. Now, the time evolution will generally take an eigenstate of the unperturbed Hamilton operator $\hat{H}_0$ to a superposition of eigenstates of the unperturbed Hamilton operator $\hat{H}_0$, i.e.

$$e^{-i\hat{H}(t-t')/\hbar}|\phi_n\rangle = \sum_k \alpha_{n\to k}(t)|\phi_k\rangle \tag{4.71}$$

with nonzero coefficients $\alpha_{n\to k}(t)$. If we make a measurement at the later time $t$ we will have a non-zero probability $|\alpha_{n\to k}|^2$ to find eigenstates $|\phi_m\rangle$ with $m \neq n$. We then say that the system has a **transition probability** $p_{n\to k}(t)$ for going from state $|\phi_n\rangle$ to state $|\phi_k\rangle$. These transitions are induced by the perturbation $\hat{V}$, which may for example be due to a laser field. Let us calculate the transition probability in lowest order in the perturbation $\hat{V}$. This is a valid approximation as long as the total effect of the perturbation is small, i.e. the probability for finding the system in the original state is close to 1. To obtain the best convergence of the perturbative expansion, we are going over to the interaction picture with respect to the unperturbed Hamilton operator $\hat{H}_0$ and then break off the Dyson series after the term linear in $\hat{H}_I(t) = e^{i\hat{H}_0 t_1/\hbar}\hat{V}(t_1)e^{-i\hat{H}_0 t_1/\hbar}$. If the initial state at time $t'$ is $|\phi_n\rangle$ then the probability amplitude for finding the system in state $|\phi_m\rangle$ at the later time $t$ is given by

$$\begin{aligned}
\alpha_{n\to m}(t) &= \langle\phi_m|\psi(t)\rangle \\
&\approx \langle\phi_m|\left(|\phi_n\rangle - \frac{i}{\hbar}\int_{t'}^{t} dt_1 e^{i\hat{H}_0 t_1/\hbar}\hat{V}(t_1)e^{-i\hat{H}_0 t_1/\hbar}|\phi_n\rangle\right) \\
&= \delta_{mn} - \frac{i}{\hbar}\int_{t'}^{t} dt_1 e^{i(\omega_m-\omega_n)t_1}\langle\phi_m|\hat{V}(t_1)|\phi_n\rangle \quad .
\end{aligned} \tag{4.72}$$

If $m \neq n$, we find

$$a_{n\to m}(t) \approx -\frac{i}{\hbar}\int_{t'}^{t} dt_1 e^{i(\omega_m-\omega_n)t_1}\langle\phi_m|\hat{V}(t_1)|\phi_n\rangle \quad , \tag{4.73}$$

and the transition probability is then given by

$$p_{n\to m}(t) = |a_{n\to m}(t)|^2 \quad . \tag{4.74}$$

**Periodic perturbation**

Let us consider the special case in which we have a periodic perturbation

$$\hat{V}(t) = \hat{V}_0 \cos\omega t = \frac{1}{2}\hat{V}_0 \left(e^{i\omega t} + e^{-i\omega t}\right) \quad . \tag{4.75}$$

Inserting this into Eq. (4.73) we find using $\omega_{mn} := \omega_m - \omega_n$

$$\begin{aligned} a_{n\to m}(t) &\approx -\frac{i}{\hbar}\int_0^t dt_1 e^{i\omega_{mn}t_1}\frac{1}{2}\langle\phi_m|\hat{V}_0|\phi_n\rangle\left(e^{i\omega t_1} + e^{-i\omega t_1}\right) \\ &= -\frac{i}{2\hbar}\langle\phi_m|\hat{V}_0|\phi_n\rangle\left(\frac{e^{-i(\omega-\omega_{mn})t}-1}{-i(\omega-\omega_{mn})} + \frac{e^{i(\omega+\omega_{mn})t}-1}{i(\omega+\omega_{mn})}\right) \end{aligned} \tag{4.76}$$

**Static perturbation** For a time independent perturbation we have $\omega = 0$. This leads to

$$a_{n\to m}(t) = -\frac{i}{\hbar}\langle\phi_m|\hat{V}_0|\phi_n\rangle\left(\frac{e^{i\omega_{mn}t}-1}{i\omega_{mn}}\right) \tag{4.77}$$

and then to

$$p_{n\to m}(t) = \frac{1}{\hbar^2}|\langle\phi_m|\hat{V}_0|\phi_n\rangle|^2\frac{\sin^2\frac{\omega_{mn}t}{2}}{(\frac{\omega_{mn}}{2})^2} \tag{4.78}$$

For sufficiently large times t this is a very sharply peaked function in the frequency $\omega_{mn}$. In fact in the limit $t \to \infty$ this function tends towards a delta-function

$$\lim_{t\to\infty}\frac{\sin^2\frac{\omega_{mn}t}{2}}{(\frac{\omega_{mn}}{2})^2} = 2\pi t\delta(\omega_{mn}) \quad . \tag{4.79}$$

For sufficiently large times ($t \gg \omega_{mn}^{-1}$) we therefore find that the transition probability grows linearly in time. We find Fermi's golden rule for time independent perturbations

$$p_{n\to m}^{\omega=0}(t) = t\frac{2\pi}{\hbar^2}|\langle\phi_m|\hat{V}_0|\phi_n\rangle|^2\delta(\omega_n - \omega_m) \quad . \tag{4.80}$$

Obviously this cannot be correct for arbitrarily large times $t$, because the transition probabilities are bounded by unity.

**High frequency perturbation** If the frequency of the perturbation is unequal to zero, we find two contributions to the transition amplitude, one with a denominator $\omega - \omega_{mn}$ and the other with the denominator $\omega + \omega_{mn}$. As the frequency $\omega$ is always positive, only the first denominator can become zero, in which case we have a resonance and the first term in Eq. (4.76) dominates. We find

$$p_{n\to m}(t) = \frac{1}{4\hbar^2}|\langle\phi_m|\hat{V}_0|\phi_n\rangle|^2 \left( \frac{\sin^2 \frac{(\omega-\omega_{mn})t}{2}}{(\frac{\omega-\omega_{mn}}{2})^2} + \frac{\sin^2 \frac{(\omega+\omega_{mn})t}{2}}{(\frac{\omega+\omega_{mn}}{2})^2} \right) \quad . \quad (4.81)$$

Again we find that for large times the transition probability grows linearly in time which is formulated as Fermi's golden rule for time dependent perturbations

$$p^{\omega\neq 0}_{n\to m}(t) = t\frac{2\pi}{4\hbar^2}|\langle\phi_m|\hat{V}_0|\phi_n\rangle|^2(\delta(\omega_n-\omega_m-\omega)+\delta(\omega_n-\omega_m+\omega)) \quad . \quad (4.82)$$

### The Zeno effect

Fermi's golden rule is not only limited to times that are not too large, but also finds its limitations for small times. In fact for times $t \ll \omega_{mn}^{-1}$ the transition probability grows quadratically in time. This is not only a feature of our particular calculation, but is a general feature of the quantum mechanical time evolution that is governed by the Schrödinger equation. This can be shown quite easily and is summarized in the following

**Theorem 50** *Given a Hamilton operator $\hat{H}$ and an initial state $|\phi(0)\rangle$, then the transition probability to any orthogonal state $|\phi_\perp\rangle$ grows quadratically for small times, i.e.*

$$\lim_{t\to 0}\frac{d}{dt}|\langle\phi_\perp|\phi(t)\rangle|^2 = 0 \quad . \quad (4.83)$$

**Proof:** Let us first calculate the derivative of the transition probability $p(t) = |\alpha(t)|^2$ with $\alpha(t) = \langle\phi_\perp|\phi(t)\rangle$ for arbitrary times t. We find

$$\frac{dp}{dt}(t) = \frac{d\alpha}{dt}(t)\alpha^*(t) + \alpha(t)\frac{d\alpha^*}{dt}(t) \quad . \quad (4.84)$$

Obviously $\alpha(t=0) = \langle \phi_\perp | \phi(0) \rangle = 0$, so that we find

$$\lim_{t \to 0} \frac{dp}{dt}(t) = 0 \quad . \tag{4.85}$$

This finishes the proof $\Box$.
The transition probability for short times is therefore

$$p(t) = p(0) + \frac{t^2}{2} p''(0) + \ldots = \frac{Ct^2}{2} + \text{higher orders in t} \tag{4.86}$$

where $C$ is a positive constant.

Theorem 50 has a weird consequence which you have encountered (in disguise) in the first problem sheet. Assume that we have a two state system with the orthonormal basis states $|0\rangle$ and $|1\rangle$. Imagine that we start at time $t = 0$ in the state $|0\rangle$. The time evolution of the system will be governed by a Hamilton operator $\hat{H}$ and after some time $T$ the system will be in the state $|1\rangle$.

Now let the system evolve for a time $T$, but, as I am very curious how things are going I decide that I will look in which state the system is after times $\frac{T}{n}, \frac{2T}{n}, \ldots, T$. The time evolution takes the initial state after a time $\frac{T}{n}$ into

$$|\phi(\frac{T}{n})\rangle = e^{-i\hat{H}\frac{T}{\hbar n}} |0\rangle \quad . \tag{4.87}$$

Now I make a measurement to determine in which of the two states $|0\rangle$ or $|1\rangle$ the system is. The probability for finding the system in state $|0\rangle$ is $p_1 = 1 - C\frac{T^2}{n^2}$ with some nonzero constant $C$. If I find the system in the state $|0\rangle$, then we wait until time $\frac{2T}{n}$ and perform the same measurement again. The probability that in all of these measurements I will find the system in the state $|0\rangle$ is given by

$$p = \left(1 - C\frac{T^2}{n^2}\right)^n \approx 1 - C\frac{T^2}{n} \quad . \tag{4.88}$$

In the limit of infinitely many measurements, this probability tends to 1. This result can be summarized as

A continuously observed system does not evolve in time.

This phenomenon has the name 'Quantum Zeno effect' and has indeed been observed in experiments about 10 years ago.

With this slightly weird effect I finish this part of the lecture and now move on to explain some features of quantum information theory, a field that has developed in the last few years only.

# Part II

# Quantum Information Processing

# Chapter 5

# Quantum Information Theory

In 1948 Claude Shannon formalised the notion of information and created what is now known as classical **information theory**. Until about five years ago this field was actually known simply as information theory but now the additional word 'classical' has become necessary. The reason for this is the realization that there is a quantum version of the theory which differs in quite a few aspects from the classical version. In these last lectures of this course I intend to give you a flavour of this new theory that is currently emerging. What I am going to show you in the next few lectures is more or less at the cutting edge of physics, and many of the ideas that I am going to present are not older than 5 years. In fact the material is modern enough that many of your physics professors will actually not really know these things very well. So, after these lectures you will know more than some professors of physics, and thats not easy to achieve as a third year student. Now you might be worried that we will require plenty of difficult mathematics, as you would expect this for a lecture that presents the cutting edge of physical science. This however, is not the case. In this lecture I taught you many of the essential techniques that are necessary for you to understand **quantum information theory**. This is again quite different from other branches of physics that are at the cutting edge of research, e.g. super string theory where you need to study for quite a while until you reach a level that allows you to carry out research. The interesting

point about quantum information theory is the fact that it unifies two apparently different fields of science, namely physics and information theory.

Your question will now be: 'What is the connection'? and 'Why did people invent the whole thing'?

The motivation for scientists to begin to think about quantum information came from the rapid development of microprocessor technology. As many of you know, computers are becoming more and more powerful every year, and if you have bought a computer 5 years ago then it will be hopelessly inferior to the latest model that is on the market. Even more amazing is the fact that computers do get more powerful, but they do not get much more expensive. This qualitative description of the development of micro-electronics can actually be quantified quite well. Many years ago, in the 1960's, Gordon Moore, one of the founders of Intel observed that the number of transistors per unit area of a micro chip doubles roughly every 18 months. He then predicted that this development will continue in the future. As it turns out, he was right with his prediction and we are still observing the same growth rate. This development can be seen in Fig. 5.1. Of course sooner or later this development has to stop because would it continue like this for another 20 years than the transistors would be so small, that they would be fully charged by just one electron and they would shrink to the size of an atom. While so far micro-electronics has worked on principles that can be understood to a large extent using classical physics it is clear that transistors that are of the size of one atom must see plenty of quantum mechanical effects. As a matter of fact we would expect that quantum mechanical effects will play a significant role even earlier. This implies that we should really start to think about information processing and computation on the quantum mechanical level. These thoughts led to the birth of quantum information theory. Nevertheless it is not yet clear why there should be an essential difference between information at a classical level and information at the quantum level. However, this is not the case.

Figure 5.1: The number of transistors per unit area of a micro-chip doubles roughly every 18 months.

A very important insight that people had is the observation that information should not be regarded as an isolated purely mathematical concept! Why is that so? You may try to define information as an abstract concept but you should never forget that information needs to

be represented and transmitted. Both of these processes are physical. An example for the storage of information are my printed lecture notes which use ink on paper. Even in you brain information is stored not in an immaterial form, but rather in the form of synaptic connections between your brain cells. A computer, finally, stores information in transistors that are either charged or uncharged. Likewise information transmission requires physical objects. Talking to you means that I am using sound waves (described by classical physics) to send information to you. Television signals going through a cable represent information transfer using a physical system. In a computer, finally, the information that is stored in the transistors is transported by small currents.

So, clearly information and physics are not two separate concepts but should be considered simultaneously. Our everyday experience is of course mainly dominated by classical physics, and therefore it is not surprising that our normal perception of information is governed by classical physics. However, the question remains, what will happen when we try to unify information theory with quantum physics. What new effects can appear? Can we do useful things with this new theory? These are all important questions but above all I think that it is good fun to learn something new about nature.

### Quantum Bits

Before I really start, I will have to introduce a very basic notion. The first one is the generalization of the classical bit. In classical information theory information is usually represented in the form of classical bits, i.e. 0 and 1. These two values may for example be represented as an uncharged transistor ('0') and a fully charged transistor ('1'), see Fig 5.2.

Note that a charged transistor easily holds $10^8$ electrons. Therefore it doesn't make much difference whether a few thousand electrons are missing. A transistor charged with $10^8$ electrons and one with $0.999 \cdot 10^8$ electrons both represent the logical value 1.

This situation changes, when we consider either very small transistors or in the extreme case atoms. Imagine an atom which stores the numbers '0' and '1' in its internal states. For example an electron in the ground state represents the value '0', while an electron in the

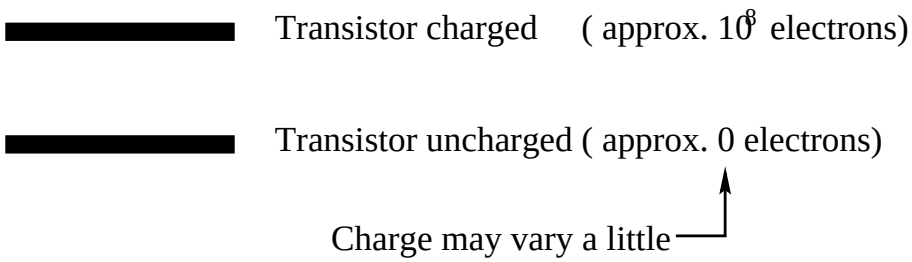


Figure 5.2: A transistor can represent two distinct logical values. An uncharged transistor represents '0' and a fully charged transistor represent '1'. The amount of charge may vary a little bit without causing an error.

excited state represents the value '1' (see Fig. 5.3). In the following we will disregard all the other energy levels and idealize the system as a two level system. Such a quantum mechanical two level system will from now on be called a **quantum bit** or shortly a **qubit**. So far this is just the same situation as in the classical case of a transistor. However, there are two differences. Firstly the atomic system will be much more sensitive to perturbations, because now it makes a big difference whether there is one electron or no electron. Secondly, and probably more importantly, in quantum mechanics we have the superposition principle. Therefore we do not only have the possibilities $'0'$ and $'1'$ represented by the two quantum states $|0\rangle$ and $|1\rangle$, but we may also have coherent superpositions between different values (Fig 5.3), i.e.

$$|\psi\rangle = a|0\rangle + b|1\rangle \tag{5.1}$$

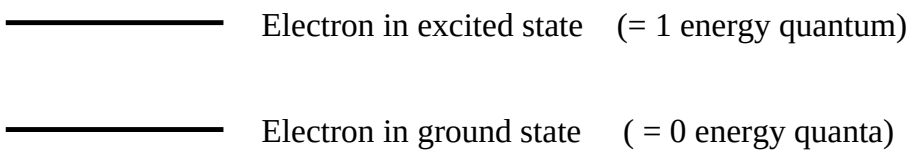


Figure 5.3:

Therefore we have the ability to represent simultaneously two values in a single quantum bit. But it comes even better. Imagine that you

are holding four qubits. Then they can be in a state that is a coherent superposition of 16 different states, each representing binary strings.

$$\begin{aligned} |\psi\rangle &= \frac{1}{4}\left(|0000\rangle + |0001\rangle + |0010\rangle + |0011\rangle \right. \\ &\quad +|0100\rangle + |0101\rangle + |0110\rangle + |0111\rangle \\ &\quad +|1000\rangle + |1001\rangle + |1010\rangle + |1011\rangle \\ &\quad \left. +|1100\rangle + |1101\rangle + |1110\rangle + |1111\rangle\right) \end{aligned} \tag{5.2}$$

Evidently a collection of $n$ qubits can be in a state that is a coherent superposition of $2^n$ different quantum states, each of which represents a number in binary notation. If we apply a unitary transformation o such a state, we therefore manipulate $2^n$ binary numbers simultaneously! This represents a massive parallelism in our computation which is responsible for the fact that a quantum mechanical system can solve certain problems exponentially faster than any classical system can do. However, it is very difficult to build such a quantum system in a way that we can control it very precisely and that it is insensitive to noise at the same time. This has is quite difficult, and there are no real quantum computers around present.

If I have enough time, I am going to explain the idea of a quantum computer in more detail, but first I would like to present some effects and applications of quantum entanglement which have actually been realized in experiments.

## 5.1 Entanglement and Bell inequalities

In the previous section we have seen that the superposition principle is one source for the new effects that can arise in quantum information theory. However, the superposition principle alone exists also in classical physics, e.g. in sound waves or electro-magnetic waves. What is completely absent from classical physics is the notion of entanglement (from the German word *Verschränkung*) of which you have heard about earlier in this lecture. Entanglement was first investigated by Schrödinger in his Gedanken experiments using cats. Later Einstein, Podolsky and Rosen proposed a famous Gedanken experiment which they used to criticise quantum mechanics. The notion of entanglement

arises from the superposition principle together with the tensor product structure of the quantum mechanical Hilbert space. States of the form

$$|\psi\rangle = \frac{1}{2}(|\downarrow\rangle + |\uparrow\rangle) \otimes (|\downarrow\rangle + |\uparrow\rangle) \tag{5.3}$$

are called product states. Such states are called disentangled as the outcome of a measurement on the first system is independent of the outcome of a measurement on the other particle. On the other hand states of the form

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|\downarrow\rangle \otimes |\downarrow\rangle + |\uparrow\rangle \otimes \uparrow\rangle) \tag{5.4}$$

are very strongly correlated. If a measurement on the first system shows that the system is in state $|\downarrow\rangle$ ($|\uparrow\rangle$) then we immediately know that the second system must be in state $|\downarrow\rangle$ ($|\uparrow\rangle$) too. If this would be all that can be said about the correlations in the state Eq. (5.4), then it would not be justified to call this states entangled simply because we can produce the same correlations also in a purely classical setting. Imagine for example, that we have two coins and that they are always prepared in a way, such that either both of them show heads or both of them show tails. Then by looking at only one of them, we know whether the other one shows head or tails. The correlations in the state Eq. (5.4), however, have much more complicated properties. These new properties are due to the fact, that in quantum mechanics we can make measurements in bases other than the $\{|\downarrow\rangle, |\uparrow\rangle\}$ basis. Any basis of the form $\{a|\downarrow\rangle + b|\uparrow\rangle, b^*|\downarrow\rangle - a^*|\uparrow\rangle\}$ can also be used. This makes the structure of the quantum mechanical state Eq. (5.4) much richer than that of the example of the two coins.

A famous example in which these new correlations manifest themselves is that of the Bell inequalities. In the rest of this section I am going to explain to you some of the ideas behind the Bell inequalities and their significance. When Bell started to think about the foundations of quantum mechanics, the work that later led to the Bell inequalities, he was interested in one particular problem. Since the discovery of quantum mechanics, physicists have been worried about the fact that quantum mechanical measurements lead to random measurement outcomes. All that quantum mechanics predicts are the probabilities for

the different possible measurement outcomes. This is quite substantially different from everything that we know from classical physics, where the measurement outcomes are not random if we have complete knowledge of all system variables. Only incomplete knowledge of the system can lead to random measurement outcomes. Does that mean that quantum mechanics is an incomplete description of nature? Are there hidden variables, that we cannot observe directly, but which determine the outcome of our measurements? In a book on the mathematical foundations of quantum mechanics John von Neumann had actually presented a proof that such theories cannot exist. Unfortunately the proof is wrong as von Neumann had made a wrong assumption. The question whether there are hidden variable theories that reproduce all quantum mechanical predictions was therefore still open. John Bell finally re-investigated the experiment that has originally been proposed by Einstein, Podolsky and Rosen and he finally succeeded in showing that there is a physically observable quantity for which local hidden variable theories and quantum mechanics give different predictions. This was an immense step forward, because now it had become possible to test experimentally whether quantum mechanics with all its randomness is correct or whether there are hidden variable theories that explain the randomness in the measurement results by our incomplete knowledge. In the following I am going to show you a derivation of Bell's inequalities.

To do this I now need to define quantitatively what I mean by correlations. The state Eq. (5.4) describes the state of two spin-$\frac{1}{2}$ particles. Assume that I measure the orientation of the first spin along direction $\vec{a}$ and that of the second spin along direction $\vec{b}$. Both measurement can only have one of two outcomes, either the spin is parallel or anti-parallel to the direction along which it has been measured. If it is parallel then I assign the value $a = 1$ ($b = 1$) to the measurement outcome, if it is anti-parallel, then I assign the value $a = -1$ ($b = -1$) to the measurement outcome. If we repeat the measurement $N$ times, each time preparing the original state Eq. (5.4) and then performing the measurement, then the correlation between the two measurements is defined as

$$C(\vec{a}, \vec{b}) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} a_n b_n \quad . \tag{5.5}$$

Now I want to show that correlations of the form Eq. (5.5) satisfy Bells inequalities, given two assumptions are being made

1. We impose locality, i.e. a measurement on the first particle has no effect on the second side when the measurements are at space-like locations (no signal can travel after the measurement on the first particle from there to the second particle before the measurement on the second particle has been carried out). This leads to probabilities that are just products of probabilities for a measurement outcome for the first particle side and the probability for an measurement outcome on the second side.

2. There are hidden variables that we cannot access directly but which influence the probabilities that are observed. This implies that all probabilities are of the form $P_A(a,\lambda)$ and $P_B(b,\lambda)$ where $\lambda$ describes the hidden variables.

Under these assumptions I want to prove that for measurements along the four directions $\vec{a}, \vec{a}', \vec{b}$ and $\vec{b}'$ we find the **Bell inequality**

$$|C(\vec{a},\vec{b}) + C(\vec{a},\vec{b}') + C(\vec{a}',\vec{b}) - C(\vec{a}',\vec{b}')| \leq 2 \ . \tag{5.6}$$

**Proof:** To see this we use the fact that for all $a_n, a'_n, b_n, b'_n \in [-1,1]$

$$|a_n(b_n + b'_n) + a'_n(b_n - b'_n)| \leq 2 \ . \tag{5.7}$$

Now we find

$$\begin{aligned} C(\vec{a},\vec{b}) &= \int d\lambda \rho(\lambda) \left[P_A(+,\lambda)P_B(+,\lambda) + P_A(-,\lambda)P_B(-,\lambda)\right. \\ &\quad \left. -P_A(+,\lambda)P_B(-,\lambda) - P_A(-,\lambda)P_B(+,\lambda)\right] \\ &= \int d\lambda \rho(\lambda) \left[P_A(+,\lambda) - P_A(-,\lambda)\right]\left[P_B(+,\lambda) - P_B(-,\lambda)\right] \\ &\equiv \int Q_A(\vec{a},\lambda) Q_B(\vec{b},\lambda) \rho(\lambda) d\lambda \end{aligned} \tag{5.8}$$

and therefore

$$\begin{aligned} |C(\vec{a},\vec{b}) + C(\vec{a},\vec{b}') + C(\vec{a}',\vec{b}) - C(\vec{a},\vec{b})| \ &\leq \\ \int |Q_A(\vec{a},\lambda)Q_B(\vec{b},\lambda) &+ Q_A(\vec{a},\lambda)Q_B(\vec{b}',\lambda) \\ &+ Q_A(\vec{a}',\lambda)Q_B(\vec{b},\lambda) - Q_A(\vec{a}',\lambda)Q_B(\vec{b}',\lambda)|\rho(\lambda)d\lambda \\ &\leq 2 \end{aligned}$$

This finishes the proof.

Therefore if we are able to explain quantum mechanics by a local hidden variable theory, then Eq. (5.6) has to be satisfied.

Now let us calculate what quantum mechanics is telling us about the left hand side of Eq. (5.6). Quantum mechanically the correlation is given by

$$C(\vec{a},\vec{b}) = \langle\psi|(\vec{a}\hat{\vec{\sigma}}) \otimes (\vec{a}\hat{\vec{\sigma}})|\psi\rangle \quad , \tag{5.9}$$

where $\hat{\vec{\sigma}} = \hat{\sigma}_x\vec{e}_x + \hat{\sigma}_y\vec{e}_y + \hat{\sigma}_z\vec{e}_z$ with the Pauli operators $\hat{\sigma}_i$. Now we can express the correlation in terms of the angle $\theta_{ab}$ between the vectors $\hat{a}$ and $\hat{b}$. We find (this is an exercise for you)

$$C(\vec{a},\vec{b}) = -\cos\theta_{ab} \quad . \tag{5.10}$$

Now we make a particular choice for the four vectors $\vec{a}, \vec{a}', \vec{b}$ and $\vec{b}'$. We

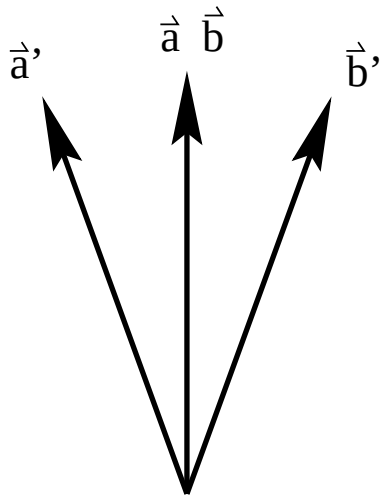


Figure 5.4: The relative orientation of the four directions $\vec{a}, \vec{a}', \vec{b}$ and $\vec{b}'$ along which the spins are measured.

chose $\vec{a}$ and $\vec{b}$ parallel and $\vec{a}'$ and $\vec{b}'$ such that all four vectors lie in one plane. Finally we chose the angles $\theta_{ab'} = \theta_{a'b} = \phi$. All this is shown in Fig. 5.4. Inserting this choice in the left hand side of Eq. (5.6), then we find

$$|1 + 2\cos\phi - \cos 2\phi| \leq 2 \quad . \tag{5.11}$$

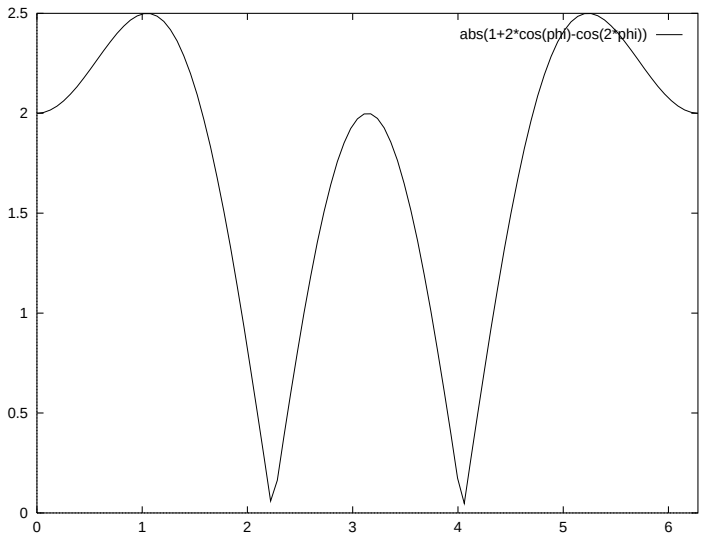


Figure 5.5: The right hand side of Eq. (5.11) is plotted. You can clearly see that it can exceed the value of 2 and achieves a maximum of 2.5.

Plotting the function on the left hand side of Eq. (5.11) in Fig. 5.5 we see that the inequality is actually violated for quite a wide range of values of $\phi$. The maximum value for the left hand side of Eq. (5.11) is given by 2.5 and is assumed for the value $\phi = \pi/3$.

Of course now the big question is whether Bell's inequalities are violated or not. Experiments testing the validity of the Bell inequalities have been carried out e.g. in Paris in 1982. The idea was very simple, but of course it was quite difficult to actually do the experiment. A central source produces photons which are in the singlet state

$$|\psi^-\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle|\downarrow\rangle - |\downarrow\rangle|\uparrow\rangle) \quad . \tag{5.12}$$

One can derive identical Bell inequalities for such a state and this state

was chosen, because it can be produced fairly easily. The way it is done, is via the decay of an atom in such a way that it always emits two photons and that the total change in angular momentum in the atom is zero. Then the two photons are necessarily in a spin-0 state, i.e. a singlet state. The two photons were then flying away in opposite directions towards to measurement apparatuses. These apparatuses were then measuring the polarization state of the two photons along four possible directions that have been chosen to give maximal violation of Bell inequalities. After each decay for each apparatus a direction was chosen randomly and independently from the other side. Finally the measurement result was noted down. This experiment found a value for the correlations of about 2.7 which is reasonably close to the quantum mechanically predicted value of 2.82. More recently more precise experiments have been carried out and the violation of Bells inequalities has been clearly demonstrated.

## 5.2 Quantum State Teleportation

The procedure we will analyse is called quantum teleportation and can be understood as follows. The naive idea of teleportation involves a protocol whereby an object positioned at a place $A$ and time $t$ first "dematerializes" and then reappears at a distant place $B$ at some later time $t + T$. Quantum teleportation implies that we wish to apply this procedure to a quantum object. However, a genuine quantum teleportation differs from this idea, because we are not teleporting the whole object but just its state from particle $A$ to particle $B$. As quantum particles are indistinguishable anyway, this amounts to 'real' teleportation. One way of performing teleportation (and certainly the way portrayed in various science fiction movies, e.g. The Fly) is first to learn all the properties of that object (thereby possibly destroying it). We then send this information as a classical string of data to $B$ where another object with the same properties is re-created. One problem with this picture is that, if we have a single quantum system in an unknown state, we cannot determine its state completely because of the uncertainty principle. More precisely, we need an infinite ensemble of identically prepared quantum systems to be able completely to deter-

mine its quantum state. So it would seem that the laws of quantum mechanics prohibit teleportation of single quantum systems. However, the very feature of quantum mechanics that leads to the uncertainty principle (the superposition principle) also allows the existence of entangled states. These entangled states will provide a form of quantum channel to conduct a teleportation protocol. It will turn out that there is no need to learn the state of the system in order to teleport it. On the other hand, there is a need to send some classical information from $A$ to $B$, but part of the information also travels down an entangled channel. This then provides a way of distinguishing quantum and classical correlations, which we said was at the heart of quantifying entanglement. After the teleportation is completed, the original state of the particle at $A$ is destroyed (although the particle itself remains intact) and so is the entanglement in the quantum channel. These two features are direct consequences of fundamental laws in information processing. I cannot explain these here as I do not have enough time, but if you are interested you should have a look at the article M.B. Plenio and V. Vedral, Contemp. Physics **39**, 431 (1998) which has been written for final year students and first year PhD students.

## 5.3 A basic description of teleportation

Let us begin by describing quantum teleportation in the form originally proposed by Bennett, Brassard, Crepeau, Jozsa, Peres, and Wootters in 1993. Suppose that Alice and Bob, who are distant from each other, wish to implement a teleportation procedure. Initially they need to share a maximally entangled pair of quantum mechanical two level systems. Unlike the classical bit, a qubit can be in a superposition of its basis states, like $|\Psi\rangle = a|0\rangle + b|1\rangle$. This means that if Alice and Bob both have one qubit each then the joint state may for example be

$$|\Psi_{AB}\rangle = (|0_A\rangle|0_B\rangle + |1_A\rangle|1_B\rangle)/\sqrt{2}\,, \tag{5.13}$$

where the first ket (with subscript A) belongs to Alice and second (with subscript B) to Bob. This state is entangled meaning, that it cannot be written as a product of the individual states (like e.g. $|00\rangle$). Note that

this state is different from a statistical mixture $(|00\rangle\langle 00| + |11\rangle\langle 11|)/2$ which is the most correlated state allowed by classical physics.

Now suppose that Alice receives a qubit in a state which is unknown to her (let us label it $|\Phi\rangle = a|0\rangle + b|1\rangle$) and she has to teleport it to Bob. The state has to be unknown to her because otherwise she can just phone Bob up and tell him all the details of the state, and he can then recreate it on a particle that he possesses. If Alice does not know the state, then she cannot measure it to obtain all the necessary information to specify it. Therefore she has to resort to using the state $|\Psi_{AB}\rangle$ that she shares with Bob. To see what she has to do, we write out the total state of all three qubits

$$|\Phi_{AB}\rangle := |\Phi\rangle|\Psi_{AB}\rangle = (a|0\rangle + b|1\rangle)(|00\rangle + |11\rangle)/\sqrt{2} \ . \tag{5.14}$$

However, the above state can be written in the following convenient way (here we are only rewriting the above expression in a different basis, and there is no physical process taking place in between)

$$\begin{aligned} |\Phi_{AB}\rangle &= (a|000\rangle + a|011\rangle + b|100\rangle + b|111\rangle)/\sqrt{2} \\ &= \frac{1}{2}\Big[|\Phi^+\rangle(a|0\rangle + b|1\rangle) + |\Phi^-\rangle(a|0\rangle - b|1\rangle) \\ &\quad + |\Psi^+\rangle(a|1\rangle + b|0\rangle) + |\Psi^-\rangle(a|1\rangle - b|0\rangle)\Big] \ , \end{aligned} \tag{5.15}$$

where

$$|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2} \tag{5.16}$$
$$|\Phi^-\rangle = (|00\rangle - |11\rangle)/\sqrt{2} \tag{5.17}$$
$$|\Psi^+\rangle = (|01\rangle + |10\rangle)/\sqrt{2} \tag{5.18}$$
$$|\Psi^-\rangle = (|01\rangle - |10\rangle)/\sqrt{2} \tag{5.19}$$

form an ortho-normal basis of Alice's two qubits (remember that the first two qubits belong to Alice and the last qubit belongs to Bob). The above basis is frequently called the Bell basis. This is a very useful way of writing the state of Alice's two qubits and Bob's single qubit because it displays a high degree of correlations between Alice's and Bob's parts: to every state of Alice's two qubits (i.e. $|\Phi^+\rangle, |\Phi^-\rangle, |\Psi^+\rangle, |\Psi^-\rangle$) corresponds a state of Bob's qubit. In addition the state of Bob's qubit

in all four cases looks very much like the original qubit that Alice has to teleport to Bob. It is now straightforward to see how to proceed with the teleportation protocol:

1. Upon receiving the unknown qubit in state $|\Phi\rangle$ Alice performs projective measurements on her two qubits in the Bell basis. This means that she will obtain one of the four Bell states randomly, and with equal probability.

2. Suppose Alice obtains the state $|\Psi^+\rangle$. Then the state of all three qubits (Alice + Bob) collapses to the following state

$$|\Psi^+\rangle(a|1\rangle + b|0\rangle) \; . \tag{5.20}$$

   (the last qubit belongs to Bob as usual). Alice now has to communicate the result of her measurement to Bob (over the phone, for example). The point of this communication is to inform Bob how the state of his qubit now differs from the state of the qubit Alice was holding previously.

3. Now Bob knows exactly what to do in order to complete the teleportation. He has to apply a unitary transformation on his qubit which simulates a logical NOT operation: $|0\rangle \rightarrow |1\rangle$ and $|1\rangle \rightarrow |0\rangle$. He thereby transforms the state of his qubit into the state $a|0\rangle + b|1\rangle$, which is precisely the state that Alice had to teleport to him initially. This completes the protocol. It is easy to see that if Alice obtained some other Bell state then Bob would have to apply some other simple operation to complete teleportation. We leave it to the reader to work out the other two operations (note that if Alice obtained $|\Phi^+\rangle$ he would not have to do anything). If $|0\rangle$ and $|1\rangle$ are written in their vector form then the operations that Bob has to perform can be represented by the Pauli spin matrices, as depicted in Fig. 5.6.

$$\begin{aligned}
&\text{Alice finds} |\Phi^+\rangle \quad \longrightarrow^0 \quad \text{Bob does nothing} \\
&\text{Alice finds} |\Phi^-\rangle \quad \longrightarrow^1 \quad \text{Bob performs } \sigma_z \\
&\text{Alice finds} |\Psi^+\rangle \quad \longrightarrow^2 \quad \text{Bob performs } \sigma_x \\
&\text{Alice finds} |\Psi^-\rangle \quad \longrightarrow^3 \quad \text{Bob performs } \sigma_z\sigma_x
\end{aligned}$$

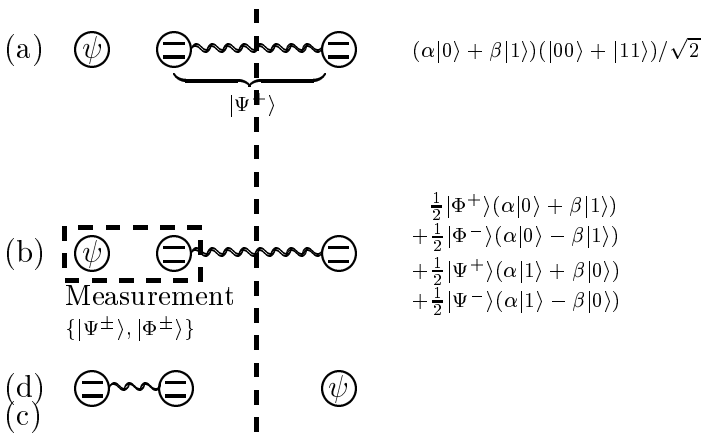


Figure 5.6:

Figure 5.6: The basic steps of quantum state teleportation. Alice and Bob are spatially separated, Alice on the left of the dashed line, Bob on the right. (a) Alice and Bob share a maximally entangled pair of particles in the state $(|00\rangle+|11\rangle)/\sqrt{2}$. Alice wants to teleport the unknown state $|\psi\rangle$ to Bob. (b) The total state of the three particles that Alice and Bob are holding is rewritten in the Bell basis Eqs. (5.16-5.19) for the two particles Alice is holding. Alice performs a measurement that projects the state of her two particles onto one of the four Bell states. (c) She transmits the result encoded in the numbers $0, 1, 2, 3$ to Bob, who performs a unitary transformation $\mathbf{1}, \sigma_z, \sigma_x, \sigma_z\sigma_x$ that depends only on the measurement result that Alice obtained but **not** on the state $|\psi\rangle$! (d) After Bob has applied the appropriate unitary operation on his particle he can be sure that he is now holding the state that Alice was holding in (a).

An important fact to observe in the above protocol is that all the operations (Alice's measurements and Bob's unitary transformations) are *local* in nature. This means that there is never any need to perform a (global) transformation or measurement on all three qubits simultaneously, which is what allows us to call the above protocol a genuine teleportation. It is also important that the operations that Bob performs are independent of the state that Alice tries to teleport to Bob. Note also that the classical communication from Alice to Bob in step 2 above is crucial because otherwise the protocol would be impossible to

execute (there is a deeper reason for this: if we could perform teleportation without classical communication then Alice could send messages to Bob faster than the speed of light, remember that I explained this in a previous lecture.

Important to observe is also the fact that the initial state to be teleported is at the end destroyed, i.e it becomes maximally mixed, of the form $(|0\rangle\langle 0| + |1\rangle\langle 1|)/2$. This has to happen since otherwise we would end up with two qubits in the same state at the end of teleportation (one with Alice and the other one with Bob). So, effectively, we would clone an unknown quantum state, which is impossible by the laws of quantum mechanics (this is the no-cloning theorem of Wootters and Zurek). I will explain this in more detail later.

For those of you who are interested in quantum information theory, here are some references for further reading:

L. Hardy, Cont. Physics 39, 415 (1998)
M.B. Plenio and V. Vedral, Cont. Physics 39, 431 (1998)
J. Preskill, http://www.theory.caltech.edu/people/preskill/ph229/