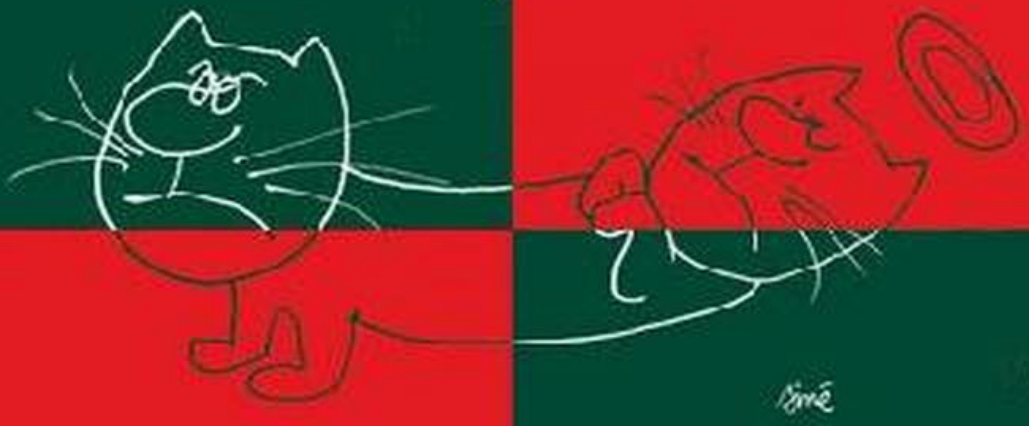
Lectures on Quantum Mechanics



Jean-Louis Basdevant

Springer

Lectures on Quantum Mechanics

Jean-Louis Basdevant

Lectures on Quantum Mechanics

Springer

Professor Jean-Louis Basdevant
Physics Department
École Polytechnique
91128 Palaiseau
France
jean-louis.basdevant@polytechnique.edu

Cover illustration: Siné, Schrödinger's cat.

Library of Congress Control Number: 2006936625

ISBN-10: 0-387-37742-5 e-ISBN-10: 0-387-37744-1
ISBN-13: 978-0-387-37742-1 e-ISBN-13: 978-0-387-37744-5

Printed on acid-free paper.

9 8 7 6 5 4 3 2 1

springer.com

To my son Nicolas

who shared this for years since he was a child,
and who taught me that people who can't laugh
are not serious people

Contents

Preface

This book is the "verbatim" transcription of the introductory lectures on quantum mechanics that I delivered for more than 25 years at the École Polytechnique. It is not a textbook. I was dragged into writing it by friends, among whom are many former students of mine. For sure, this text is obviously less complete than a textbook.

The argument that eventually convinced me to write this book is that the textbooks I had written on the subject, both in French and in English, were terribly deprived of life, action, thoughts, and the questioning that I always liked to put in my narrative account of the ideas and applications of the subject. The human aspect of the experimental investigations and of the ensuing discovery of basic principles made the lectures lively (besides the fact that the mind needs to rest a few minutes following a difficult argument). I always thought that teaching science is incomplete if it does not incorporate the human dimension, be it of the lecturer, of the audience, or of the topic to which it is devoted.

What is true is that the students at the École Polytechnique, who were all selected after a stiff entrance examination, and whose ambitions in life were diverse – in science, in industry, in business, in high public office – all had to follow this introductory physics course. As a consequence, the challenge was to try to get them interested in the field whatever their future goals were. Of course, quantum mechanics is an ideal subject because one can be interested in it for a variety of reasons, such as the physics itself, the mathematical structure of the theory, its technological spinoffs, as well as its philosophical or cultural aspects. So the task was basically to think about the pedagogical aspects, in order to satisfy audiences that went up to 500 students during the last 10 years.

I must say a few words about the content of this book. First, of course, my lectures evolved quite a lot in 25 years. Actually, they were never the same from one year to the next. Minds evolve; students' minds as well as mine. Science evolves; during that period, there appeared numerous crucial experimental and technological steps forward. The experimental proof of the

violation of Bell's inequalities and quantum entanglement is perhaps the most fundamental and notorious in the history of science and knowledge. There are many other experiments, on decoherence, quantum information, the appearance of quantum dots in advanced technologies, the discovery of Bose–Einstein condensation, and enormous progress in astrophysics and cosmology.

So each lecture itself must be considered as a superposition of texts and topics, which I could not have covered completely in about an hour and a half. I used to make selections according to my mood, to latest experimental results, and to the evolution of the students' minds in mathematics, in physics, and in regard to the world they were facing. After each lecture, on a weekly basis, students went to classes in groups of 20, where the real "serious" applications and exercises were performed, be it in order to understand the following week's lecture. The first lecture always consisted of a general description of contemporary physics and of the various courses that students were offered in their curriculum. I have reproduced an example in the first chapter.

I have deliberately omitted in this book many physical or technical questions that were treated in the smaller-group classes. I published two books. One is a textbook: *Quantum Mechanics*, J.-L. Basdevant and Jean Dalibard, Heidelberg: Springer-Verlag, 2002 (new edition in 2005). The other one is a collection of problems and their solutions: *The Quantum Mechanics Solver*, J.-L. Basdevant and Jean Dalibard, Heidelberg: Springer-Verlag, 2000 (completely revised in 2005). All of these problems concern contemporary experimental or theoretical developments, some of which had appeared in the specialized literature a year or so before we gave them as written examinations. Needless to say that, if the second of these books is somewhat unusual, there are dozens of excellent books on quantum mechanics, some masterpieces among them, which I often prefer to mine.

I thank Jean Dalibard, who is now my successor, and Philippe Grangier for their constant help during the last 10 to 15 years. They are, in particular, responsible for part of the text on quantum entanglement and Bell's inequalities, of which they are worldwide-known specialists.

I want to pay tribute to the memory of Laurent Schwartz, and I thank Jean-Michel Bony. Both had the patience to explain to me with an incredible profoundness and clarity the mathematical subtleties of quantum mechanics. This enabled me to eliminate most of the unnecessary mathematical complications at this stage and to answer the questions of my more mathematically minded students. Indeed, if quantum mechanics was a very rich field of investigation for mathematicians, it is really the physics that is subtle in it.

Jean-Louis Basdevant
Paris, April 2006

1

Praise of physics

Why do physics? Laurent Schwartz, the man I admired most, liked the question: what's the use of doing mathematics? "It's very simple," he said,

> mathematicians study $\mathcal{L}^p$ spaces, negligible sets, and representable functors. One must certainly do mathematics. Because mathematics allows to do physics. Physics allows to make refrigerators. Refrigerators allow to keep lobsters, and lobsters are useful for mathematicians who can eat them and therefore be in a good mood to do $\mathcal{L}^p$ spaces, negligible sets, and representable functors. It is obviously useful to do mathematics.

Why do physics? I've often thought about that question.

1.1 The interplay of the eye and the mind

First, there are intellectual reasons. Physics is a fascinating adventure between the eye and the mind, that is, between the world of phenomena and the world of ideas. Physicists look at Nature and ask questions to which they try and imagine answers.

For instance: why do stars shine? It's important. The sun is an ordinary star, similar to 80% of the 200 billion stars of our galaxy. But it is unique and incomparable, because it is our star. In mass, the sun is made of 75% hydrogen and 25% helium (actually a plasma of electrons and nuclei). Its parameters are

$$\text{radius } R = 700,000 \text{ km} , \quad \text{mass } M = 2\ 10^{30} \text{ kg} ,$$

$$\text{power (luminosity) } L = 4\ 10^{23} \text{ kW}, \text{ surface temperature } T = 6000 \text{ K} .$$

One mustn't overestimate the power of the sun. We are much more efficient. If you calculate the power-to-mass ratio, the sun has a score of 0.2 mW/kg which is very small. We consume on the average 2400 kilo-calories per day; that is, 100 watts, 25% of which is used by the brain. Our brain has

therefore a power of 25 watts, it is consequently 10,000 times more powerful than the sun for a given mass! One always tells kids they are bright, but without explaining why, and where that can lead them.

So, why does the sun shine? One usually thinks that it shines because of the powerful thermonuclear reactions that take place inside it. But that is not true! I now show you that, contrary to common prejudices, it is gravitation that makes stars shine and that thermonuclear reactions cool them permanently.

1. Stars shine because they are thot and any hot body radiates energy.
2. They are hot because of gravity. Stars are huge masses of gas, mainly hydrogen, which are strongly compressed by their own weight. This brings them to high temperatures. Stars are self-gravitating systems in equilibrium under their own weight.
3. OK, but you may object that a hot compressed gas loses energy by radiating. If it loses energy, then it contracts and it cools down.
4. Well, the amazing thing is that, on the contrary, a self-gravitating gas does contract when it loses energy, but its temperature increases!

This is understandable. If the size of a self-gravitating system decreases, the gravitational attraction increases and, in order to maintain equilibrium, the centrifugal force must increase, therefore the components must rotate more and more rapidly.

Because temperature reflects motion (i.e., kinetic energy of the constituents of a gas), if the particles move faster, the temperature increases. Therefore, if a self-gravitating system contracts, its temperature increases.

This is quite easy to formulate. A star such as the sun can be represented by an ideal gas of $N \sim 10^{57}$ particles, at an average temperature $\langle T \rangle$. The temperature itself varies from 15 million degrees in the center to 6000 degrees on the surface.

• The gravitational potential energy of a sphere of mass M and radius R is proportional to Newton's constant, to the square of the mass, and to the inverse of the radius

$$E_G = -\gamma \frac{GM^2}{R},$$

where γ is a dimensionless constant of order 1 ($\gamma = 3/5$ if the mass distribution is uniform in the sphere). It is negative because one must give energy to the system in order to dissociate it.

• In a self-gravitating system, the total kinetic energy E_{kin} of the orbiting particles, that is, the internal energy U of the gas, is equal to half of the absolute value of the potential energy: $U = E_{kin} = 1/2|E_G|$. This is called the virial theorem, which is obvious for a circular orbit around a massive center, and which can be generalized with no great difficulty to a large number of arbitrary trajectories. It follows directly by assuming that the time derivative of the moment of inertia is a constant $d^2(\sum m_i r_i^2)/dt^2 = 0$.

• Therefore, the total energy of the star is

$$E = E_{kin} + E_G = -\frac{1}{2}\gamma\frac{GM^2}{R} \ .$$

Again, it is negative because the system is bound. One must bring energy in order to dissociate it.

• On the other hand, the average temperature $\langle T\rangle$ of the star is related to the mean kinetic energy of the constituents by Boltzmann's relation

$$\frac{3}{2}Nk\langle T\rangle = E_{kin} = \frac{1}{2}\gamma\frac{GM^2}{R}, \text{ with } N \sim 10^{57} \ .$$

• When the star radiates, it loses energy. Its energy decreases and becomes more negative, therefore its radius decreases, it is compressed, and its temperature increases. When the star loses energy it radiates more and more strongly.

Therefore, stars shine because of gravitation. Since its formation in a molecular cloud, the sun, whose present mass and radius we know, has lost a gravitational energy of $\Delta E \simeq 10^{41}$ joule; its average temperature is of the order of $\langle T\rangle \simeq 3$ million degrees, which is quite acceptable.

Now, we must think! Paleontologists teach us the following:

• The "blue-green algae" or cyanobacteria, who are responsible for the birth of life on earth because they manufactured the oxygen in the atmosphere, existed 3.5 billion years ago.

• Our cousin, the Kenyapithecus, lived 15 million years ago, and our ancestors Lucy, an Australopithecus afarensis, as well as Abel in Chad, lived 3.5 million years ago, and Orrorin (the ancestor of the millennium) lived 6 million years ago.

• Dinosaurs lived 200 million years ago; they ate a lot of greenery.

Therefore, the sun must have been stable during all that time. It must have had approximately the same power (the weather stayed roughly the same) and the same external temperature (the sun radiates in the visible part of the spectrum where photosynthesis takes place, which allows vegetables to grow).

Now, if the sun has been stable, we can evaluate roughly when it started to shine. If it has had the same power L for a long time, we can evaluate the time that it took to get rid of the energy ΔE, that is, a time $t = \Delta E/L \sim 10$ million years. The sun started shining 10 million years ago. Therefore, the sun is only 10 million years old.

Consequently, we have just proven scientifically that dinosaurs never existed; they were just invented to make *Jurassic Park*. The Kenyapithecus was just invented to give us a superiority complex. Because the sun did not shine at that time!

Of course, maybe, in order to save energy, the Creator turned on the sun once in a while, just when it was necessary.

Because it is not my purpose here to give rise to religious vocations, we must find something else. Actually, the answer is right in front of us. Suppose

there is a source of energy in the sun, and that at some temperature something ignites in the gas. The combustion releases energy. It increases the energy of the gas, which becomes less negative. Therefore the gas expands, which is understandable. But if its radius increases, then its temperature decreases!

If a self-gravitating gas loses energy, its temperature increases; if it gains energy, its temperature decreases. It has a negative specific heat. And that's great. A combustion stabilizes the star's temperature. An excess of combustion cools the gas and slows down combustion. Conversely, an insufficient combustion rate heats the gas and revives combustion. The combustion energy is calmly evacuated at constant temperature. The system is self-regulated. The energy we receive is indeed due to thermonuclear reactions, but as long as the combustion lasts, the star evacuates that energy in stable conditions. That is exactly what we need for blue-green algae, dinosaurs, Lucy, and Orrorin.

So, our star is stable, but for how long? As long as the fuel is not exhausted. With the mass and power of the sun, one can check that if the combustion were chemical, for instance,

$$2H \to H_2 + 4,5\,\text{eV}, \text{ that is, 2 eV per proton,}$$

the available energy would be 10^{38} joule; the lifetime of the sun would be at most 30,000 years, which is much too short. On the contrary, nuclear fusion reactions such as

$$4\,p \to^4 He + 27\,\text{MeV}, \text{that is, } \simeq 7,000,000 \text{ eV per proton,}$$

is a million times more energetic, which leads us to roughly ten billion years. And we have made in three pages a theory of the sun which is not bad at all in first approximation!

The conclusion is that stars shine because of gravitation, which compresses them and heats them. Nuclear reactions, which should make them explode, simply allow them to react against gravitational collapse. They cool down the stars permanently and give them a long lifetime. The sun has been shining for 4.5 billion years and will continue to do so for another 5 billion years.

That is an example of the confrontation of a physicist's ideas with the observed world. And that's what is interesting in physics. If the ideas we have do not correspond to what we see, we must find other ideas. One cannot change Nature with speeches.

In physics, one can make mistakes but one cannot cheat.

One can do lots of things with speeches. The story says that some governor found that the number π was too complicated. So he decided that from then on, in his State, π would be equal to 3 (π version 3.0), which is much simpler for everyone. Well, that doesn't work! One can observe that if π were equal to 3, four inches of tires would be missing on bicycles, which would be uncomfortable, and five inches of stripes would be missing on a colonel's hat, which would be inelegant.

Politicians learn to make speeches and scientists learn to use their intelligence. It is a radically different way of thinking. The two methods happen to be useful in practice: there exist scientists who can explain their findings, and there exist intelligent politicians.

1.2 Advanced technologies

There are many other reasons to learn physics, of course. Our world is filled with advanced technologies such as the Internet, GPS, nanotechnologies, optoelectronics, and so on. Many of these new technologies come from the results of fundamental research obtained in the last 10 or 20 years, sometimes in very recent years.

Fig. 1.1. Forest of microlasers, each of which is a pile of pancakes of alternating slices of GaAs and GaAl semiconductors. The diameter of each element is 0.5 micrometers, the height is 7 micrometers.

Figure 1.1 shows the details of a sample of contemporary microelectronics. This was made in the 1990s. It consists of a forest of microlasers each of which is a pile of pancakes of alternating slices of gallium–arsenide and gallium–aluminum semiconductors. We come back to such devices. These components have numerous applications in infrared technologies. Infrared sensors are used as temperature sensors for night vision, on automobiles to see pedestrians at night, in rescuing operations in the ocean, to measure the temperature of the earth and of the ocean from satellites, in telecommunications with fiber optics, and so on.

What is really amazing is the size. The size of each of these elements is that of a small bacteria, a fraction of a micrometer. The thickness of the slices is of the order of 10 nanometers, the size of a virus, the smallest living object. In order to imagine the order of magnitude, if instead of lasers one had written letters (which is quite possible, even though it may seem ridiculous) one could write and read on 1 cm^2 of silicon, the complete works of Sigmund Freud, Carl von Clausewitz, Karl Marx, Shakespeare, Snoopy, Charlie Brown, and Anaïs Nin (which may be useful during a boring lecture).

These components are the optoelectronic components of the new generation where light is used instead of electricity. Laser photons replace electrons. They collect and transmit information directly on the components of integrated circuits. They are called quantum components because in order to conceive, to manufacture, and to use such components, one cannot bypass quantum physics.

Micro- and nanotechnologies are undergoing tremendous progress at present. In electronics, one of the present world records consists of a transistor that is 18 nm long, a hundred times smaller than the smallest present transistors. One could put three billion such transistors on a dime. It is the physical limit due to the Heisenberg inequalities. One builds automated microsystems that possess the three functions of being sensors, of processing information, and of activating a reaction or a response. Such systems are found in all sectors of technology, from electronic equipment of cars up to medicine, and including telecommunications, computers, and space technologies.

One can multiply the number of such examples. What is true is that whatever one's own perspectives are, being an executive, an engineer, or a scientist in any domain, one must be familiar with such developments, be it only to position oneself in front of them. It is useless to try and know everything, but one must be capable of inventing and acting. Bill Gates, the richest man in the world, made his fortune because he was able to use these developments; quantum mechanics accounts for at least 30% of each of his dollars.

1.3 The pillars of contemporary physics

In order to understand contemporary physics, three fundamental links are necessary: quantum mechanics, statistical mechanics, and relativity.

Quantum mechanics, which is dealt with in this book, is the complete and fundamental theory of structures and processes at the microscopic scale, that is, atomic, molecular, or nuclear scales. It is the fundamental and inescapable field. All physics is quantum physics.

The first success of quantum mechanics is to explain the structure of matter, atoms and molecules. But it is in the interaction of atoms and molecules with radiation that one finds the greatest progress, both fundamental and technological, in recent years.

1.3.1 Mysteries of light

Light has always been considered as a physical phenomenon of its own, the great mystery. It is our first tool to explore the world, to probe the cosmos as well as the infinitely small. In physics, light delivered simultaneously the two great discoveries of the 20th century: relativity with Michelson, Einstein, Lorentz, and Poincaré in 1905, and quantum physics with the black-body theory of Planck in 1900 and Einstein's photon in 1905.

The nature of light has always been a mysterious and fundamental question. The first theory of light originated from the importance given to light rays. Just look at Figure 1.2. This drawing seems quite natural and ordinary,

Fig. 1.2. Child's painting.

not at all scientific. Fifty percent of children draw the sun that way. But 50% is extraordinary, because you have never seen those light rays attached to the sun. This child cannot explain why he drew them, but for everybody their presence is perfectly normal.

In nature, one can see light rays only under special circumstances, when light is partially screened by clouds or trees. And the fact that light rays are straight, and that they materialize the perfect straight lines of geometry was always considered as fundamental.

For thousands of years, a sacred character was attributed to light rays, as one can see in Figure 1.3. In Egyptian as well as in Christian culture, light rays are a medium through which the beyond becomes accessible to humans.

In the 18th century, Newton decided that light was made of corpuscles, because only particles can travel along straight lines. However, since the end of the 17th century, interference and diffraction phenomena were known and the 19th century saw the triumph of wave optics. Nobody could imagine the incredible answer of quantum theory. Einstein understood in 1905 that light was both wavelike and corpusclelike. Quantum optics, that is, the quantum description of electromagnetic radiation, also plays a decisive role in modern science and technology. The interaction between radiation and light has produced laser physics. Lasers beams are the modern legendary light rays.

The manipulation of cold atoms with laser beams is one of the highlights of present fundamental research. There are numerous practical applications such as CD and DVD records, inertial controlled fusion, optoelectronics, gyrolasers, and others. Intensive work is carried out on optical computers.

Fig. 1.3. Left: Stele of Taperet (around 900–800 B.C.) Taperet worships the sun god Horakhty whose rays are materialized by lily flowers of all colors. (Le Louvre Museum, Paris.) Right: Il Sodoma, Saint Sebastian (1526). (Galleria Pitti, Florence.)

1.3.2 Fundamental structure of matter

Elementary particle physics started a bit more than one century ago with the discovery of the electron by J. J. Thomson in 1897. It tries to answer two questions:

- What is the world made of?
- How does the world work?

In one century, one has found a nearly complete answer. At present, we possess a simple theory of the Universe, called the Standard model, in which a small number of elementary constituents of matter, quarks and leptons, interact through a simple set of forces. And that theory explains all natural phenomena!

In October 1989, an extraordinary event happened. A measurement, done in the CERN LEP collider in Geneva, allowed us to count the number of different constituents of matter. There are 24 of them.

The validity of the Standard model is constantly verified experimentally more and more accurately. The next to last element, the top quark, was identified in 1995. The last one, the τ neutrino, was observed directly in 2001. One expects to identify the Higgs boson, a field quantum responsible for the mass of particles, in the future Large Hadron Collider facility. Many physicists consider the Standard model to be very close to the end of the story in the infinitely small structure of matter, and, for the moment, there is no experimental evidence against that. It is simply a problem of esthetics and a semi-metaphysical problem, namely the whereabouts of the big bang.

Matter is made of atoms. In 1910, Rutherford discovered that atoms are made of tiny but heavy nuclei bound to electrons by electromagnetic forces. In the 1930s, people showed that nuclei also have an internal structure. They

are systems of nucleons (protons and neutrons), bound by nuclear forces of small range and large intensity. Then, in the 1960s, people understood that nucleons are not elementary either. They have an internal structure: they are systems of three quarks. There are two sorts of quarks, the u (up) quark of charge $+2/3$ and the d (down) quark of charge $-1/3$. The proton is a (uud) system, and the neutron a (udd) system. Quarks are imprisoned against each other by "gluons".

What is amazing in the Standard model is that apparently quarks and leptons (electrons, neutrinos, etc.) are experimentally pointlike. "After" them, there is nothing else. Electrons and quarks are elementary down to 10^{-18} m. They are the true elements of matter.

Actually, this end of the story is a problem. The model works too well! Pointlike objects are not consistent with what we know from quantum field theory or from general relativity. At very short distances, it seems that the notion of particles must be replaced by some other concept: superstrings, which are extended objects. This is one of the major problems of fundamental physics. This problem is related to something we have not yet mastered, unifying general relativity, which is primarily a geometrical theory, with quantum mechanics which is basically nongeometrical. In this problem, we might find the answer to fascinating questions such as: why is the dimensionality of space equal to three? The answer is probably that actually there are several other dimensions but that these cannot be seen with the naked eye. Like a bug on a straw, it seems that the bug moves up and down on a one-dimensional space, the straw, but the bug itself knows that it can also turn around along the surface of the straw, and its world is two-dimensional.

Nuclear physics (i.e., the physics of atomic nuclei) is a beautiful and complex fundamental field of research, but it is also an engineering science that plays a considerable role in our societies.

It has many aspects. In medicine, nuclear magnetic resonance imaging, as well as the various applications of radioactivity, and proton and heavy ion therapy, are revolutions in medical diagnosis and therapy. It is needless to emphasize the problems of energy in the world. It is a fact that in order to dismantle a nuclear plant, it takes 50 years, and in order to launch a new nuclear option (in fusion or in fission) it will take 30 or 40 years. In any case, we are concerned with that question.

1.4 The infinitely complex

Now, it is very nice to know the laws of physics at the microscopic scale, but we must some day turn back to the physical world at our scale, namely macroscopic physics. When we eat a pound of strawberry pie, we don't think we're eating half a pound of protons, half a pound of neutrons, and a little overweight of electrons. It's perfectly true, but it's silly, it's perverse, and it's disgusting.

Statistical physics studies the global and collective behavior of large numbers of particles or systems whose individual properties are known. It is a great discovery of the last decades that one cannot reconstruct everything from the very beginning, that is, microscopic laws. As soon as one deals with large numbers of constituents, there appear new phenomena, new singularities or regularities that are specifically macroscopic. These are related to the number of constituents rather than to their specific nature. Examples are:

- Collective effects, phase transitions
- Shapes, ordered structures
- Irreversibility, life and death

This kind of problem (i.e., physics of the infinitely complex world) is one of the most fascinating fields of physics at present. To understand it, to dominate it, will have a considerable impact not only in physics, but in biology where reproducible ordered structures are fundamental, to some extent in economics, and maybe some day in sociology. The most fascinating system is the brain itself.

At this point, there appears a much simpler and more relevant answer to the question of what is the use of doing physics. Physics is fun; it is amusing.

Take a simple example. The fact that water freezes at 0°Celsius is a very ancient scientific observation. Everyone knows that. At school, that property is used to define water: "Water is a colorless tasteless liquid, it is used to wash, some people even drink it, and it freezes at zero degrees C!"

But, one day, we learn physics. We learn that water is a liquid made of H_2O molecules that wander around at random. Ice is a crystal where the same molecules H_2O are well organized in a periodic structure.

That's really an amazing phenomenon! Why on earth do those molecules decide at 0° to settle down in an ordered structure? It is a mystery! We all know how difficult it is, after a break, to put in an ordered state a number of objects or beings whose natural tendency is to be dispersed.

Therefore, because we have learned some physics, we discover a very deep aspect in a very familiar fact: the freezing of water. And that is when we make progress.

But, in order to do that, one must learn to observe and ask oneself questions about reality. Creativity is much more important than knowledge or equations, and it is fundamental to develop it and to preserve it. Physics, and in particular experimental physics, is an excellent field for that operation.

Materials

Physics of condensed matter, as opposed to corpuscular physics, is a broad domain common to physics, to mechanics, to chemistry, and to biology.

Materials have perhaps the most important role in the evolution of science and technology, including semiconductors, steels, concretes, composite materials, glasses, polymers, paints, and so on. Practically all the important

breakthroughs of the progress of mankind are associated with the discovery and the use of new materials: think of stones, flint, bronze 10,000 years ago, iron, more recently aluminum and aeronautics, silicon, electronics and computer science.

Up to the 1970s, it was customary to differentiate between the mechanical properties of solids, that is, metallurgy and electrical properties. But thanks to quantum physics and statistical physics, materials science has become a unified theory, because we can understand it from its microscopic aspect.

Solids are aggregates of atoms or molecules that are bound by the electrons of crystalline bonds. These electrons form a more or less hard cement that determines the mechanical properties, resistance, hardness, and plasticity. And it is, in turn, the physics of these electrons that determines the electrical and thermal properties. All these properties are intimately connected.

At first, it is difficult to appreciate the importance and the depth of such a global synthetic understanding. Metallurgy was for a long time purely empirical. By manipulating such and such a mixture, one used to obtain such and such a result; knowledge was transmitted by word of mouth. Sometimes it was great, such as in Syria in the 13th century. There was a problem in the weapons industry for making swords. Iron is a resistant material, but it is soft and iron swords got bent easily. On the other hand, carbide is hard, but it breaks easily. Damascus steel consisted of alternating sheets of iron and carbide. This allowed them to make swords that were both hard and resistant (sometimes physics isn't that funny; it would have been much more fun if the result had been soft and fragile). It was a revolution in weaponry, and it is very clever from the modern point of view; in fact this is an example of composite materials.

The best composite materials that people try to imitate are biological composites such as bones or shells. These associate the hardness of limestone apatite, which is fragile but hard, with the resistance of biological collagen.

For modern purposes, one must conceive a material directly in view of the function it should have, namely the desired mechanical, electrical, chemical, and optical properties. And this is done more and more systematically.

In recent years, there has been a technological breakthrough with what are known as smart materials, for instance, materials with shape memory. A piece of material can have some shape (think of a metal wire) that we can change. The surprise is that a smart material recovers its initial shape if it is heated. This does not occur with just any material. The alloys with shape memory are metal alloys (for instance, nickel–titanium) that undergo a phase transition between two crystalline structures, martensite and austenite, called a martensitic transition. Above the transition temperature, the structure is a compact face centered cubic austenite; below it is a less compact centered cubic crystal. One can give a material the shape one wants above the transition temperature. It holds that shape below the transition point, but one can change this shape by a plastic deformation. If after that change, one heats the material, it recovers its original shape because in the martensite phase, there

are domains with an fcc arrangement that "remember" the initial shape and convey the structure to the entire material.

The industrial issues are huge. The applications of such materials are found in many different domains such as opening up satellite antennas, bone or tooth prosthesis, and heart and blood vessel surgery. One can crumple pieces of smart material at usual temperature (20°C to 25°C) and insert them in a blood vessel. Once they have reached their destination they open up and take their functional shape at the temperature of the human body, 37°C. This also gives an explanation of the "magic power" of magicians or crooks who became famous by winding keys or forks at a distance by "caresses."

There exist in addition hysteresis phenomena. One can "educate" such materials and construct artificial "muscles" that can transform heat into work. Again, industrial issues are huge.

1.5 The Universe

To end this brief panorama of physics, one should say a few words about astrophysics. The three basic fields – quantum mechanics, statistical physics, and relativity – are deeply connected in astrophysics and cosmology, the history of the Universe.

Nuclear astrophysics gives us the clue as to how stars work, how old they are, and how they evolve. The sun is a complex object, with permanent activity, spectacular solar flares, and surface volcanism. It emits matter at millions of kilometers and at millions of degrees, which is difficult to understand inasmuch as the surface temperature is 6000 K. It is in stars that heavy elements are synthesized by thermonuclear fusion reactions. Hence, nuclear physics allows us to give a life, a scenario, to the cosmos, which is a very special theater in which one cannot perform any experiment, not even applaud.

Finally, there is a major question, perhaps the most fascinating: are we alone in the Universe? Are there other, extraterrestrial, thinking beings in the Universe? More and more extrasolar planets are being discovered, around other stars. How can we know whether they are inhabited? We give a partial answer to that question at the end of the book. Because all that, stars, extraterrestrials, and so on are full of quantum mechanics, which we now discuss.

2

A quantum phenomenon

If you ask someone in the street to state a physics formula, the odds are that the answer will be $E = mc^2$. Nevertheless, the formula $E = h\nu$, which was written in the same year 1905 by the same Albert Einstein concerns their daily life considerably more.

In fact, among the three great scientific events of the beginning of the 20th century, 1905 with the special relativity of Einstein, Lorentz, and Poincaré, 1915 with Einstein's general relativity, an extraordinary reflection on gravitation, space, and time, and 1925 with the elaboration of quantum mechanics, it is certainly the last that has had the most profound impact on science and technology.

The first, and only, Nobel prize for relativity was awarded in 1993 to Taylor and Hulse for the double pulsar. Nobel prizes for quantum mechanics can hardly be counted (of the order of 120) including Einstein's for the photon in 1921. That reflects discoveries which have had important consequences. About 30% of the gross national product of the United States comes from byproducts of quantum mechanics.

Quantum mechanics is inescapable. All physics is quantum physics, from elementary particles to the big bang, semiconductors, and solar energy cells.

It is undoubtedly one of the greatest intellectual achievements of the history of mankind, probably the greatest of those that will remain from the 20th century, before psychoanalysis, computer science, or genome decoding.

This theory exists. It is expressed in a simple set of axioms that we discuss in chapter 6. Above all, this theory works. For a physicist, it even works too well, in some sense. One cannot determine its limits, except that during 10^{-43} seconds just after the big bang, we don't know what replaced it. But afterwards, that is, nowadays, it seems unbeatable.

However, this theory is subtle. One can only express it in mathematical language, which is quite frustrating for philosophers. Knowing mathematics is the entrance fee to the group of the happy few who can understand it, even though, as we show, the core of these mathematics is quite simple. It is the physics that is subtle. We show how and why quantum mechanics is still a

subject of debate as to its interpretation and its intellectual content. In some sense, mankind has made a beautiful and successful intellectual construction that escapes human understanding to some extent.

The discovery of quantum mechanics could have happened by analyzing a variety of physical facts at the end of the 19th century. The notion of quanta was proposed in 1900 by Max Planck. Planck had found semi-empirically a remarkable formula to explain a problem that fascinated people, the spectrum of black-body radiation. The frequency distribution of radiation inside an oven at temperature T depends only on the temperature, not on the nature or shape of the oven. It is a universal law. Planck obtained the good result

$$u(\nu) = \frac{8\pi h}{c^3} \frac{\nu^3}{e^{(h\nu/kT)} - 1} \quad , \tag{2.1}$$

where ν is the frequency, T the temperature, and k Boltzmann's constant, by assuming that radiation of frequency ν can exchange energy with the inner surface of the oven only by discrete quantities that are integer multiples of an elementary energy quantum $h\nu$,

$$\Delta E = nh\nu \quad . \tag{2.2}$$

Planck understood that the constant h in the above formula, which now bears his name and whose value is

$$h \approx 6,62\,10^{-34}\,\text{j.s} \quad ,$$

is a fundamental constant as is the velocity of light c in relativity and Newton's constant G in gravitation. For technical simplicity, we mainly use the reduced Planck constant

$$\hbar \equiv \frac{h}{2\pi} \approx 1,05\,10^{-34}\text{j.s} \quad .$$

Planck's formula works remarkably well. The direct verification would require us to be inside an oven. We have the great luck to live inside the cosmic background radiation of the big bang, which cooled down as the Universe expanded. The temperature of that radiation is at present 3 K. Its observation and its more and more precise measurement (Figure 2.1) is perhaps the best observational evidence in favor of the big bang theory, as well as of Planck's formula (we come back to this in the last chapter).

Planck's quanta were somewhat mysterious, and it was Einstein who made a decisive step forward in 1905, the same year as Brownian motion theory and special relativity. By performing a critique of Planck's ideas, and for reasons due to equilibrium considerations (i.e., entropy) Einstein understood that the quantized aspect is not limited to the energy exchanges between radiation and matter, but to the electromagnetic field itself. Light, which was known to be a wave propagation phenomenon since the beginning of the 19th century, must also present a particlelike behavior. Light of frequency ν must be carried by particles, photons as the chemist Gilbert called them in 1926, of energy

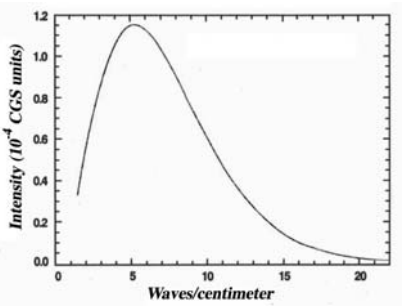


Fig. 2.1. Wave-number distribution of the cosmic background radiation measured in 1992 by the COBE satellite. The agreement between Planck's formula at a temperature $T = 2.728$ K lies within the line. (Photo credit: Mather et al., *Astrophys. J.*, **420**, 439, (1994). http://lambda.gsfc.nasa.gov/product/cobe/firas_image.cfm.)

$$E = h\nu \quad , \tag{2.3}$$

and momentum $\boldsymbol{p} = \hbar\boldsymbol{k}$, where $\boldsymbol{k}$ is the wave vector $k = 2\pi/\lambda$, as was proven experimentally by Compton in 1921.

In that respect, Einstein understood an essential feature of quantum theory, the so-called "dual" manifestations of the properties of light, which appear to be both wavelike and particlelike. In the course of his work, Einstein found the explanation of the photoelectric effect, which was one of the first experimental confirmations of his ideas. Of course, such ideas were considered revolutionary or even iconoclastic because they seemed to contradict Maxwell's equations which were a great triumph of the 19th century.

At the same time, atomic spectroscopy was considered one of the great enigmas of physics. The third breakthrough, which derives in some respect from Einstein's ideas, came in 1913 from Niels Bohr.

There are three parts in Bohr's ideas and results:

- He postulated that matter is also quantized and that there exist discrete energy levels for atoms, which was verified experimentally by Franck and Hertz in 1914.
- He postulated that spectral lines which had been accumulated during the 19th century, came from transitions between these energy levels. When atoms absorb or emit radiation, they make a transition between two discrete energy levels, and the positions of spectral lines are given by the difference
$$\nu_{nm} = \frac{|E_n - E_m|}{h} \ . \tag{2.4}$$
- Finally, Bohr constructed an empirical model of the hydrogen atom that works remarkably well and gives the energy levels E_n of this atom as
$$E_n = -\frac{m{q_e}^4}{2(4\pi\varepsilon_0)^2\hbar^2 n^2} \quad , \tag{2.5}$$

where n is a positive integer. With that formula, the wavelengths $\lambda = c/\nu_{nm}$ of spectral lines coincide with experiment to one part in a thousand. Bohr's formula (2.5) expresses the famous "Rydberg constant" of spectroscopists in terms of fundamental constants, which impressed people, in particular, Einstein.[1]

So we are facing three similar formulae, $E = h\nu$. The first (2.2) is an assumption about the interaction of radiation and matter, the second (2.3) has to do with radiation itself, and the third (2.4) is a property of atoms, namely matter.

Bohr's success was fantastic, but it was too easy. Actually one realized later on that it was a piece of luck. But this easy result generated an obscure prequantum era, where people accumulated recipes with fluctuating results deprived of any global coherence.

2.1 Wave behavior of particles

The synthetic and coherent formulation of quantum mechanics was performed around 1925. It is due to an incredible collective work of talented people such as Louis de Broglie, Schrödinger, Heisenberg, Max Born, Dirac, Pauli, and Hilbert, among others. Never before, in physics, had one seen such a collective effort to find ideas capable of explaining physical phenomena.

We are now going to discover some of the main features on a simple concrete experiment that shows the wavelike behavior of particles. This is symmetric in some respect to the particlelike behavior of light. We show in particular that the behavior of matter at atomic scales does not follow what we expect from daily "common sense." It is impossible to explain it with our immediate conceptions.

In order to understand quantum mechanics, one must get rid of prejudices and ideas that seem obvious, and one must adopt a critical intellectual attitude in the face of experimental facts.

2.1.1 Interferences

We recall interference phenomena in wave physics, optics, or acoustics, in the simple case of Young slit fringes.

One sends a light beam on a screen pierced with two slits, and one observes the variation of the intensity of light on another screen as a function of the distance x to the center.

The two slits act as secondary sources in phase, and the amplitude of the wave at a point C of the screen is the algebraic sum of the amplitudes issued from each of them.

[1] The $1/n^2$ behavior was known since 1886 and Balmer's empirical discovery.

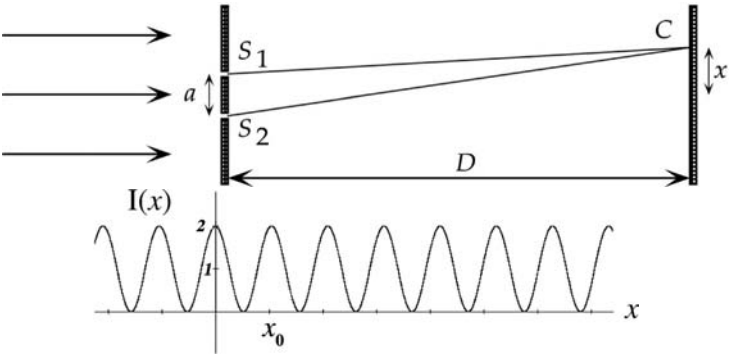


Fig. 2.2. Sketch of a Young two-slit interference experiment.

If the two waves are in phase, the amplitude is twice as large. If they are out of phase by π the amplitude vanishes; there is no luminous energy at that point. Naturally, there exist all intermediate cases.

In other words, the amplitude at some point is the sum of amplitudes reaching that point,

$$\text{Amplitude at C} : \; A_C = A_1 + A_2 \; , \quad \text{Intensity} : \; I(x) = |A_C|^2 \quad . \tag{2.6}$$

The amplitudes emitted by the two slits add up, the intensity is the square of that sum and it presents a periodic variation, the distance of fringes being $x_0 = \lambda D/a$.

2.1.2 Wave behavior of matter

We turn to the wave behavior of matter.

In 1923, Louis de Broglie made the bold but remarkable assumption that to any particle of mass m and of velocity $\boldsymbol{v}$ there is an "associated" wave of wavelength

$$\lambda = \frac{h}{p} \quad ; \tag{2.7}$$

$\boldsymbol{p} = m\boldsymbol{v}$ is the momentum of the particle and p its norm.

Louis de Broglie had many reasons to propose this. In particular he had in mind that the discrete energy levels of Bohr might come from a stationary wave phenomenon. This aspect struck the minds of people, in particular that of Einstein, who was enthusiastic.

How can one verify such an assumption? One way is to perform interference and diffraction experiments. The first experimental confirmation is due to Davisson and Germer in 1927. It is a diffraction experiment of an electron beam on a nickel crystal.

It is more difficult to perform a Young double-slit interference experiment with electrons. However, a group of Japanese physicists from Nippon Electronics (NEC) performed in 1994 a beautiful interference experiment of cold atoms in Young slits. Neon atoms are initially trapped in stationary laser waves (so-called optical molasses). They are then released and undergo free fall across a two-slit device. The slits are 2 μm large, they are 6 μm apart. The scale in Figure 2.3 is distorted.

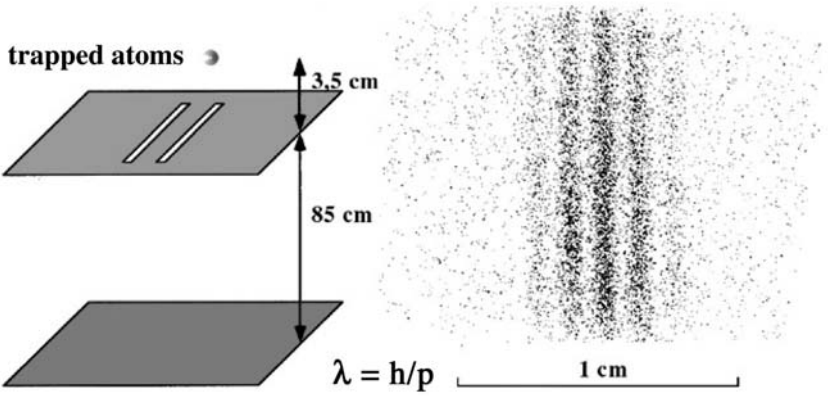


Fig. 2.3. Double slit Young interference experiment performed with neon atoms cooled down to a milliKelvin (left part). Each point of the Figure (right part) corresponds to the impact of an atom on the detector. Interference fringes are clearly visible.

What do we observe in Figure 2.3? The distribution of impacts of atoms on the detecting plate is the same as the optical intensity in the same device. The fringes are at the same positions provided Louis de Broglie's relation is satisfied $\lambda = h/p$. (Of course, one must take care of the uniform acceleration in this particular setup.)

The same phenomenon can be observed with any particle: neutrons, helium atoms, or hydrogen molecules, always with the same relation between the wavelength and the velocity. The present record is to perform interferences with large molecules such as fullerenes, that is, C_{60} molecules.

Therefore matter particles exhibit a wave behavior with a wavelength given by de Broglie's formula.

2.1.3 Analysis of the phenomenon

Now, a number of questions are in order:

- What is this wave?
- And why is this result so extraordinary?

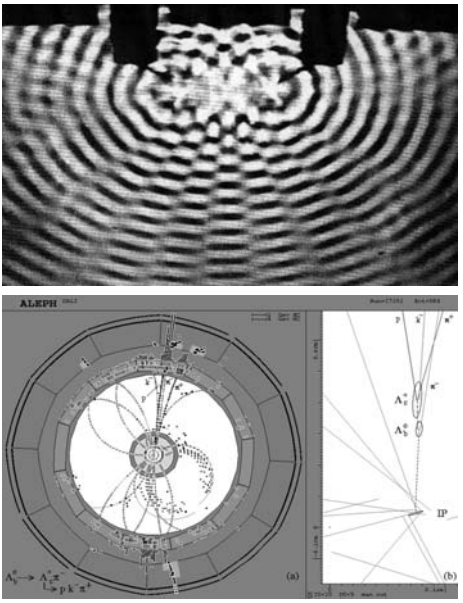


Fig. 2.4. Top: two source interferences on the surface of water; the radial lines are nodes of interferences. Bottom: tracks of particles in the Aleph detector of LEP at CERN.

It is extraordinary because atoms are known to be particles. An atom has a size of the order of an Angstrom (0.1 nm) and it is pointlike at the scales of interest (μm or mm). With a counter, one can measure whether an atom has arrived at some point with as large an accuracy as one wishes. When an atom is detected, it has a well-defined position; it does not break up into pieces; it is point-like.

But a wave fills all space. A wave, on the surface of water, is the whole set of deformations of that surface on all its points.

So, what is a particle? Is it a pointlike object or is it spread out in the entire space? A simple glance at Figure 2.4 shows that we are facing a conceptual contradiction.

How can we escape this contradiction? Actually, the phenomenon is much richer than a simple wave phenomenon; we must observe experimental facts and use our critical minds.

Because atoms are particles, we can send them individually, one at a time, and all in the same way.

This proposition is perfectly decidable; it is feasible experimentally. We can set up the device so that it releases atoms one after the other and that they are all released in the same way.

2.2 Probabilistic nature of quantum phenomena

2.2.1 Random behavior of particles

What do we observe? Actually, we can guess it in Figure 2.3.

- Each atom has a well-defined impact; indeed an atom does not break into pieces.
- But the positions of the impacts are distributed at random. In other words, to the same initial conditions, there correspond different impacts.

In other words, atoms, or particles in general, have a random behavior. Each atom arrives where it wants, but the whole lot is distributed with a probability law similar to the intensity observed in optics or acoustics:

$$P(z) \propto I_{(optical)}(\lambda = h/p) \quad .$$

Therefore, there is a second difference with classical physics: to identical initial conditions, there correspond different final conditions.

But, one can object that random, or probabilistic, phenomena exist in classical physics, such as playing dice, or heads and tails, and so on.

True, but the big problem is that this is by no means a classical probabilistic phenomenon, as in usual probability theory. Why is that?

2.2.2 A nonclassical probabilistic phenomenon

If we block one of the slits, the atoms will pass through the other one and their distribution on the detector shows no sign of any interference. If we block the other slit, the distribution is approximately the same, except for a small shift (1 μm/1 mm) Now let's make a logical argument and perform the critique of what we say:

1. We send the atoms one by one. These are independent phenomena; atoms don't bother each other; they do not act on each other's trajectory.
2. Each atom has certainly gone through one of the slits.
3. We can measure which slit each atom went through. There exist techniques for this; send light on the slits, put counters, and so on. This is possible.
4. If we perform this measurement, we can separate the outgoing atoms in two samples, those that have passed through the first slit, and those that have passed through the second one. And we know where each atom arrived.
5. For those that passed through the first slit, everything is as if the second slit were blocked, and vice versa. Each sample shows no interference.

Now, we have two independent samples, and we can bring them together. Classically, the result we would obtain by opening the two slits should be the sum, the superposition of the two distributions such as (2.5). But not at all!

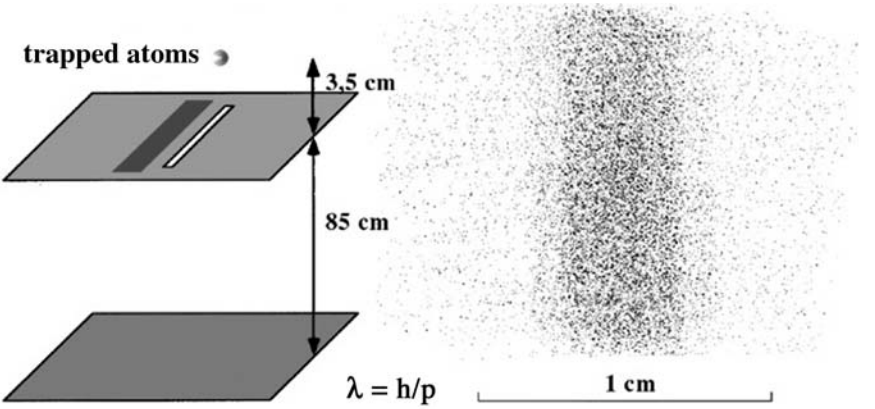


Fig. 2.5. Same experiment as in Figure 2.3 but opening only one slit. The interference fringes disappear and one observes a diffraction pattern (this figure is not experimental).

It's even worse! Opening a second slit (i.e., giving an extra possibility for the atoms to reach the detector) has prevented the atoms from arriving at certain points. That's really incredible to be able to stop people from entering your house by opening another door!

We must admit that the usual logical ideas of probability theory do not apply. We cannot explain the phenomenon in classical terms. It is a non-classical probabilistic phenomenon.

2.3 Conclusions

At this point, it seems we are at a logical dead end. How can we find our way? Our argument, however logical it may seem, leads to wrong conclusions. There is something we haven't thought about. Because physics is consistent. The answer is experimental. What actually happens is the following:

1. If we measure by which slit each atom passed, we can indeed make the separation and indeed we observe the sum of two distributions such as in Figure 2.5. Therefore we no longer observe interferences, these disappear. It is another experiment!
2. Conversely, if we do observe interferences, it is not possible to know through which slit each atom passed. We can talk about it, but we can't do anything with it.
   Knowing by which slit an atom has passed in an interference experiment is a proposition that has no physical meaning; it is undecidable. It is perfectly correct to say that an atom passed through both holes at the same time, which seems paradoxical or absurd classically.

What was wrong was to assume implicitly that, at the same time, we could measure by which slit each atom passed and observe interferences. We assumed that without checking it.

We can draw certain conclusions:

• First, *a measurement perturbs the system.* If we do not measure by which slit they pass, the atoms are capable of interfering. After we perform this measurement, they are in another state where they are no longer capable of interfering. They have been perturbed by the measurement.

• Second and consequently, *there is no trajectory* in the classical sense. If we observe an atom in an interference experiment, we know when and where it was emitted and where and when it was detected, but we cannot say where it was in the meantime.

However, these two ideas seemed obvious in classical physics. The fact that we can make a measurement as accurate as we wish without affecting the system is an old belief of physics. Physicists used to say that they just needed to improve the measuring apparatus. Quantum physics tells us that there is a absolute lower bound to the perturbation that a measurement produces.

The notion of a trajectory, namely that there exists a set of points by which we can claim and measure that a particle has passed at each moment, is as old as mankind. Cavemen knew that intuitively when they went hunting. It took centuries to construct a theory of trajectories, to predict a trajectory in terms of initial conditions. Newton's classical mechanics, celestial mechanics, ballistics, rests entirely on that notion, but its starting point is beaten up by the simple quantum phenomenon we just examined.

Classically, we understand the motion of a particle by assuming that, at each moment one can measure the position of a projectile, that the collection of the results consists of a trajectory, and that we can draw a reproducible conclusion independent of the fact that we measure the positions at any moments. We learn these ideas as if they were obvious, but they are wrong. More precisely: in order to penetrate the quantum world, one must get rid of such ideas. Figure 2.6, or analogous ones, is completely wrong in quantum mechanics.

Of course, one mustn't go too far. These are very good approximations in the classical world. If a policeman stops you on a freeway saying you were driving at 80 miles an hour, the good attitude is to claim, "Not at all! I was driving peacefully at 35 mph on the little road under the bridge, and your radar perturbed me!" Unfortunately, he won't believe you even if he knows some physics. Because it is Planck's constant $\hbar$ that governs such effects. However, in quantum driving one must change the rules. Changing the rules consists of constructing the theory of all that.

Fig. 2.6. Stroboscopic picture of the free fall of an apple which then bounces on the floor. This is a good example of the a priori representation of an intuitive phenomenon that is wrong in quantum mechanics. (William McLaughlin, "The resolution of Zeno's paradoxes," *Sci. Amer.*, 1994.)

2.4 Phenomenological description

The interference phenomenon would be very complicated to explain if we did not have the luck that it so closely resembles usual interference, with, in addition, a simple formula for the wavelength $\lambda = h/p$.

So, let's try and use the analogy with wave physics in order to formalize Louis de Broglie's idea. Here, we should be able to explain the interference experiment in the following way:

- The behavior of an atom of velocity v and momentum $p = mv$ in the incoming beam corresponds to that of a monochromatic plane wave

$$\psi_{\text{incident}} = e^{-i(\omega t - \mathbf{p}\cdot\mathbf{r}/\hbar)}, \quad \boldsymbol{k} = \boldsymbol{p}/\hbar,\ \lambda = 2\pi/k = h/p\ ; \tag{2.8}$$

  which has the good wave vector $\boldsymbol{k} = \boldsymbol{p}/\hbar$ and the good wavelength.
- After the two slits, the behavior is that of the sum of two waves each of which has been diffracted by a slit

$$\psi_{\text{outgoing}}(x) = \psi_1 + \psi_2 \quad , \tag{2.9}$$

  which would describe, respectively, the behavior of the atom if it passed through one of the slits, the other one being blocked. We can calculate the phase shift of these waves at any point because we know the wavelength.
- Finally, the probability for an atom to reach some point C of the detector is simply the modulus squared of that sum

$$P(C) = |\psi_C|^2 \quad . \tag{2.10}$$

We just follow the same argument as for usual interferences.

We now have an answer to one of our questions above; what is the physical meaning of these waves?

In usual wave physics, one manipulates electromagnetic or acoustic wave amplitudes which add up and whose modulus squared gives intensities, that is, energy densities.

Our quantum waves are *probability amplitudes*. Their modulus squared gives us probabilities, or probability densities.

One does not work directly with probabilities but with these intermediate tools, these probability amplitudes that add up.

The interference experiment gives us the wavelength, but not the frequency ω of the waves. Louis de Broglie made a good choice by assuming that this frequency is related to the energy of the particles in the same way as for Einstein's photons

$$\omega = E/\hbar, \quad \text{that is,} \quad \nu = E/h \quad , \tag{2.11}$$

where $E = p^2/2m$ is the kinetic energy of the atoms. This leads to the complete structure of *de Broglie waves*:

$$\psi_{\text{incident}} = e^{-(i/\hbar)(Et - \mathbf{p}\cdot\mathbf{r})}, \text{ where } E = p^2/2m \, , \tag{2.12}$$

which is the probability amplitude for the presence of a particle at point $\mathbf{r}$ and time t of a particle of momentum $\mathbf{p} = m\mathbf{v}$.

Notice that because the kinetic energy and the momentum are related by $E = p^2/2m$, one can find with this expression a wave equation, which is satisfied whatever the value of the momentum p. Indeed, if we take the time derivative on one hand, and the Laplacian on the other, we obtain

$$\frac{\partial}{\partial t}\psi_{\text{incident}} = -\frac{iE}{\hbar}\psi_{\text{incident}}, \quad \text{and} \quad \Delta\psi_{\text{incident}} = -\frac{p^2}{\hbar^2}\psi_{\text{incident}} \, ,$$

therefore, because $E = p^2/2m$, we have the wave equation

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\Delta\psi \quad , \tag{2.13}$$

which is nothing but the Schrödinger equation for a free particle.[2]

Of course, we are not completely finished. For instance, atoms have a particlelike behavior that is obscure in all that. But we're getting closer.

[2] It is surprising that de Broglie didn't think of writing this equation, or its relativistic equivalent (because he used the relativistic energy-momentum relation $E^2 = (p^2c^2 + m^2c^4)$).

3

Wave function, Schrödinger equation

In the first chapter, we described an interference experiment of atoms which, as we have understood, is both a wave and a probabilistic phenomenon.

We now want to construct the theory of this experiment. More generally, we want to find the quantum theory of the simplest problem of classical mechanics; the nonrelativistic motion of a particle of mass m in a field of force.

This is called wave mechanics. It is due to de Broglie and to Schrödinger. We generalize it later on.

We do not want to say what the nature of an atom or an electron is; we simply want to determine their behavior in a field of force. In celestial mechanics, one does not worry about the nature of planets. They are considered as points whose motion we can calculate.

3.1 Terminology and methodology

3.1.1 Terminology

Before we start, we must agree on the meaning of words and on the methodology. We cannot avoid using ordinary language. Words are necessary. But words can also be traps when discussing phenomena that are so new and unusual. We constantly use the following words: *physical system, state, physical quantities*.

The foundation of physics is experimental observation and the measurement process that consists of characterizing aspects of reality, namely what we observe, by numbers. These aspects of reality are elaborated into concepts of *physical quantities* (for instance, velocity, energy, electric intensity, etc.).

In given circumstances, we say that a *physical system* (i.e., an object pertaining to reality) is in a certain *state*. The state of the system is "the way the object is" (i.e., the particular form in which its reality can manifest itself). That is what we are interested in. We want to know the state of an atom in space, not its internal structure, which we study later. We possess some more

or less complete knowledge of a state if we perform some set of measurements of physical quantities on the system.

We give a name to the system. We can call it a particle, an atom, or an electron. But a particle is simply for the moment an object with a well-defined mass m and electric charge q, and which preserves any internal structure it may have in the experiments of interest. We are not interested for the moment in the possible internal degrees of freedom of the particle.

3.1.2 Methodology

The construction of our theory has the following elements that we describe in the classical case of Newtonian mechanics of a massive particle in a field of force that we assume to derive from a potential energy $V(\boldsymbol{r})$.

1. We must first describe the state of the system. This means associating with this state a mathematical representation that defines it from an operational point of view. In Newton's theory, the state of a massive particle in space is described at time t by six numbers; its position $\boldsymbol{r}$ and velocity $\boldsymbol{v}$, or its momentum $\boldsymbol{p} = m\boldsymbol{v}$.
2. Then, we must know the law that governs the time evolution of the state of the system when it is placed in given conditions; that is, we must be able to predict the state at time t, given the state at time $t = 0$. In Newton's theory, it is the fundamental law of dynamics $d\boldsymbol{p}/dt = \boldsymbol{f}$, that allows us to calculate the trajectory.
3. Next, we must know the laws that enable us to calculate the results of measurements of physical quantities, laws that transform the mathematical representation of the state of a system into measurable numbers. In Newton's theory, physical quantities are functions of the state variables $\boldsymbol{r}$ and $\boldsymbol{p}$.
4. Finally, we must address a question that is absent in Newtonian theory. In what does the measurement process result? What do we know after a measurement?

In this chapter, we study the first two questions. We examine the two others in the next chapter.

In quantum mechanics, there is no direct intuitive link between physical concepts and their mathematical representation as we can find in classical physics. For instance, any child knows that instantaneous velocity and acceleration exist because in a car there is an object, the speedometer, whose indications correspond to the various physical feelings of the child. When later on, the child learns the mathematical notion of derivatives, it is quite natural to associate that notion with the physical concept that has become familiar.

So, what we do is to place ourselves in the position of de Broglie, Schrödinger, Einstein, and Born, after they have investigated all alternative possibilities.

• We give the principles of the theory.
• We check that they are consistent, and that they account for observed phenomena; we will then understand how the theory works.
• Finally, we say a few words about how such ideas came to the minds of people. This is useful for understanding the theoretical scheme and approaching it in a human way. Principles were not dictated by any superior being. It is human scientists who fought with reality in order to elaborate them.

Later on, we shall be capable of understanding that the mathematical structure which is perhaps the most important in quantum mechanics is the first of the four elementary operations, addition. But it takes some time to fully appreciate that.

3.2 Principles of wave mechanics

3.2.1 The interference experiment

In the previous chapter, we have described the experimental aspects of quantum interferences of particles on the specific case of atoms.

A beam of particles of given momentum $\boldsymbol{p}$ is sent on a diffracting setup. this gives rise to an interference pattern similar to light interferences. The impact point of a given atom can be anywhere with a probability law proportional to the intensity of light fringes.

We attempted to make an empirical description using de Broglie waves, which are probability amplitudes that add and whose modulus squared gives a probability

$$\psi_{\text{incident}} = e^{-(i/\hbar)(Et-\mathbf{p}\cdot\mathbf{r})}, \text{ where } E = p^2/2m \; . \tag{3.1}$$

These waves obey the free particle Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\Delta\psi \quad . \tag{3.2}$$

We now go back to a deductive presentation.

3.2.2 Wave function

1. **Description**
   Schrödinger and Born understood in 1926 that the complete description of a particle in space at time t is performed with a complex wave function $\psi(\boldsymbol{r}, t)$, whose physical interpretation is that the probability $dP(\boldsymbol{r})$ to find the particle in a vicinity d^3r of the point r is given by

$$dP(\boldsymbol{r}) = |\psi(\boldsymbol{r}, t)|^2 \, d^3r \quad . \tag{3.3}$$

This is indeed a probabilistic description, but it is a nonclassical one. One cannot describe a quantum system with probabilities only. The wave function is a probability amplitude (it is necessarily complex). Its modulus squared gives the probability density of finding the particle at point r at time t and, quite naturally, the integral of this quantity over all space is equal to one:

$$\int |\psi(\boldsymbol{r},t)|^2 \, d^3r = 1 \quad , \tag{3.4}$$

which is obvious but fundamental (the particle must surely be somewhere).

2. **Probabilistic interpretation**
The meaning of this description is the following. We prepare N atoms independently, in the same state, so that, when each of them is measured, they are described by strictly the same wave function. Then the result of a position measurement is for each of them as accurate as we wish (limited by the accuracy of the measuring apparatus) but is not the same for all. The set of impacts is distributed in space with the probability density $|\psi(\boldsymbol{r},t)|^2$.
Quite generally, one can plot the histogram of the distribution. The set of N measurements is characterized by an expectation value $\langle x \rangle$ and a root mean square dispersion (or simply dispersion, for short) Δx,

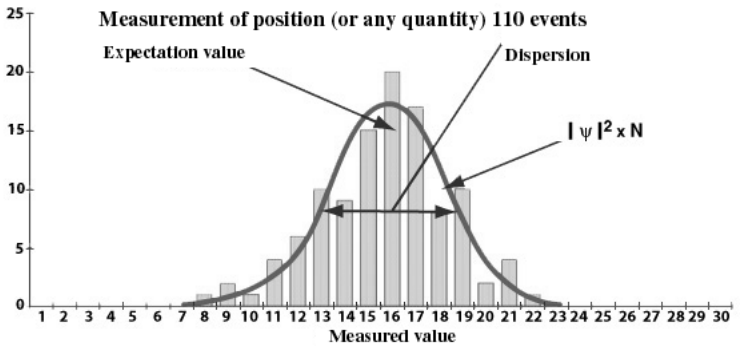


Fig. 3.1. Simple example of a histogram of the distribution of a position measurement for particles prepared all in the same state ψ.

$$\langle x \rangle = \int x \, |\psi(\boldsymbol{r},t)|^2 \, d^3r \; ; \tag{3.5}$$

similarly, the square of the dispersion $(\Delta x)^2$ is by definition

$$(\Delta x)^2 = \langle x^2 \rangle - (\langle x \rangle)^2 = \langle (x - \langle x \rangle)^2 \rangle \quad . \tag{3.6}$$

It is a theorem of probability theory that the probability of finding a result within a few times Δx of the value $\langle x \rangle$ is close to one. If the accuracy δx

of the measuring apparatus is not as fine as the dispersion Δx we can say the particle has a well-defined position.
A probabilistic description can perfectly well accommodate the description of a "pointlike" object.

3.2.3 Schrödinger equation

In 1926, Schrödinger discovered that when the particle is placed in a field of force that derives from a potential $V(\boldsymbol{r})$, the time evolution of the wave function, therefore of the state of the particle, is given by the partial differential equation

$$i\hbar \frac{\partial}{\partial t}\psi(\boldsymbol{r},t) = -\frac{\hbar^2}{2m}\,\Delta\psi(\boldsymbol{r},t) + V(\boldsymbol{r})\,\psi(\boldsymbol{r},t) \quad . \tag{3.7}$$

It is the same as previously (3.2) for a free particle. The incorporation of the forces lies in the second term on the right-hand side.

1. The Schrödinger equation is the wave equation. It replaces the fundamental law of dynamics ($m\boldsymbol{\gamma} = \boldsymbol{f}$).
2. It is a partial differential equation of first order in time. Therefore, if the state $\psi(\boldsymbol{r},0)$ is known at some initial time $t = 0$, the equation determines the state $\psi(\boldsymbol{r},t)$ at any further time t.
3. One cannot prove this equation (any more than one can prove $m\boldsymbol{\gamma} = \boldsymbol{f}$). Its justification is that it works: when one calculates, it gives the correct results.
   When he tried to apply de Broglie's idea to the hydrogen atom, Schrödinger wrote several other equations. These were more clever, in his mind, because they had a relativistic structure. But, if they gave the correct $1/n^2$ term, they did not give the proper relativistic corrections, which were small but known. His wife said that for three weeks, he had been in a terrible mood. Then, he surrendered and abandoned relativity, saying that "God decided that things were so." Actually, when he did this, he did not know about the electron spin, which contributes in an important way to these corrections. That is a piece of luck. Most of quantum physics is nonrelativistic, therefore comparatively simple.
4. Nearly all quantum mechanical problems consist of solving that equation. We come back to it constantly.

At this point, our real problem is to become familiar with this new concept of a wave function. We want to understand its structure, its properties. We want to understand why we need a whole function in order to describe the state of a particle, whereas for Newton six numbers were sufficient.

3.3 Superposition principle

Consider our empirical analysis of atom interferences and de Broglie waves. Everywhere the atom is free, and the potential vanishes (except on the screen which atoms cannot cross).

Starting from Louis de Broglie's choice for the frequency ω, we had deduced the Schrödinger equation in the absence of forces, with $V = 0$. Conversely, the de Broglie waves are particular solutions of the Schrödinger equation for $V = 0$. These two principles are the mathematical formulation of de Broglie's idea. Schrödinger's real contribution lies in the incorporation of forces.

The previous analysis seems correct; it will account for experimental observation provided one condition is satisfied. This condition is that after the slits the wave function at a point C is indeed the sum

$$\psi_C = \psi_A + \psi_B \tag{3.8}$$

of the wave function which would describe the atom if it went through A (B being blocked) and of the wave function which would describe the atom if it went through B (A being blocked), each of which would give diffraction distributions showing no interference fringes.

Here, we have in front of us the most fundamental thing of all this chapter. Wave functions have the right to add up. That is the fundamental property of wave functions.

We promised *addition*; there it is!

More generally, let ψ_1 and ψ_2 be two wave functions, then the combination

$$\Psi = \alpha\psi_1 + \beta\psi_2 \quad , \tag{3.9}$$

where α and β are complex numbers (one can produce a phase shift or an attenuation of a wave) belong to the family of wave functions. This sum is a possible wave function, and this first principle is called the *superposition principle*.

The superposition property is completely connected with the interference phenomenon. It is much more important than the formula $\lambda = h/p$; any other formula would not change the fundamentals of the experiment.

In other words, two states have the right to add up in order to construct a third one.

That is the fundamental notion. What is amazing with mathematicians is that when they see simple things such as that, they understand immediately underlying structures. They tell us that the set of wave functions $\{\psi(\boldsymbol{r}, t)\}$ is a complex vector space. If one imposes the normalization condition (3.4), $\int |\psi|^2 = 1$, it is what one calls a Hilbert space, the space of square integrable functions. This property, which expresses interferences, is much more important than the concept of a wave function itself.[1]

[1] Such a situation is frequently encountered. Zebras have the property of running fast. This property is more important than the concept of a zebra itself. Indeed, if

Does this theory account for Young slit interferences?

Yes, because the Schrödinger equation is linear and any linear combination of solutions is a solution. If one formulates the problem of sending a plane wave from the left of the slits, one can prove that in the vicinity of the axis and, at large enough distances, the usual interference formulae apply. It is a complicated mathematical problem, but it has a well-defined solution. (Note that the problem of interferences on the surface of water presented in the previous chapter, Figure 2.4, is much simpler mathematically, because the two sources are independent. Here, one must take into account that it is the same plane wave that is incident on the slits.)

3.4 Wave packets

However, de Broglie waves are not wave functions. In fact they are not normalizable.

This is not a very difficult problem. It is frequently encountered in wave physics: a plane wave does not exist in practice. It is a useful idealization that makes calculations easy, but physically it would fill all space at all times. A physical wave is always localized in space at a given moment, and localized in time at a given place. A wave is never exactly monochromatic; there is always some dispersion in frequency and in wavelength.

3.4.1 Free wave packets

The representation of a realistic physical situation is a linear superposition of monochromatic plane waves, each of which is a particular solution of the Schrödinger equation, of the form

$$\psi(\boldsymbol{r},t) = \int \varphi(\boldsymbol{p})\, e^{-(i/\hbar)(Et-\mathbf{p}\cdot\mathbf{r})}\, \frac{d^3p}{(2\pi\hbar)^{3/2}}\,, \quad \text{where } E = \frac{p^2}{2m}\,, \tag{3.10}$$

such that all these waves interfere destructively outside some region of space. This is called a *wave packet*. In this formula, we have introduced the constant $(2\pi\hbar)^{3/2}$ for convenience, and the complex function $\varphi(\boldsymbol{p})$ is arbitrary (it determines ψ) except that the expression exists and we want it to be properly normalized (we want that $\int |\psi|^2 = 1$).

This expression satisfies the Schrödinger for a free particle. One proves in mathematics that it is the general solution of the free Schrödinger equation.

---

it didn't exist, zebras would all have been eaten by lions and the notion of zebra would be of no interest.

3.4.2 Fourier transformation

However, we want the resulting wave function to be normalizable. How can we do this without having to check each time?

The tool that answers this question is the *Fourier transformation*. The Fourier transformation is one of the most important mathematical structures discovered in the 19th century. It has numerous applications in mathematics, in electronics, in physics, in chemistry, and so on.

For our purpose, we only need to have in mind one definition and three consequences.

One morning in 1812, Fourier found out that he could solve many problems by transforming a function g of a real variable k into another function f of a real variable x with the formula

$$f(x) = \frac{1}{\sqrt{2\pi}} \int e^{ikx}\, g(k)\, dk \quad . \tag{3.11}$$

• The inverse transformation, which allows us to get back g knowing f, is given by

$$g(k) = \frac{1}{\sqrt{2\pi}} \int e^{-ikx}\, f(x)\, dx \quad . \tag{3.12}$$

The similarity between the two expressions (3.11) and (3.12) is such that we can say that f and g are Fourier transforms of each other.

• The second property is that the Fourier transformation is what is called an isometry. If $f_1(x)$ and $f_2(x)$ are, respectively, Fourier transforms of $g_1(k)$ and $g_2(k)$, then we have the Parseval–Plancherel theorem:

$$\int f_1^*(x)\, f_2(x)\, dx = \int g_1^*(k)\, g_2(k)\, dk \quad . \tag{3.13}$$

Of course in all this, we assume that all expressions exist and behave properly.

• The third property is that the more the support of $|g(k)|^2$ is concentrated (around some value k_0), the larger is the support of $|f(x)|^2$ (and vice versa). In particular, if we normalize f and g to one, that is $\int |f|^2 dx = 1$ and $\int |g|^2 dk = 1$, we can consider $|g(k)|^2$ and $|f(x)|^2$ as probability laws for the variables k and x, respectively, and if we consider the resulting expectation values and dispersions

$$\langle k \rangle = \int k\, |g(k)|^2\, dk\,; \quad (\Delta k)^2 = \langle k^2 \rangle - \langle k \rangle^2, \tag{3.14}$$

and similarly for $\langle x \rangle$ and Δx in terms of f, the product of the dispersions Δx and Δk is constrained by the inequality

$$\Delta x\, \Delta k \geq 1/2 \quad . \tag{3.15}$$

3.4.3 Shape of wave packets

Coming back to wave packets (3.10), the constant $\hbar$ is present in order to have dimensionless quantities because $\boldsymbol{p}.\boldsymbol{r}$ has the dimension of an action.

We see that ψ is the Fourier transform of $\varphi(\boldsymbol{p})\, e^{-iEt/\hbar}$ which implies that

$$\int |\psi(\boldsymbol{r},t)|^2\, d^3r = \int |\varphi(\boldsymbol{p})|^2\, d^3p \quad . \tag{3.16}$$

Therefore, $\int |\psi|^2 = 1$ if and only if $\int |\varphi|^2 = 1$. In the choice of φ above, we only need to have $\int |\varphi|^2 = 1$; φ is otherwise arbitrary.

We see that

1. If φ is very concentrated around some value p_0 the wave function will be close to a monochromatic plane wave in a large region of space. It is therefore a realistic wave function for a beam of atoms. A simple case is represented in Figure 3.2 If we look at the wave function closely, it

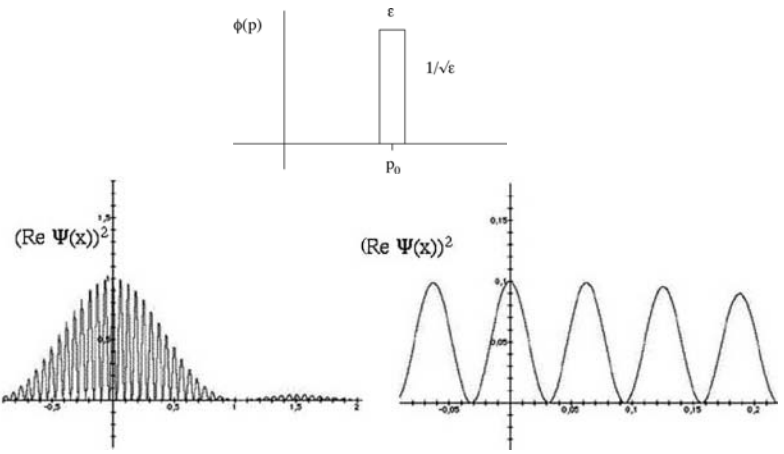


Fig. 3.2. Wave packet obtained with a square form of $\phi(p)$ localized in the vicinity of a value p_0 (top). The two figures at the bottom show the probability distribution (actually the square of the real part) at two scales which differ by a factor of 10.

resembles a plane wave; if we look at it from far away, it may seem to be a concentrated distribution.

2. Conversely, it suffices to exchange the two functions in order to obtain the wave function of an atom whose position is well localized in space; that is, ψ is concentrated near r_0, as can be seen in the symmetric Figure 3.3.

3.5 Historical landmarks

Now, how and when did these ideas emerge?

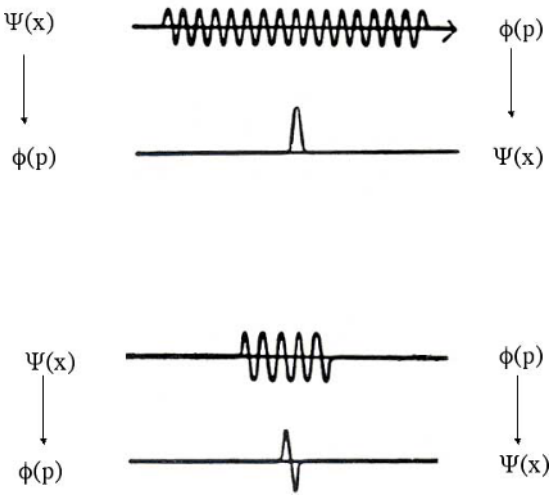


Fig. 3.3. Real part of the wave function $\psi(x)$ corresponding to a localized function $\phi(p)$ in the vicinity of p_0 and vice versa, in two cases.

In 1923, Louis de Broglie, who initially studied history, proposed his wave hypothesis. His thesis is a beautiful text with five chapters. The Science Faculty of the Sorbonne was a little bit embarrassed. Was that brilliant or foolish?

• Paul Langevin gave de Broglie's thesis to Einstein in April 1924, and Einstein was very enthusiastic for several reasons.

• First of all, the wave hypothesis could lead to energy quantization as a stationary wave problem, provided one used proper mathematics. This quantization could be achieved without abandoning the continuity of physical laws. Bohr kept on saying that discontinuity was a fundamental property of matter, Einstein was shocked by such an idea.

• In addition, Einstein managed to explain problems of statistical physics with the wave assumption. He advertised the idea, saying that it was much more than an analogy with photons; it was very deep.

• At the beginning, Schrödinger found that the idea was elegant, but he was skeptical because of relativity. It was Einstein who urged Schrödinger to work on the subject. Actually, Schrödinger called that the wave theory of Einstein and de Broglie. He acknowledged that without Einstein he would not have done anything.

• It was Schrödinger who introduced the Greek letter ψ for the wave function, which has become a tradition. However, he made a mistake in its interpretation.

• It was Max Born (a mathematician, who had the chair of theoretical physics in Göttingen) who gave the probabilistic interpretation at the end of 1926. What happened is that, using the Schrödinger equation, Born calculated the scattering of electrons on nuclei. He noticed that the square of the wave

function $|\psi|^2$ gave the value of the measured intensity. At the same time, the development of Geiger–Müller counters provided an alternative measurement technique. Instead of measuring electric current intensities, one could count numbers of electrons; in other words, one did statistics. Hence the interpretation of $|\psi|^2$ as a probability density.

This probabilistic interpretation achieved the synthesis of two "complementary" aspects of the behavior of particles (electrons or atoms). In the same experiment, an atom behaves as a wave when it interacts with the Young slits, and as a particle when it is detected (a detector and a diffracting system do not act in the same way).

That is the real problem. An atom is not schizophrenic; it is neither a wave nor a particle in the classical sense. It is a well-defined quantum object. In given experimental situations, this quantum object may behave as a Newtonian particle or as a wave, which seems contradictory to those who possess only classical concepts.

3.6 Momentum probability law

So, everything seems to work well up to now. But one can go much further.

3.6.1 Free particle

We said that the wave function ψ describes completely the state of the particle at time t. Therefore, it must contain the information on the particle's velocity or momentum. But we do not see that information at the moment.

So, let's invent. The Fourier transformation is fantastic because it enables us to invent. Now, we use mathematics just as in the case of instantaneous velocity and derivatives.

In fact, the properties of the Fourier transformation suggest that the probability law of momentum is given by $|\varphi(\boldsymbol{p})|^2$; that is,

$$dP(\boldsymbol{p}) = |\varphi(\boldsymbol{p})|^2 \, d^3p \quad . \tag{3.17}$$

This can be proven;[2] we won't do this here. We have an idea and we wish to see if it is plausible and consistent, and if it works.

1. $|\varphi|^2$ is nonnegative, and, because $|\psi|^2$ is normalized to one, then $|\varphi|^2$ is also normalized to one, its integral is equal to one. Therefore it has all the properties of a probability law.
2. There is a one-to-one correspondence between φ and ψ, which satisfies our requirement that the information on the momentum is contained in the wave function. If our idea is true, the Fourier transformation extracts the information on the momentum of the particle from the wave function.

[2] See J.-L. Basdevant and J. Dalibard, *Quantum Mechanics*, Chapter 2, Section 6.

3. De Broglie's assumption says that the wave function of a free particle of well-defined momentum p_0 should be a monochromatic plane wave. In order to get close to that, one must find a function φ that is large in the vicinity of p_0 and small elsewhere as in Figure 3.2.

Now, we see that we can use a mathematical notion in order to represent a physical idea. The Fourier transform of the wave function is the probability amplitude for the momentum of the particle (again, this can be proven).

That's a fantastic physical result! If the particle is localized in momentum p, then it is wavelike in position x. If it is localized in position, then it is wavelike in momentum, as one can see in Figure 3.3. The space of positions and the space of momenta are reciprocal spaces, conjugate of each other. In order to describe the state of the particle, we can use either of them, in a symmetrical way.

3.6.2 General case

In fact, not only is this true, but it is general. For a particle placed in an arbitrary potential, if we consider the Fourier transform of the wave function,

$$\varphi(\boldsymbol{p},t) = \int \psi(\boldsymbol{r},t)\, e^{-(i/\hbar)(\mathbf{p}\cdot\mathbf{r})}\, \frac{d^3r}{(2\pi\hbar)^{3/2}} \quad , \tag{3.18}$$

then $\varphi(\boldsymbol{p},t)$ is the probability amplitude for the momentum,

$$dP(\boldsymbol{p}) = |\varphi(\boldsymbol{p},t)|^2\, d^3p \quad . \tag{3.19}$$

The free particle case is a particular case for the behavior in time.

Now, we can calculate various expectation values of the momentum or of functions of the momentum.

3.7 Heisenberg uncertainty relations

Here, we face an extraordinary discovery, a central result, and, at the same time, a tragedy.

Indeed, it is a consequence of Fourier analysis and of equation (3.15) that, whatever the wave function is, whatever the state of the system, the dispersions on measurements of positions and momenta along the same axis always satisfy the inequalities

$$\Delta x\, \Delta p_x \geq \frac{\hbar}{2}\,, \quad \Delta y\, \Delta p_y \geq \frac{\hbar}{2}\,, \quad \Delta z\, \Delta p_z \geq \frac{\hbar}{2}\,, \tag{3.20}$$

which are called the Heisenberg uncertainty relations.

Here is a true physical constraint on a wave packet. If we compress it in one variable, it expands in the other! If it is compressed in position (i.e., localized) then it must must be spread out in wavelength. If it is compressed in

momentum, it is spread out in space.

What is the physical meaning of Heisenberg's inequalities?

- Once more, suppose we prepare N systems in the same state. For half of them, we measure their positions x; for the other half, we measure their momenta p_x. Whatever way we prepare the state of these systems, the dispersions obey these inequalities.
- These are intrinsic properties of the quantum description of the state of a particle.
- Heisenberg uncertainty relations have nothing to do with the accuracy of measurements. Each measurement is done with as great an accuracy as one wishes. They have nothing to do with the perturbation that a measurement causes to a system, inasmuch as each particle is measured only once.
- In other words, the position and momentum of a particle are defined numerically only within limits that obey these inequalities. There exists some "fuzziness" in the numerical definition of these two physical quantities. If we prepare particles all at the same point, they will have very different velocities. If we prepare particles with a well-defined velocity, then they will be spread out in a large region of space.
- Newton's starting point must be abandoned. One cannot speak simultaneously of x and p. The starting point of classical mechanics is destroyed.

Some comments are in order.

A plane wave corresponds to the limit $\Delta p = 0$. Then Δx is infinite. In an interference experiment, the beam, which is well defined in momentum, is spread out in position. The atoms pass through both slits at the same time. We cannot "aim" at one of the slits and observe interferences.

The classical limit (i.e. how does this relate to classical physics) can be seen in a variety of ways that are more or less equivalent. One possibility is that the orders of magnitude of x and p are so large that $\hbar/2$ is not a realistic constraint. This is the case for macroscopic systems. Another possibility is that the accuracy of the measuring devices is such that one cannot detect the quantum dispersions Δx and Δp. We show later on how one recovers quantitatively the laws of classical mechanics themselves.

3.7.1 Size and energy of a quantum system

Our third comment is that now we can do a lot of physics. In fact, uncertainty relations enable us to estimate orders of magnitude of various effects without solving complicated equations.

In the lowest energy state of a quantum system, its "ground state," that is, when it is no longer excited, the product $\Delta x\, \Delta p$ is of the order of $\hbar$,

$$\Delta x\, \Delta p_x \sim \hbar \quad . \tag{3.21}$$

This allows us to estimate quickly orders of magnitude. Of course we will always miss a numerical factor of order 1. This factor can only be obtained by solving equations, but we will know the orders of magnitude.

For instance, there is a relation between the size and the energy of a quantum system.

Consider a particle that is bound to a fixed center, for instance an electron around a proton in an atom, or nucleons (neutrons and protons) in a nucleus. We assume the center is fixed at position $\boldsymbol{r_0}$; that is, we neglect quantum effects of the center which is very massive. The particle has some wave function. In the center of mass system, we can choose the origin of coordinates so that we have by assumption $\langle \boldsymbol{r} \rangle = 0$, $\langle \boldsymbol{p} \rangle = 0$.

The dispersion Δx is of the order of r_0, the size of the system. The square of the momentum dispersion Δp is equal to $2m$ times the kinetic energy

$$\Delta p^2 = \langle p^2 \rangle = 2m \langle E_{\text{kin}} \rangle \quad .$$

There is therefore a relation between the kinetic energy and the size of the system

$$\langle E_{\text{kin}} \rangle \simeq \frac{\hbar^2}{2mr_0^2} \quad .$$

The smaller the system is, the larger its kinetic energy.

For an external electron in an atom, the size is of the order of an Angstrom, and we obtain a kinetic energy of a few eV.

In a nucleus, we lose a factor of 2000 because of the mass, but the size gives a factor of 10^{10} larger; we therefore obtain energies of the order of tens of MeV. We do obtain the correct orders of magnitude, because the kinetic energy, binding energy, and potential energy are of the same order of magnitude unless there exists a pathology.

This also explains why, in order to probe matter at short distances, one must use large energies. It is necessary to use powerful particle accelerators.

3.7.2 Stability of matter

More important is that uncertainty relations allow us to prove the stability of matter, which is one of the greatest contradictions of classical physics. In the world of Newton and Maxwell, matter should be unstable and the world should collapse.

This is an inevitable consequence of the theories of Newton and Maxwell. Consider the very simple case of a hydrogen atom; one electron orbiting around a proton in the Coulomb potential $V(r) = -q_e^2/4\pi\epsilon_0 r$. Suppose the orbit is circular, for simplicity, of radius r. Mechanical equilibrium implies $mv^2/r = q_e^2/4\pi\epsilon_0 r^2$, and the energy of the electron is therefore

$$E = \frac{p^2}{2m_e} + V(r) = -\frac{1}{2}\frac{q^2}{4\pi\epsilon_0 r} \ .$$

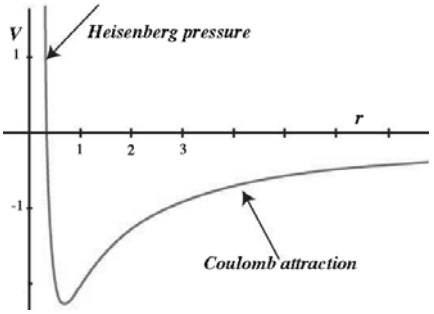


Fig. 3.4. Effective interaction potential due to the interplay between Coulomb attraction and quantum repulsion owing to uncertainty relations in an atom.

This energy is not bounded from below; the more the radius decreases the more the energy decreases. Now, in its circular motion, the electron is accelerated. The consequence of Maxwell's equations is that it must then radiate and lose energy. Therefore, from the classical point of view, matter is unstable. The electron should radiate continuously and it should collapse on the nucleus ($r \to 0$) by radiating an infinite amount of energy.

It is perhaps the most serious problem of classical physics, although people did not realize that because the electron and the nucleus were found rather late.

Uncertainty relations preserve us from this catastrophic fate and suppress this inconsistency. Let $\langle r \rangle$ be the average distance of the electron and the proton, which we consider fixed (the recoil is negligible). The Coulomb energy is of the order of $q_e^2/4\pi\epsilon_0\langle r\rangle$. If we use the order of magnitude (3.21), the kinetic energy is $E_k \geq \hbar^2/2m_e\langle r\rangle^2$. Therefore, the total energy is of the order of

$$E \geq \frac{\hbar^2}{2m_e\langle r\rangle^2} - \frac{q^2}{4\pi\epsilon_0\langle r\rangle} \ . \tag{3.22}$$

This quantity is bounded from below. Its minimum is attained for $\langle r\rangle = 4\pi\epsilon_0\hbar^2/(m_e q^2) \sim 0.53\ 10^{-10}$ m, which is bounded from below, and leads to

$$E_{\min} \sim -\frac{m_e}{2\hbar^2}\left(\frac{q^2}{4\pi\epsilon_0}\right)^2 = -13.6 \text{ eV} \ .$$

This is a fundamental result. The uncertainty relations put a lower bound on the average distance of the electron and the proton, as well as on their potential energy, and their total energy. This explains the stability of matter.

The above argument is not rigorous; it can be made rigorous because one can prove other forms of the uncertainty relations, in particular that for any system we have $\langle p^2\rangle \geq \hbar^2\langle 1/r\rangle^2$. If we apply this to $\langle E\rangle = \langle p^2\rangle/2m - (q_e^2/4\pi\epsilon_0)\langle 1/r\rangle$ the argument becomes rigorous: $\langle 1/r\rangle$ is bounded from above. The uncertainty relations create a "Heisenberg pressure," which reacts against

the fact that the electron and proton come too close to each other. Consequently, this results in an equilibrium situation and the quantum impossibility that matter collapses.

This term has the behavior of a centrifugal force, but it does not come from angular momentum. We show later on that, in its quantum ground state, the electron of the hydrogen atom has zero angular momentum and does not radiate.

3.8 Controversies and paradoxes

Heisenberg was a young assistant of Max Born and, as early as 1924, he had elaborated his own version of quantum mechanics which seems at first very different from what we are doing (we come back to this point). It was only in 1927 that he stated his uncertainty principle (the proof was given later). It is one of the most fundamental elements of quantum mechanics. It is compulsory for the consistency of quantum mechanics. If an experiment contradicts Heisenberg's relations, all of quantum mechanics is destroyed.

3.8.1 The 1927 Solvay Congress

The uncertainty principle set off terrible debates. Some people were astonished; some others were enthusiastic. And there was the tragedy we mentioned above. The year 1927 was a breaking off for one of the founders of quantum mechanics, and not the least, Albert Einstein, one of the greatest physicists in history. In 1905 he had originated the theory of Brownian motion, he had invented special relativity, and he had invented the photon, for which he was awarded the Nobel prize in 1921. In 1915, he had constructed general relativity, a most beautiful theory. He had made tremendous contributions to statistical physics and to quantum physics. In 1917, he understood the existence of stimulated emission, a keypoint for laser physics. At the fifth Solvay Congress, in Brussels in October 1927, he rose up and revolted against the probabilistic aspect of quantum mechanics, and against uncertainty relations which he disliked profoundly. Einstein did not understand what was going on.

Einstein's revolt concerned two points. One is the notion of a *complete* description of reality. He thought that a complete description is possible in principle, but that the probabilistic description is simply quicker to handle and more convenient. The other reason is the notion of *determinism*: same causes produce same effects.

Einstein said the famous: "God does not throw dice!" Actually, Einstein's words are in a letter he wrote to Max Born on December 4th, 1926,[3] "The theory produces a good deal but hardly brings us closer to the secret of the

[3] "Die Theorie liefert viel, aber dem Geheimnis des Alten bringt sie uns kaum näder. Jedenfalls bin ich überzeugt, dass *der* nicht würfelt."

Old One. I am at all events convinced that *He* does not play dice." Because the theory works and gives good results, it must be some intermediate step: there must exist a more complete underlying theory, involving, for instance, "hidden variables"[4] to which we do not yet have any access and over which we average things.

And, at the Solvay meeting, Einstein refused and reacted abruptly. This was due partly to the dogmatic attitude of Niels Bohr. So Einstein attacked; in particular he attacked uncertainty relations. Uncertainty relations are, in that respect, similar to the Carnot principle and perpetual motion in thermodynamics. Einstein accumulated counterexamples. But, of course, his counterexamples were far from being obvious. One had to work hard to disprove them.

3.8.2 The EPR paradox

In 1935, Einstein proposed a famous paradox: the EPR – Einstein, Podolsky, and Rosen – paradox which we shall describe in chapter 13.

In that "gedanken experiment," by considering a couple of correlated particles, he showed how because of momentum conservation, one could in principle know as accurately as one wished both the position and the momentum of one of the particles, which "beats" uncertainty relations.

3.8.3 Hidden variables, Bell's inequalities

In another version, proposed 20 years later by David Bohm, with spin, it is a genuine puzzle. This goes so far that physicists addressed the question as to whether there exists a more complete theory involving "hidden variables" of which we would have some ignorance.

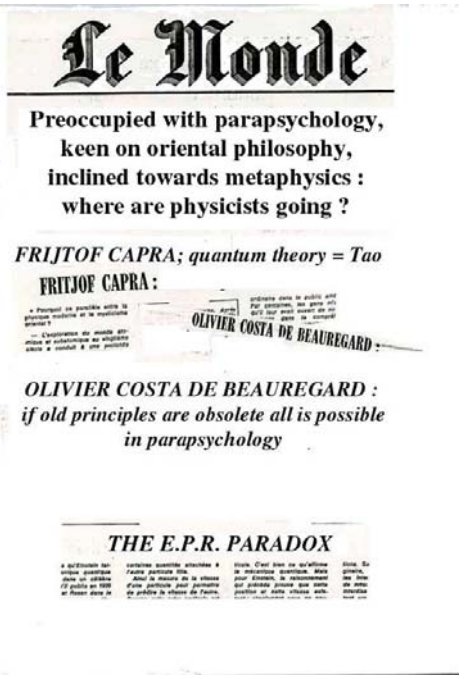
After all, when we play cards, we are blind to the identity of each card that is dealt. But if one used sophisticated enough devices, one could certainly tell the difference between the ace of spades and the seven of hearts. In order to play end enjoy it, we forbid ourselves to have a more complete knowledge which exists in principle.

The amazing thing is that John Bell, in 1965, was able to show that this assumption leads to *quantitative measurable consequences* that are in opposition to the predictions of quantum mechanics. We come back to this in chapter 13. And, since the 1980s, a series of experiments has been performed, among which are those of Alain Aspect and his group in Orsay, to try to see whether one could falsify quantum mechanics.

[4] Actually Einstein never used that word.

3.8.4 The experimental test

This attracted the attention of the general public and the press. There are philosophical considerations attached to that problem, and even more as we can see on two documents.


Le Monde

Preoccupied with parapsychology, keen on oriental philosophy, inclined towards metaphysics : where are physicists going ?

FRIJTOF CAPRA; quantum theory = Tao

FRITJOF CAPRA :

OLIVIER COSTA DE BEAUREGARD :

OLIVIER COSTA DE BEAUREGARD : if old principles are obsolete all is possible in parapsychology

THE E.P.R. PARADOX

Fig. 3.5. Excerpts from *Le Monde* of October 24, 1979. Titles and headlines have been translated into English.

In 1979, at the Cordoba Colloquium, there were many discussions on parapsychology, levitation, oriental philosophy, and other bold considerations. This was reported by the newspaper *Le Monde* (considered as leftist and antigovernmental at that time) as one can see in Figure 3.5. All of that was based on the Einstein–Podolsky–Rosen paradox.

But exactly three years later, in 1982, in the same newspaper which in the meantime had become rightist because it remained antigovernmental, the phraseology changed. One can see the words: "Experiment might ruin Einstein's hopes," "God probably plays dice," and so on. Why did that change of phraseology occur? Because exactly in the meantime, in May 1981, there


11781 DERNIÈRE ÉDITION

Le Monde

Fondateur : Hubert Beuve-Méry Directeur : André Laurens

EXPERIMENT MIGHT RUIN EINSTEIN'S HOPES

God probably plays dice

The three Orsay experiments

Faster than light!

Fig. 3.6. Excerpts from *Le Monde* of December 15, 1982. Titles and headlines have been translated into English.

had been a major event in France.[5] In fact the experiments of Alain Aspect and others were published and they were negative! Quantum mechanics holds, even though it hurts some "common sense." We must face experimental results and revise our way of thinking. It is an extraordinary experiment which we discuss in chapter 13.

So, as one can see, not only is quantum mechanics useful, but it can be used for many purposes. And the same newspaper, which has the reputation of being a very serious source of information, can also say things that change with time.

[5] There was also a minor one, that is, the election of the Socialist François Mitterrand as President. President Mitterrand was given the name "God" in a TV puppet show.

4

Physical quantities

What we have done up to now on wave mechanics seems complicated at first sight. But when one comes back to it, one realizes that indeed there are new words, complicated formulae, but the number of important things to remember is not that large.

Actually, everything can be summarized as follows:

1. The superposition principle says that the state of a particle in space at time t is described by a wave function $\psi(\boldsymbol{r}, t)$. This wave function has a probabilistic interpretation. We have said, but not used yet, that the wave function belongs to a vector space. This fundamental property expresses interference phenomena. We show in a moment that it is not trivial at all.
2. The time evolution of the wave function is given by the Schrödinger equation.
   This theoretical scheme explains interference phenomena of particles through the form of de Broglie waves.
3. By studying wave packets and the properties of the Fourier transformation, we have learned a lot.
   We have understood that a wave packet can be very close locally to a monochromatic plane wave, but still be normalizable. In other words, if one looks at it from a distance, it can look well localized in space.
   It also has good velocity. The group velocity of a wave packet is equal to the velocity of a classical particle. This was one of the successes of Louis de Broglie. More generally, it's a good exercise on the Fourier transformation to show that the time derivative of the expectation value of a wave packet is equal to the expectation value of the momentum divided by the mass
   $$\frac{d}{dt}\langle x\rangle = \frac{\langle p\rangle}{m} \quad . \tag{4.1}$$
4. We have understood that the probability amplitude for the momentum of the particle is given by the Fourier transform of the wave function. In other words, the Fourier transformation extracts in the wave function the information on the momentum of the particle.

5. Finally, using again Fourier analysis, we have ended up with the Heisenberg uncertainty relations, which are a cornerstone of quantum mechanics.

4.1 Statement of the problem

We now finish our program: we examine the question of physical quantities and their measurement.

We discuss the following questions simultaneously. The first is: what does one find by measuring a physical quantity on a system whose state ψ is known?

The second is: of what does the measurement process consist? We discuss this question by using one of Einstein's counterexamples. That is the point on which the philosophical considerations we mentioned, and Einstein's objections to quantum mechanics, bear.

Next, we examine a physical quantity that plays a central role, the energy. And we finish this chapter by describing the celebrated Schrödinger cat problem which shows the kind of debate one can have on the interpretation of quantum mechanics.

Here we are dealing with a difficult subject. One does not understand everything immediately, but things come quickly when one manipulates the theory. We really face the problem that the relation between physical concepts and their mathematical representation is not obvious in quantum mechanics.

What we are going to say on physical quantities would be terribly abstract and puzzling if we did not have the Fourier transformation at our disposal. But in order to progress we need it. We will digest it afterwards. It is difficult for us, but not for the people who were constructing quantum mechanics. We will soon understand why.

4.1.1 Physical quantities

Let's start from the beginning. Physical quantities are things corresponding to aspects of the reality of a system that can be measured, that is, they can be characterized by numbers. In classical physics, considering a particle, the result of the measurement of a quantity A is a number a and this single number at time t.

In quantum mechanics, what we said concerning position and momentum holds for any physical quantity. If one uses the same apparatus as in classical physics, and if ψ is the state of the system, quite generally, the result of a measurement on one system is some value a_i. But if we measure that quantity on a large number N of identical systems, all in the same state ψ, we obtain N values that are not all the same in general. There is some probability law for this quantity.

Consequently, the complete result of a measurement of a quantity A on the state ψ is the whole set of issues a_i and associated probabilities p_i, or rather the probability amplitudes $\alpha(a_i)$. This set may be a continuous set, as

for the position of a particle; it may also be a discrete set, as for the energy levels of an atom. The theory must therefore be able to define laws that allow us to calculate these numbers starting with the wave function ψ:

$$\psi(\boldsymbol{r},t) \longrightarrow \text{ numbers } \quad \{a_i,\, \alpha(a_i)\,, p(a_i) = |\alpha(a_i)|^2\} \quad .$$

Here, we are interested in only one combination of these numbers, the expectation value of the result

$$\langle a\rangle = \sum_i a_i p(a_i), \text{ in the discrete case, or } \langle a\rangle = \int a\, p(a)\, da \quad . \tag{4.2}$$

We say a few words on the issues, namely the possible results of a measurement, but all this, issues and amplitudes, is simpler when we have Hilbert space analysis at our disposal.

Notice something rather obvious intuitively: it is an experimental result that

- The issues only depend on the nature of the system, not on its state. For instance, the energy levels are different for an electron in an atom and for a proton in a nucleus.
- It is the probability amplitudes that depend on the state of the system.

We must therefore transform a function into numbers.

4.1.2 Position and momentum

However, one can object that we already know a lot. We know the probability laws for the position and for the momentum. Therefore we can calculate a whole lot of expectation values. For instance, we can calculate the expectation value of the potential energy of the particle. At each point, the potential energy of the particle would be $V(r)$. The expectation value, averaged over the probability law, is given by the integral

$$\langle V\rangle = \int V(r)|\psi(\boldsymbol{r},t)|^2 d^3r \quad . \tag{4.3}$$

Similarly for any function of the momentum, for instance, the kinetic energy

$$\langle E_k\rangle = \int \frac{p^2}{2m}|\varphi(\boldsymbol{r},t)|^2 d^3p \quad . \tag{4.4}$$

True, but there are two objections:

- First of all, for what concerns the momentum, thanks to the beautiful properties of the Fourier transformation, one can calculate the expectation value of any function of the momentum. But it is slightly frustrating. We must take the wave function, calculate an integral,

$$\varphi(\boldsymbol{p},t) = \int \psi(\boldsymbol{r},t)\, e^{-(i/\hbar)\mathbf{p}\cdot\mathbf{r}}\, \frac{d^3r}{(2\pi\hbar)^{3/2}} \quad , \tag{4.5}$$

and then start over again, and calculate another integral. Altogether, this means calculating a triple integral.

- And, more seriously, there exist physical quantities that are functions of both the position $\boldsymbol{r}$ and the momentum $\boldsymbol{p}$. For instance, the total energy is $E = p^2/2m + V(r)$ and, even worse, the angular momentum is $\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p}$.

There, we don't know what to do. Even if we can calculate the expectation value of the energy by adding the two expressions (4.3) and (4.4),

$$\langle E \rangle = \langle E_c \rangle + \langle V \rangle \quad , \tag{4.6}$$

we don't know the probability law for the energy. It's even worse for the angular momentum $\boldsymbol{L}$. What are its probability laws and its expectation value?

Do we need to invent a new type of Fourier transform each time and be amazed? One would never believe it. Even though, in the end, it is indeed what will happen; the structure of the Fourier transformation is quite general.

We must have a general principle. We must be able to calculate everything directly in terms of the wave function ψ that contains all the physical information on the system.

4.2 Observables

The principle that we have to understand and learn can be stated in a single sentence. But it is a sentence that seems at the beginning completely abstract and far away from physics.

> To each physical quantity A, there corresponds an observable $\hat{A}$, a mathematical object that is a linear Hermitian operator acting on the space of wave functions, such that the expectation value of the result of a measurement of the quantity A on the system in a state ψ is given by the number:
>
> $$\langle a \rangle = \int \psi^*(\boldsymbol{r},t)\, [\hat{A}\, \psi(\boldsymbol{r},t)]\, d^3r \quad . \tag{4.7}$$

For the moment, we don't understand a word, except that a linear operator is a linear mapping of the space onto itself. For instance we can multiply the function by x, take its derivative with respect to x, or take its Laplacian (provided these operations are legal).

That observables are Hermitian is inessential for the moment. We show later on that it comes from the fact that measured quantities are real numbers.

Therefore, our principle says that there exist linear Hermitian operators (which remain to be defined) that perform the operation of transforming a function into numbers using the above prescription.

The observable $\hat{A}$ is the tool that extracts from the wave function the information about the quantity A. It transforms the wave function into another function; we multiply by the complex conjugate of the wave function, integrate, and there is the desired number.

4.2.1 Position observable

For what concerns space properties, this is completely trivial. It simply consists of developing the modulus square and in writing

$$< x >= \int \psi^*(\boldsymbol{r},t)\, x\, \psi(\boldsymbol{r},t) d^3r\,,\ < V >= \int \psi^*(\boldsymbol{r},t) V(\boldsymbol{r}) \psi(\boldsymbol{r},t) d^3r, \quad (4.8)$$

which indeed have good form. The position and potential energy observables are the multiplication of the wave function by x, y, and z, or by $V(\boldsymbol{r})$.

4.2.2 Momentum observable

But for momentum it is less obvious, except that the form

$$< p_x >= \int \varphi^*(\boldsymbol{r},t)\, p_x\, \varphi(\boldsymbol{r},t) d^3p \quad (4.9)$$

has good structure.

In fact, if we remember the definition of the Fourier transform, its inversion, and the Plancherel theorem (3.13), we see that we simply have to find a function whose Fourier transform is $p_x\varphi$. But that is very simple; we only need to take the derivative

$$\frac{\partial}{\partial x}\psi(\boldsymbol{r},t) = \int \frac{ip_x}{\hbar}\varphi(\boldsymbol{p},t)\, e^{(i/\hbar)\mathbf{p}\cdot\mathbf{r}}\, \frac{d^3p}{(2\pi\hbar)^{3/2}} \quad , \quad (4.10)$$

or

$$\frac{\hbar}{i}\frac{\partial}{\partial x}\psi(\boldsymbol{r},t) = \int p_x\, \varphi(\boldsymbol{p},t)\, e^{(i/\hbar)\mathbf{p}\cdot\mathbf{r}}\, \frac{d^3p}{(2\pi\hbar)^{3/2}} \quad . \quad (4.11)$$

Therefore, $(\hbar/i)\partial\psi/\partial x$ is the Fourier transform of $p_x\varphi$

$$\frac{\hbar}{i}\frac{\partial}{\partial x}\psi \ \longleftrightarrow p_x\varphi \quad ,$$

and thanks to Plancherel, this is becoming really interesting and nontrivial.

We obtain

$$< p_x >= \int \varphi^*(\boldsymbol{r},t)\, p_x\, \varphi(\boldsymbol{r},t) d^3p = \int \psi^*(\boldsymbol{r},t)\, \left(\frac{\hbar}{i}\frac{\partial}{\partial x}\psi(\boldsymbol{r},t)\right) d^3r \quad . \quad (4.12)$$

In other words, in our principle, the observable $\hat{p}_x$ associated with the x component of the momentum is

$$\hat{p}_x = \frac{\hbar}{i}\frac{\partial}{\partial x} \quad , \tag{4.13}$$

and, because this holds for other components, the vector observable $\hat{\boldsymbol{p}}$ is simply

$$\hat{\boldsymbol{p}} = \frac{\hbar}{i}\boldsymbol{\nabla} \quad . \tag{4.14}$$

And there we win! We don't have to calculate the Fourier transform. We know how to express $\langle p_x \rangle$ or the expectation value of any function of the momentum directly in terms of the wave function.

For instance, the expectation value of the kinetic energy, $E_k = p^2/2m$ in equation (4.4) is

$$\langle E_k \rangle = \int \psi^*(\boldsymbol{r}, t)\,(\frac{-\hbar^2}{2m})\Delta\psi(\boldsymbol{r}, t)\,d^3r \quad . \tag{4.15}$$

4.2.3 Correspondence principle

Now, we are nearly home. We just have to fix the form of other observables. In order to do this, we simply remember classical physics and use a principle that ensures, as we show later on, that in the classical limit one recovers the classical equations. This is called the *correspondence principle*.

In classical mechanics, physical quantities are functions $A(\boldsymbol{r}, \boldsymbol{p})$ of the position and momentum variables. The correspondence principle consists of choosing in quantum mechanics the same functions of the position and momentum observables.

To the quantity $A(\boldsymbol{r}, \boldsymbol{p})$ there corresponds the observable $\hat{A} = A(\hat{\boldsymbol{r}}, \hat{\boldsymbol{p}})$. For instance,

$$\hat{E}_c = \frac{-\hbar^2}{2m}\Delta\,,\ \hat{\boldsymbol{L}} = \frac{\hbar}{i}\boldsymbol{r} \times \boldsymbol{\nabla}\,,\ \text{that is,}\quad \hat{L_z} = \frac{\hbar}{i}(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}) \quad .$$

4.2.4 Historical landmarks

Why is this principle of physical quantities so difficult to accept for us, whereas it caused no difficulty for the people who built quantum mechanics? Because we are missing a link (or some mathematics) and because history is not a logical sequence but an arborescence.

In 1926, there were actually two versions of quantum mechanics. There was wave mechanics (in 1923 de Broglie's hypothesis, in 1926 Schrödinger's equation). But there was also matrix mechanics, which had been constructed

since 1924 in a most inspired way by Heisenberg, and had been developed by Heisenberg, Born, and Jordan.

Furthermore, both versions gave good results. So people quarreled. Are electrons or atoms waves or matrices? That was a terrible question. Physicists discussed and argued about fundamental questions, but not this question, which is basically technical.

Born had declared Schrödinger's works, namely finding energy levels as stationary wave problems, as being of "unsurpassed greatness in theoretical physics." In 1926, Schrödinger had given a series of talks in Copenhagen. There, he, who was a man of great distinction and class, had been literally attacked by Niels Bohr, rigid and dogmatic, to such an extent that he said, "If one must stick to that idea of quantum jumps, I really regret getting involved in all this business!"

Niels Bohr had answered him: "But, Herr Schrödinger, you shouldn't, given all the publicity that your works have given to our ideas." One can imagine the atmosphere.

But Schrödinger and Dirac, the little Mozart of quantum mechanics, were wise people and they brought back peace, as we show in Chapter 6, because they knew some mathematics. Independently, at the end of 1926, they showed that the two approaches were equivalent, and they saw this principle emerge in a natural way. There was no problem. They had both the physical concepts and the mathematical tools.

Wave mechanics was interested primarily in the state of a system; matrix mechanics concentrated primarily on physical quantities. For our information, the importance of the Fourier transform, which has been very useful for us, was pointed out quite late, in 1927, by Darwin.

4.3 A counterexample of Einstein and its consequences

At this stage, in order to progress, it is useful to analyze one of Einstein's counterexamples. This will allow us to address the question, "OK, God plays dice, but is it a full time job?" In other words, are there situations where indeterminism disappears and God ceases to throw dice?

Einstein did not like the probabilistic aspect of the theory. So he attacked the root of the problem. Indeterminism is inscribed in the uncertainty relations, so why not break them to pieces, once and for all.

The first counterexample, "gedanken experiment," of Einstein is the following. Between two diaphragms, one puts a device made of two cogwheels that let pass particles of a well-defined velocity (as in Fizeau's measurement of the velocity of light). The second diaphragm has a dimension δz as small as one wishes.

The cogwheel device ensures that particles which pass through the second diaphragm have a well-defined longitudinal velocity v_{long}. But, in order for

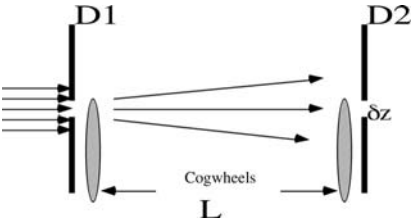


Fig. 4.1. Gedanken experiment of Einstein in order to beat Heisenberg's uncertainty relations.

the particles to pass through the diaphragm, it is necessary for them to have a transverse velocity v_t less than $v_t \leq v_{long}\delta z/L$.

We can make the size of the diaphragm δz as small as we wish. Therefore, immediately after the particles have passed the diaphragm, we are sure of their positions along z, up to δz, and we know their transverse velocity v_t up to $v_{long}\delta z/L$.

We therefore beat Heisenberg!

• "Nein, nein, nein! Not at all!" said Heisenberg to Einstein.

• "Warum nicht? Why not?" asked Einstein.

• "Look at the result of the experiment. If you measure the velocity v_z after particles have passed through the diaphragm, for a large number of particles you will not find only one value up to $v_{long}\delta z/L$, but a whole lot of possible values."

• "Why?" asked Einstein.

• "Because of the uncertainty relations! These particles are emitted by a small hole and have a small spreading δz in position along z. Therefore, they have a large spreading in momentum $\Delta p_z \geq \hbar/2\delta z$."

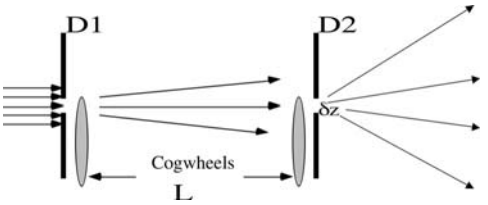


Fig. 4.2. Velocity distribution of the particles after they pass through the second diaphragm, owing to the uncertainty relations.

• "You're making fun of me," said Einstein. "If you use uncertainty relations to prove uncertainty relations, something's going wrong in your mind!"

• "OK, you prefer de Broglie waves," said Heisenberg, "but it's exactly the same result. The wave diffracted by a small hole is more and more dispersed in the wave vector as the hole gets smaller. Uncertainty relations also exist in usual wave physics. They express globally one aspect of diffraction. It is only

because we apply them to particles on which we have some a priori knowledge that we are shocked."

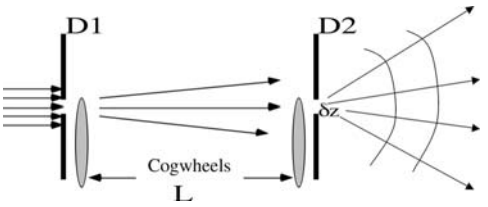


Fig. 4.3. Diffraction of a wave by a small hole.

The only thing we know with certainty is that after passing through the second diaphragm, the particles have passed through a hole of size δz, therefore there is a dispersion in momentum Δp_z which is larger than $\hbar/2\delta z$ in the statistical sense (i.e., if we measure a large number of particles).

This is what Heisenberg explained to Einstein: the longitudinal and transverse velocities belong to the past; talking about them is history or philosophy. One cannot do anything with that. And Heisenberg added, "It is a matter of personal belief, without any operational content, to decide whether the past history of a particle has a physical reality. It is something one cannot measure." Heisenberg came close to convincing Einstein just before the 1927 Solvay meeting.

4.3.1 What do we know after a measurement?

But we have learned a lot with this counterexample.

We know something for sure after the particle exits. It has passed through the hole. Therefore, its position is well-defined (up to δz).

But, to place a hole and to verify that a particle has passed through it is simply one of the many ways to measure the position of that particle. We could use any measuring apparatus and measure any physical quantity; the situation would be the same.

1. Therefore, we know something for sure after a measurement on a single system. We have found some well-defined measured value (up to δz in this particular case). Therefore, if we redo the measurement immediately afterwards (immediately means that the wave function hasn't had time to evolve appreciably) we will find the same result z_0 up to δz, and only that result. Therefore, in performing that measurement, we have obtained information on the state of the particle after the measurement process.
   A measurement on a single system provides us with some information on the state of the system after the measurement. This can also be considered as preparing the state of the system.

2. However, can we check that Einstein was partially right and that the velocity was well-defined by his device?
   Yes, of course; we must put a speedometer before the diaphragm. For N particles, the velocity is well-defined, but then a position measurement will show that the particles are delocalized; that is, the position dispersion Δz is very large before the diaphragm. And we can simply move the hole around in all possible positions, which will give us the position probability before particles pass through the hole, and then apply the Fourier transformation (which in full rigor means making an assumption on the phase).
   It is by a measurement on N identical systems that the measurement gives us information on the state before the measurement.
3. We can imagine the origin of many philosophical questions at the time, and still now: what is physical information? What is reality? Does reality exist independently from the fact that one observes and measures? Does a tree falling in a forest make noise even if no one is there to listen to it?

4.3.2 Eigenstates and eigenvalues of an observable

Now, obviously, this analysis implies much more. Let's pursue this argument.

1. Just after a position measurement, there is no longer any determinism on the position. A further position measurement will give us a known value; the value of the position is sure and well-defined. The state is such that God has stopped playing dice!
2. The completely general consequence is that, for any physical quantity A, there must exist particular states such that the result of a measurement is sure and well-defined.
3. What are the corresponding particular wave functions? This is formalized using the notions of eigenfunctions and eigenvalues of the observable $\hat{A}$.

A function ψ_α is an eigenfunction of $\hat{A}$ if the application of $\hat{A}$ on this function gives the same function multiplied by a number a_α which is the corresponding eigenvalue

$$\hat{A}\psi_\alpha(\boldsymbol{r},t) = a_\alpha \psi_\alpha(\boldsymbol{r},t) \quad . \tag{4.16}$$

With this definition, the theorem is very simple.

Theorem 1. *If ψ is an eigenfunction of $\hat{A}$, then the result of a measurement of A is certain and equal to the corresponding eigenvalue.*

Proof. The expectation value of A is a_α, and the expectation value of A^2 is a_α^2, therefore the dispersion vanishes, and there is no uncertainty. Indeed, we have

$$\langle a \rangle = \int \psi_\alpha^* \hat{A} \psi_\alpha \, d^3r = a_\alpha \int |\psi_\alpha|^2 \, d^3r = a_\alpha \quad , \tag{4.17}$$

$$\text{but} \quad \langle a^2 \rangle = \int \psi_\alpha^* \hat{A}^2 \psi_\alpha \, d^3r = a_\alpha^2 \int |\psi_\alpha|^2 \, d^3r = a_\alpha^2 \quad ,$$

$$\text{therefore} \quad (\Delta a)^2 = \langle a^2 \rangle - \langle a \rangle^2 = 0 \quad \text{QED} \quad .$$

Later on, we show that the converse is true. Therefore the result of a measurement is well-defined if and only if the wave function ψ is an eigenfunction of $\hat{A}$. We can also say that "the value of A is well-defined."

Careful: all of this is easier to formulate and to understand with discrete probabilities than with continuous probabilities.

Hence, the answer to the question, "What are the possible results of a measurement?" Obviously, these are the eigenvalues of the corresponding observable. If we are sure that in a further measurement, we will find the value a_i we have already found, that means that the number a_i found previously is one of the eigenvalues of $\hat{A}$.

4.3.3 Wave packet reduction

Now, the last consequence and not the least.

The measurement, which we can decide to perform at some instant, is an irreversible act that changes the wave function completely. Before the measurement of a given quantity, the particle had some wave function that led to a variety of possible results in a measurement of that quantity. After the measurement, it has another wave function for which the result is unique.

This is completely irreversible! Once we have made the measurement, we cannot decide that we haven't made it and come back to the previous wave function. In other words, a measurement, which we are free to perform or not, instantaneously modifies the wave function in all space. Because of the very intuitive example presented above, this is called the *reduction of the wave packet*. Before, it was spread out in space; just after the measurement it is very concentrated. And this wave packet reduction postulate is essential in the formalization of quantum mechanics.

It is understandable that there lies the source of serious problems which still last in the interpretation of quantum mechanics. There appears to be an instantaneous phenomenon at a distance. Isn't that in contradiction to relativity?

In fact, this is the point on which Einstein constructed his EPR paradox a few years later. No signal can travel faster than light; it is not possible to change the wave function instantaneously at a distance. However, this is what experiments on Bell's inequalities have shown to occur. Something does happen instantaneously at a distance. Experiment says that the predictions of quantum mechanics are right. But no information can travel faster than light because, in order to observe that, one must phone other people at other places to know what happened there.

We now have all the elements of the theory. Notice that the language is getting closer and closer to that of linear algebra: we have spoken of linear

mappings, we used the words Hermitian, eigenvalues, and indeed, in a while, quantum mechanics becomes more and more similar to matrix calculus.

4.4 The specific role of energy

In the following, we study various systems, and various physical quantities. But one of these quantities has a special role and we encounter it constantly because it has considerable importance. That is energy.

4.4.1 The Hamiltonian

How did Schrödinger test his equations?

- He was looking for wave equations. The problem was to find energy levels as the result of a stationary wave problem. Therefore, Schrödinger looked for stationary solutions of his equations (he actually wrote several before ending up with the good one), solutions of the form

$$\psi(\boldsymbol{r},t) = \phi(\boldsymbol{r})e^{-i\omega t} \quad . \tag{4.18}$$

  Inserting this into the Schrödinger equation, the term $e^{-i\omega t}$ factorizes and we obtain a time-independent equation

$$\frac{-\hbar^2}{2m}\Delta\phi(\boldsymbol{r}) + V(\boldsymbol{r})\phi(\boldsymbol{r}) = \hbar\omega\phi(\boldsymbol{r}) \quad . \tag{4.19}$$

- Schrödinger knew a lot of mathematics. He knew that the physically acceptable solutions (i.e., for us square integrable functions) of such a problem form a discrete set $\{\phi_n(\boldsymbol{r}), \hbar\omega_n\}$. Therefore, the energy levels are quantized

$$E_n = \hbar\omega_n \quad .$$

- But all of this must be consistent. If we are seeking energy levels of atoms, we are looking for particular solutions such that the energy is well-defined.

Now, let's consider a particle in a potential $V(r)$. Classically its total energy is $E = p^2/2m + V$, and the corresponding observable is therefore

$$\hat{H} = \frac{\hat{p}^2}{2m} + V = -\frac{\hbar^2}{2m}\Delta + V \quad . \tag{4.20}$$

The equation

$$-\frac{\hbar^2}{2m}\Delta\psi(r,t) + V\psi(r,t) = E\psi(r,t)\,, \text{ that is, } \hat{H}\psi(r,t) = E\psi(r,t), \tag{4.21}$$

is nothing but the eigenvalue equation for the energy. It is the same as (4.19), except that the time variable does not play any role.

The energy observable $\hat{H}$ (4.20) is called the *Hamiltonian* of the system.

How is it that Hamilton, who lived from 1805 to 1865, is involved in quantum mechanics, although he existed one century before its discovery? Well, he deserves it! The story is fascinating. Although he did not know quantum mechanics, Hamilton had understood that the structure of classical mechanics and quantum mechanics is the same.

He had first shown that geometrical optics was a limit of wave optics for small wavelengths. He was fascinated by variational principles, in particular by the similarity between Fermat's principle in optics and the least action principle of Maupertuis.

In 1828, Hamilton wrote the following extraordinary observation, "Newtonian mechanics is only one of the possible mechanics. It corresponds to the same limit as geometrical optics compared to wave optics, whereas geometrical optics is only an approximation." Nobody noticed, even Hamilton himself. The great mathematician Felix Klein pointed that out with some regrets in 1891. Of course there was no experiment at that time which showed that this idea could have any application. Planck's constant appeared nowhere.

4.4.2 The Schrödinger equation, time and energy

We observe a most remarkable property. The Schrödinger can be written in the form

$$i\hbar\frac{\partial}{\partial t}\psi = \hat{H}\psi. \tag{4.22}$$

This is the true Schrödinger equation. It is "simpler" than the other one. Naturally it boils down to the same thing in the present context. But it is more general, as we show later. It is valid for any system and, when we generalize quantum mechanics, the wave function ψ is replaced by another object, a vector in Hilbert space, the Hamiltonian $\hat{H}$ is an operator in that space, and the Schrödinger equation is simply (4.22).

If we look at it from that point of view, it tells us some thing remarkable. What determines the time evolution of the state of a system is the energy observable!

Caution! The information on the energy of a system is contained in the wave function, and $\hat{H}$ is the tool that enables us to extract it. Nevertheless, there exists a relation between two fundamental but equally mysterious physical concepts: energy and time.

We must be careful. Because people think that scientists are partially idiotic and that they know only complicated things such as mathematics, in order to make you feel comfortable at parties, they will ask you about Fermat's theorem, black holes, or what was there before the big bang. But take care, if they ask you what are time or energy, it's far better to shift the conversation to football, old Europe, or tornadoes.

The word energy is everywhere. Time is one of the most difficult physical concepts to define. I've looked it up in the *Oxford Concise Dictionary*, in order to see simple definitions.

• Energy is "a body's power of doing work by virtue of stresses resulting from its reaction to other bodies."

• Time is "The progress of continued existence viewed as affecting persons or things."

That's superb, but we haven't really made much progress.

We don't know what time is. Does time simply exist?

• The past no longer exists.

• The future does not yet exist.

• The present instant is of measure zero, in the mathematical sense, as soon as it's arrived, it's already gone.

In the *Confessions*, Saint Augustine wrote: "What is time? If no one asks me, I know. As soon as someone asks the question, and I want to explain it, I no longer know."

It's true that we don't know what time is, no more than what energy is (as opposed to other concepts such as velocity, flux, wavelength, and so on). What is extraordinary is to see that, if we don't know what time and energy are, there is a well-defined relation between the two concepts. It is in front of us in quantum mechanics. It also exists in Hamilton's analytical mechanics.

4.4.3 Stationary states

In order to see that in a more concrete way, we use the notions introduced above.

Consider an isolated system, that is, a system whose potential energy does not depend on time.

We have just seen that states with well-defined energies, eigenstates of energy (energy levels of atoms as for Schrödinger) for which $\Delta E = 0$, have no uncertainty in the energy.

They have a particularly simple time dependence

$$\psi_n(\boldsymbol{r}, t) = \phi_n(\boldsymbol{r}) e^{-iE_n t/\hbar} \quad . \tag{4.23}$$

It is then remarkable that in such cases $|\psi|^2$ does not depend on time.

The position probability law, the expectation of the position, does not depend on time. The system does not move! No physical quantity changes!

Such states are called *stationary states*. The functions $\phi_n(r)$ satisfy the eigenvalue equation (4.19), that is,

$$\hat{H}\phi_n(\boldsymbol{r}) = E_n \phi_n(\boldsymbol{r}) \quad ,$$

in the space variables only. This equation is called the time-independent Schrödinger equation.

Therefore, if a system has a well-defined energy, it cannot evolve. For that system, time does not exist. No evolution, no motion can occur.

If energy is well-defined, there are no oscillations for a pendulum, no Kepler motion along an orbit, and so on. The planetary model of atoms is wrong. Electrons don't orbit around the nucleus.

If the energy is well-defined, the system is frozen in time. That's a rather frightening observation.

4.4.4 Motion: Interference of stationary states

What causes motion? (This seems to be a question for Greek philosophers in antiquity.)

In order for motion to appear, the wave function must be a linear superposition of stationary states in interference. For instance, one can easily check that a superposition of two stationary states of different energies E_1 and E_2

$$\psi = \lambda\varphi_1 e^{-iE_1 t/\hbar} + \mu\varphi_2 e^{-iE_2 t/\hbar} \tag{4.24}$$

has a position probability $|\psi|^2$ in which the crossed term depends on time. However, for such a system $\Delta E \neq 0$, the energy is not well-defined.

Motion occurs, and time exists, only if energy is not well-defined, and if God plays dice with energy. Then, there are given nonvanishing probabilities of finding E_1, E_2, and so on. In order for motion to appear, the system must be an interference between states of different energies. That's a rather fascinating relation between time and energy.

Actually, we have just written a fundamental technical result, useful for all that follows. Suppose that at $t = 0$ the wave function is a given superposition of states with well-defined energies

$$\psi(r, t = 0) = \sum_n c_n \, \phi_n(r) \quad , \tag{4.25}$$

where the $\phi_n(r)$ are the energy eigenfunctions. Then its time evolution can be written directly, without solving any equation:

$$\psi(r, t) = \sum_n c_n \, \phi_n(r) e^{-iE_n t/\hbar} \quad . \tag{4.26}$$

One can readily check that this expression satisfies the Schrödinger equation (4.22) with the boundary condition (4.25). The evolution is known; motion is known. We show later on that it is a fundamental theorem of Hilbert space analysis that any wave function can be written in that way.

Therefore, the solution of the evolution problem in quantum mechanics, for an isolated system, always involves finding its energy eigenfunctions and eigenvalues, that is, solving the time-independent Schrödinger equation or, equivalently, the eigenvalue problem of the Hamiltonian. The time evolution follows immediately.

4.5 Schrödinger's cat

Fig. 4.4. Dirac, Heisenberg, and Schrödinger on their way to Stockholm. (All rights reserved.)

To end this chapter, we show one of the most famous paradoxes on the problem of measurement in quantum mechanics, in particular on the superposition of states and wave packet reduction. It's necessary for our culture; people also like to talk about it at parties. It's not that important if one does not understand everything at the beginning.

That's something that always makes me sad. Schrödinger was really an extraordinary man.

The photograph (Figure 4.4) shows Schrödinger, Dirac, and Heisenberg on their way to Stockholm to receive the Nobel Prize. From left to right, one can see Heisenberg's mother, Ms. Schrödinger, the mother of Dirac (who liked his mother very much), Heisenberg, and Schrödinger. One notices that Schrödinger was somewhat fancy. He was a very pleasant and friendly man.

But he had an enormous defect: he didn't like cats. I like cats very much.

4.5.1 The dreadful idea

In 1935, Schrödinger had the following monstrous idea. One puts a cat in a steel chamber with a diabolic device consisting of a single atom of radioactive chlorine 39 which decays with a mean life of 60 minutes into argon 39. This means that the probability for the atom to decay within an hour is 50%. If it decays, it emits an electron that is detected by a Geiger–Müller counter, which itself operates an apparatus, a hammer that falls on a capsule of cyanide, which kills the cat instantly. It's revolting.

Now, one observes the cat after an hour; the horrible questions are the following:

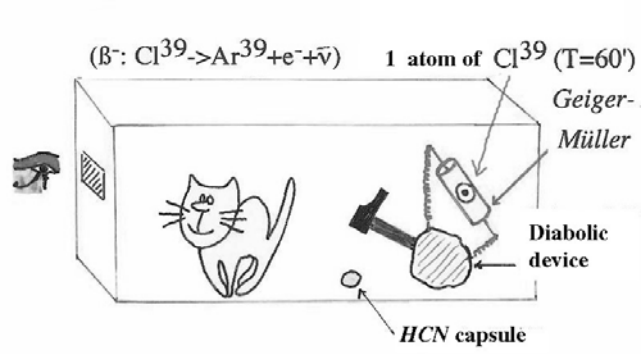


Fig. 4.5. Diabolic device imagined by Schrödinger.

1. What killed the cat?

2. When did the cat die? (Of course, if the cat is still alive, we redo the experiment an hour later, which is horrendous.)

The atom, more exactly its nucleus, is a quantum object. Its state is not described by a wave function of the type we have seen, but by another mathematical object with the same name ψ. The atom has a 50% probability of having decayed after one hour, therefore its quantum state after an hour is

$$\psi(\text{atom}) = \frac{1}{\sqrt{2}}(\psi(\text{nondecayed atom}) + \psi(\text{decayed atom})) \quad . \tag{4.27}$$

It is an equal weight quantum superposition. As long as one has not measured whether it has decayed, the atom is in a superposition of the two states, nondecayed atom and decayed atom.

But, because of the diabolic device, the entire system, including the cat, has become a quantum system. And, the cat is alive if the atom has not decayed; it is dead if the atom has decayed.

To describe a cat requires a very complicated wave function, with some 10^{27} variables. But that's inessential. The state of the cat can be inferred directly from the state of the atom. It is a linear superposition of the states {live cat} and {dead cat}, both at the same time, in interference

$$\psi(\text{cat}) = \frac{1}{\sqrt{2}}(\psi(\text{live cat}) + \psi(\text{dead cat})) \quad . \tag{4.28}$$

The cat is both alive and dead. It is an abominable state, quite inconceivable and very difficult to represent, as one can see in Figure 4.6!

So, coming back to our outrageous questions: what killed the cat? When did the cat die?

- Cyanide? No, it's simplistic; it was in a capsule that must be broken. Similarly, the hammer must be released.

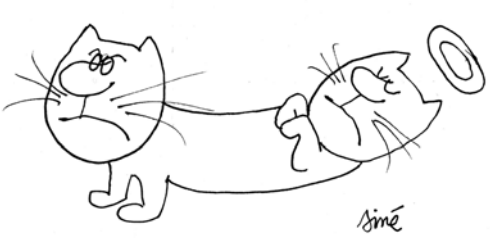


Fig. 4.6. Siné: Schrödinger's cat.

- The decay? Maybe, but the atom is a quantum system. Its state is a superposition; the superposition must cease; that is, there must be a wave packet reduction in order to know whether the atom has decayed. Then it will be possible to accuse the atom.
- Therefore, we must have an measuring instrument in order to know whether the atom has decayed.
- The horror is that the porthole and the cat form an instrument that enables us to see whether the atom has decayed.
- Therefore, one must observe the cat in order to reduce the wave packet of the atom!
- It is by observing the cat that we reduce the wave packet of the cat.
- It is when we observe the cat that we commit the atrocity of killing it, destroying any hope for the poor animal.

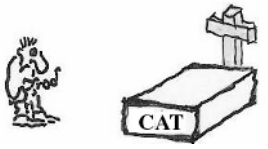


Fig. 4.7. Consequence of the observation of the cat.

- Before we looked at it, the cat was in a much more profitable state (4.28)!

Wigner had suggested that observation is a transcendental act of conscience that destroys the superposition. He thus addressed the question of the role of the observer in quantum physics. But one can reply that one can put a camera with shape recognition that types whether the cat is dead or alive and puts the answer in an envelope. One reads the letter a year after. If Wigner was right, the transcendental act of conscience would go backwards in time, because we exert it a year later.

Actually, in this disgusting example, what is shocking is not so much the wave packet reduction but the superposition. The quantum superposition of states, which seems natural when we apply it to objects deprived of souls, such as atoms or electrons, becomes very disconcerting if we apply it to familiar objects, as in Figure 4.8.

Fig. 4.8. The superposition principle applied to familiar objects. (Courtesy of Siné.)

4.5.2 The classical world

In order to settle the matter, I gave that as a written examination to my students. Other students are lucky to escape such torture. The answer is that it is quite conceivable to manufacture paradoxical states such as the superposition {live and dead cat}, that is, paradoxical macroscopic states.

But such states are extremely vulnerable and fragile. They imply a coherence, or a conspiracy of the 10^{27} particles, which destroys itself in an incredibly short time because of the interaction with the environment. This is called "decoherence" theory; it is quite fashionable these days.

Consequently, one must point out that the so-called "macroscopic world", the world of large objects, is not identical to the world of objects that follow classical physical laws. The classical world is the world of large objects or systems that are furthermore stable under quantum fluctuations in their interactions with the external world.

I tried to explain that to my own cat, but either he didn't understand or he just didn't give a damn about it, as one can see in Figure 4.9.

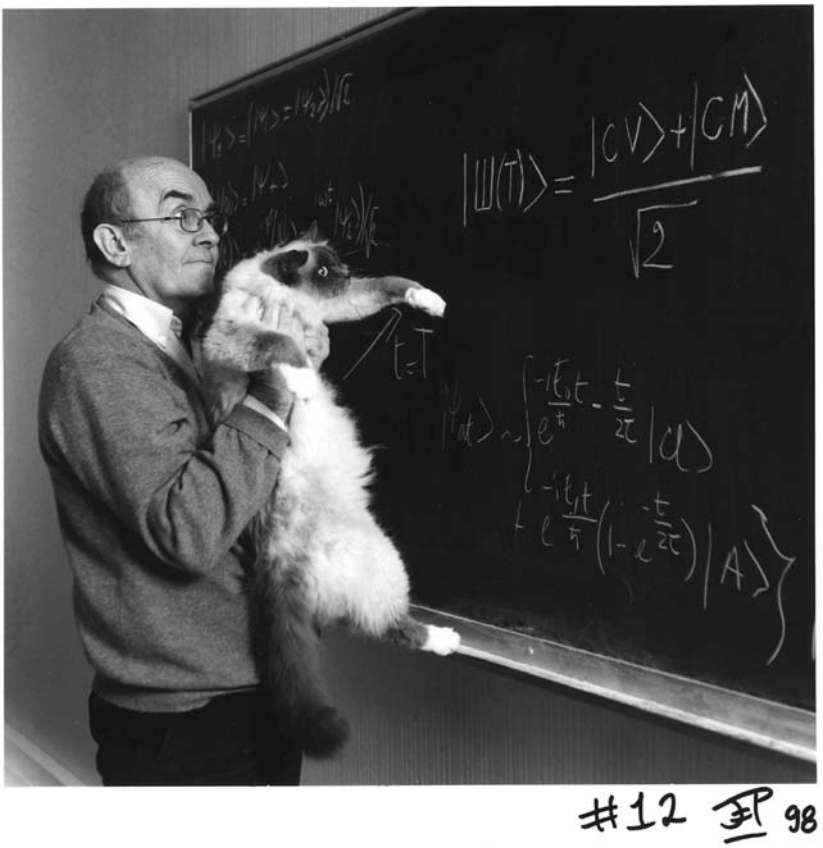


Fig. 4.9. Reaction of a contemporary cat to Schrödinger's perversity. (Photograph by Jacques Paultre.)

5

Energy quantization

We are now in a position to solve quantum mechanical problems in order to see how the theory works. We are interested in the motion of particles in simple potentials and in energy quantization.

We are going to do three things:

- First, explain the position of the problem.
- Next, see in two simple cases the origin of quantization of energy levels.
- And, finally, we study an example that is basically as simple as these two but much more subtle in its consequences. This is a model of the ammonia molecule NH_3, on which we really do quantum mechanics. We discover the tunnel effect, one of the most important quantum effects. By generalizing our results, this leads us to modern applications and nanotechnologies.

5.1 Methodology

The solution of such a problem consists of solving the Schrödinger equation. The Hamiltonian of a particle in a potential V is

$$\hat{H} = -\frac{\hbar^2}{2m}\Delta + V(\boldsymbol{r}) \quad . \tag{5.1}$$

We must impose boundary conditions on the wave function. One could, of course, choose a wave function at $t = 0$ and calculate its evolution. But the methodology is slightly more general. There are some common points with classical mechanics. First of all, energy is conserved; it is a constant of the motion. We therefore study solutions of given energy. We have seen that these are stationary states which have a very simple time-dependence

$$\psi_n(\boldsymbol{r}, t) = \phi_n(\boldsymbol{r})\, e^{-iE_n t/\hbar} \ . \tag{5.2}$$

The functions ϕ_n are the eigenfunctions of the Hamiltonian $\hat{H}$ and E_n are the corresponding eigenvalues

$$\hat{H}\phi_n(\boldsymbol{r}) = E_n\phi_n(\boldsymbol{r}) \quad . \tag{5.3}$$

5.1.1 Bound states and scattering states

In classical physics, one differentiates between two regimes in the motion of a particle in a potential according to the value of the energy.

If the energy is greater than the value of the potential at infinity, $E \geq V(\infty)$, the system is in a scattering state. The particle goes to infinity as t tends to infinity, and extracts itself from the field of force.

If the energy is smaller than the value of the potential at infinity, $E < V(\infty)$, we are dealing with a bound state. The particle is on an orbit. At any time its position remains confined in a finite region of space.

The same distinction exists in quantum mechanics, but going from one regime to the other is not as simple as in classical physics, both physically and technically. In classical physics, we are always interested in the trajectory. In quantum mechanics, the physical quantities of interest are not the same in the two cases.

The general problem of three-dimensional scattering, which is of great practical importance particularly in nuclear or particle physics, is technically complicated and we do not treat it in this book.

Here, we are interested only in bound states.

A bound state of well-defined energy is defined by the fact that its wave function satisfies the eigenvalue equation of the Hamiltonian (5.3) and that it is square integrable. In other words it is a "good" wave function

$$\int |\psi_n(\boldsymbol{r})|^2 \, d^3r = 1 \quad . \tag{5.4}$$

This condition is essential and corresponds to the fact that classically the particle is confined in a finite region of space.

Therefore, we are interested in finding the set $\{\phi_n, E_n\}$ of eigenfunctions and corresponding eigenvalues of the Hamiltonian, with this boundary condition (5.4).

In mathematics, one can prove that this is a discrete set. This is the origin of energy quantization: the E_n are the energy levels of the system. That is what Schrödinger knew.

The most general bound state is a linear superposition of stationary bound states. It evolves in a nontrivial way: a wave packet moves around in the potential. In the limit of large energy values, this motion becomes the classical motion.

If we relax the normalization condition (5.4), there exist solutions of the equation

$$\hat{H}\psi(\boldsymbol{r}) = E\psi(\boldsymbol{r}) \quad , \tag{5.5}$$

for a continuous set of values of the energy E. Such solutions must be properly interpreted, and they correspond to scattering states. Their linear superpositions are wave packets that move in space.

5.1.2 One-dimensional problems

Here, we are concerned with simple one-dimensional problems. For instance, a marble on a rail. It is much simpler technically, and we treat three-dimensional problems later. The eigenvalue equation is then an ordinary second-order differential equation

$$-\frac{\hbar^2}{2m}\psi_n''(x) + V(x)\psi_n(x) = E_n\psi_n(x) \quad . \tag{5.6}$$

We wish to determine the functions $\{\psi_n(x)\}$ and the numbers $\{E_n\}$ (we use the symbol ψ because there is no ambiguity) with the normalization condition

$$\int |\psi_n(x)|^2\, dx = 1 \quad . \tag{5.7}$$

In addition, we consider simple potentials for which analytical solutions exist, in order to become familiar with the physics.

5.2 The harmonic oscillator

5.2.1 Harmonic potential

A first example is the harmonic potential

$$V(x) = \frac{1}{2}m\omega^2 x^2 \quad . \tag{5.8}$$

This corresponds classically to the sinusoidal motion, of frequency ω, of a particle elastically bound to a center (here $x_0 = 0$). It is called a harmonic oscillator. It only has bound states ($V(\infty) = \infty$).

The eigenvalue equation is

$$-\frac{\hbar^2}{2m}\,\frac{d^2\psi(x)}{dx^2} + \frac{1}{2}m\omega^2 x^2\,\psi(x) = E\,\psi(x) \quad . \tag{5.9}$$

The harmonic oscillator has numerous applications. For motions of small amplitude around its equilibrium position, any system has harmonic oscillations.

Eigenvalues, eigenfunctions

This equation is a classic of 19th century mathematics. We turn to dimensionless quantities

$$\varepsilon = \frac{E}{\hbar\omega},\ y = \frac{x}{\sqrt{a}}, \quad \text{with } a = \frac{\hbar}{m\omega} \quad , \tag{5.10}$$

and the eigenvalue equation becomes

$$\frac{1}{2}\left(y^2 - \frac{d^2}{dy^2}\right)\phi(y) = \varepsilon\,\phi(y) \quad . \tag{5.11}$$

The square integrable solutions $\psi(x) = \sqrt{a}\ \phi(x/\sqrt{a})$, were found by Charles Hermite:

$$\phi_n(y) \ = c_n\ e^{y^2/2}\ \frac{d^n}{dy^n}\left(e^{-y^2}\right) \ = c_n\ e^{-y^2/2}\ H_n(y) \quad , \tag{5.12}$$

where $H_n(y)$ is a polynomial of degree n, called a Hermite polynomial. It contains only even (resp., odd) powers of y if n is even (resp., odd), and the normalization constant is $c_n = \pi^{-1/4}\,2^{-n/2}\,(n!)^{-1/2}$. The corresponding eigenvalues are

$$\varepsilon_n = n + \frac{1}{2},\ n \text{ nonnegative integer} \ . \tag{5.13}$$

The differential equation (5.11) has solutions for all positive values of ε. However, in general, these solutions increase at infinity as $\simeq e^{+y^2/2}$. It is only for the set of values (5.13) of ε that the solutions are square integrable.

5.2.2 Energy levels, eigenfunctions

The energy levels of the one-dimension harmonic oscillator are therefore equal-spaced

$$E_n = (n + \frac{1}{2})\,\hbar\omega = (n + \frac{1}{2})\,h\nu \quad . \tag{5.14}$$

Planck had correctly guessed the term $nh\nu$. Energy exchanges of the oscillator in transitions between levels occur for integer multiples of $\hbar\omega = h\nu$,

$$\Delta E_{nn'} = (n - n')h\nu \quad .$$

But Planck had no means (and no reason) to guess the presence of the constant $\hbar\omega/2$, which is called the *zero point energy*, and which can be measured. This constant is essential in order to satisfy uncertainty relations. In fact the classical equilibrium state of the oscillator consists of being at rest at the minimum of the potential, in other words, when its velocity is zero and its position is at the minimum of the potential. This would be contrary to Heisenberg's uncertainty relations. The zero-point energy is inevitable and measurable.[1]

This result, which can be extended easily to three dimensions, has numerous applications. It allows us to understand the vibration spectra of molecules, and the specific heats of solids (molecules in a crystal vibrate around their equilibrium position). It is a basic tool in field quantization and in relativistic quantum physics.

[1] This energy may manifest itself in what is called the "vacuum energy" of the Universe in cosmology.

The eigenfunctions are real and orthogonal

$$\int \psi_n^*(x)\,\psi_{n'}(x)\,dx = \delta_{n,n'} \quad . \tag{5.15}$$

From the definition of Hermite functions (5.12), one obtains the action of the operators x and d/dx on $\psi_n(x)$:

$$\sqrt{\frac{2}{a}}\; x\;\psi_n(x) = \sqrt{n+1}\;\psi_{n+1}(x) + \sqrt{n}\;\psi_{n-1}(x) \tag{5.16}$$

$$\sqrt{2a}\;\frac{d}{dx}\psi_n(x) = \sqrt{n}\;\psi_{n-1}(x) - \sqrt{n+1}\;\psi_{n+1}(x) \quad . \tag{5.17}$$

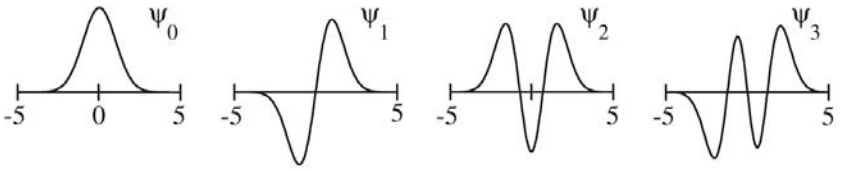


Fig. 5.1. The first four Hermite functions (abscissa: $xa^{1/2}$); $\psi_0(x)$ is a Gaussian; $\psi_1(x)$ is this Gaussian multiplied by $x(2a)^{1/2}$, and so on.

5.3 Square well potentials

5.3.1 Square potentials

Even simpler models consist of what are known as "square potentials", that is, potentials which are piecewise constant.

The solutions are, piecewise, exponential or sinusoidal according to the sign of $V - E$ (E is a number we want to determine).

1. In regions where $V = V_0$ and $E - V_0 > 0$, the wave functions have a sinusoidal behavior $\psi(x) \propto e^{\pm ikx}$ with $k = \sqrt{2m(E - V_0)/\hbar^2}$.
2. When $E - V_0 < 0$, the wave functions have an exponential behavior $\psi(x) \propto e^{\pm Kx}$ with $K = \sqrt{2m(V_0 - E)/\hbar^2}$.

In order to obtain the bound states of such systems, we require as usual that the wave functions be square integrable. But in addition, the wave functions must be continuous as well as their first derivative at the points of discontinuity of the potential. This prescription is a consequence of the Schrödinger equation. It can be proven easily in distribution theory. One can understand

this by considering a step in the potential as the limit of a continuous function (for instance, $V \sim \lim_{\lambda \to 0} \tanh(x/\lambda)$) and by reductio ad absurdum.

This procedure allows us to determine the values of the energy E, as we show a simple case.

In modern microelectronic technologies, such potentials have many applications.

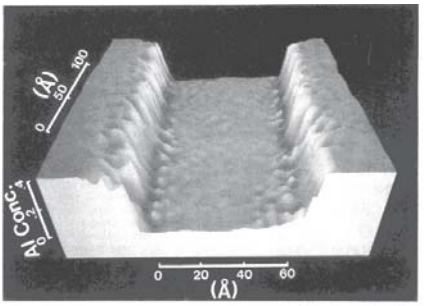


Fig. 5.2. Sandwich of AlGaAs–GaAs–AlGaAs. The central slice of GaAs has a width of 6 nm. On the vertical axis, the aluminum concentration is indicated. The shape corresponds to the variation of the potential as "seen" by a conduction electron (i.e., the electrostatic potential averaged over one period of the crystal lattice). (Photograph due to Abbas Ourmazd, ATT Bell Labs.)

In "sandwiches" of alternating thin layers of semiconductors (Ga As and Ga Al As), shown in Figure 5.2, one can manufacture 20 to 50 Å wide quantum wells (we showed an example of this in the first chapter). The quantum confinement of electrons in such domains has opened a new era in electronics and computer components.[2] These are also used in optoelectronics, because the corresponding transitions between electron levels ($\Delta E \sim 50$ to 200 meV) are in the infrared part of the spectrum.

At the end of this chapter, we show how one can obtain such a picture with quantum tunneling microscopes.

What is important in such models is to understand the physics, that is, the qualitative results. Exact calculations can then be given safely to computers.

5.3.2 Symmetric square well

Energy quantization is a theorem, but let's see how it occurs. We consider the energy levels of a particle in a symmetric potential well of depth V_0 and width $2a$, centered at $x = 0$.

[2] M.A. Reed, Quantum boxes, *Sci. Amer.*, Jan. 93 ; L.L. Chanz and L. Esaki, Semiconductor Quantum Heterostructures, *Phys. Today*, **45**, p. 36 (1992).

We choose the origin of energies at the bottom of the potential well, so that the energy E, which we want to determine, is the kinetic energy of the particle inside the well.

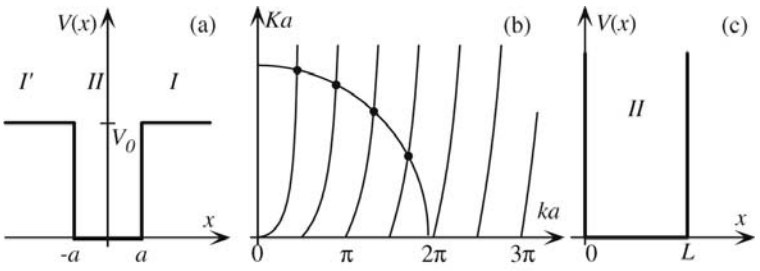


Fig. 5.3. Square well potential: (a) shape of the potential; (b) graphical solution giving the energy levels; (c) limit of an infinitely deep well.

We are only interested in bound states, namely states for which $0 \leq E < V_0$ which, in classical physics correspond to a particle that is confined inside the well. We assume the energy is not sufficient for the particle to jump out of the well. We set

$$k = \sqrt{\frac{2mE}{\hbar^2}}, \text{ and } K = \sqrt{\frac{2m(V_0 - E)}{\hbar^2}} \quad , \tag{5.18}$$

and we have

$$k^2 + K^2 = \frac{2mV_0}{\hbar^2} \quad . \tag{5.19}$$

It is straightforward to solve the Schrödinger equation. The wave functions are exponentials on the left and on the right (regions I' and I) and sinusoids in the middle (region II).

An important simplification comes from the symmetry of the problem $V(x) = V(-x)$. One can indeed classify the solutions in two categories: symmetric (or even) solutions, and antisymmetric (or odd) solutions. In fact, because the Hamiltonian is symmetric $\hat{H}(x) = \hat{H}(-x)$, if we change x into $-x$, we obtain for any solution $\psi(x)$

$$\hat{H}(x)\psi(x) = E\psi(x), \quad \text{and} \quad \hat{H}(x)\psi(-x) = E\psi(-x) \quad .$$

In other words, if $\psi(x)$ is a solution of the Schrödinger equation, then $\psi(-x)$ is also a solution for the same eigenvalue of the energy. Therefore, $\psi(x) \pm \psi(-x)$ is either a solution for the same value of the energy, or it is identically zero. This is a particular case of a very important property in quantum mechanics. To the invariance laws of the Hamiltonian, there correspond symmetry properties of the solutions. (This can also be obtained directly here by performing

the calculation. We could have made the same remark on the harmonic oscillator.)

The symmetric (ψ_S) and antisymmetric (ψ_A) solutions have the form

$$\begin{aligned}
&\psi_S : \ (I') \quad \psi(x) = B\, e^{Kx} \quad , (II)\ \psi(x) = A\cos kx \ , \ (I)\ \psi(x) = B\, e^{-Kx} \ , \\
&\psi_A : \ (I') \quad \psi(x) = -D\, e^{Kx} \ , \ (II)\ \psi(x) = C\sin kx \ , \ (I)\ \psi(x) = D\, e^{-Kx} \ ,
\end{aligned} \tag{5.20}$$

where the constants A, B, C, and D are determined by the continuity of ψ and ψ' at $x = \pm a$. Notice that in (5.20) we have omitted terms that increase exponentially at infinity because of the normalizability condition.

We must join these expressions at $x = \pm a$. When V is continuous, the equation handles the problem by itself. If not, as is the case here, we impose, as we have said, that ψ and ψ' are continuous at $x = \pm a$.

We therefore obtain

$$\begin{aligned}
&\psi_S : \ A\cos ka = Be^{-Ka} \ , \text{ and } \quad kA\sin ka = KCe^{-Ka} \\
&\psi_A : \ C\sin ka = De^{-Ka} \ , \text{ and } \quad kC\cos ka = -KDe^{-Ka} \ ,
\end{aligned} \tag{5.21}$$

and, if we take the ratios,

$$k\tan ka = K \ , \text{ for } (\psi_S), \quad \text{and } \ -k\cot ka = K \ , \text{ for } (\psi_A) \quad . \tag{5.22}$$

These two relations between k and K must be completed by the definition (5.19) which can be written as

$$k^2a^2 + K^2a^2 = \frac{2mV_0a^2}{\hbar^2} \quad . \tag{5.23}$$

In the (ka, Ka) plane, this is the equation of a circle. We must therefore find the intersections of this circle with the curves $Ka = ka\tan ka$ and $Ka = -ka\cot ka$ as represented in Figure 5.3b. Suppose the width $2a$ of the well is given, these intersections k_n form a finite set, and they correspond alternatively to even and odd solutions. The number of solutions, that is, bound states, increases with V_0 (deeper well). There is only one bound state if V_0 is less than

$$\frac{a\sqrt{2mV_0}}{\hbar} < \frac{\pi}{2} \quad \text{or} \quad V_0 < \frac{\pi^2\hbar^2}{8ma^2} \quad . \tag{5.24}$$

The energy levels $E_n = \hbar^2k_n^2/2m$ are quantized. This quantization is not a consequence of the continuity conditions, which simply enable us to calculate the eigenvalues E_n, but of the normalizability that eliminates exponentially increasing terms in (5.20).

The solutions can be classified by increasing values of the energy E, according to the number of nodes of the wave functions (Sturm–Liouville theorem). The lowest energy state is called the ground state.

Notice an essential difference from classical mechanics: the particle has a nonvanishing probability to be in the classically forbidden regions, (I) and (I'), where its kinetic energy would be negative and is therefore energetically illegal.

However, it does not propagate in those regions; it can penetrate them, but it bounces off them. The wave function decreases exponentially with a mean penetration distance of $1/K$. This is analogous to the skin effect in electromagnetism. We can see how the "classical limit" appears: $1/K \to 0$, if we make $\hbar \to 0$ or if the mass is large $m \to \infty$.

As we said, quantum wells such as the one represented in Figure 5.2 have allowed decisive improvements in advanced technologies, in particular infrared technologies such as shown in Figure 5.4.

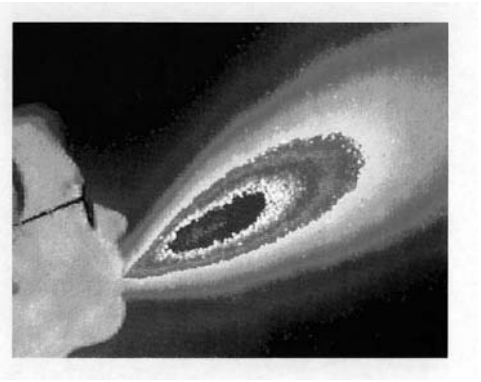


Fig. 5.4. Temperature distribution in a breath (the original picture is in artificial colors). The background is black, the breath comes out of the mouth at body temperature and it expands at room temperature. (Courtesy of Emmanuel Rosencher.)

5.3.3 Infinite well, particle in a box

Another simple, but interesting, limiting case is that of an infinitely deep well $V_0 = \infty$. The particle is confined inside a region that we place for convenience between $x = 0$ and $x = L$ as in Figure 5.3c.

In this limit of the previous case, the wave function vanishes, $\psi(x) = 0$, outside the interval $[0, L]$. The continuity conditions are then different: only the wave function is continuous (this can easily be seen by taking the limit $V_0 \to \infty$ on the above solutions).

The normalized eigenfunctions of the Hamiltonian are then

$$\psi_n(x) = \sqrt{\frac{2}{L}}\, \sin(\frac{n\pi x}{L}) \quad \text{n integer} > 0 \quad . \tag{5.25}$$

The corresponding energy levels are

$$E_n = n^2 \frac{\pi^2 \hbar^2}{2mL^2} \quad \text{n integer} > 0 \quad . \tag{5.26}$$

Quantization appears here as a very simple stationary wave phenomenon.

This calculation can easily be generalized to three dimensions, that is, the case of a particle in a box. If we consider a rectangular box of sides (a, b, c), the solutions can be factorized:

$$\Psi_{n_1,n_2,n_3}(\boldsymbol{r}) = \frac{\sqrt{8}}{\sqrt{abc}} \sin(\frac{n_1 \pi x}{a}) \sin(\frac{n_2 \pi y}{b}) \sin(\frac{n_3 \pi z}{c}) \, , \tag{5.27}$$

$$E = E_{n_1,n_2,n_3} = \frac{\hbar^2 \pi^2}{2m} \left(\frac{n_1^2}{a^2} + \frac{n_2^2}{b^2} + \frac{n_3^2}{c^2} \right) \, . \tag{5.28}$$

This very simple result has numerous applications. Many systems can be approximated by infinite wells: molecules in a gas, neutrons inside a neutron star, conduction electrons in a metal, and others.

5.4 Double well, the ammonia molecule

Consider now a similar problem, one on which we really do quantum mechanics and discover unexpected results. This is the case of a symmetric double potential well. At the beginning it is similar to the infinite potential well, but with a potential barrier in the middle: in other words, an infinite well containing two wells of width a centered respectively at $\pm b$, and separated by a barrier of height V_0 and width $\Delta = 2b - a$.

5.4.1 The model

Consider a concrete example, the ammonia molecule NH_3.

In its lowest energy states, this molecule has the shape of a pyramid with the nitrogen atom at the top and the three hydrogen atoms at the base, on an equilateral triangle. It is a very complex object made of 14 particles (10 electrons and 4 nuclei) and there are many possible motions of this system. However, the lowest energy motions correspond to the global displacement of the planar triangle of hydrogen atoms, which we call collectively a "particle of mass m," with respect to the nitrogen atom along the symmetry axis of the molecule.

When the abscissa x of this plane varies along $x > 0$, the potential energy of the system has a minimum that corresponds to a classical equilibrium configuration. However, the molecule can invert itself, as an umbrella, and there exists another symmetric stable configuration for a negative value of x. The potential energy is symmetric with two minima and a maximum in

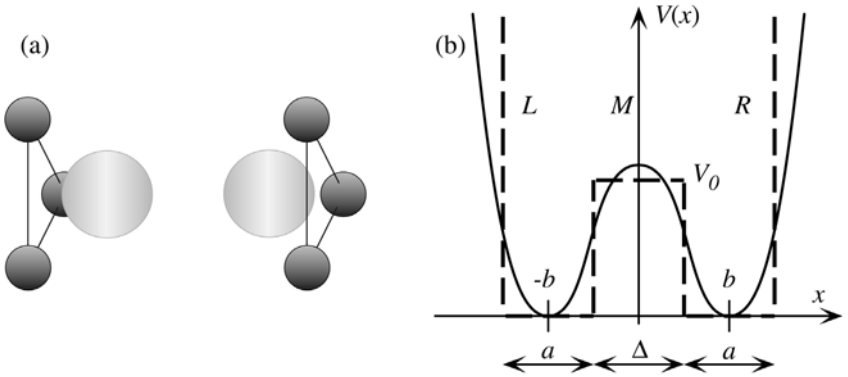


Fig. 5.5. The ammonia molecule: (a) the two classical configurations; (b) the actual molecular potential energy (full line) and the simplified model (dashed line) that describes the reversal of the molecule.

between. This maximum corresponds to an unstable configuration where the four atoms are in the same plane.

These two configurations of classical equilibrium are not physically equivalent because the molecule possesses an intrinsic angular momentum, and that one can define the "right" and the "left," as in an umbrella.

This model can be solved numerically, but it suffices to study the square well potential $V(x)$ drawn in a dashed line in Figure 5.5b. This potential consists of two wells of width a centered at b and $-b$, respectively, and separated by a barrier of height V_0 and width $\Delta = 2b - a$.

We ask the following question. For $E > V_0$ there are periodic oscillations from left to right. But when $E < V_0$, for a given value of the energy E, there are always two possible classical configurations of same energy, one in the left-hand side well, the other in the right-hand side well.

In particular, there are two positions of equilibrium, one on the left, the other on the right; both have the same energy.

What is the quantum situation?

5.4.2 Stationary states, the tunnel effect

From the calculational point of view, the problem is strictly analogous to what we just did. We consider the energy levels of a particle of mass m such that classically the particle cannot cross the potential barrier: $E < V_0$.

As previously, we define

$$k = \sqrt{\frac{2mE}{\hbar^2}} \quad \text{and} \quad K = \sqrt{\frac{2m(V_0 - E)}{\hbar^2}} \quad . \tag{5.29}$$

The problem has the symmetry ($x \leftrightarrow -x$). We can classify the solutions according to their parity or symmetry.

The solutions are sinusoidal in the regions L and R and exponential in the middle region M. The wave functions must vanish for $x = \pm(b + a/2)$, and the eigenfunctions of the Hamiltonian are of the form

$$\begin{aligned}
\psi(x) &= \pm\lambda \sin k(b + a/2 + x) && \text{L region} \\
&= \lambda \sin k(b + a/2 - x) && \text{R region} \\
\psi(x) &= \mu \ \cosh Kx \quad \text{Symmetric solution} && \text{M region} \\
\psi(x) &= \mu \ \sinh Kx \quad \text{Antisymmetric solution} && (5.30)
\end{aligned}$$

The two lowest energy solutions are represented in Figure 5.6. The ground state is symmetric; the first excited state is antisymmetric.

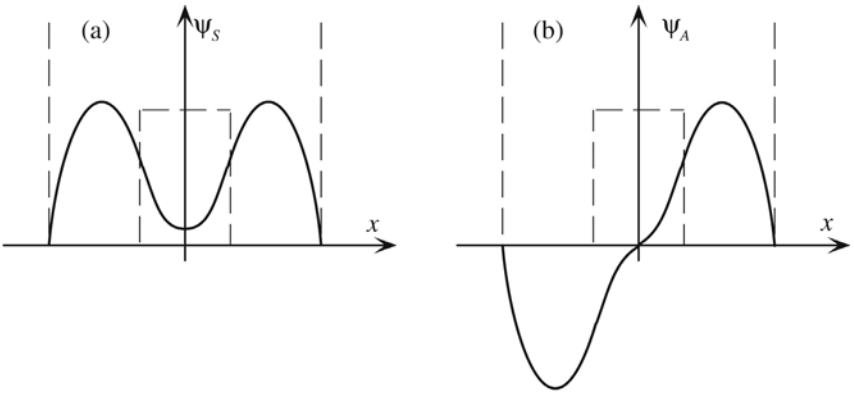


Fig. 5.6. (a) Symmetric solution and (b) antisymmetric solution of lowest energy in the double square well potential which is a model of the ammonia molecule.

We observe that the wave functions exist in the classically illegal middle region $E < V_0$, which is not surprising. It comes from the exponential of the simple symmetric well which is cut off at a finite distance. This results in a symmetrization or antisymmetrization of the wave functions.

Therefore, in all stationary states, the particle has the same probability of being on the right and on the left.

So it is a two-well problem, but, because the particle has a non-vanishing probability to be in the classically forbidden middle region, these two wells are coupled by the *tunnel effect*. Things resemble a classical situation where one would have dug a quantum tunnel in the intermediate potential barrier, in order to allow the particle to communicate between the two wells.

5.4.3 Energy levels

The continuity of the function and of its derivative at points $x = \pm(b - a/2)$ yields the conditions:

$$\tan ka = -\frac{k}{K}\coth K(b-a/2) \quad \text{for a symmetric solution } \psi_S \ ,$$
$$\tan ka = -\frac{k}{K}\tanh K(b-a/2) \quad \text{for an antisymmetric solution } \psi_A \ .$$

This, together with the condition $k^2 + K^2 = 2mV_0/\hbar^2$ gives transcendental equations that can be solved numerically.

However, in order to understand the physics of the problem in a simple manner, we assume the orders of magnitude are such that: $V_0 \gg E$, that is, $K \simeq \sqrt{2mV_0/\hbar^2}$ = constant and $K\Delta \gg 1$, which is quite reasonable in the specific case of the NH_3 molecule. Under such conditions, we end up with

$$\tan ka \simeq -\frac{k}{K}\left(1 \pm 2e^{-K\Delta}\right) \quad , \tag{5.31}$$

where the + sign corresponds to ψ_S, and the − sign to ψ_A. With this equation, we can calculate the quantized values of ka. These values appear on the graph in Figure 5.7 as the positions of the intersections of the successive branches of $y = \tan ka$ with the two straight lines $y = -\varepsilon_A ka$ and $y = -\varepsilon_S ka$. These intersections are located in the vicinity of $ka \sim \pi$. The two constants ε_A and ε_S are

$$\varepsilon_A = \frac{1}{Ka}\left(1 - 2e^{-K\Delta}\right) \ , \quad \varepsilon_S = \frac{1}{Ka}\left(1 + 2e^{-K\Delta}\right) \ . \tag{5.32}$$

They are close to each other, and such that $\varepsilon_A < \varepsilon_S \ll 1$, because $Ka \gg ka \sim \pi$.

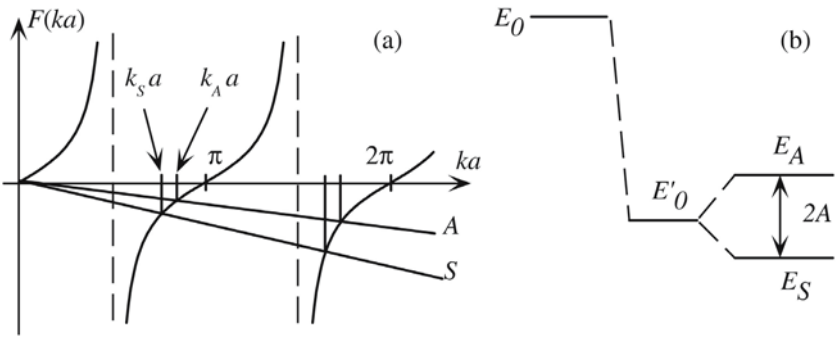


Fig. 5.7. (a) Graphical determination of the energy levels in the double well; (b) the two first levels are lower than the ground state of a single well centered at L or R ($E_0 \to E'_0$), and there is a splitting by tunneling between these two levels ($E'_0 \to E_A$ and E_S).

For K infinite, V_0 infinite, we recover two independent wells. The particle is in one of them. The energies are $E_n = n^2\pi^2\hbar^2/2ma^2$. The order of magnitude of E_S and E_A is $E_0 = \pi^2\hbar^2/2ma^2$.

For finite K (V_0 finite), the two levels are shifted downwards to E_0', which is intuitively understandable. The two wells communicate by quantum tunneling and the particle "sees" an effective well broader than a. This lowering is accompanied by a splitting in two sublevels as can be seen in Figure 5.7. The symmetric state is more tightly bound.

In our approximation ($K \gg k$, $Ka \gg 1$), we obtain

$$k_S \sim \frac{\pi}{a(1+\varepsilon_S)}, \quad k_A \sim \frac{\pi}{a(1+\varepsilon_A)}, \tag{5.33}$$

with ε_S and $\varepsilon_A \ll 1$. Using (5.31) and (5.33), we obtain a mean energy $E_0' = (E_A + E_S)/2$

$$E_0' \simeq \frac{\hbar^2\pi^2}{2ma^2}\left(1 - \frac{2}{Ka}\right). \tag{5.34}$$

The splitting $E_A - E_S$ of the two levels by the tunnel effect is

$$E_A - E_S \equiv 2A \simeq \frac{\hbar^2\pi^2}{2ma^2}\left[\frac{1}{(1+\varepsilon_A)^2} - \frac{1}{(1+\varepsilon_S)^2}\right] \simeq \frac{\hbar^2\pi^2}{2ma^2}\,\frac{8e^{-K\Delta}}{Ka}, \tag{5.35}$$

or

$$A \equiv \frac{E_A - E_S}{2} \simeq \frac{\hbar^2\pi^2}{2ma^2}\,\frac{4}{Ka}e^{-K\Delta} \quad . \tag{5.36}$$

Because $K \simeq \sqrt{2mV_0}/\hbar$ in this approximation, the splitting tends to zero exponentially when the width Δ, or the height V_0, of the potential barrier increases.

In fact, in all its consequences, namely the splitting $E_A - E_S$ and the value of the wave function in the middle region, the tunnel effect is proportional to $e^{-K\Delta}$ where $\Delta = 2b - a$ and $K \sim \sqrt{2}m(V_0)/\hbar^2$ (or the square of this for probabilities).

We also notice that $A \to 0$ extremely rapidly in the limit $\hbar \to 0$. This is in contrast with the polynomial level spacing, in $\hbar$, of the harmonic oscillator or potential well levels. Here, the splitting is exponential in $-1/\hbar$.

We therefore observe a first difference with the classical case. There are indeed two lowest lying energy states, but classically they have the same energy (E_0), whereas here they are split. This qualitative feature is general, for any symmetric double well of arbitrary shape.

5.4.4 Wave functions

The two corresponding wave functions, which are shown in Figure 5.6, are such that the probability of finding the particle on the left or on the right is the same. If the particle is in a well-defined energy state, it has the same probability of being on either side. It is both on the right and on the left.

That result is really contrary to classical observations. Classically, in its two lowest energy states, the particle is either on the right or on the left. Here it is both on the right and on the left at the same time.

Is that a real difference? That's not really convincing for the moment, because of the statistical interpretation. If, classically, we fix the condition that the energy be minimum, then it is natural to find half of the particles on the right and the other half on the left. We must find something else.

Something else means doing true quantum physics, taking into account the time parameter.

5.4.5 Inversion of the molecule

How can we put the particle on the right- or left-hand sides in quantum mechanics? We must prepare it in a state where its wave function is concentrated on the right or on the left. But we know how to do that! We just need to look at the wave functions in Figure 5.6. We just have to take the sum and the difference of these two eigenfunctions, according to the superposition principle,

$$\psi_R = (\psi_S + \psi_A)/\sqrt{2} \quad \text{and} \quad \psi_L = (\psi_S - \psi_A)/\sqrt{2} \quad . \tag{5.37}$$

The resulting wave functions describe states where almost all the probability is concentrated on one side only, on the left for ψ_L and on the right for ψ_R. (Actually the residual probability is of the order of $\approx e^{-2K\Delta}$, which is a very small quantity.) These two states correspond to the "classical" configurations, molecule on the right and molecule on the left (Figure 5.8).

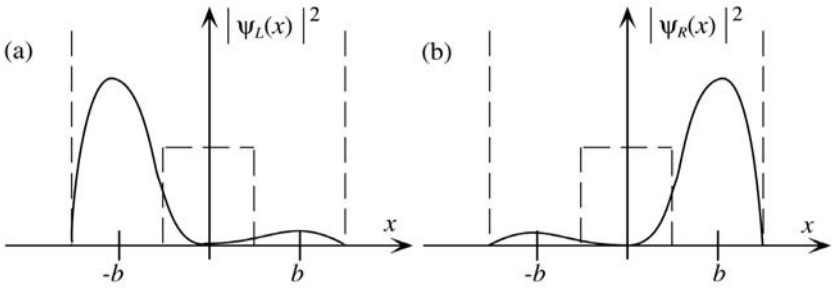


Fig. 5.8. Classical configurations of the ammonia molecule.

This observation is interesting. We can phrase it in the following way. ψ_R is a linear superposition of states with a well-defined energy that interferes destructively on the left. Similarly, ψ_L is a linear superposition of states with a well-defined energy that interferes destructively on the right. In this way, interferences are really simple; there is nothing more from the mathematical point of view.

We can invert these relations and say that states of well-defined energies, which are on both sides, are linear interfering superpositions of "contradictory" classical states. Classically, one is either on the right or on the left.

Nobody is both on the right and on the left: well nearly nobody, some people once in a while, but very few.

Here, it seems perfectly natural. But that is what seemed shocking with the cat! What was inconceivable for a cat seems perfectly natural here. Notice that we had forgotten a possible state $\psi_-(\text{cat}) = (\psi(\text{live cat}) - \psi(\text{dead cat}))\sqrt{2}$. Let's be honest, the second elementary operation, subtraction, is just as noble as the first one.

Now we can apply what we know on time evolution. The state ψ_R is not a stationary state; the system does not have a well-defined energy: $p(Es) = p(Ea) = 1/2$. Therefore, it must evolve in a non-trivial way with time. Its time evolution is

$$\begin{aligned}\psi(x,t) &= \frac{1}{\sqrt{2}}\left(\psi_S(x)\, e^{-iE_S t/\hbar} + \psi_A(x)\, e^{-iE_A t/\hbar}\right) \\ &= \frac{e^{-iE_S t/\hbar}}{\sqrt{2}}\left(\psi_S(x) + \psi_A(x) e^{-i\omega t}\right)\ , \end{aligned} \tag{5.38}$$

where we have introduced the *Bohr frequency* $\hbar\omega = E_A - E_S = 2A$ of the system.

Now, we are facing a truly astonishing and unpredictable thing. At time $T = \pi/\omega$ the particle is on the left! Its wave function is ψ_L, up to a phase factor,

$$\psi(x,T) = \frac{e^{-iE_S T/\hbar}}{\sqrt{2}}\left(\psi_S(x) - \psi_A(x)\right).$$

This phenomenon is really contrary to classical mechanics where, if at $t = 0$ the particle is at rest in one of the wells, it stays there!

In quantum mechanics, there is a permanent oscillation between the two wells at the frequency $\omega = 2A/\hbar$. The wave function and the probability density "flow" periodically from one well to the other. The particle keeps on shifting from one side to the other. That is truly non-classical.

In the specific case of NH_3, this phenomenon is called the "inversion of the ammonia molecule." If, at $t = 0$, we prepare it in a classical configuration, it reverses itself periodically at the Bohr frequency.

The inversion frequency can be measured very accurately. In the lowest energy state, the splitting is $2A \approx 10^{-4}$ eV, hence a frequency $\nu = 24$ GHz, a wavelength $\lambda = 1.25$ cm, and a period $\tau = 4.2\,10^{-11}$ s.

In fact, NH_3 possesses an electric dipole moment. The center of gravity of positive charges is different from the center of gravity of negative charges (the nitrogen atom attracts electrons more strongly). This electric dipole moment $\boldsymbol{D}$ is inverted when the molecule reverts. This produces the emission or absorption of a radio wave at that frequency. (Actually, at the microscopic level, a molecule can absorb or emit a photon of energy $h\nu \approx 10^{-4}$ eV under a transition between the two states. It is only on a macroscopic sample that the electric dipole argument is valid.) That is a characteristic line of NH_3, a fingerprint of that molecule which is used in radioastronomy to detect ammonia

in the interstellar medium. We come back later to an important application, the ammonia maser.

It was totally impossible to predict the order of magnitude of that frequency qualitatively, by dimensional analysis, before we understood the mechanism of the tunnel effect and before we found the exponential (5.36).

5.5 Illustrations and applications of the tunnel effect

This brings us to a series of concluding remarks on the tunnel effect that is one of the most important phenomena of quantum mechanics.

5.5.1 Sensitivity to the parameters

The exponential dependence of the splitting is fundamental. An exponential varies very rapidly. This same effect, this same mechanism explains phenomena whose orders of magnitude are incredibly different.

An interesting case is NH_3 and similar molecules ND_3, PH_3, AsH_3. This is treated in detail by Townes and Schawlow[3] who give the form of realistic potentials in this kind of physics. It is instructive to go from NH_3 to AsH_3:

$$\begin{aligned}
NH_3 &: V_0 = 0,25 \text{ eV} , \ b = 0,4 \text{ Å} : \quad \nu_0 = 2,4\ 10^{10} \text{ Hz} , \\
ND_3 &: V_0 = 0,25 \quad m_d = 2m_p , \ b = 0,4 \text{ Å} : \quad \nu_0 = 1\,600 \text{ Hz} , \\
PH_3 &: V_0 = 0,75 \text{ eV} , \ b = 1 \text{ Å} : \quad \nu_0 = 2,4\ 10^{10} \text{ Hz} , \\
AsH_3 &: V_0 = 1,50 \text{ eV} , \ b = 2 \text{ Å} : \quad \nu_0 = 1,6\ 10^{-8} \text{ Hz} .
\end{aligned}$$

A change by a factor of 6 in V_0 and 5 in the size b produces a spectacular decrease of 18 orders of magnitude in the inversion frequency between NH_3 and AsH_3. For AsH_3, the frequency is one inversion in two years, which is not measurable. One can only calculate it theoretically. In other words, AsH_3, which seems quite similar to NH_3 from a chemical point of view, behaves as a classical structure from the point of view we have studied here, simply because the arsenic atom is 5 times larger than nitrogen.

There are a variety of similar situations where, by quantum tunneling, a process occurs through a nonclassical transition across a potential barrier. The first physical phenomenon that was explained in this way was the alpha decay of nuclei. This was understood by Gamow in 1928. Many other phenomena have the same origin, such as catalysis, the formation of interstellar molecules on interstellar dust, nuclear fusion and fission, and so on.

[3] C.H. Townes and A.L. Schawlow, *Microwave Spectroscopy*, Chapter 12. New York: McGraw-Hill (1955).

5.5.2 Molecular structure

Valence electrons

Similar examples are provided by electrons in molecules. One can develop an analogous formalism for a variety of cases and extend the previous results to different double-well situations. Consider, for instance, the case of two identical atoms located at a distance Δ of each other. An external electron "sees" a double well as shown in Figure 5.9 (for simplicity, we assume the atoms are at a given distance that we can vary). We choose the origin of energies such that $V \to 0$ for $x \to \infty$. If Δ is sufficiently large, one can safely consider that $V \sim 0$ halfway between the atoms.

In order to get from one atom to the other, an electron in an energy level $E_0 < 0$ must cross a potential barrier of height $-E_0$ and width Δ. We can calculate the typical time T it takes to get from one atom to the other.

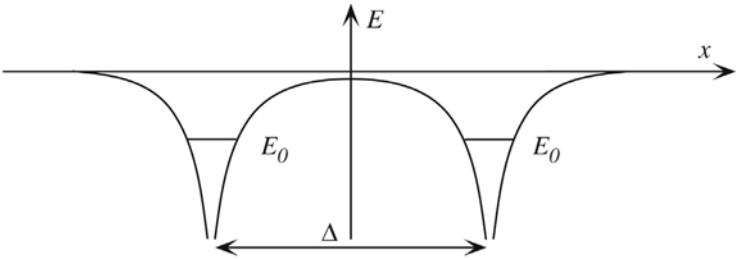


Fig. 5.9. Double potential well as seen by an electron when two atoms are separated by a distance Δ.

Suppose the kinetic energy E_k of the electron in one well is of the order of $|E_0|$ (for the hydrogen atom, this is exact, owing to the virial theorem). In the exponential of the tunnel effect, we have $K = \sqrt{2m|E_0|}/\hbar$. For an electron bound in an atom, one has in good approximation $Ka \sim 1$. The exponential dependence of the oscillation frequency in terms of the parameter $K\Delta$ remains true. We can use (5.36) in the form $A \sim E_k e^{-K\Delta} \sim |E_0| e^{-K\Delta}$, up to a numerical factor of order one.

In a molecule or in a crystal, the distances of atoms are of the order of 0.1 nm. In a gas at usual temperatures and pressures, they are roughly ten times larger ($\sim$ 3 nm). The binding energies of valence electrons in an atom are of the order of a few eV. One then finds:

$$\begin{aligned} \text{Molecule:} \quad & \Delta = 0.2 \text{ nm} \;\; |E_0| = 4 \text{ eV} \; A = 1\text{eV} \; T = 10^{-15} \text{ s} \,, \\ \text{Gas:} \quad & \Delta = 3.0 \text{ nm} \;\; |E_0| = 4 \text{ eV} \; A = 10^{-12}\text{eV} \; T = 10^{-3} \text{ s} \,. \end{aligned}$$

The time to get from one atom to the other is very small for valence electrons in a molecule or in a solid. These electrons are completely delocalized in the

molecular structure. Conversely, this phenomenon is completely negligible in gases. In fact, owing to thermal motion, two molecules in a gas remain at a distance of, say, 3 nm for a length of time smaller than 10^{-10} s. Quantum tunneling oscillations have a period of 10^{-3} s and they cannot occur appreciably on such a small time scale. Therefore, the idea that in a gas each electron belongs to a given molecule is quite acceptable.

Molecular binding

Similarly, in this result, one can find a starting idea for the explanation of *molecular binding*, which cannot be understood classically.

Consider the simplest molecule, the H_2^+ ion, made of two protons and one electron. The decrease $E_0 \to E_0'$ explains that it is more favorable energetically for the electron to have an equal probability to be on both protons rather that being bound to one of them and let the other live its own life. Classically, the moon belongs to the earth. It could belong to the planet Mars, but it has chosen the earth and it sticks to that choice.

However, quantum mechanically, the fact that the electron belongs to both protons at the same time stabilizes it. This effect increases as the distance between protons decreases. However, a compromise must be found, owing to the Coulomb repulsion of protons. An equilibrium situation results. After doing the calculations, one can prove that the H_2^+ ion is bound.

If, in addition, one takes into account spin and the Pauli principle, the splitting and the fact that the symmetric state is more tightly bound $E_S < E_A$ accounts for the chemical covalent bond. In other words, quantum tunneling is responsible for our existence!

Potential barriers

One can find exercises of the following type. A particle of energy E hits a potential barrier of width Δ and height $V_0 > E$. Classically, the particle should bounce back. What is the tunneling probability?

For an electron and with atomic orders of magnitude $E = 1\text{eV}$, $V_0 = 2\text{eV}$, $\Delta \approx 0{,}1$ nm, we obtain a probability of $p = 80\%$ that the particle crosses the barrier which is completely anticlassical.

In the same conditions, a proton $m_p \sim 2000 m_e$ has a probability of crossing the barrier of $p \approx 10^{-19}$ because of the mass effect. In other words, saying that nuclei, protons and neutrons, have well-defined positions in an atom or a molecule makes sense.

But, at nuclear scales $V_0 = 2E \approx 10$ MeV and $\Delta \approx 1$ fm, the probability is 80%, and a proton is delocalized in a nucleus.

5.6 Tunneling microscopy, nanotechnologies

An important practical application is the construction by Binnig and Rohrer in the 1980s of the scanning tunneling microscope (STM, 1986 Nobel prize).

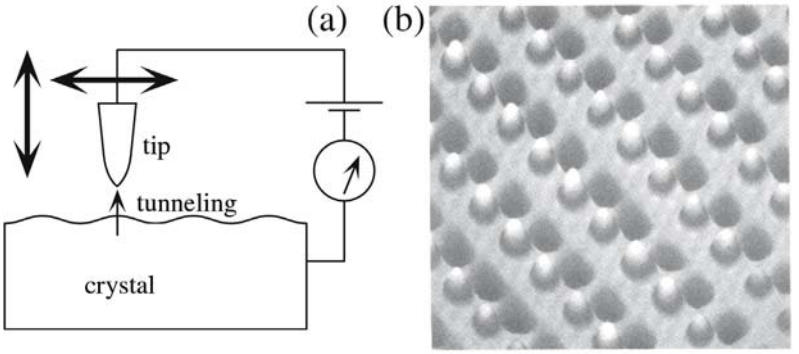


Fig. 5.10. (a) Principle of a tunneling microscope. A thin tip is moved in the vicinity of a solid surface with piezoelectric transductors. One adjusts the distance of the tip to the surface in such a way that electric current due to tunneling between the surface and the tip is constant. This provides a mapmaking of the electron density distribution (actually the electrostatic potential) at the surface of the crystal. An example is shown in (b), where one can see a surface of In Sb. The Sb atoms appear to be raised. The actual size of the sample in the figure is ∼3 nm. (After Y. Liang et al., *J. Vac. Sci. Technol.* **B9**, 730, (1991).)

A conducting tip is moved along a surface at a very short distance $\Delta \approx 10$Å. A potential difference is applied, and the electrons pass from the surface to the tip by the tunnel effect. The current is extremely sensitive to the distance (actually to the electrostatic potential V_0). That way, one can detect incredibly fine details ≈0.01 nm, and one can make a map of the surface. An example is shown in Figure 5.10.

5.6.1 Nanotechnologies

With the tunnel effect one gets near to science fiction by inventing *nanotechnologies*, that is, creating operation techniques at the nanometer scale, at distances comparable to the size of the smallest living systems, viruses.

On September 29, 1989, D. Eigler, research engineer at IBM, was able to manipulate individual atoms on a metal surface. He picked them with a tip and put them on another site of the surface (the device is similar to that shown in Figure 5.10).

Eigler first managed to write the letters IBM with 35 xenon atoms on a nickel substrate. One year later, he was able to construct an electronic switch whose moving part was made of a single atom (5 nm high)!

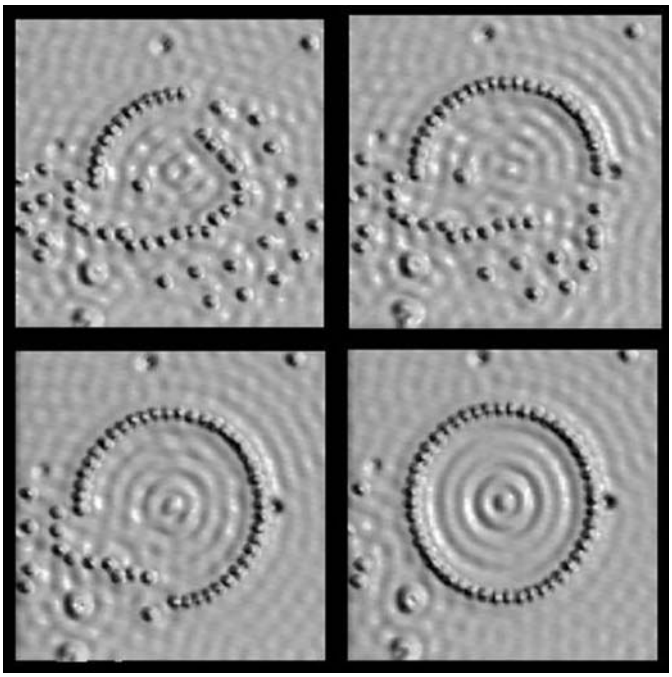


Fig. 5.11. Progressive construction of a "coral reef" made of 48 Fe atoms on an nickel crystalline surface by atom manipulation.

In Figure 5.11, one can see the progressive construction of a "coral reef" of 48 iron atoms on a nickel substrate by atom manipulation. Figure 5.12 shows a side view of the corral reef. The "waves" correspond to the surface density of electrons trapped inside the structure. In other words we see the ground state of de Broglie waves in a circular two-dimensional well directly (this is a consequence of the Pauli principle; we see the probability densities of higher excitations because the lower ones are already occupied). We could have considered this problem in Section 5.3 above.

5.6.2 Classical limit

These developments are truly impressive. However, what is even more striking perhaps is what one cannot do with the tunnel effect.

Indeed, quantum mechanics opens the possibility of fantastic dreams. One can get to another galaxy by quantum tunneling and see extraterrestrials. One can win a bike race by crossing mountains by quantum tunneling without any fear of being dope-tested, or, more seriously, see a piece of dust cross a tulle curtain by quantum tunneling. In principle, quantum mechanics gives us the possibility to do all that. Unfortunately, the probability is very small. One can, as an exercise, calculate the probabilities and one ends up with the most

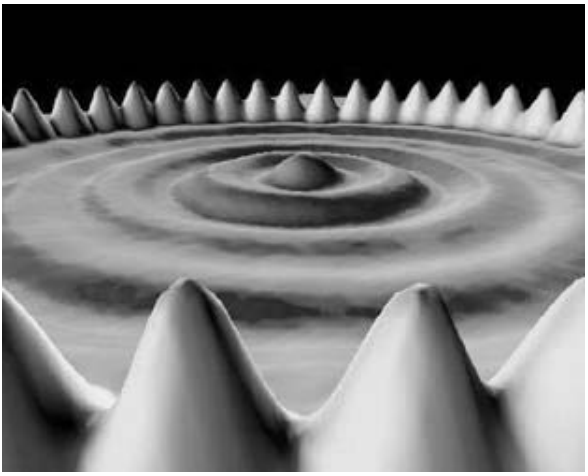


Fig. 5.12. Side view of the electron "lake" inside the coral reef of the previous picture. The waves are stationary de Broglie waves.

extravagantly small numbers one can imagine $p \sim \exp(-10^{(30\text{ to }65)})$. Such numbers are impossible to write in the binary system! It is undoubtedly much more promising, in order to win the Nobel prize, to buy a cat. The probability that a cat will type the unified theory of the Universe, or Shakespeare's works by walking at random on the keyboard of a computer without making any mistake is enormously larger: $p \sim \exp(-10^{(4\text{ to }6)})$.

All of this is meaningless for simple reasons. There exist other terms in the Hamiltonian of a piece of dust that allow it to cross a curtain: someone can do the cleaning; there can be a small hole in the curtain, and so on. But it is fascinating to see that an effect which is so important for our existence cannot manifest itself at our scale.

6

Principles of quantum mechanics

We are now going to put some order in our ideas and do theoretical physics. During the years 1925 to 1927, quantum mechanics took shape and accumulated successes. But three persons, and not the least, addressed the question of its structure. In Zürich there was Erwin Schrödinger, in Göttingen David Hilbert, and in Cambridge Paul Adrien Maurice Dirac.

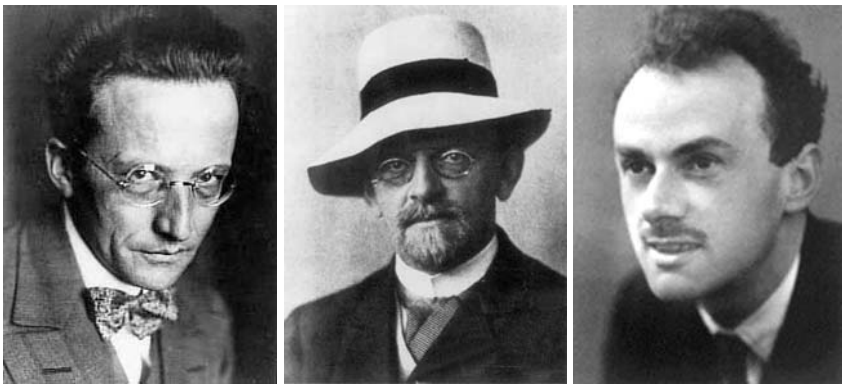


Fig. 6.1. Schrödinger, Hilbert, and Dirac at the end of the 1920s. (All rights reserved.)

Hilbert was 65; he was considered the greatest living mathematician since the death of Poincaré. Schrödinger was 40. Dirac was a young 23-year-old student in Cambridge, simply brilliant.

Their reflections, which we can follow, generated the general principles of quantum mechanics. Here, we do the following.

- First we write a more concise and general formulation of what we have done by using the formalism of Hilbert space and the notations due to Dirac.

- Then, we prove in a simple manner some results we have already guessed on observables, in order to state the general principles of quantum mechanics.
- This allows us to discover at last Heisenberg's matrices and the "matrix mechanics" that had been elaborated by Heisenberg, Born, Jordan, and Pauli between 1924 and 1925.
- In the course of this, we understand quite simply how Schrödinger and Dirac understood the equivalence of the two approaches in 1926.
- Finally, we illustrate these principles on a quantum phenomenon that is directly visible. It is the only one: the polarization of light.

From the point of view of Hilbert, Schrödinger, and Dirac, what were the problems? First, there were two versions of quantum mechanics, and one had to impose some order.

Even without that, our theory is nice and appealing, but it is not very esthetically pleasing. First, it is restricted to the motion of a particle in space, and it must be generalized. Also, it is somewhat ambiguous. In wave mechanics, one describes the state of a particle in space at time t by a wave function $\psi(\boldsymbol{r}, t)$. An unpleasant feature is that the description is not unique. The Fourier transform of the wave function $\varphi(p, t)$ is a completely equivalent description of the state of the particle. We can perfectly well state our principles using $\varphi(p, t)$. For instance, the position observable x is then

$$\hat{x} = i\hbar \frac{\partial}{\partial p_x} \quad .$$

To get a clear insight of this consists of doing theoretical physics. We now speak about mathematics. We do not want to do mathematics in the sense of taking care of rigor and such, but we want to use mathematics and to see how this allows us to understand some physics.

Actually, the only difficulty is the language. After one has become familiar with the language, things become much simpler; life is easier. But learning a new language is always difficult at the beginning.

6.1 Hilbert space

Consider two wave functions and their Fourier transforms. Because of Plancherel's theorem, the two following integrals are equal,

$$\int \psi_1^*(\boldsymbol{r}, t)\psi_2(\boldsymbol{r}, t)\, d^3r \; = \int \varphi_1^*(\boldsymbol{p}, t)\varphi_2(\boldsymbol{p}, t)\, d^3p \quad . \tag{6.1}$$

When mathematicians see such properties, they understand the underlying structures. Indeed these integrals are scalar products. The extraordinary idea of people such as Banach, Hilbert, and Fréchet was to consider functions as vectors or points in vector spaces, and to use a geometric language in order to solve problems of analysis.

What happens here is quite similar to what happens in ordinary geometry. A given vector can be represented by different sets of coordinates in different reference systems. But the lengths, the angles, namely scalar products, are independent of the specific reference system. We show that there are, in fact, many representations of the state of a system. Each one has its own advantages.

6.1.1 Two-dimensional space

Before entering the core of the subject, we can recall some simple notions about Hermitian spaces, that is, complex vector spaces of finite dimension. These notions are useful in what follows.

In two dimensions, for simplicity, one can represent a vector u by the column matrix of its components. The conjugate vector is the line matrix $\bar{u}$ where we transpose and take the complex conjugate of each coordinate. We denote $\langle v|u\rangle$ as the Hermitian scalar product of u and v:

$$u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \ , \ \bar{u} = (u_1^*, u_2^*) \ , \ \langle v|u\rangle = v_1^* u_1 + v_2^* u_2 \quad . \tag{6.2}$$

This scalar product is positive definite and the norm $\|u\|$ of a vector is defined by

$$\|u\|^2 = \langle u|u\rangle \quad . \tag{6.3}$$

The Hermitian conjugate $M^\dagger$ of a matrix M is obtained by transposing and taking complex conjugates of numbers $M_{ij}^\dagger = (M_{ji})^*$.

A matrix is said to be Hermitian if it is equal to its Hermitian conjugate $M = M^\dagger$. The eigenvalues of a Hermitian matrix are real. The corresponding eigenvectors, normalized to one, form an orthonormal basis, or a Hermitian basis of the space.

6.1.2 Square integrable functions

Let us now consider square integrable functions that are of interest in quantum mechanics. A complex function $f(x)$ of a real variable x is said to be square integrable if it satisfies

$$\int_{-\infty}^{\infty} |f(x)|^2 \, dx < \infty \quad . \tag{6.4}$$

In mathematics, one calls this set of functions $\mathcal{L}^2(\mathcal{R})$, so we write $f \in \mathcal{L}^2(\mathcal{R})$.

Square integrable functions form a complex vector space (any linear combination of square integrable functions is square integrable). The extension to three dimensions (or three variables) $g(x, y, z)$ is straightforward; the corresponding space is denoted $\mathcal{L}^2(\mathcal{R}^3)$.

At this point, we consider the results of Charles Hermite in 1860. In fact, considering complex functions, Hermite defined an *Hermitian scalar product* of two functions f and g by

$$\langle g|f\rangle = \int g^*(x) f(x)\, dx \quad . \tag{6.5}$$

This is linear in f and antilinear in g, and it possesses the Hermitian symmetry

$$\langle g|f\rangle = \langle f|g\rangle^* \quad . \tag{6.6}$$

This allows us to define the norm $\|f\|$ of the function f by

$$\|f\|^2 = \int |f(x)|^2\, dx \quad . \tag{6.7}$$

Algebraically, it is exactly the same as in finite-dimensional spaces, as considered above. It is convergence, that is, topological properties, that is different.

Now, Hermite made a very remarkable discovery. Without being aware of that, he studied the quantum harmonic oscillator eigenvalue problem

$$(x^2 - \frac{d^2}{dx^2})\varphi_n(x) = \varepsilon_n \varphi_n(x)$$

or

$$\hat{h}\varphi_n(x) = \varepsilon_n \varphi_n(x) \quad \text{with} \quad \hat{h} = (x^2 - \frac{d^2}{dx^2}) \quad , \tag{6.8}$$

and he found all the square integrable solutions $\{\varphi_n(x), \varepsilon_n\}$,

$$\varphi_n(x) \ = \gamma_n \ e^{x^2/2} \frac{d^n}{dx^n} \left(e^{-x^2}\right) \ , \quad \varepsilon_n = 2n+1 \ , \quad \text{n integer} \geq 0 \quad . \tag{6.9}$$

These functions are normalized to one ($\|\varphi_n\| = 1$) if

$$\gamma_n = \pi^{-1/4}\, 2^{-n/2}\, (n!)^{-1/2} \quad . \tag{6.10}$$

The Hermite functions $\varphi_n(x)$ are orthonormal (i.e., orthogonal and normalized to one) as one can check; they form a free orthonormal set.

But Hermite found a most remarkable property. All square integrable functions can be expanded on the set of Hermite functions,

$$\forall f \in \mathcal{L}^2(\mathcal{R}) \ , \quad f(x) = \sum_{n=0}^{\infty} C_n \varphi_n(x) \quad , \tag{6.11}$$

where the components C_n of f are

$$C_n = \langle \varphi_n | f \rangle \quad , \tag{6.12}$$

because the $\varphi_n(x)$ are orthogonal and normalized. In other words the Hermite functions are a *complete set*, called a Hilbert basis of $\mathcal{L}^2(\mathcal{R})$.

That is the great discovery. Square integrable functions form a *Hilbert space* which has the three properties of being a complex vector space, where a Hermitian positive definite scalar product is defined, and that contains Hilbert bases (the $\varphi_n(x)$ are an example of such a basis for the space of square integrable functions in one variable $\mathcal{L}^2(\mathcal{R})$).

This space is infinite-dimensional. For mathematicians it is more interesting to study than a two-dimensional space, but psychologically, for us it is essentially similar. We are not concerned with topological properties (even though we mention some of them, and they play an important role in more elaborate quantum mechanical problems). The algebraic rules are the same as in finite-dimensional spaces.

Therefore, in a Hilbert basis, f is entirely determined by the set of its components

$$f(x) \Longleftrightarrow \{C_n\} \quad .$$

We show that one can "forget" about the elements of the basis $\varphi_n(x)$ which are simply kept in a catalogue, and work directly with the components that define the vector $f \in \mathcal{L}^2(\mathcal{R})$.

For instance, consider a function g, whose expansion is

$$g(x) = \sum_{n=0}^{\infty} C_n \varphi_n(x) \quad . \tag{6.13}$$

The scalar product $\langle g|f\rangle$ is expressed simply in terms of the components of f and g, $C_n = \langle \varphi_n|f\rangle$ and $B_n = \langle \varphi_n|g\rangle$, as

$$\langle g|f\rangle = \sum B_n^* C_n \quad , \tag{6.14}$$

and the norm of f is given by

$$\|f\|^2 = \sum |C_n|^2 \quad , \tag{6.15}$$

which is simply the extension of the Pythagorean theorem. (In infinite dimensions, mathematicians call that the Bessel–Parseval theorem.) If f is normalized to 1, we have

$$\sum \|C_n\|^2 = 1.$$

One thing that one does not fully appreciate at first is that the geometrical properties of a Hilbert space are very similar to those of a Euclidian space: the Pythagorean theorem and the triangle inequality hold in both.

The result of the application of $\hat{h}$ (equation (6.8)) to f is

$$\hat{h}f(x) = \sum C_n \varepsilon_n \varphi_n(x) \quad . \tag{6.16}$$

Hence, the scalar product of $\hat{h}f$ and f (or, more generally, any function g) is

$$\langle f|\hat{h}f\rangle = \sum \varepsilon_n |C_n|^2 \quad . \tag{6.17}$$

We therefore can use a geometrical language in order to speak about problems of analysis.

6.2 Dirac formalism

Coming back to quantum mechanics, Hilbert and Dirac understood in 1927 that the wave function ψ and its Fourier transform φ are simply two representations of the same unique mathematical object, a vector in Hilbert space. And they were able to give a clear formulation of the theory with these ideas. Dirac invented notations that have been adopted by mathematicians.

From now on, we say that the state of a system is described a any time t by a state vector

$$|\Psi(t)\rangle\,, \quad \text{which belongs to a Hilbert space } \mathcal{E}_H\ , \tag{6.18}$$

and which Dirac calls a "ket". The functions ψ and φ are simply particular representations of this vector. For a particle in three-dimensional space R^3, the Hilbert space is the space of square integrable functions in three variables $\mathcal{L}^2(R^3)$, but this can be generalized to any system.

6.2.1 Notations

Vectors

1. The *vectors* are denoted by the symbol: $|(name)\,\rangle$.
   In (6.18), we have indicated the presence of the time variable. Of course, there exist fixed vectors, such as the elements of a basis that can be denoted $|\varphi_n\rangle$ or $|n\rangle$.
2. The *Hermitian scalar product of* ψ_1 and ψ_2 is denoted
   $$\langle\psi_2|\psi_1\rangle = \langle\psi_1|\psi_2\rangle^* \quad . \tag{6.19}$$
   It is not commutative, it has Hermitian symmetry, and it is linear on the right and antilinear on the left.
3. This scalar product is positive definite which allows us to define the *norm* $\|\psi\|$ of a vector by
   $$\|\psi\|^2 = \langle\psi|\psi\rangle \quad . \tag{6.20}$$
   The norm of a vector ψ vanishes if and only if ψ is the null vector. The norm of a state vector is always equal to one.
4. At the beginning, it is useful to have a little dictionary, where one can see that Dirac's notations are obviously quicker to use.
5. The elements of the *dual* space are denoted
   $$\langle\psi(t)| \in \mathcal{E}_H^* \tag{6.21}$$
   and called *bras* (this is justified owing to a theorem of F. Riesz). This notation comes from the scalar product. A rule in Dirac's formalism is the contraction of products. When a bra $\langle u|$ is on the left of a ket $|v\rangle$

Dictionary	
Dirac formalism	Wave functions
$\|\varphi\rangle$	$\varphi(\boldsymbol{r})$
$\|\psi(t)\rangle$	$\psi(\boldsymbol{r},t)$
$\langle\psi_2\|\psi_1\rangle$	$\int \psi_2^*(\boldsymbol{r})\psi_1(\boldsymbol{r})\,d^3r$
$\|\psi\|^2 = \langle\psi\|\psi\rangle$	$\int \|\psi(\boldsymbol{r})\|^2\,d^3r$
$\langle\psi_2\|\hat{A}\|\psi_1\rangle$	$\int \psi_2^*(\boldsymbol{r})\hat{A}\psi_1(\boldsymbol{r})\,d^3r$
$\langle a\rangle = \langle\psi\|\hat{A}\|\psi\rangle$	$\int \psi^*(\boldsymbol{r})\hat{A}\psi(\boldsymbol{r})\,d^3r$

Fig. 6.2. Dirac vs. wave functions dictionary.

the expression contracts in a "bracket", that is, the number $\lambda = \langle u|v\rangle$. In other words, the bra "eats" the ket and gives this number.
We encounter other examples of this contraction rule, which comes from the tensor structure of quantum mechanics.

6. *Hilbert basis.* A Hilbert basis is a free, orthonormal and *complete* set of vectors
$$\{|n\rangle\} : \quad \langle n|m\rangle = \delta_{nm}.$$
Any vector $|\psi\rangle$ can be expanded on this basis as
$$|\psi\rangle = \sum_n C_n|n\rangle, \text{ with } C_n = \langle n|\psi\rangle \quad . \tag{6.22}$$
C_n is the component of $|\psi\rangle$ along $|n\rangle$. All Hilbert spaces possess Hilbert bases.

6.2.2 Operators

Consider now linear operators, that is, linear mappings of the space onto itself. We will keep the same notation $\hat{A}$ as before:
$$|\chi\rangle = \hat{A}|\psi\rangle\,, \quad (\psi\,,\chi) \in \mathcal{E}_H \quad . \tag{6.23}$$

1. We are interested in the following numbers, which are the scalar products of (6.23) with some other vector:
$$\langle\psi_2|\chi_1\rangle = \langle\psi_2|(\hat{A}|\psi_1\rangle) \quad . \tag{6.24}$$
One can show that this product is *associative*; that is, one can define the action of $\hat{A}$ in the dual space

$$\langle\psi_2|\chi_1\rangle = \langle\psi_2|(\hat{A}|\psi_1\rangle) = (\langle\psi_2|\hat{A})|\psi_1\rangle = \langle\psi_2|\hat{A}|\psi_1\rangle \quad . \tag{6.25}$$

We call this scalar product the *matrix element* of $\hat{A}$ between ψ_1 and ψ_2. This is the same as in finite-dimensional spaces. In ordinary matrix calculus, this expression would be of the type $\bar{v}Mu$, where M is a matrix.
In order to be rigorous, mathematicians have another way of writing: $\langle\psi_2|\hat{A}\psi_1\rangle$, which does not show the useful associativity. Of course, they are right to do so in their work. One cannot bypass rigor. In infinite-dimensional spaces there exist operators whose "domain" (i.e., the set of functions on which they act safely and produce square integrable results) is not the entire space. For instance, if we multiply a square integrable function by x, the result is not always square integrable. Here, we do not worry about such a question.
Again, the basic theme in all this is that we act as if we were in finite-dimensional spaces. To a large extent the topological questions can be put aside for what concerns us at the present level.

2. *Adjoint operators.* The adjoint operator $\hat{A}^\dagger$ of an operator $\hat{A}$ can be defined by the relation

$$\langle\psi_2|\hat{A}|\psi_1\rangle = \langle\psi_1|\hat{A}|\psi_2\rangle^* , \quad \forall\psi_1, \psi_2 \in \mathcal{E}_H \quad . \tag{6.26}$$

It is the same definition as for matrices: we transpose and take the complex conjugate.

3. *Self-adjoint operators.* An operator $\hat{A}$ is said to be self-adjoint or Hermitian if $\hat{A} = \hat{A}^\dagger$, or, equivalently,

$$\langle\psi|\hat{A}|\psi\rangle \text{ is real for all vectors } \psi \text{ of } \mathcal{E}_H \quad . \tag{6.27}$$

But we already know this expression. If we look at that above dictionary, (6.27) is the *expectation value* of the quantity A. Now results of measurements, in particular expectation values, are real numbers. Therefore, *observables are Hermitian operators* $\hat{A} = \hat{A}^\dagger$ as we announced in (4.7).

Theorem 2. *Observables are Hermitian operators.*

4. *Commutation of observables.* We can see easily with the correspondence principle that observables do not commute in general. The product of $\hat{A}$ and $\hat{B}$ is not the same as the product of $\hat{B}$ and $\hat{A}$. One defines the *commutator* $[\hat{A}, \hat{B}]$ of two operators as

$$[\hat{A}, \hat{B}] \equiv \hat{A}\,\hat{B} - \hat{B}\,\hat{A} \quad . \tag{6.28}$$

One can check the fundamental commutation relation

$$[\hat{x}, \hat{p}_x] = i\hbar\hat{I} \quad , \tag{6.29}$$

where $\hat{I}$ is the identity operator.
We show later on that commutation relations between observables play a fundamental role. They allow us, in particular, to derive uncertainty relations for any couple of physical quantities.

6.2.3 Syntax rules

Before we come back to observables and measurement results, we make two observations on syntax rules in Dirac's formalism.

1. *Contraction of products.* In this formalism, expressions are products of terms. Such products can be contracted. We have seen this for the scalar products; it is true for any expression (as we have said, this comes from the tensor structure of the theory).
   With this rule, we discover a special kind of operators of the form

$$|u\rangle\langle v| \quad . \tag{6.30}$$

   This object is a linear operator because if we apply it to a ket $|\psi\rangle$, the bra $\langle v|$ eats the ket $|\psi\rangle$, which gives the number $\lambda = \langle v|\psi\rangle$, and one ends up with the vector $\lambda|u\rangle$.
2. *Hermitian conjugate* of an expression. Just as in finite-dimensional spaces, one must transpose and take the conjugate of each term. In other words, we reverse the order of the factors, and we change the kets into bras and vice versa, the operators in their adjoints and the numbers in their complex conjugates. The Hermitian conjugate of $\lambda|\varphi\rangle\langle\psi|\hat{A}^\dagger\hat{B}$ is $\lambda^*\hat{B}^\dagger\hat{A}|\psi\rangle\langle\varphi|$.

6.2.4 Projectors; decomposition of the identity

Projector. Consider a Hilbert basis $\{|n\rangle\}$. The operator

$$\hat{P}_n = |n\rangle\langle n| \tag{6.31}$$

is the projector on the basis vector $|n\rangle$ (this remark applies to any vector that is normalized to one). Indeed we have

$$\hat{P}_n^2 = \hat{P}_n \text{ and } \hat{P}_n|\psi\rangle = C_n|\psi\rangle \quad ,$$

where C_n is the component (6.22) of $|\psi\rangle$.

One can define a projector $\hat{P}_\nu$ on a subspace $n \in \{\nu\}$ by

$$\hat{P}_\nu = \sum_{n\in\{\nu\}} |n\rangle\langle n| \quad . \tag{6.32}$$

If we extend this to the entire space, we obtain the important *closure relation*, also called the decomposition of the identity

$$\sum_n |n\rangle\langle n| = \hat{I} \quad . \tag{6.33}$$

6.3 Measurement results

We can now understand that, in a measurement of a quantity A,

• the possible results of the measurement are the eigenvalues a_n of the observable $\hat{A}$, and

• the probability of finding the result a_n is the modulus square of the scalar product of the state vector with the corresponding eigenvector:

$$p(a_n) = |\langle\varphi_n|\psi(t)\rangle|^2 \quad .$$

6.3.1 Eigenvectors and eigenvalues of an observable

We denote $|\varphi_n\rangle$ the eigenvectors of $\hat{A}$ and a_n the corresponding eigenvalues, that is, by definition

$$\hat{A}|\varphi_n\rangle = a_n|\varphi_n\rangle \quad . \tag{6.34}$$

We constantly use the following theorems, well known in finite-dimensional spaces:

Theorem 3. *The eigenvalues a_n of an Hermitian operator are real.*

The proof is straightforward. If we multiply (6.34) on the left by $\langle\varphi_n|$, we obtain

$$\langle\varphi_n|\hat{A}|\varphi_n\rangle = a_n\langle\varphi_n|\varphi_n\rangle \quad .$$

The left-hand side is real because $\hat{A}$ is Hermitian, and $\langle\varphi_n|\varphi_n\rangle$ is real and positive. Therefore, a_n is a real number.

Theorem 4. *The eigenvectors corresponding to different eigenvalues are orthogonal.*

We multiply (6.34) on the left by another eigenvector $\langle\varphi_m|$; we obtain

$$\langle\varphi_m|\hat{A}|\varphi_n\rangle = a_n\langle\varphi_m|\varphi_n\rangle = a_m\langle\varphi_m|\varphi_n\rangle \quad .$$

In the last expression, $\hat{A}$ acts on the left on $\langle\varphi_m|$. Therefore, we obtain

$$(a_n - a_m)\langle\varphi_m|\varphi_n\rangle = 0 \quad , \tag{6.35}$$

so that either $a_n = a_m$, or, if $a_n \neq a_m$, $\langle\varphi_m|\varphi_n\rangle = 0$.

We can therefore choose a set of eigenvectors $\{\varphi_n\}$ that are orthonormal.

Let us mention a minor and inessential technical difficulty in this context. It can happen that to the same eigenvalue a_n there correspond several independent eigenvectors $|\varphi_{n,k}\rangle$

$$\hat{A}|\varphi_{n,r}\rangle = a_n|\varphi_{n,r}\rangle\,, \quad r = 1, \ldots, k \quad . \tag{6.36}$$

In such a case, one says that the eigenvalue a_n is degenerate with a degeneracy of order k. Notice that the projector on the eigensubspace of a_n, of dimension k, is

$$\hat{P}_n = \sum_{r=1}^{k} |\varphi_{n,r}\rangle\langle\varphi_{n,r}| \quad , \tag{6.37}$$

where we assume we have chosen the $|\varphi_{n,k}\rangle$ to be orthonormal

$$\langle\varphi_{n',k'}|\varphi_{n,k}\rangle = \delta_{nn'}\delta_{kk'} \quad .$$

6.3.2 Results of the measurement of a physical quantity

Theorem 5. *The result of the measurement of a quantity A on a system is certain (i.e., with probability one) if and only if the system is in an eigenstate of the observable $\hat{A}$.*

We have already seen in chapter 4 that if ψ is an eigenvector of $\hat{A}$ then the result of the measurement is certain: the dispersion vanishes; $\Delta a = 0$. The converse is easy to prove. Consider the norm of the vector $|\chi\rangle = (\hat{A} - \langle a\rangle\hat{I})|\psi\rangle$, where $\hat{I}$ is the identity. We obtain

$$\|(\hat{A} - \langle a\rangle\hat{I})|\psi\rangle\|^2 = \langle\psi|(\hat{A} - \langle a\rangle\hat{I})^2|\psi\rangle = \langle\psi|\hat{A}^2|\psi\rangle - \langle a\rangle^2 = \Delta a^2 \ . \tag{6.38}$$

If the dispersion is zero $\Delta a^2 = 0$, the norm of the vector $(\hat{A} - \langle a\rangle\hat{I})|\psi\rangle$ vanishes. This vector is therefore the null vector and $\hat{A}|\psi\rangle = \langle a\rangle\,|\psi\rangle$. Therefore, if $\Delta a = 0$, $|\psi\rangle$ is necessarily an eigenvector of $\hat{A}$ with eigenvalue $\langle a\rangle$.

Theorem 6. *The result of a measurement of a quantity A on a single system is one of the eigenvalues a_n of the observable $\hat{A}$.*

In chapter 4, we made the following remarks:

- For a given system and a given quantity, nothing restricts the accuracy of the measurement. One finds some value with the accuracy allowed by the measuring instruments.
- By consistency, if, immediately after the measurement, we perform another measurement of the same quantity on the system that has already been measured, we will find the same answer with probability one. (This is a postulate on the consistency of physics; no experiment has ever contradicted this.)
- Therefore, the measurement on a system is actually a means to prepare this system in a new state for which we know the value of A exactly ($\Delta a = 0$).
- Owing to theorem 5, the system is then necessarily in an eigenstate of $\hat{A}$, and therefore the value found previously is an eigenvalue of $\hat{A}$.
- Obviously, after the measurement, the state vector of the system is in the eigenspace corresponding to the eigenvalue that has been obtained.

6.3.3 Probabilities

What is the probability of finding the result a_n by measuring A on a system whose state is ψ? There, the answer comes from geometry.

In fact, the question is to know "how much" of the eigenvector $|\psi_n\rangle$ does the state vector contain. By the superposition principle, we can understand that if it is 100% $|\psi_n\rangle$ the probability is equal to 1, and if it does not have any component along $|\psi_n\rangle$, if it doesn't contain it at all, the probability is zero; one will never find the value a_n.

How can we evaluate this probability? The answer is that the probability is the modulus squared of the component of the state vector $|\psi\rangle$ along the normalized eigenvector $|\varphi_n\rangle$, or, equivalently, of the scalar product of $|\psi\rangle$ and $|\varphi_n\rangle$,

$$p(a_n) = |\langle\varphi_n|\psi(t)\rangle|^2 \quad . \tag{6.39}$$

The scalar product $\langle\varphi_n|\psi(t)\rangle$ is the probability amplitude $\alpha(a_n)$ to find a_n.

Do we really find that by some magic inspiration? No; it is practically written above in (6.17) in a particular case.

Let's go back to Section 6.1.2, to the equations (6.15) and (6.17), and to the case of the harmonic oscillator. The $\{\varphi_n\}$ are the eigenstates of the energy: $\hat{h}\varphi_n(x) = E_n\varphi_n(x)$. The components $C_n = \langle\varphi_n|\psi\rangle$ of the state vector on this basis are these scalar products. The normalization of the state vector (i.e., the probabilistic interpretation) and the Pythagorean theorem tell us that

$$\sum_{n=0}^{\infty} |C_n|^2 = 1 \quad , \tag{6.40}$$

as should be satisfied by a probability law. The expectation value of the energy (6.17) is indeed given in terms of the possible issues E_n by

$$\langle E\rangle = \sum_{n=0}^{\infty} E_n|C_n|^2 = 1 \quad . \tag{6.41}$$

This shows that in this case the $|C_n|^2 = |\langle\varphi_n|\psi\rangle|^2$ are the probabilities of finding E_n. Indeed, it is a theorem, which is not difficult to prove for a finite set, that if we know the issues of a probability law and its moments (i.e., $\langle E^k\rangle$ for all values of the integer k) we know the probabilities (obviously we can calculate the expectation value of any power of $\hat{h}$ and obtain the value of $\langle E^k\rangle$).

And that is completely general! Above, we never referred to the specific form of $\hat{h}$, of φ_n, or of the values E_n. We simply used the fact that the $\{\varphi_n\}$ form a Hilbert basis, and that they are the eigenvectors of $\hat{h}$.

6.3.4 The Riesz spectral theorem

This property relies on the *spectral theorem* of Frederic Riesz, which is a fundamental theorem of Hilbert space analysis.

Theorem 7. ***Spectral theorem.*** *The set* $\{|\varphi_n\rangle\}$ *of eigenvectors of an Hermitian operator* $\hat{A}$ *forms a Hilbert basis of the space.*

This is well known in finite-dimensional spaces and matrix calculus.

We have deliberately stated it in a mathematically incorrect way for an infinite dimensional space. The true statement is: To any self-adjoint operator, there corresponds a decomposition of the identity and a spectral decomposition.

In other words, if we forget about possible degeneracies for simplicity,

1. Any vector $|\psi\rangle$ can be decomposed on the basis$\{|\varphi_n\rangle\}$ of the eigenvectors of $\hat{A}$,

$$\forall\, |\psi\rangle\ , \quad |\psi\rangle = \sum_n C_n |\varphi_n\rangle\ , \quad \text{with } C_n = \langle\varphi_n|\psi\rangle \quad . \tag{6.42}$$

   If $|\psi\rangle$ is a normalized state vector, then

$$\langle\psi|\psi\rangle = \sum_n |C_n|^2 = 1 \quad . \tag{6.43}$$

2. The decomposition of the identity, or closure relation, is

$$\hat{I} = \sum_n |\varphi_n\rangle\langle\varphi_n| \quad . \tag{6.44}$$

3. The operator $\hat{A}$ has a spectral decomposition; that is,

$$\hat{A} = \sum_n a_n |\varphi_n\rangle\langle\varphi_n| \quad . \tag{6.45}$$

4. Therefore

$$\hat{A}|\psi\rangle = \sum_n C_n\, a_n |\varphi_n\rangle\ , \quad \text{and} \quad \langle a\rangle = \sum_n a_n\, |C_n|^2 \quad . \tag{6.46}$$

We can read in (6.43) and (6.46) that the numbers $\{|C_n|^2\}$ are the probabilities of finding the results a_n, as previously.

As stated in Theorem 7, this "theorem" isn't quite true; why? It is true in spirit, but not in form; it lacks rigor. In fact, there exist pathologies. For instance, there exist operators which, when they are applied to some vectors, "push them out" of the Hilbert space, such as x or d/dx. Their eigenvectors do not belong to the Hilbert space, but to another space called the space of *eigendistributions*.

The eigenfunctions of $\hat{p}_x$ are not square integrable, because they are plane waves $\propto e^{ip_0 x/\hbar}$. But the amazing fact is that one can nevertheless expand any square integrable function on this set, this is simply the Fourier transformation $f(x) = (2\pi)^{-1/2} \int g(p) e^{ipx/\hbar}\, dp$. One can expand a square integrable function

on a continuous set of functions that does not belong to the Hilbert space, a continuous basis that belongs to another space. Of course, it is quite feasible to write quantum mechanics in a rigorous way, but it is tedious and too complicated at our stage, and it does not bring anything new physically. It suffices to be aware of that.

6.3.5 Physical meaning of various representations

At this point, things are becoming physically interesting.

In a given Hilbert basis, it is obvious that the state vector $|\psi\rangle$ is completely determined by the set of its components $\{C_n\}$,

$$|\psi\rangle \leftrightarrow \{C_n = \langle\varphi_n|\psi\rangle\} \quad .$$

which we can write as a column vector, the corresponding bra being the conjugate line vector.

This representation of the state vector is completely equivalent to the wave function $\psi(\boldsymbol{r}, t)$.

Therefore, there are not only two, but an infinite number of equivalent representations of the state of the system. What is their physical meaning?

In the basis of the eigenstates of the Hamiltonian, the interpretation of this representation is simple and crystal clear: the C_n's are the amplitudes to find E_n in an energy measurement. Therefore,

- The representation $\psi(r, t)$ is more convenient if we are interested in the properties of the particle in space,
- Its Fourier transform $\varphi(p, t)$ is more convenient if we are interested in its momentum properties,
- And the components $\{Cn\}$ in the basis of energy eigenstates are more convenient if we are interested in the energy of the particle.

But, owing to Riesz's theorem, this can be done with any physical quantity, for instance, the angular momentum, which we examine later on and which also has discrete eigenvalues. This can be thought of as a "generalization" of the properties of the Fourier transform.

6.4 Principles of quantum mechanics

We are now able to state the general principles of quantum mechanics. Up to a technical detail, which we discuss below, these are the following three principles.

6.4.1 The principles

I. Superposition principle

With each physical system is associated a Hilbert space $\mathcal{E}_H$. The state of the system is defined at any instant by a vector $|\psi(t)\rangle$ of $\mathcal{E}_H$ normalized to one.

Comment. This means that any linear superposition of state vectors $|\psi\rangle = \sum C_i|\psi_i\rangle$, with C_i complex such that $|\psi\rangle$ is normalized, is a possible state vector. Notice that the convention $\|\psi\| = 1$ leaves an indetermination. A state vector is defined up to an arbitrary phase factor $e^{i\delta}$. However, the relative phases of different states of the system are essential. If $|\psi_1'\rangle = e^{i\delta_1}|\psi_1\rangle$ and $|\psi_2'\rangle = e^{i\delta_2}|\psi_2\rangle$, the superposition of states $C_1|\psi_1'\rangle + C_2|\psi_2'\rangle$ is different from the superposition $C_1|\psi_1\rangle + C_2|\psi_2\rangle$.

II. Physical quantities

1. To each physical quantity A there corresponds a linear Hermitian operator $\hat{A}$ in $\mathcal{E}_H$: $\hat{A}$ is the observable corresponding to the quantity A.
2. Let $|\psi\rangle$ be the state of the system when a measurement of A is performed. Whatever the state $|\psi\rangle$ is, the only possible results of the measurement are the eigenvalues a_n of the observable $\hat{A}$.
3. Denoting by $\hat{P}_n$ the projector of the subspace associated with the eigenvalue a_n, the probability of finding the value a_n in a measurement of A is
$$\mathcal{P}(a_n) = \|\psi_n\|^2 \text{ where } \quad |\psi_n\rangle = \hat{P}_n|\psi\rangle \ . \tag{6.47}$$
This reduces to (6.39) in the absence of degeneracies, but takes into account all cases in a geometrical way.
4. Immediately after a measurement of A that has given the value a_n, the system is in a new state $|\psi'\rangle$:
$$|\psi'\rangle = \frac{|\psi_n\rangle}{\|\psi_n\|} \quad . \tag{6.48}$$

Comments. Relation (6.47) can be written in the equivalent forms:

$$\mathcal{P}(a_n) = \langle\psi|\hat{P}_n|\psi\rangle = |\langle\psi|\psi_n\rangle|^2 \ . \tag{6.49}$$

Principle II.2 is called the *principle of quantization*, and Principle II.3 is the *principle of spectral decomposition*. Principle II.4 is the *principle of wave*

packet reduction. It is the quantitative form of the fact that a measurement perturbs the system.

III. Evolution in time

Let $|\psi(t)\rangle$ be the state of a system at time t. As long as no measurement is performed on the system, its evolution in time is given by the Schrödinger equation:

$$i\hbar \frac{d}{dt}|\psi(t)\rangle = \hat{H}|\psi(t)\rangle \ , \tag{6.50}$$

where $\hat{H}$ is the observable energy, or the Hamiltonian of the system.

The state vector $\psi(t)$ depends on time and evolves in Hilbert space according to this first-order ordinary differential equation.

6.4.2 The case of a continuous spectrum

This is the case for the position and momentum variables. In such a case, the only prediction that makes sense is to find a result inside some range $[a, a+da[$. The discrete probability law (6.47) is replaced by a continuous law. In the case of the position observable x, the law is

$$\mathcal{P}(x)\,dx = |\psi(x)|^2\,dx \quad , \tag{6.51}$$

where $\psi(x)$ is the wave function of chapter 3, and similarly for the momentum variable p.

6.4.3 Interest of this synthetic formulation

This formulation has many advantages. It is general. It is precise mathematically speaking. It exhibits the important features of the theory.

It is instructive to compare what we just did with the history of Maxwell's equations. On October 27, 1864, Maxwell presented his memoir on the unification of electricity and magnetism to the Royal Society.

Maxwell used 283 symbols to write his equations. Here is a sample:

$$\text{Magnetic Force } (\alpha,\ \beta,\ \gamma) \begin{cases} \frac{d\gamma}{dy} - \frac{d\beta}{dz} = 4\pi p' \\ \frac{d\alpha}{dz} - \frac{d\gamma}{dx} = 4\pi q' \\ \frac{d\beta}{dx} - \frac{d\alpha}{dy} = 4\pi r' \end{cases} .$$

"In these equations for the electromagnetic field, we have introduced twenty variable quantities," said Maxwell, who added "Between these twenty quantities, we have found twenty equations. Therefore these equations are sufficient to determine all the quantities involved provided we know the conditions of the problem."

Of course, we can write the equations in a simpler way by using vectors and vector analysis which exhibits the rotation invariance of the equations.

$$\boldsymbol{\nabla}\cdot\boldsymbol{B}=0\,,\ \boldsymbol{\nabla}\cdot\boldsymbol{E}=\frac{\rho}{\varepsilon_0}\,,\ \boldsymbol{\nabla}\times\boldsymbol{E}=-\frac{\partial\boldsymbol{B}}{\partial t}\,,\ c^2\boldsymbol{\nabla}\times\boldsymbol{B}=\frac{j}{\varepsilon_0}+\frac{\partial\boldsymbol{E}}{\partial t}\quad .$$

This way of writing uses only 59 symbols and it leads easily to results such as the propagation equation in vacuum

$$\left(\frac{\partial^2}{c^2\partial t^2}-\Delta\right)\boldsymbol{E}=0\quad .$$

However, relativistic invariance is the fundamental property that underlies Maxwell's equations and when a theorist writes them, they appear as

$$\partial_\mu \mathrm{F}^{\mu\nu}=j^\nu\quad , \tag{6.52}$$

where relativistic invariance is explicit, and one uses only 8 symbols.

Of course, when it comes to constructing the antenna of a satellite, we must come back to more concrete quantities and recall that in (6.52) there exist a number of implicit conventions ($\varepsilon_0=c=1$, $A_\mu=(\varphi,\boldsymbol{A})$, $F^{\mu\nu}=\partial^\mu A^\nu-\partial^\nu A^\mu$, $j_\mu=(\rho,\boldsymbol{j})$, and so on). But the proof of intermediate results is much simpler.

6.5 Heisenberg's matrices

We can now understand what Heisenberg's matrices are.

6.5.1 Matrix representation of operators

In a Hilbert basis, we have seen that a state vector can be written as a column vector made of its components.

In a Hilbert basis $\{|\varphi_n\rangle\}$, any linear operator $\hat{A}$ has simply a matrix representation.

Consider a vector $|\psi\rangle$ and its expansion

$$|\psi\rangle=\sum_n C_n|\varphi_n\rangle\quad . \tag{6.53}$$

If we apply a linear operator $\hat{A}$ on this, we obtain a vector $|\chi\rangle$ with components $\{B_n\}$

$$\hat{A}|\psi\rangle=|\chi\rangle=\sum_n B_n|\varphi_n\rangle\quad . \tag{6.54}$$

Multiplying (6.54) on the left by the basis vector $|\varphi_n\rangle$, we obtain

$$B_n=\langle\varphi_n|\hat{A}|\psi\rangle=\sum_m\langle\varphi_n|\hat{A}|\varphi_m\rangle C_m\quad , \tag{6.55}$$

where we have inserted the expansion (6.53) of $|\psi\rangle$.

We obtain the matrix relation between the coefficients $\{B_n\}$ and $\{C_n\}$,

$$B_n = \sum_m A_{n,m} C_m \quad , \tag{6.56}$$

where the matrix elements $A_{n,m}$ are given by

$$A_{n,m} = \langle \varphi_n | \hat{A} | \varphi_m \rangle \quad . \tag{6.57}$$

This is the expected result: the observable $\hat{A}$ is represented in this basis by the matrix $(A_{n,m})$, which acts on the line or column vectors above.

6.5.2 Matrices X and P

It is interesting to write the matrix representations of the operators x and p_x in the basis of the harmonic oscillator eigenstates. In order to do this, we can use the recursion relations (5.16) of the Hermite functions (chapter 5).

We obtain:

$$\hat{x} \quad \Rightarrow \quad \sqrt{\frac{\hbar}{2\,m\omega}} \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & \dots \\ \sqrt{1} & 0 & \sqrt{2} & 0 & \dots \\ 0 & \sqrt{2} & 0 & \sqrt{3} & \dots \\ 0 & 0 & \sqrt{3} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{pmatrix} , \tag{6.58}$$

$$\hat{p} \quad \Rightarrow \quad -i\sqrt{\frac{m\omega\hbar}{2}} \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & \dots \\ -\sqrt{1} & 0 & \sqrt{2} & 0 & \dots \\ 0 & -\sqrt{2} & 0 & \sqrt{3} & \dots \\ 0 & 0 & -\sqrt{3} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{pmatrix} . \tag{6.59}$$

In this basis, the matrix of the Hamiltonian is of course diagonal.

Here are examples of Heisenberg's matrices!
But how did he end up there?

6.5.3 Heisenberg's thoughts

In 1924–1925, Heisenberg was a young assistant of Max Born in Göttingen (he was 23). He did things that his contemporaries had difficulty understanding.

Heisenberg was an amazing person. It is difficult to understand his way of thinking. He was both very simple but infused with Northern philosophy.

He admired Niels Bohr who was a propagandist of positivism and of Kierkegaard's principle, "Any new experimental field can only be analyzed with concepts of its own. It is not possible to use concepts and principles used previously in other contexts."

And Göttingen was a fabulous place. The philosopher Husserl had been there. He was the founder of phenomenology, "One must come back from

discourses and opinions to facts," that is, describe what one observes in the simplest possible language before trying to interpret it. Husserl had left his student Wittgenstein, a great specialist of language, who said that "It is the language which must adapt to facts and not the reverse. If one attempts to adapt the interpretation of a phenomenon with a language which is already formed and filled with a priori's, one is bound to draw wrong conclusions on the nature of things."

So, Heisenberg said to himself, "I cannot talk about the position x and velocity v of an electron in an atom. I can only talk about what I am able to see, the positions and the intensities of spectral lines."

In the classical theory of radiation, a fundamental consequence of Maxwell's equations and of relativity is that if a charged particle is accelerated, it radiates energy. The simplest example is a charged dipole with a sinusoidal motion $x = a\cos(\omega t)$. It radiates and loses energy. At large distances, the field decreases not as $1/r^2$ as for a fixed charge, but as $1/r$. Therefore, the energy radiated through a sphere of radius R is independent of R; its flux is conserved. The radiated amplitude is proportional to the acceleration $E \propto x'' \propto a\omega^2$, and the total radiated power is therefore proportional to $P \sim a^2\omega^4$. An arbitrary periodic motion of period $T = 2\pi/\omega$ can be expanded in a Fourier series:

$$x = \sum_n a_n e^{in\omega t} \ ,$$

and the resulting radiation occurs at frequencies $\nu_n = n\omega/2\pi$ with intensities proportional to $I_n \propto n^4\omega^4 a_n^2$.

Now, said Heisenberg, in atoms there are two indices. The atomic frequencies follow the rule

$$\nu_{nm} = (E_n - E_m)/h \quad .$$

So, Heisenberg introduced "quantum amplitudes" or "quantum quantities" with two indices, and a standard time behavior

$$A \Longrightarrow A_{nm} e^{-i(E_n - E_m)/\hbar} \quad .$$

(There is a technical difference between the Schrödinger picture, where the state vector evolves and the observables are fixed, and Heisenberg's where the state vector is fixed, but the observables evolve in time.) The positions of spectral lines are $\nu_{nm} = (E_n - E_m)/h$ and their intensities are proportional to $|A_{nm}|^2$.

Using such ideas, Heisenberg developed a theory that worked quite well. However, he realized that in order for the product of two quantum quantities to have a proper time-dependence and to be expressed in terms of the two quantities, he needed to invent a "symbolic multiplication," for instance,

$$A^2 \to \sum_k A_{nk} e^{-i(E_n - E_k)/\hbar} A_{km} e^{-i(E_k - E_m)/\hbar} = (\sum_k A_{nk} A_{km})\ e^{-i(E_n - E_m)/\hbar}$$

in other words,

$$(A^2)_{nm} = \sum_k A_{nk} A_{km} \quad .$$

Max Born was intrigued. He was a mathematician, and that reminded him of his youth. Those are just the rules of matrix calculus! Born was very excited.

The matrices of Heisenberg are there in front of us in (6.58). What seems incredible is that in 1925 mathematicians knew of the existence of matrices, as examples of noncommutative algebras, but they considered them as very formal objects, and they weren't used to working with matrices in practice.

So, Born was very excited: "We must absolutely go further in Heisenberg's ideas! In his works, there is an underlying structure to be explored!" He publicized Heisenberg's work widely.

On July 17, 1925, Born took the train from Göttingen to Hanover and, in the same compartment, he met Pauli. He was excited and said to Pauli, "You know, Heisenberg's quantities are matrices! Do you want to collaborate?"

Pauli, who was very young but very famous because of his monumental treatise on relativity, which he had written in 1922 when he was 22, was a tough character. Most of the time, he found other people's ideas either stupid or obvious. So he replied to Born , "You're going to spoil Heisenberg's beautiful physical ideas with your futile mathematics." They quarreled.

In the same compartment, there was a young and shy mathematics assistant, Pascual Jordan, who said to Born, once the train had arrived and Pauli had left, "I have worked on matrices. Maybe I can help you." On the evening of the next day, Born and Jordan established the fundamental commutation relation between the matrices (X) and (P)

$$(X)(P) - (P)(X) = \frac{h}{2\pi} iI \quad ,$$

where I is the unit matrix, that Born called the fundamental equation of the quanten mechanik,[1] that is, the mechanics specific to quanta. This led to three fundamental articles of Born, Heisenberg, and Jordan.

One must say that in the meantime, Pauli, who was pragmatic, had reconsidered his opinion on matrices, and he had calculated the energy levels of the hydrogen atom with matrices in 1925 before Schrödinger's calculation.[2] Pauli managed to calculate all eigenvalues of an infinite matrix, and some effects that were not measured before the 1980s!

Now, Born, Heisenberg, and Jordan went to see Hilbert, who was the great mathematics professor in Göttingen, and questioned him about matrices. "It's very formal," said Hilbert. "The only case when I observed it was useful in practice is in eigenvalue problems related to differential equations

[1] The formula $pq - qp = ih/(2\pi)$ is carved on the gravestone of Born and his wife in Göttingen.

[2] In order to do this, Pauli used the $SU(2) \times SU(2)$ symmetry of the hydrogen atom.

with boundary conditions." Born, Heisenberg and Jordan retired politely, thinking that poor Hilbert did not understand the issue. Six months later, Hilbert had fun saying, "If these arrogant youngsters had listened to me, they would have found the Schrödinger equation six months before him." Indeed the Schrödinger equation was greeted as the great step forward because it enabled the calculation of energy levels as a stationary wave problem, in opposition to quantum restrictions on classical trajectories, as advocated by Bohr.

We come back to Dirac in chapter 8. He appeared in the same summer of 1925; he was just 23. We have done, backwards, the work of unification done by Schrödinger at the end of 1926 and independently by Dirac at the beginning of 1927. Schrödinger knew a lot of mathematics; he had worked on eigenvalue problems. Dirac was simply a young genius.

This brought to an end the quarrels between the pros of wave mechanics and the pros of matrix mechanics. On the other hand, this opened a new field of research to mathematicians. In 1927 Hilbert and von Neumann laid the mathematical foundations of quantum mechanics where one can find the first steps of the theory of distributions developed by Laurent Schwartz in 1946.

6.6 The polarization of light, quantum "logic"

In order to illustrate all this and to show once more that the most important mathematical structure of quantum mechanics is addition, we examine a quantum phenomenon that is the only one directly visible.

Light waves are transverse and they possess a polarization that describes the behavior of the electric field in the plane transverse to the electric field.

There are several types of polarizations. The light coming out of a projector is nonpolarized; it is in a statistical mixture of polarization states. A polarizer, for instance a polaroid, filters linear polarization along its optical axis (it can be an anisotropic medium that absorbs light whose polarization is perpendicular to the axis). In Figure 6.3, left, this axis is assumed to be horizontal.

If we place another polarizer, also called an analyzer, at an angle θ with the first one, the transmitted intensity is proportional to $I \propto \cos^2\theta$ (Figure 6.3, middle). At $\theta = 45°$ the intensity is half of that which came out of the first polarizer. If the axis of the analyzer is vertical, at $\theta = 90°$, no light comes out (Figure 6.3, right).

Classically, this is how Fresnel understood and explained the phenomenon. It seems to be elementary geometry. Notice that it happens in the same manner independently of the wavelength.

But, actually, this is purely a quantum phenomenon; it can and it must be described exactly by this quantum state formalism. (This can be done in classical optics as Stokes understood in the 19th century.) In fact:

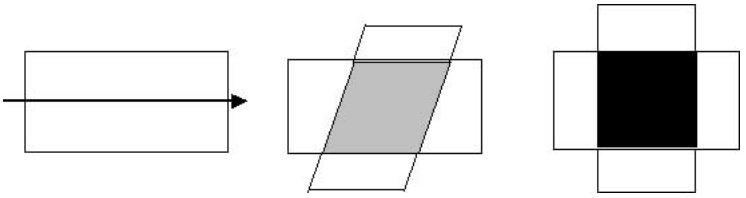


Fig. 6.3. Outgoing light from a horizontal polarizer (left). Intensity across an analyzer at an angle θ (middle); extinction if the analyzer is at 90° of the first one (right).

1. Light is composed of photons.
2. Photons are elementary particles; they cannot be broken into pieces.

So, let's discuss the matter in terms of photons.[3] When a photon impinges on a polaroid, either it goes through it or it doesn't; there is no other choice for it.

Of course, with a macroscopic source, a lot of photons are produced. A light beam of 1 watt carries $\sim 10^{18}$ photons per second. At an angle θ, it is a fraction $\cos^2\theta$ of these photons that comes out of the analyzer. In other words, each photon has a probability of $\cos^2\theta$ to get through.

This is even clearer if we cross the polarizer and analyzer at a right angle $\theta = 90°$. Nothing gets through. States of orthogonal polarizations are incompatible; there is a zero probability that a photon in the horizontal polarization state can be found in the vertical polarization state. Now we can observe an

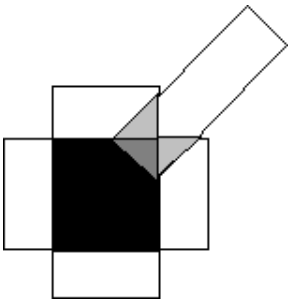


Fig. 6.4. Reappearance of light issuing from two crossed polarizers if a third polarizer at some angle is inserted between them.

amazing phenomenon. If we insert a third polarizer at some non-zero angle,

[3] Polarization of light comes from the fact that the photon is a "spin one" particle. It is a pointlike massless particle that carries an intrinsic angular momentum whose projection on the direction of propagation is either $+\hbar$ or $-\hbar$. The reader will understand that such odd properties are outside the scope of this book.

say 45°, between the crossed polarizers, the light reappears (Figure 6.4) although we have inserted an absorbing object that can only reject all photons polarized perpendicular to its axis! (Actually, is it really the only thing it can do? No! It's not a triviality to say that it is also able to let photons pass if their polarization is parallel to its axis.)

We know the solution. We must describe polarization states of photons in a two-dimensional Hilbert space. In this space, we can choose as basis states the states of horizontal and vertical linear polarization, which we denote

$$|\rightarrow\rangle \quad \text{and} \quad |\uparrow\rangle \quad . \tag{6.60}$$

If the photon is in the state $|\rightarrow\rangle$ it passes through the horizontal polarizer with probability 1. If it is in the state $|\uparrow\rangle$, it is absorbed by this polarizer, but it passes through a vertical polarizer with probability one. By definition, these states are orthogonal $\langle\uparrow|\rightarrow\rangle = 0$ (Figure 6.3).

We denote $|\theta\rangle$ as the state of a photon polarized linearly along a direction at an angle θ with the horizontal axis ($0 \leq \theta < \pi$). This state is a linear combination of the basis states (6.60), as is the orthogonal state $|\theta + \pi/2\rangle$:

$$|\theta\rangle = \cos\theta|\rightarrow\rangle + \sin\theta|\uparrow\rangle\,, \quad |\theta + \pi/2\rangle = -\sin\theta|\rightarrow\rangle + \cos\theta|\uparrow\rangle\ . \tag{6.61}$$

For the particular value $\theta = \pi/4$, we have (this does not restrict the generality of our argument)

$$|\nearrow\rangle = \frac{1}{\sqrt{2}}(|\rightarrow\rangle + |\uparrow\rangle)\,, \quad |\nwarrow\rangle = \frac{1}{\sqrt{2}}(-|\rightarrow\rangle + |\uparrow\rangle)\ ; \tag{6.62}$$

this relation can be inverted:

$$|\rightarrow\rangle = \frac{1}{\sqrt{2}}(|\nearrow\rangle + |\nwarrow\rangle)\,, \quad |\uparrow\rangle = \frac{1}{\sqrt{2}}(-|\nearrow\rangle + |\nwarrow\rangle)\ . \tag{6.63}$$

The explanation of the observations (Figure 6.3) is that the probability for a horizontally polarized photon to get through a polarizer at an angle θ is

$$p(\rightarrow, \theta) = |\langle\theta|\rightarrow\rangle|^2 = \cos^2\theta \tag{6.64}$$

as announced before; that is, $p(\rightarrow, 45°) = 1/2$.

We now come to the observation of Figure 6.4. After crossing the polaroid successfully at 45°, by the principle of wave packet reduction, the photon is necessarily in the state $|\nearrow\rangle$, which can be decomposed according to (6.62). In this new state, it is in both states $|\rightarrow\rangle$ and $|\uparrow\rangle$. It is therefore natural to observe that it can cross the vertical polarizer with probability 1/2, whereas this was forbidden in the absence of the intermediate polarizer. If this latter polarizer is at an angle θ, the probability of finding that a photon crosses the entire setup is

$$p(\rightarrow, \theta, \uparrow) = \cos^2\theta \sin^2\theta \quad . \tag{6.65}$$

Notice that, although all we have done here is ordinary Euclidian geometry in two dimensions, it is necessary for the space to be complex, thatis, Hermitian, in order to describe all the pure polarization states. There exist states with complex components such as

$$|\Psi_{L,R}\rangle = \frac{1}{\sqrt{2}}\left(|\rightarrow\rangle \pm i|\uparrow\rangle\right) \ . \tag{6.66}$$

One can check that such states keep the same form under an arbitrary rotation of the linear polarization basis states. These states correspond to left and right circular polarized states (more generally, elliptic polarization states).

Quantum "logic"

This allows us to understand better the difference between classical and quantum logics, namely the difference between "or" and "and."

The polaroids are filters that let photons pass if their polarization is along their axis and eject them otherwise.

Let's use a metaphor (after all, at the end of such a chapter, we deserve it!). Let's use other words. Instead of horizontal and vertical, let's say ladies and gentlemen. A horizontal polaroid is a filter that only allows ladies to pass. Similarly, a vertical one only lets gentlemen pass. Of course, if one places the two filters one after the other, no one passes. And, if we want to know how many gentlemen are in a population, it suffices to put a vertical polarizer and to measure the outgoing intensity.

Let's say that a polaroid at 45° allows smokers to get through, and a polaroid at 135° does so for the nonsmokers. No smoker is a nonsmoker.

Now assume we are quantum kids trying to understand set theory. We first make a selection of ladies with a horizontal polarizer. Then, in this new sample, we select those who smoke with a polarizer at 45°, as in Figure 6.5 middle.

In classical logic, the succession of the two filters allows the passage of people who are both ladies and smokers, that is, the intersection of the sets {ladies} and {smokers}. We can in fact check that none of them is a nonsmoker.

But see what happens if we look at who they are! We observe with great horror that, by placing a vertically oriented polarizer, half of the people we had selected initially get through (Figure 6.5 right). In other words half of these ladies are gentlemen! In quantum mechanics there is no way out but to conclude that half of the ladies who smoke are in fact gentlemen!

Therefore, there appears to be a very different way of conceiving logic for quantum children when they play. If they play with cubes and spheres each of which can be either blue or red, it's a hard task, if not an impossible one, to try to find the intersections of sets.

Whatever you do in quantum mechanics, smokers are always both ladies and gentlemen, ladies are always both smokers and nonsmokers.

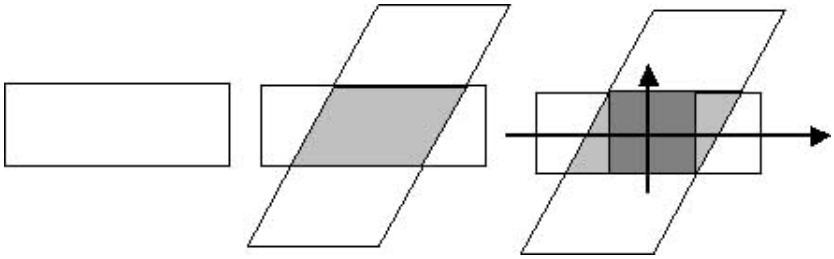


Fig. 6.5. Difference between the quantum and and the classical or: outgoing light after crossing a series of polarizers. Left, only one polarizer, middle, a second polarizer at 45 degrees, right a third polarizer on top of the second, perpendicular to the first one.

All that is simply the mechanism of the superposition principle and the reduction of the wave packet.

7

Two-state systems

In the previous chapter, we saw how Heisenberg's matrix mechanics arose in 1924–1925. What we want to do here is to come back to the problem of the NH_3 molecule, seen in chapter 5 and to do matrix mechanics on this particular case, in a similar way to what we did in wave mechanics by considering the simple problem of the motion of a particle in space.

This will allow us to become familiar with Dirac's formalism and to treat problems in a much simpler way, mathematically speaking, than with wave functions. We show that a great deal of quantum mechanics can be performed with very little mathematics, namely two-dimensional matrices.

We have seen a first example of this with the polarization of the photon in the previous chapter. In other cases, the link with physics might have been obscure if we had started quantum mechanics in this way. We might have had problems in relating simple calculations with actual physical phenomena if we hadn't already treated the problem of the NH_3 molecule.

This leads us to a series of applications. Here, we consider the behavior of the ammonia molecule in an electric field. This enables us to understand the mechanism of masers and to see some applications such as atomic clocks and the tests of the predictions of relativity on time.

Our last section is devoted to a remarkable phenomenon, which is impossible to imagine with classical concepts: neutrino oscillations. It is only after having understood quantum oscillations in two-state systems that one can understand this amazing discovery where an elementary particle in vacuum can change into another one periodically.

7.1 The NH_3 molecule

We recall what we found on NH_3. We had made a model of the inversion motion of this molecule by a symmetric double well whose minima correspond to the two classical equilibrium configurations. We saw that the lowest energy level is actually split in two sublevels by the tunnel effect:

- $E_S = E_0 - A_0$, corresponds to a symmetric wave function ψ_S, and
- $E_A = E_0 + A_0$, corresponds to an antisymmetric wave function ψ_A.

We constructed states corresponding to the "left and right" classical configurations as linear superpositions of these stationary states. And we understood the inversion motion of the molecule between these states.

In the transitions between the two energy levels, the molecule emits or absorbs radiation at the Bohr frequency $\nu = 24$ GHz, related to the splitting. This emission can easily be detected and measured inasmuch as the molecule has an electric dipole moment.

Therefore, one can measure directly the splitting $E_S - E_A$. It is even measured so accurately that it is nonsense to try to calculate it theoretically from first principles.

7.2 "Two-state" system

This is where we follow Heisenberg.

We study physical processes that involve states of the NH_3 molecule which are linear combinations of only the two lowest energy states

$$|\psi\rangle = a|\psi_S\rangle + b|\psi_A\rangle.$$

In other words, we are interested in physical situations where the state vector $|\psi\rangle$ of the molecule remains in a two-dimensional subspace of the Hilbert space, which is infinite-dimensional. This type of situation is called a two-state system or a two-level system, that is, $d(\mathcal{E}_H) = 2$. There are an infinite number of states, but all are linear combinations of two of them. We are going to do quantum mechanics in the case of a two-dimensional Hilbert space where the mathematics are simple.

Of course, this is conceivable mathematically.

- Can it be achieved physically?
- Does it have any physical interest?

Yes! It is indeed Heisenberg's starting point! In NH_3, one knows everything in principle, but it's an awfully complicated problem. It is a system of 14 particles, 10 electrons, and 4 nuclei, with pairwise Coulomb interactions. The Hamiltonian is

$$\hat{H} = \sum_{n=1}^{4} \frac{p_n^2}{2M_n} + \sum_{i=1}^{10} \frac{p_i^2}{2m_e} - \sum_{n=1}^{4}\sum_{i=1}^{10} \frac{q_e q_n}{4\pi\varepsilon_0 |\boldsymbol{r}_n - \boldsymbol{r}_i|} + \frac{q_e^2}{2} \sum_{i=1}^{10}\sum_{j=1}^{10} \frac{1}{4\pi\varepsilon_0 |\boldsymbol{r}_i - \boldsymbol{r}_j|} + \frac{1}{2} \sum_{n=1}^{4}\sum_{m=1}^{4} \frac{q_n q_m}{4\pi\varepsilon_0 |\boldsymbol{r}_n - \boldsymbol{r}_m|} \quad , \tag{7.1}$$

where $(\boldsymbol{r}_i, \boldsymbol{p}_i)$ are the positions and momenta of electrons, and (q_n, M_n), $(\boldsymbol{r}_n, \boldsymbol{p}_n)$ are the charges, masses, positions, and momenta of the nuclei.

Our potential model allowed us to understand qualitatively the tunnel effect, without which the radiowave of $\nu = 24$ GHz would be a mystery, because the energy splitting is tiny.

But the general problem of finding these wave functions in 42 variables is much too complicated. And, above all, it is completely uninteresting. The description of the positions of these 14 particles is of no interest.

However, we know for sure, owing to the spectral theorem,

1. That there exists a set of energy levels E_n and corresponding eigenfunctions ψ_n. (It is an imaginary catalogue. We cannot write these functions, but we only need to know they exist; we can talk about them even though we do not know them.)

2. And we know the $\{E_n\}$ experimentally.

For instance, there exists a whole part of the spectrum due to the inversion motion discussed in chapter 5. It is a series of levels whose energies are higher and higher and where the splitting increases with the energy until it fades away, as depicted in Figure 7.1.

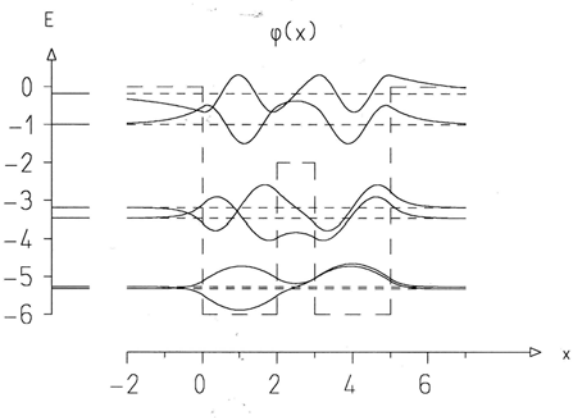


Fig. 7.1. Energy levels and corresponding wave functions in a double well. The abscissa axis is taken, for a given wave function, at the value of the corresponding energy level. (Courtesy of Denis Gratias.)

The following orders of magnitude are of interest. The lowest energy splitting is $2A_0 \approx 10^{-4}$ eV, the splitting between this doublet and the following one is $E_1 - E_0 \approx 0,12$ eV, and the splitting of the latter doublet (E_1) is $2A_1 \approx 4\,10^{-3}$ eV.

If we denote this imaginary catalogue $\{|\psi_n\rangle\}$, we can represent the most general state of the NH_3 molecule as

$$|\psi\rangle = \sum_n a_n |\psi_n\rangle \quad , \tag{7.2}$$

or by an infinite column vector whose components a_n are the probability amplitudes of finding the system in each energy level

$$|\psi\rangle \ : \ \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \\ \vdots \end{pmatrix} \quad \text{with probabilities } P(E_n) = |a_n|^2$$

by definition of the $\{a_n\}$.

This is what Heisenberg does, and it's more realistic than a description in space by a wave function of the 14 particles because

1. one can measures the values E_n, and
2. one can control much more easily the energy of such a complex system than all these positions.

It is not deeper than a wave function representation, but it is not less deep for the moment.

Now, is it possible to construct physically states that are only superpositions of the two lowest energy states?

Yes, of course, because we control the energy. For instance, in a gas at temperature T, we know that the ratio of populations of molecules in energy states E_1 and E_2 is given by Boltzmann's factor $n(E2)/n(E1) = e^{-(E2-E1)/kT}$. Therefore, inserting the above values,

• At 100 K: $N_A \approx N_S$, $n(E1)/n(E0) \approx 6\ 10^{-7}$ which is a small probability, one makes the approximation to consider it to be zero;

• If that's not enough, we can go down to 50 K, $p \approx 3\ 10^{-13}$, or to 25 K, $p \approx 10^{-25}$ (not a single molecule in 22.4 liters).

In order for the first excited level E_1 to play a role and for $p(E1)$ to be nonnegligible, one must reach temperatures of $T \approx 1300$ K.

Temperature gives us a cutoff on the finite number of significant components of the state vector.

We know this type of situation. Gravity holds us on the ground. If we navigate, we live in a effective two-dimensional world in first approximation. If we climb mountains or fly in a spacecraft, then we must worry about the third dimension.

7.3 Matrix quantum mechanics

7.3.1 Vectors

Consider a two-dimensional subspace of E_H generated by the set $\{|\psi_S\rangle\ ,\ |\psi_A\rangle\}$ which forms a basis of the subspace. We can represent an arbitrary state $|\psi\rangle$,

by a two-component vector,

$$|\psi\rangle = a|\psi_S\rangle + b|\psi_A\rangle\,, \quad \text{with } P(E_S) = |a|^2,\ P(E_A) = |b^2| \quad . \tag{7.3}$$

In a matrix representation, this is written as

$$|\psi_S\rangle \,:\, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |\psi_A\rangle \,:\, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad |\psi\rangle \,:\, \begin{pmatrix} a \\ b \end{pmatrix} \quad ,$$

$$\text{for instance,} \quad |\psi_{R/L}\rangle \,:\, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm 1 \end{pmatrix} \quad .$$

7.3.2 Hamiltonian

What is the expression of the Hamiltonian? Of course, not the full Hamiltonian of the molecule, but its restriction to the subspace of interest. The basis states are by definition energy eigenstates of $\hat{H}$ which is a diagonal matrix, and, in the subspace, it is the 2×2 matrix:

$$\hat{H} = \begin{pmatrix} E_0 - A & 0 \\ 0 & E_0 + A \end{pmatrix}. \tag{7.4}$$

The Schrödinger equation is

$$i\hbar \frac{d}{dt}|\psi(t)\rangle = \hat{H}|\psi(t)\rangle\,, \quad \text{with} \quad |\psi(t)\rangle = \begin{pmatrix} \alpha(t) \\ \beta(t) \end{pmatrix} \quad . \tag{7.5}$$

This gives us two uncoupled equations whose solutions are

$$|\psi(t)\rangle = e^{-i(E_0 t/\hbar)} \begin{pmatrix} a\, e^{i\omega t/2} \\ b\, e^{-i\omega t/2} \end{pmatrix} \quad , \tag{7.6}$$

where we have introduced the Bohr frequency $\omega = 2A/\hbar$.

7.3.3 Observables

In fact, in this subspace, any linear operator, any observable, is a 2×2 matrix. The "restriction" of an observable to the subspace is the first 2×2 block of the infinite matrix on the basis of our imaginary catalogue.

Now, how are we going to do calculations? In wave mechanics, we have a differential equation and a potential. What should we do here?

Simply make models of observables (i.e., 2×2 matrices) instead of making models of potentials as in wave mechanics.

Eigenvalues and eigenvectors of a Hermitian 2 × 2 matrix

Before we continue, it is useful to recall a few formulae on 2×2 matrices. Any Hermitian 2×2 matrix can be written as

$$\hat{A} = \begin{pmatrix} a & c\, e^{i\phi} \\ c\, e^{-i\phi} & b \end{pmatrix} \quad , \tag{7.7}$$

where a, b, c are real and ϕ is a phase.

The eigenvalues of $\hat{A}$ are

$$\lambda_{\pm} = \frac{1}{2}\left(a + b \pm \sqrt{(a-b)^2 + 4c^2}\right) \quad , \tag{7.8}$$

corresponding to eigenvectors

$$|\psi_+\rangle = \begin{pmatrix} \cos\theta \\ \sin\theta\, e^{-i\phi} \end{pmatrix} , \quad |\psi_-\rangle = \begin{pmatrix} -\sin\theta \\ \cos\theta\, e^{-i\phi} \end{pmatrix} , \quad \text{with } \tan 2\theta = \frac{2c}{(a-b)} \ . \tag{7.9}$$

Position observable

The following observable plays a central role here. We can define a "position" observable, or rather an observable "disposition of the particle with respect to the center" by the matrix

$$\hat{X} = x_0 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \tag{7.10}$$

where x_0 is a length parameter, grosso modo equal to the positions of the minima of the double well. The observable $\hat{X}$ has eigenvalues $\pm x_0$ and eigenvectors

$$|\psi_{\pm}\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm 1 \end{pmatrix} \ = \ |\psi_{R/L}\rangle \quad , \tag{7.11}$$

that is, the "classical" configurations of chapter 5.

This is indeed the restriction of the position observable $\hat{x}$ in the previous sense. It has the same structure as the first 2×2 block of equation (6.58) in the harmonic oscillator basis.

Strictly speaking, it is not the position observable but rather the disposition with respect to the center. Our assumptions forbid us to talk about the position with greater accuracy than the half-width of each well. This is the consequence of the Heisenberg uncertainty relations. If we prepare a wave function whose dispersion in x is much smaller than $\Delta x = a/2$ (i.e., the half width of one of the wells) then one cannot remain in the two-dimensional subspace. One must attain the levels E_1. We need more eigenfunctions in order to localize the particle better as one can understand in Figure 7.1.

7.3.4 Examples

We can do a series of simple exercises. Consider the state

$$|\psi(t)\rangle = \begin{pmatrix} a\, e^{i\omega t/2} \\ b\, e^{-i\omega t/2} \end{pmatrix} , \quad \text{with } \hbar\omega = 2A \quad . \tag{7.12}$$

The expectation value of X in this state is

$$\langle X\rangle = \langle\psi(t)|\hat{X}|\psi(t)\rangle = 2x_0 \mathcal{R}e(a^* b\, e^{-i\omega t}) \quad ; \tag{7.13}$$

that is, for $|\psi(0)\rangle = |\psi_R\rangle$ with $a = b = 1/\sqrt{2}$,

$$\langle X\rangle = |\langle\psi_R|\hat{X}|\psi(t)\rangle|^2 = x_0 \cos\omega t \quad , \tag{7.14}$$

and a probability

$$P_t(X = +x_0) = |\langle\psi_R|\psi(t)\rangle|^2 = \cos^2(\frac{\omega t}{2}) \quad , \tag{7.15}$$

which shows simply the inversion of the NH_3 molecule.

For the moment, there is nothing really new, but calculations are very simple. We have got rid of the potential and of wave functions. All the interesting physics is in the value of the parameter A which is given by experiment, and in the parameter x_0 which defines the size of the molecule.

This is a model. In order to improve it, we must increase the size of the Hilbert subspace, namely take more terms into account in the expansion (7.2).

7.3.5 Basis of classical configurations

It is interesting to make a change of basis and in particular to express vectors and operators in the alternative basis of classical equilibrium configurations $\{|\psi_R\rangle, |\psi_L\rangle\}$. We then express quantum effects in terms of classical situations.

The Hamiltonian is not diagonal in this basis, as opposed to X:

$$\hat{H} = \begin{pmatrix} E_0 & -A \\ -A & E_0 \end{pmatrix} , \quad \hat{X} = x_0 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad . \tag{7.16}$$

The off-diagonal terms are transition terms that allow transitions $L \leftrightarrow R$, the inversion of the molecule. If $A = 0$, the two classical configurations have the same energy.

The energy levels are the eigenvalues of H. The calculation is straightforward: we indeed obtain the eigenvalues and eigenvectors we know

$$\begin{aligned} E_- &= E_0 - A\,, \quad |\psi_S\rangle = (|\psi_R\rangle + |\psi_L\rangle)/\sqrt{2}, \\ E_+ &= E_0 + A\,, \quad |\psi_A\rangle = (|\psi_R\rangle - |\psi_L\rangle)/\sqrt{2} \quad . \end{aligned}$$

7.3.6 Interference and measurement

At this point, it is instructive to apply the principles. Suppose we start with an energy eigenstate, say $|\psi_S\rangle$,

$$|\psi_S\rangle = \frac{1}{\sqrt{2}}(|\psi_R\rangle + |\psi_L\rangle).$$

If we measure X, we can find $\pm x_0$ with probabilities $1/2 = (1/\sqrt{2})^2$.

Suppose the measurement has given the result $+x_0$; the state after the measurement is

$$|\psi_R\rangle = \frac{1}{\sqrt{2}}(|\psi_S\rangle + |\psi_A\rangle) \quad . \tag{7.17}$$

If we measure X again immediately afterwards, before the oscillation is appreciable, we find $+x_0$ with probability 1; the state after the measurement is $|\psi_R\rangle$.

Now, suppose that, on this new state $|\psi_R\rangle$, we measure not X but the energy E which we are sure was $E = E_S$ when we started. One can read in (7.17) that we do not always find E_S but the two possibilities E_S and E_A, each with a probability of $1/2$.

We see in this case how the measurement has perturbed the system. At the beginning, the state was $|\psi_S\rangle$; at the end it is a mixture of $|\psi_S\rangle$ and $|\psi_A\rangle$ in interference, for which $\langle E\rangle = (E_S + E_A)/2$.

All of this results from the superposition principle on one hand and the filtering of which a measurement consists.

We remark that, consequently, a position measurement implies a minimum energy exchange with the system. Here, on the average, the exchange of energy is equal to A.

7.4 NH_3 in an electric field

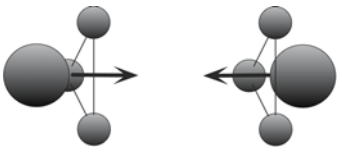


Fig. 7.2. The two classical configurations of the molecule NH_3 and the associated electric dipole.

Let us come back to NH_3. In order to make those energy or position measurements, we must control the energy in a more refined way than thermal motion. The temperature has allowed us to be in a subspace; now we want to manufacture individual states of a given energy.

We have seen that the NH_3 molecule has an electric dipole moment. We can make this electric dipole interact with an electric field. How can we describe this problem?

Classically, if a system with an electric dipole moment $\boldsymbol{D}$ is placed in an electric field $\boldsymbol{\mathcal{E}}$ it acquires a potential energy $W = -\boldsymbol{\mathcal{E}} \cdot \boldsymbol{D}$. If we assume that $\boldsymbol{\mathcal{E}}$ and $\boldsymbol{D}$ are parallel, this becomes

$$W = -\mathcal{E}\, D\ . \tag{7.18}$$

What happens if we place the NH_3 molecule in an electric field $\boldsymbol{\mathcal{E}}$? What is the quantum potential energy $\hat{W}$, which is related to the observable $\hat{D}$, according to the correspondence principle, by

$$\hat{W} = -\mathcal{E}\, \hat{D} \quad . \tag{7.19}$$

The answer is simple. In the particular configurations of interest here, $\hat{D}$ is simply proportional to $\hat{X}$, that is,

$$\hat{D} = q_0 \hat{X} = \begin{pmatrix} 0 & d_0 \\ d_0 & 0 \end{pmatrix}\ , \quad \hat{W} = \begin{pmatrix} 0 & -d_0\mathcal{E} \\ -d_0\mathcal{E} & 0 \end{pmatrix} \quad , \tag{7.20}$$

where q_0 is an effective charge and d_0 is an electric dipole moment that is measured experimentally: $d_0 \approx 3 \times 10^{-11}$ (eV)/(V/m) $= 5 \times 10^{-30}$ C.m.

In other words, if we measure X and we find $\pm x_0$ with some probabilities, a measurement of D will give $\pm d_0$ with the same probabilities. And $\hat{W}$ is the product of $\hat{D}$ times the number $\mathcal{E}$, the value of the electric field.

The only difficulty, here, is to accept that a good model for the observable $\hat{D}$, within our assumptions, is to be proportional to $\hat{X}$.

The potential energy observable $\hat{W}$ is simply the product of the observable $\hat{D}$ by the numerical value of the applied electric field.

The only real justification for this choice is that it works very well.

From then on, things are quite simple. The Hamiltonian of the molecule in a field is the sum of the free Hamiltonian (7.4) and the potential energy (7.20).

7.4.1 Uniform constant field

In a uniform constant field, if we set $\eta = d_0\mathcal{E}$, the Hamiltonian is

$$\hat{H} = \begin{pmatrix} E_0 - A & -\eta \\ -\eta & E_0 + A \end{pmatrix} \quad . \tag{7.21}$$

Finding the energy levels and the corresponding eigenstates amounts to diagonalizing this matrix.

Turning to (7.8) and (7.9), the eigenvalues and eigenvectors of $\hat{H}$ are

$$E_- = E_0 - \sqrt{A^2+\eta^2} \quad |\psi_-\rangle = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} , \tag{7.22}$$

$$E_+ = E_0 + \sqrt{A^2+\eta^2} \quad |\psi_+\rangle = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} \text{ with } \tan 2\theta = \eta/A \ . \tag{7.23}$$

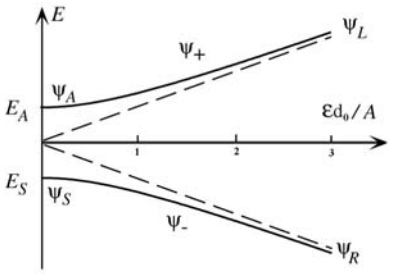


Fig. 7.3. Energy levels of a NH_3 molecule in an electric field.

The variation of these energy levels with the applied electric field $\mathcal{E}$ is represented in Figure 7.3. The validity of the result rests on the condition that E must not be large enough to reach the levels E_1. Otherwise one would have to take care of higher levels (4-, 6-, etc. level systems). The value of d_0 leaves a lot of space.

7.4.2 Weak and strong field regimes

It is interesting to consider two limits: the weak field and the strong field limits. The borderline between the two domains, weak and strong fields, is grosso modo $\mathcal{E} \sim A/d_0 \approx 1,7\,10^6$ V/m for NH_3. The quantity d_0^2/A is called the polarizability of the molecule. It is large for NH_3 in its ground state because the splitting A_0 is small.

We notice that if we consider excited levels E_1, $A_1 \approx 40\,A_0$, the polarizability is much smaller. The borderline is then around $7\,10^7$ V/m. For usual fields of 10^6 V/m, the polarization of the states E_1 is completely frozen, these states do not participate in energy exchanges through an electric field; our starting assumption is again consistent.

Weak field

In the weak field regime, such that $\mathcal{E} \ll d_0/A$ or $\theta \ll 1$, the levels and eigenstates are to lowest order in $\mathcal{E}$:

$$E_\mp \simeq E_0 \mp \left(A + \frac{d_0^2\mathcal{E}^2}{2A} \right) , \tag{7.24}$$

$$|\psi_-\rangle \simeq |\psi_S\rangle + \frac{d_0\mathcal{E}}{2A}|\psi_A\rangle , \quad |\psi_+\rangle \simeq |\psi_A\rangle - \frac{d_0\mathcal{E}}{2A}|\psi_S\rangle \ . \tag{7.25}$$

This is understandable. In the absence of a field, the molecule has a symmetric probability and $\langle D \rangle = 0$. The effect of the field is to polarize the molecule, which acquires a mean electric dipole moment proportional to the field $\langle D \rangle \approx \pm d_0^2 \mathcal{E}/A$, hence a quadratic result in $\mathcal{E}$.

Strong field

For strong fields, $\mathcal{E} \gg d_0/A$ or $\theta \simeq \pi/4$, the effect of the field dominates over the tunnel effect; the molecule is completely polarized. The eigenstates are close to the classical configurations $|\psi_{R/L}\rangle$ with $\langle D \rangle = \pm d_0$ and the energies are $E \pm d_0 E$; the response to the field is linear as for a classical dipole.

There is a competition between two effects:

• The tunnel effect tends to symmetrize the molecule, which results in a vanishing dipole moment $\Rightarrow \langle D \rangle = 0$.

• The field pulls the molecule toward the classical configurations $|\psi_{R/L}\rangle$, where it has a dipole moment $\langle D \rangle = \pm d_0$.

7.4.3 Other two-state systems

We have made progress on the ammonia molecule. All this could have been done with wave functions, but it would soon have been complicated. Here,the calculation of energy levels amounts to diagonalizing a matrix.

We come back below to the ammonia maser.

We insist that this is simply an example, because the same mathematics, finite dimensional matrix calculus, applies to many effects. Some of them are exactly finite-dimensional such as

• The spin 1/2 of particles such as electrons, protons, quarks, and so on

• The polarization of the photon

• The physics of "strange" neutral mesons $K_0 \rightleftharpoons \bar{K}_0$, and "beautiful" mesons $B_0 \rightleftharpoons \bar{B}_0$

• The universality of weak interactions and the mixing matrix of d, s, b quarks

• The quantum oscillations of neutrinos, which we show below.

It can also be an approximate model as for lasers, the chemical bond, and nuclear magnetic resonance.

7.5 The ammonia molecule in an inhomogeneous field

We come back to the case of the ammonia molecule. With the tools we have developed, we can understand the principle of the maser, which has been a revolution in the physics of radiowaves, in telecommunications, and in astrophysics. We do not prove technical results here. They are intuitive, and we come back to them in the similar, but simpler, case of the electron spin in chapter 11.

7.5.1 Force on the molecule in an inhomogeneous field

How can we separate energy eigenstates of the molecule? How can we prepare a sample of NH_3 in only one quantum state?

Consider for definiteness the weak field case. The eigenstates are close to $|\psi_S\rangle$ and $|\psi_A\rangle$ with energies

$$E_{\mp} = E_0 \mp \sqrt{A^2 + d_0^2 \mathcal{E}^2} \simeq E_0 \mp A \mp \frac{d_0^2 \mathcal{E}^2}{2A}. \tag{7.26}$$

The last term is simply the potential energy $V_{\mp}$ of the molecule in the field. We see that it is different according to the internal quantum state of the molecule.

We prepare a molecular beam. Suppose, first that all molecules are in the state $|\psi_S\rangle$. If these molecules which are "big" classical objects cross a region where there is an inhomogeneous field, their energy depends on the point where they are, therefore a force will act on them

$$F_- = -\nabla V_- = +\frac{d_0^2}{2A} \nabla \mathcal{E}^2. \tag{7.27}$$

Similarly, if they are all in the state $|\psi_A\rangle$ the force acting on them is

$$F_+ = -\nabla V_+ = -\frac{d_0^2}{2A} \nabla \mathcal{E}^2. \tag{7.28}$$

The force is different according to the internal state of the molecule. The two forces actually have opposite signs.

Therefore, the molecular beam will follow a different path according to whether the internal state of the molecules is $|\psi_S\rangle$ or $|\psi_A\rangle$. What is quantum mechanical is the internal state of the molecule, not the motion of its center of gravity.

What happens if the molecule is in a superposition of the states $|\psi_S\rangle$ and $|\psi_A\rangle$? This problem is slightly more complicated. The state of a molecule must be represented by a vector such as we have done above, but the components of this vector depend on the position $\boldsymbol{R}$ of the center of gravity of the molecule,

$$\begin{pmatrix} \Psi_1(\boldsymbol{R}, t) \\ \Psi_2(\boldsymbol{R}, t) \end{pmatrix} \quad .$$

The probabilistic interpretation is that $|\Psi_1|^2$ is the probability density of being at point $\boldsymbol{R}$ in the internal state $|\psi_S\rangle$, and similarly $|\Psi_2|^2$ is the probability density of being at point $\boldsymbol{R}$ in the state $|\psi_A\rangle$.

This can be proven in the similar, but simpler, case of the electron with its spin. One proves that the expectation values of the positions of the wave packets $\Psi_1(\boldsymbol{R}, t)$ and $\Psi_2(\boldsymbol{R}, t)$ each evolve according to Newton's laws with, respectively, the potentials (7.27) and (7.28). All happens as if different forces acted on these components.

Therefore an inhomogeneous field allows to perform the following operations:

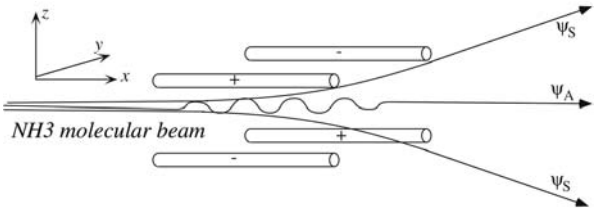


Fig. 7.4. Stabilization of the beam $|\psi_+\rangle$ and divergence of the beam $|\psi_-\rangle$ in an electric quadrupole field ($\mathcal{E}^2 \propto y^2 + z^2$).

1. One can perform a *spatial separation* of the molecules according to their internal state. This device is a filter that selects the states $|\psi_S\rangle$ and $|\psi_A\rangle$. Notice also that in a given field one can select any real linear superposition of $|\psi_S\rangle$ and $|\psi_A\rangle$.
2. Here we face an incredible phenomenon. The first time physicists saw this, in the experiment of Stern and Gerlach (chapter 11), they thought it was an experimental proof of the quantization of trajectories advocated by Bohr. There are only two quantum trajectories whereas classically, if the electric dipole moments were oriented at random there should be a continuous set of impacts on a screen.
3. But, if we select in space the states $|\psi_S\rangle$ and $|\psi_A\rangle$, that means we measure their energies. If a molecule arrives above, it has an internal quantum energy E_S; if a molecule arrives below, it has energy E_A. The respective numbers of them give the probabilities that in the incoming beam the molecules are in the states $|\psi_S\rangle$ and $|\psi_A\rangle$; we obtain information on the initial state.
4. This apparatus is a concrete example of a quantum mechanical measuring apparatus. It transfers internal quantum degrees of freedom into classical space properties.
5. It is also a device to prepare the molecules in the states $|\psi_S\rangle$ or $|\psi_A\rangle$, or in linear superpositions of them (it is sufficient to vary the value of the field $\mathcal{E}$ in the device).
   Here we can at last speak of the measurement problem in quantum mechanics. We see that in order to make the separation, one must have, for a given velocity, a minimal length, therefore a minimal time. A measurement is never instantaneous or pointlike. A measurement always has a finite extension in space and time.
   A little calculation (again with spin 1/2) shows that the condition for the measurement to be possible is universal. One must have the inequality

$$\Delta E \, \Delta t > \hbar \quad , \tag{7.29}$$

   where ΔE is the transverse kinetic energy acquired by the molecules and Δt the time they spend in the inhomogeneous field zone. This is called the "time-energy uncertainty relation" by analogy with the previous ones.

Actually, there a big difference because time t is not an observable (even though it can be measured).

7.5.2 Population inversion

Actually, in the case of NH_3, if one uses a molecular beam and an electric quadrupole field, the beam $|\psi_A\rangle$ is stable and can be focused, whereas the beam $|\psi_S\rangle$ is unstable and gets dispersed. What is interesting is that by this technique (a molecular beam, a diaphragm, an inhomogeneous field), one can perform what is called a *population inversion*. One can select all molecules in the state Ψ_A. One breaks the $(A) - (S)$ thermal equilibrium there was in the initial beam.

This is only one example of a technique of population inversion. There are many others.

We notice that in order for this population inversion to occur, it is not necessary to align the axis of a molecule along the field. Whatever the value of $\mathcal{E}$, the ψ_A beam is focused and ψ_S is dispersed.

7.6 Reaction to an oscillating field, the maser

In order to make a maser, we will force the molecules in the state $|\psi_A\rangle$ to give back their energy $2A$ by making a transition to the state $|\psi_S\rangle$. We must force them because spontaneously, these molecules do fall back in the state $|\psi_S\rangle$, but they do it very slowly. The mean time to do this transition is of the order of one month, which is much too long for our purpose.

In quantum physics, a system can absorb a photon of energy $h\nu$ and reach an excited state; it can fall back in the initial level by emitting a photon spontaneously. However, there exists a third mechanism which was understood by Einstein as soon as 1917, called stimulated emission. If the excited system is placed in an electromagnetic field properly tuned to the Bohr frequency, it can undergo a transition to the lower state very rapidly. This is done by a resonance mechanism.

Technically, this exercise is not much more difficult than the previous one, but it is a new problem. We describe the position of the problem and the result. Again, the calculation itself is done later on with spin 1/2 and magnetic resonance. We place the molecule in an oscillating field

$$\mathcal{E} = \mathcal{E}_0 \cos \omega t \quad ,$$

and we set $\eta = d_0 \mathcal{E}_0$. The Hamiltonian is

$$\hat{H} = \begin{pmatrix} E_0 - A & -\eta \cos \omega t \\ -\eta \cos \omega t & E_0 + A \end{pmatrix} \quad . \tag{7.30}$$

We see the difference with previous problems. The Hamiltonian H now depends explicitly on time. The system is not isolated, one cannot speak of

stationary states. We must solve the Schrödinger equation. If we write $|\psi(t)\rangle = a(t)|\psi_S\rangle + b(t)|\psi_A\rangle$, we must solve the equation $i\hbar(d/dt)|\psi(t)\rangle = \hat{H}|\psi(t)\rangle$ in order to determine the evolution of the system.

If we write the state vector of a molecule as

$$|\psi(t)\rangle = \begin{pmatrix} a(t) \\ b(t) \end{pmatrix} \quad , \tag{7.31}$$

the Schrödinger equation is a first-order differential system:

$$i\hbar\dot{a} = (E_0 - A)a - \eta b \cos\omega t \ , \tag{7.32}$$

$$i\hbar\dot{b} = (E_0 + A)b - \eta a \cos\omega t \ . \tag{7.33}$$

Setting $a(t) = e^{-i(E_0-A)t/\hbar}\alpha(t)$ and $b(t) = e^{-i(E_0+A)t/\hbar}\beta(t)$, we obtain:

$$2i\dot{\alpha} = -\omega_1 \beta \left(e^{i(\omega-\omega_0)t} + e^{-i(\omega+\omega_0)t}\right) \quad , \tag{7.34}$$

$$2i\dot{\beta} = -\omega_1 \alpha \left(e^{-i(\omega-\omega_0)t} + e^{i(\omega+\omega_0)t}\right) \quad . \tag{7.35}$$

This set of coupled equations involves three frequencies:

$$\omega, \quad \omega_0 = \frac{2A}{\hbar}, \quad \text{and } \omega_1 = \frac{\eta}{\hbar} = \frac{d_0\mathcal{E}_0}{\hbar} \quad . \tag{7.36}$$

Starting with state $|\psi_A\rangle$ at $t = 0$, $a(0) = 0$ and $b(0) = 1$, we want to calculate the probability $|a(t)|^2$ of finding the system in state $|\psi_S\rangle$ at time t.

Physically, this differential system corresponds to forced oscillations with a resonance phenomenon at $\omega = \omega_0$.[1] There is no analytic solution, however, one can obtain a good approximation in the vicinity of the resonance $\omega \sim \omega_0$, if we neglect terms that oscillate rapidly in time $e^{\pm i(\omega+\omega_0)t}$. This leads to an exactly soluble problem, for which we can give the solution.

Near the resonance, the transition probability $P_{A\to S}(t)$ that at time t the molecules undergo a transition to the state $|\psi_S\rangle$ under the influence of the oscillating field, and that they release their energy $2A = E_A - E_S$ is given by

$$P_{A\to S}(t) \simeq \frac{\omega_1^2}{(\omega-\omega_0)^2 + \omega_1^2} \sin^2\left(\sqrt{(\omega-\omega_0)^2 + \omega_1^2}\,\frac{t}{2}\right) \quad . \tag{7.37}$$

This formula is due to Rabi. As can be seen in Figure 7.5a, the probability $P_{A\to S}(t)$ oscillates in time between 0 and a maximum value $P_{\max}$ given by

$$P_{\max} = \frac{\omega_1^2}{(\omega-\omega_0)^2 + \omega_1^2} \quad .$$

When the frequency ω of the applied field is varied (Figure 7.5b), the maximum probability $P_{\max}$ has a characteristic resonant behavior, with a maximum equal to 1 at the resonance, that is, for $\omega = \omega_0$. The width at half maximum of the resonance curve is ω_1.

[1] Actually, there are two resonances at $\omega = \omega_0$ and $\omega = -\omega_0$, but the two values are equivalent for what concerns us here.

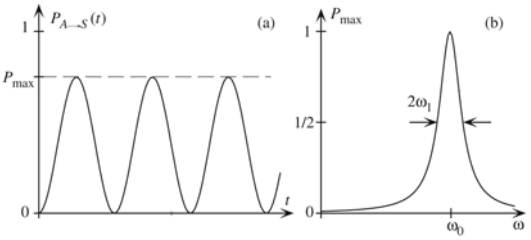


Fig. 7.5. Rabi oscillations: (a) Probability of finding the molecule in the state $|\psi_S\rangle$ as a function of time. (b) Resonance curve showing the maximum transition probability as a function of the external field frequency ω.

If the frequency of the external field is tuned in the vicinity of the resonance, $|\omega - \omega_0| \ll \omega_1$, practically all the molecules will release their energy $2A$ at the time $T = \pi/\omega_1$. This energy emission occurs in the form of an electromagnetic radiation of frequency $\nu = \omega_0/2\pi = 24$ GHz. This is called stimulated emission. The smaller ω_1, the narrower the resonance curve of Figure 7.5b is, and the greater the time to obtain that emission.

Therefore, if spontaneously the molecules decay in ≈1 month, with this phenomenon one can force them to do it quickly, and to beat the natural time constant of the process. This happens whatever the intensity of the exciting field, provided it is properly tuned.

7.7 Principle and applications of the maser

In practice, a NH_3 maser (microwave amplification by stimulated emission of radiation), which was invented by Townes in 1951, works in the following way, represented in Figure 7.6. One starts with a molecular beam of velocity v coming from an oven at a temperature of 100 K. One then separates the molecules in the state $|\psi_A\rangle$ by an electric quadrupole field. The beam then enters a high-frequency cavity where there is a field $\mathcal{E}_0 \cos \omega t$ and whose length L is adjusted[2] so that $L/v = T = (2n+1)\pi/\omega_1$. The outgoing molecules are in the state $|\psi_S\rangle$ and they have released their energy $2A$ in the cavity in the form of an electromagnetic radiation of frequency ω_0. In the setup made by Townes in 1951, a beam of 10^{14} molecules s^{-1} gave a power of 10^{-9} W on a single frequency 24 GHz with a width of 3000 Hz, a quality factor of $Q = 10^7$.

This maser effect, which is based on two elements, population inversion (which breaks the thermodynamic equilibrium) and stimulated emission which

[2] It is not necessary that L be exactly adjusted to the correct value; the transition probability is appreciable if one doesn't have the bad luck of falling on the unfavorable values $T = 2n\pi/\omega_1$. In practice a feedback device adjusts the length of the cavity so that the signal is maximum.

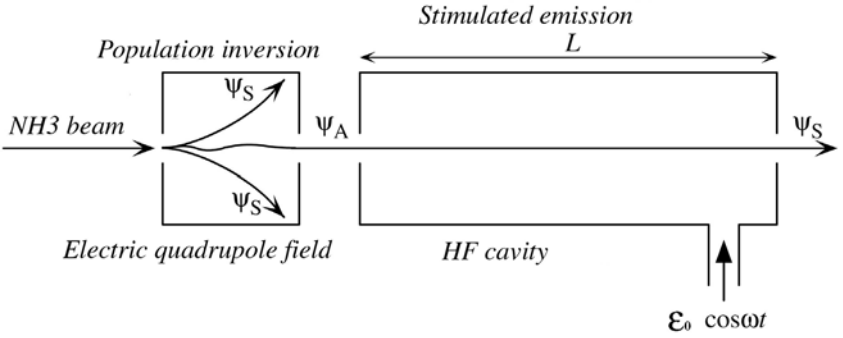


Fig. 7.6. Sketch of a NH_3 maser device.

generates the quick and coherent transition, was discovered by Townes in 1951 (1964 Nobel prize). It has numerous applications.

Lasers (light) work in a different frequency range, with a more sophisticated technique of population inversion, optical pumping. They are treated as three- or four-state systems. There is no basic difference from what we have seen otherwise.

There are essentially three types of applications of such devices.

7.7.1 Amplifiers

One can amplify in a selective way, and without any background noise, a very weak signal. Hence there are very important applications in radioastronomy in order to study the interstellar medium, as we show in chapter 14.

Initially, Townes used a molecular beam of 10^{14} molecules per second. Nowadays one uses solid-state masers, such as a crystal of ruby (a crystal of Al_2O_3 with Cr^{+3} ions of a concentration of $\sim 0,05\,\%$). This allows gains of 36 dB. Such masers were used in 1965 by A. Penzias and R.W. Wilson when they discovered the cosmic background radiation at 3 Kelvin, one of the major observational proofs in favor of the big bang, which we show in chapter 14.

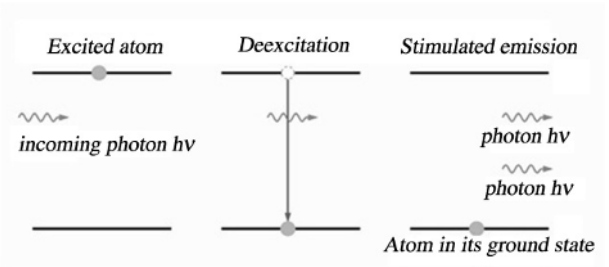


Fig. 7.7. Chain reaction of a tuned photon on a set of excited atoms.

The mechanism of the amplification is simple if one visualizes it in terms of photons. The population inversion is achieved on a macroscopic number of atoms, such as in a crystal of ruby. As shown in Figure 7.7, a photon coming from the interstellar medium will induce the transition of a first atom by stimulated emission. This results in two tuned photons both of which can in turn generate stimulated emission on other atoms, leading to four photons, and so on. The chain reaction yields the amplification.

The device presented in Figure 7.6 is not an amplifier but an emitter or an oscillator, because the outgoing intensity is independent of the incoming intensity. In order to see the system act as an amplifier, one must calculate its response to an incoherent signal spread out in frequency. Our calculation only concerns a monochromatic coherent field.

7.7.2 Oscillators

In Figure 7.6, we actually see an oscillator. A field as small as one wishes can be self-produced in the cavity; one evacuates the electromagnetic wave of frequency 24 GHz, which results in a very monochromatic output wave.

7.7.3 Atomic clocks

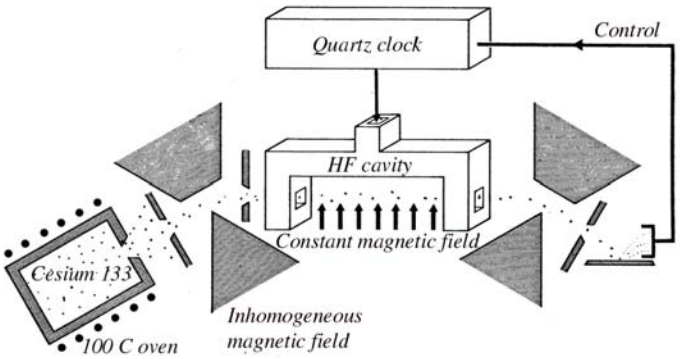


Fig. 7.8. Sketch of an atomic clock. (Courtesy of Patrick Bouchareine.)

As oscillators, masers have allowed the construction of atomic clocks, which are the present standard time-keeping devices. Such devices use a jet of cesium atoms (isotope ^{133}Cs). The ground state of cesium is split by the hyperfine interaction, which we describe later on. The physical origin of this effect is magnetic instead of being electric. It results from the magnetic interaction of the nuclear spin and the spin of the valence electron, and a splitting in two

levels $|g_1\rangle$ and $|g_2\rangle$, of energies E_1 and E_2. Since 1967, the definition of the hertz is given by the fact that the ^{133}Cs hyperfine frequency $\nu_{12} = (E_2 - E_1)/h$ is equal to 9,192,631,770 Hz. It is a considerable improvement compared to astronomical definitions.

In order to transform the above device into an atomic clock, we prepare a jet of cesium atoms in the state $|g_1\rangle$. These atoms cross a cavity inside which one injects an electromagnetic wave of frequency ν, and one adjusts ν in order to maximize the number of outgoing atoms in the state $|g_2\rangle$. The frequency ν is thus locked at ν_{12}. One can measure a time interval by simply counting the number of oscillations during that time interval. The principle is shown in Figure 7.8. Atomic clocks were first developed in the 1950s by Zacharias at the National Institute of Standards and Technology (NIST). The clock NIST-7 together with ten others in the world thus serves as a "time keeper."

Present clocks have a relative accuracy of 10^{-15}, so that the time standard is the most accurate of all standards. This explains that, in 1983, the definition of the meter was changed by fixing the value of the velocity of light to an integer value: $c = 299,792,458$ ms^{-1}.

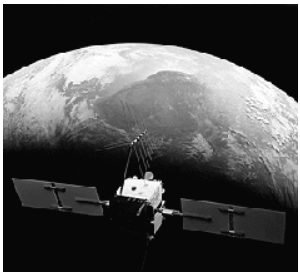


Fig. 7.9. Satellite of the GPS Navstar system. There are 24 of them orbiting around the Earth.

Such accuracy is essential both in applied physics and in a variety of technical devices, such as positioning and navigation, be it terrestrial with the GPS system, or on satellites. They are used to guide ships, they equip airplanes. Since 1990, they have appeared in automobiles.

Perhaps more amazing is their use in space navigation. The recent mission of the Cassini spacecraft, launched in 2000, saw, on January 14, 2005, the Huygens probe land on the planet Titan. Titan is a satellite of Saturn that is similar to the Earth because of its atmosphere, though it is much colder (in some sense it is an Earth in a refrigerator). This has yielded an impressive number of results and will continue to do so. In the final phase, the probe had to guide itself by its own means and land properly (the time of propagation of a signal to the earth is one hour). (See http:/www.esa.int/SPECIALS/Cassini-Huygens/index.html.)

7.7.4 Tests of relativity

As soon as atomic clocks appeared, people thought of directly verifying the predictions of relativity on time, in particular the twin paradox. In this "paradox" the twin who has traveled comes back younger. His proper time is shorter than the time of the other twin who stayed on Earth. His clock must therefore be late compared to the clocks that remained on Earth,

$$t = \frac{t_0}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad \Delta t = t - t_0 \simeq t_0 \frac{v^2}{2c^2} , \tag{7.38}$$

where t is the time measured by the twin who stayed on Earth, and t_0 the time of the twin who traveled.

In 1975, Alley sent 4 atomic cesium clocks for a 15-hour flight on an airplane and compared them with 11 clocks that stayed on Earth. Those clocks should have been late. Not at all; they they had gained 47 10^{-9} s!

In fact, there is an effect of the opposite sign due to gravitation and predicted by general relativity:

$$\left(\frac{\Delta t}{t}\right)_{GR} = \frac{\Delta \Phi}{c^2} , \quad \left(\frac{\Delta t}{t}\right)_{SR} = -\frac{v2}{2c^2} ,$$

where Φ is the gravitational potential. In the conditions of flight of Alley's clocks, the result is the sum of

$$\left(\frac{\Delta t}{t}\right)_{GR} = 53\ 10^{-9}\, s , \quad \left(\frac{\Delta t}{t}\right)_{SR} = -6\ 10^{-9}\, s .$$

The predictions of relativity are verified here with an accuracy of 1%.

These effects were measured in 1979 by R.F.C. Vessot and collaborators.[3] A hydrogen maser was sent to an altitude of 10,000 km by a Scout rocket, and the variation in time of its frequency was made as the gravitational potential increased (algebraically). There are many corrections, in particular due to the Doppler effect of the spacecraft and to the Earth's rotation. It was possible to test the predictions of general relativity on the variation of the pace of a clock as a function of the gravitational field with a relative accuracy of 7×10^{-5}. This was done by comparison with atomic clocks, or masers, on Earth. Up to now, it has been one of the best verifications of general relativity. The recording of the beats between the embarked maser and a test maser on Earth is shown in Figure 7.10. (These are actually beats between signals that are first recorded and then treated in order to take into account all physical corrections.) A simple calculation shows that with this accuracy, one must specify the exact

[3] R.F.C. Vessot, M.W. Levine, E.M. Mattison, E.L. Blomberg, T.E. Hoffman, G.U. Nystrom, B.F. Farrel, R. Decher, P.B. Eby, C.R. Baugher, J.W. Watts, D.L. Teuber, and F.D. Wills, "Test of Relativistic Gravitation with a Space-Borne Hydrogen Maser", Phys. Rev. Lett. **45**, 2081 (1980).

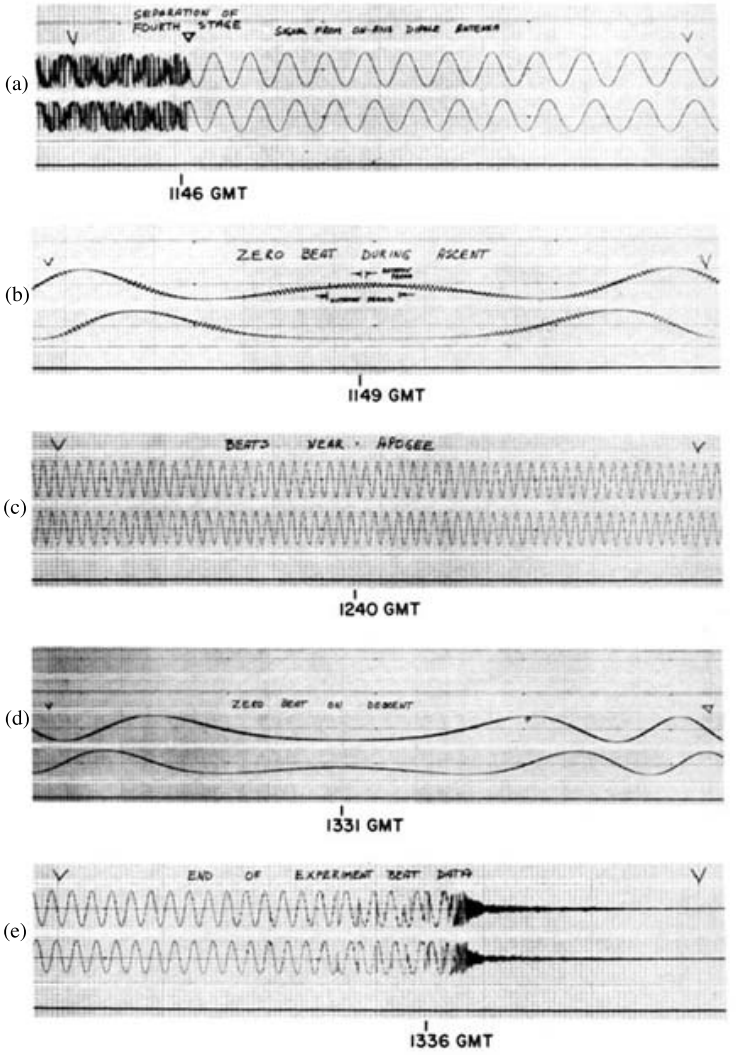


Fig. 7.10. Beats between a maser onboard the spacecraft launched by a Scout rocket and a maser on Earth at various instants in GMT. (a) Signal of the dipole antenna; the pointer shows the delicate moment when the spacecraft separated from the rocket (it was important that the maser onboard had not been damaged by vibrations during takeoff). During this first phase, the special relativity effect due to the velocity is dominant. (b) Time interval of "zero beat" during ascent when the velocity effect and the gravitational effect, of opposite signs, cancel each other. (c) Beat at the apogee, entirely due to the gravitational effect of general relativity. Its frequency is 0.9 Hz. (d) Zero beat at descent. (e) End of the experiment. The spacecraft enters the atmosphere and the maser onboard ceases to work. (Courtesy of R.F.C. Vessot.)

altitude of a clock up to one meter or so, in order to know one's time. Time is longer at the top of a conference hall than at the bottom. Students know this very well.

7.8 Neutrino oscillations

To finish this chapter, we consider a quantum oscillation effect that is completely contrary to common intuition. This is an oscillation between two or more types of pointlike elementary particles. The effect consists of the fact that if we prepare a particle in a vacuum, it can transform spontaneously and periodically into one or several types of different elementary particles. In other words, we face a phenomenon of "successive periodic hermaphroditism" between elementary particles.[4]

7.8.1 Lepton families

In beta radioactivity or, more generally, in weak interactions, the electron is always associated with a neutral particle, the neutrino ν_e. There exists in Nature another particle, the μ lepton, or muon, discovered in 1937, whose physical properties seem to be completely analogous to those of the electron, except for its mass $m_\mu \simeq 200\, m_e$. The muon has the same weak interactions as the electron, but it is associated with a different neutrino, the ν_μ.

A neutrino beam produced in an accelerator can interact with a neutron (n) in a nucleus and give rise to the reactions

$$\nu_e + n \to p + e \quad \text{and} \quad \nu_\mu + n \to p + \mu \ , \tag{7.39}$$

whereas the reactions $\nu_e + n \to p + \mu$ or $\nu_\mu + n \to p + e$ are never observed. The reactions (7.39) are used in practice in order to detect neutrinos.

Similarly, a π^- meson can decay via the modes

$$\pi^- \to \mu + \bar{\nu}_\mu \text{ (dominant mode), and } \pi^- \to e + \bar{\nu}_e \ , \tag{7.40}$$

whereas $\pi^- \to \mu + \bar{\nu}_e$ or $\pi^- \to e + \bar{\nu}_\mu$ are never observed. This is how one can produce neutrinos abundantly (it is easy to produce π mesons). In (7.40), we have introduced the antiparticles $\bar{\nu}_\mu$ and $\bar{\nu}_e$. There is a (quasi) strict symmetry between particles and their antiparticles, so that, in the same way as the electron is associated with the neutrino ν_e, the antielectron, or positron, e^+ is associated with the antineutrino $\bar{\nu}_e$. One can observe the "charge-conjugate" reactions of (7.39) and (7.40)

$$\bar{\nu}_e + p \to n + e^+ \, , \quad \bar{\nu}_\mu + p \to n + \mu^+ \, , \quad \text{and} \quad \pi^+ \to \mu^+ + \nu_\mu \quad . \tag{7.41}$$

[4] Successive hermaphroditism exists in zoology, for instance, in several fish species.

In all what follows, what we say about neutrinos holds symmetrically for antineutrinos.

The experimental discovery of the electron antineutrino is due to Cowan and Reines in 1956. They operated near the Savannah River nuclear reactor, and they observed the reaction

$$\bar{\nu}_e + p \to n + e^+ , \quad \text{and not} \quad \bar{\nu}_e + n \to p + e \quad . \tag{7.42}$$

The $\bar{\nu}_e$ antineutrinos came from the many beta decays of the type $n \to p + e + \bar{\nu}$ in the reactor. (The second reaction in (7.42) does not occur because the electron is associated with its neutrino and not with the antineutrino $\bar{\nu}_e$ which is the partner of the antielectron e^+.)

In 1975, a third lepton, the τ, was discovered, beside the electron and muon. It is much more massive, $m_\mu \simeq 3500\, m_e$, it is associated with its own neutrino ν_τ, and it obeys the same physical laws as the two lighter leptons, except for mass effects. Since the 1990s, measurements at the LEP colliding ring in CERN have shown that these three neutrinos ν_e, ν_μ, ν_τ (and their antiparticles) are the only ones of their kinds (at least for masses less that 100 GeV/c^2).

For a long time, particle physicists believed that neutrinos were massless particles, as is the photon. In any case, their masses (multiplied by c^2) are considerably smaller than the energies involved in experiments where they are observed. Therefore, many experimental limits on these masses are consistent with zero. However, both theoretical and cosmological arguments suggested that this might not be the case ($m \neq 0$). The proof that neutrino masses are not all zero is a great discovery of the 1990s.

For this discovery, the 2002 Nobel prize was awarded to Raymond Davis, Jr., a pioneer of this physics, and to Masatoshi Koshiba who led experiments performed in Japan with detectors particularly well adapted to that kind of physics. The Japanese experiments have gathered an impressive amount of results.

The mass differences of neutrinos can be measured by a quantum oscillation effect which we explain below. The idea is that the "flavor" neutrinos ν_e, ν_μ, and ν_τ, which are produced or detected experimentally are not eigenstates of the mass, but rather linear combinations of mass eigenstates ($|\nu_1\rangle, |\nu_2\rangle, |\nu_3\rangle$), with masses m_1, m_2, and m_3.

The neutrinos observed on Earth have various origins. They can be produced in accelerators, in nuclear reactors, and also in the atmosphere by cosmic rays, or in thermonuclear reactions inside stars, in particular the core of the sun, and in supernovae explosions.

7.8.2 Mechanism of the oscillations; reactor neutrinos

Consider first oscillations between two types of neutrinos, the ν_e and the ν_μ. This simple case allows us to understand the underlying physics of the general case.

In all what follows, we assume that if m is the neutrino mass and p and E its momentum and energy, the mass is so small, compared to its energy (divided by c^2), that the energy of a neutrino of mass m and momentum p is

$$E = \sqrt{p^2c^2 + m^2c^4} \simeq pc + \frac{m^2c^4}{2pc} \quad , \tag{7.43}$$

and the neutrino propagates to very good approximation at the velocity of light c.

Let $\hat{H}$ be the Hamiltonian of a free neutrino of momentum p, which we assume to be well defined. We note $|\nu_1\rangle$ and $|\nu_2\rangle$ the two eigenstates of $\hat{H}$:

$$\hat{H}|\nu_j\rangle = E_j|\nu_j\rangle \ , \quad E_j = pc + \frac{m_j^2c^4}{2pc} \ , \quad j = 1, 2 \ .$$

m_1 and m_2 are the two masses of the states $|\nu_1\rangle$ and $|\nu_2\rangle$, and we assume $m_1 \neq m_2$.

The oscillations of freely propagating neutrinos come from the following quantum effect. If the physical states of the neutrinos that are produced (reactions (7.40)) or detected (reactions (7.39)) are not $|\nu_1\rangle$ and $|\nu_2\rangle$, but linear combinations of these:

$$|\nu_e\rangle = |\nu_1\rangle \cos\theta + |\nu_2\rangle \sin\theta \ , \quad |\nu_\mu\rangle = -|\nu_1\rangle \sin\theta + |\nu_2\rangle \cos\theta \ , \tag{7.44}$$

where θ is a mixing angle to be determined. These linear combinations of energy eigenstates will oscillate in time and will lead to measurable effects.

We analyze the data obtained with nuclear reactors. The average energy of the (anti-)neutrinos produced in reactors is $E = 4$ MeV, with a dispersion of the same order.

At time $t = 0$, one produces a neutrino of momentum p in the state $|\nu_e\rangle$. This state is

$$|\nu(0)\rangle = |\nu_e\rangle = |\nu_1\rangle \cos\theta + |\nu_2\rangle \sin\theta \quad . \tag{7.45}$$

At time t, we therefore have

$$|\nu(t)\rangle = |\nu_1\rangle \cos\theta \exp(-i\ E_1 t/\hbar) + |\nu_2\rangle \sin\theta \exp(-i\ E_2 t/\hbar) \quad .$$

The probability amplitude to find this neutrino in the state $|\nu_\mu\rangle$ at time t is

$$a_\mu = \langle\nu_\mu|\nu(t)\rangle = \cos\theta \sin\theta \, (-\exp(-i\ E_1 t/\hbar) + \exp(-i\ E_2 t/\hbar)) \quad .$$

Hence, the probability to detect a neutrino ν_μ at time t

$$P_\mu = |a_\mu|^2 = \sin^2 2\theta \sin^2((E_1 - E_2)t/2\hbar) \quad .$$

We have

$$E_1 - E_2 = \frac{(m_1^2 - m_2^2)c^4}{2pc} \simeq \frac{\Delta m^2 c^4}{2E} \quad ,$$

where $E = pc$ is the energy of the neutrino beam.

One defines the *oscillation length* L_{12} by

$$L_{12} = \frac{4\pi\hbar pc}{(m_1^2 - m_2^2)c^3} \simeq \frac{4\hbar E}{\Delta m^2 c^3} \quad , \tag{7.46}$$

where $\Delta m^2 \equiv (m_1^2 - m_2^2)$.

For an energy $E = pc = 4$ MeV and a mass difference $\Delta m^2 c^4 = 10^{-4}$ eV^2, we obtain an oscillation length $L_{12} \simeq 100$ km.

Therefore, the probability P_μ to detect a neutrino in the state $|\nu_\mu\rangle$ at time t is

$$P_\mu = \sin^2 2\theta \sin^2 \left(\frac{\pi ct}{L_{12}}\right) \quad . \tag{7.47}$$

The probability P_e to find the system in the state $|\nu_e\rangle$ is of course

$$P_e = 1 - P_\mu = 1 - \sin^2 2\theta \sin^2 \left(\frac{\pi ct}{L_{12}}\right) \quad . \tag{7.48}$$

If the detector is at a distance ℓ from the production point, the probabilities are therefore

$$P_\mu = \sin^2 2\theta \sin^2 \left(\frac{\pi \ell}{L_{12}}\right) , \quad P_e = 1 - P_\mu \quad .$$

We notice that a ν_μ energy of only 4 MeV is below the threshold of the reaction $\nu_\mu + n \to p + \mu$. Therefore one cannot measure the ν_μ flux in this reaction with reactor neutrinos, owing to energy conservation.

Results

Several experiments on neutrinos produced by nuclear energy plants have been performed in particular in Chooz and in Bugey in France. The most recent data come from the KamLAND collaboration, in Japan. The results are given in Figure 7.11.

For a maximum mixing $\theta = \pi/4$, we have $P_\mu = \sin^2(\pi\ell/L_{12})$.

- In all experiments except KamLAND, the distance is smaller than 1 km. Therefore, in all of these experiments $P_\mu = \sin^2(\pi\ell/L_{32}) \leq 10^{-3}$, and $P_e = 1 - P_\mu$. The oscillation effect is not detectable if the estimate $|\Delta m^2|\, c^4 \sim 10^{-4}$ eV^2 is correct.
- In order to see an effect, one must have $P_\mu \geq 0.1$; that is, $\sin^2(\pi\ell/L_{12}) \geq 0.1$ or $\ell \geq 10$ km. The typical distances necessary in order to observe this phenomenon are of the order of the oscillation length.

The KamLAND experiment, which was performed in 2002, and published in January 2003, is a collaboration between Japanese, American, and Chinese physicists. The detector is a spherical 1000-m^3 container filled with liquid

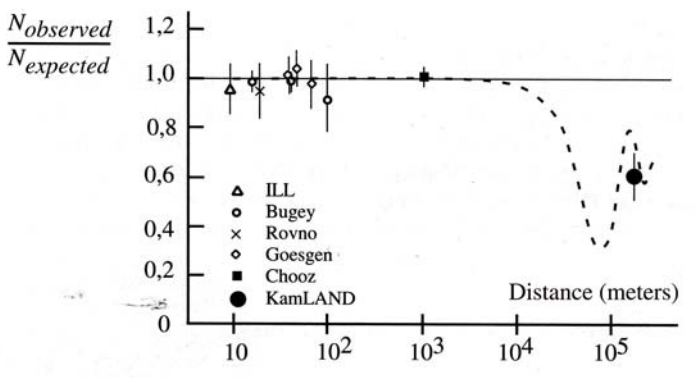


Fig. 7.11. Ratio between the numbers of observed electron neutrinos $\bar{\nu}_e$ and those expected in the absence of oscillations as a function of the distance ℓ to the reactor. The dashed curve is a best fit to solar neutrino data. All experiments except the most recent one, KamLAND, give a probability consistent with 1. Notice that the oscillations are quite rapidly damped. This comes from the dispersion in energy of the beams of reactor neutrinos, which is impossible to control. This damping can be calculated and it is taken into account in practice.

scintillator (an organic C-H liquid). The name means KAMioka Liquid scintillator Anti-Neutrino Detector.[5] The experiment consisted of measuring the neutrinos coming from all the (numerous) reactors in Japan and neighboring countries, which amounts grosso modo to taking an average distance of $\ell = 180$ km. The experiment gives the result $P_e = 0.61 \pm 0.1$.

For $\Delta m^2 c^4 = 10^{-4}$ eV^2, $\sin 2\theta = 1$, and $\ell = 180$ km, we obtain $P_e = 0.63$ in excellent agreement with the experiment.

Putting together these data and the results of numerous experiments performed on solar neutrinos, the physicists of Kamland came to the following results,

$$\Delta m_{12}^2 c^4 = (7.1 \pm 0.4) \times 10^{-5}\,, \quad \text{eV}^2 \qquad \tan^2 \theta_{12} = 0.45 \pm 0.02 \quad .$$

For these values, one obtains a central value $P_e = 0.49$, which is also in agreement with the measurement of Figure (7.11).

7.8.3 Successive hermaphroditism of neutrinos

The great lines of the general problem of three neutrino species are the following.

[5] KamLAND Collaboration, *Phys. Rev.* **90**, p. 021802 (2003); see also http:/kamland.lbl.gov/.

We name $|\nu_\alpha\rangle$, $\alpha = e, \mu, \tau$ the "flavor" neutrinos and $|\nu_i\rangle$, $i = 1, 2, 3$ the mass eigenstates. These two bases are related to each other by the Maki–Nagawaka–Sakata (MNS) matrix $\hat{U}$,

$$|\nu_\alpha\rangle = \sum_{i=1}^{3} U_{\alpha i} |\nu_i\rangle \ , \quad \hat{U} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \quad . \tag{7.49}$$

This matrix is unitary ($\sum_i U^*_{\beta i} U_{\alpha i} = \delta_{\alpha\beta}$) and it can be written as

$$\hat{U} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13} e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13} e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad ,$$

where $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$. The complete experimental solution of the problem would consist of measuring the three mixing angles $\theta_{12}, \theta_{23}, \theta_{13}$, the phase δ, and the three masses m_1, m_2, m_3.

The Super-Kamiokande experiment, performed in 1998, consists of detecting "atmospheric" neutrinos. Such neutrinos are produced in the collision of high-energy cosmic rays with nuclei in the atmosphere at high altitudes. In a series of reactions, $\pi^\pm$ mesons are produced abundantly, and they decay through the chain:

$$\pi^- \to \mu^- + \bar{\nu}_\mu \ , \quad \text{followed by} \quad \mu^- \to e^- + \bar{\nu}_e + \nu_\mu \ , \tag{7.50}$$

and an analogous chain for π^+ mesons. The neutrino fluxes are detected in an underground detector by the reactions (7.39) and (7.41).[6]

The Super-Kamiokande experiment consists of varying the time of flight of the neutrinos by measuring selectively the direction they come from, as indicated in Figure 7.12. This is a difficult but feasible measurement.

The neutrinos coming from above ($\cos\alpha \sim 1$) have traveled a distance equal to the atmospheric height plus the depth of the detector, whereas those coming from the bottom ($\cos\alpha \sim -1$) have crossed the diameter of the Earth (13,400 km). Given the weakness of the interaction of neutrinos with matter, one can consider that the neutrinos propagate freely on a measurable distance between a few tens of km and 13,400 km.

The neutrino energies are typically 4 GeV in this experiment.

One defines three oscillation lengths

$$L_{ij} = \frac{4\pi\hbar p}{(m_j^2 - m_i^2)c^2} \simeq \frac{4\pi\hbar E}{(m_j^2 - m_i^2)c^3} \quad . \tag{7.51}$$

The oscillation lengths are proportional to the energy. We can therefore scale the previous results by a factor of 1000 from 4 meV to 4 GeV. We see that for

[6] The detector contains 50,000 tons of water, and 11,500 photomultipliers detect the Čerenkov radiation of the final electrons or muons.

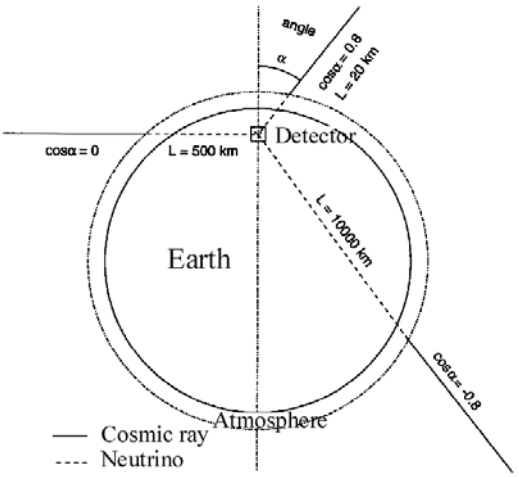


Fig. 7.12. Production of atmospheric neutrinos in collisions of cosmic rays with terrestrial atmospheric nuclei. The underground detector measures the flux of electron and muon neutrinos as a function of the zenithal angle α. (Figure due to André Rougé.)

$\Delta\, m_{12}^2\, c^4 = 10^{-4}\,\mathrm{eV}^2$ we have $L_{12} = 100,000$ km, and for $\Delta\, m_{23}^2\, c^4 = 10^{-3}\,\mathrm{eV}^2$ we have $L_{23} = 10,000$ km.

The condition $\sin^2(\pi d/L) \geq 0,1$ is equivalent to $d \geq 0,1\,L$ or $d_{12} \geq 10,000$ km in the first case, and $d_{23} \geq 1000$ km in the second case. At those energies the oscillations $\nu_e \rightleftharpoons \nu_\mu$ seen previously cannot be seen on terrestrial distances, whereas if $\Delta\, m_{23}^2\, c^4 = 10^{-3}\,\mathrm{eV}^2$, terrestrial distance scales allow us to observe "$(1 \leftrightarrow 3)$" oscillations.

The data are consistent with the fact that one observes a $\nu_\mu \leftrightarrow \nu_\tau$ oscillation, no $\nu_e \leftrightarrow \nu_\tau$ oscillation , and no $\nu_e \leftrightarrow \nu_\mu$ oscillations.

The complete results published by the Super-Kamiokande experiment are

$$|\Delta m_{23}^2| = 2,5 \times 10^{-3}\, eV^2/c^4\,, \quad \theta_{23} = \pi/4\,,\ \theta_{13} = 0 \quad .$$

In 1998, the first uncontroversial observation of the oscillation $\nu_\tau \rightleftharpoons \nu_\mu$ was announced in Japan by the Super-Kamiokande experiment (Y. Fukuda et al., *Phys. Rev. Lett.* **81**, 1562 (1998)). About 60 ν_τ's were also detected, but this number is too small to give further information. An accelerator experiment confirmed the results afterwards (K2K collaboration, M.H. Ahn et al., *Phys. Rev. Lett.* **90**, 041801 (2003)).

Very many experimental results come from solar neutrinos, which we have not dealt with here. This problem is extremely important, but somewhat too complex for our purpose. The pioneering work is due to Davis in his 1964 celebrated paper (R. Davis Jr., *Phys. Rev Lett.* **13**, 303 (1964)). Davis operated on a ^{37}Cl perchlorethylene detector and counted the number of ^{37}Ar atoms

produced. In 25 years, his overall statistics has been 2200 events, in other words, one atom every 3 days.

In 1991, the SAGE experiment done with gallium confirmed the deficit (A. I. Abasov et al., *Phys. Rev Lett.* **67**, 3332 (1991) and J. N. Abdurashitov et al., *Phys. Rev Lett.* **83**, 4686 (1999)).

In 1992, the GALLEX experiment, using a gallium target in the Gran Sasso also confirmed the solar neutrino deficit (P. Anselmann et al., *Phys. Lett.*, **B285**, 376 (1992)).

In 2001, the Sudbury Neutrino Observatory (SNO) gave decisive experimental results on solar neutrinos (Q.R. Ahmad et al., *Phys. Rev. Lett.*, **87**, 071307 (2001) and **89**, 011301 (2002)). See also M.B. Smy, *Mod. Phys. Lett.*, **A 17**, 2163 (2002).

The amazing phenomenon in this example is that quantum oscillations are not restricted to "geometrical" structures such as the NH_3 molecule. Oscillations appear between elementary pointlike particles of different species!

We notice that in the case of neutrinos, quantum oscillation phenomena can only be observed at macroscopic distances because of the smallness of mass differences. It is amusing that phenomena that are so far from our usual intuition can only be observed at very large distance scales.

8

Algebra of observables

The main subject that concerns us here is called the algebra of observables, and the arrival of the unusual person who was Paul Adrien Maurice Dirac on the scene of quantum mechanics during the summer of 1925.

8.1 Commutation of observables

What is the problem? Our axioms may be very elegant and profound, but they seem to be deprived of substance. Within this framework, how does one treat a particular case? What are the Hamiltonian and the observables for a given system? In wave mechanics, we have operators $\hat{p}_x = -i\hbar(\partial/\partial x)$, $\hat{H} = p^2/2m + V$ and we solve differential equations, but what should we do here? It is true that for the ammonia molecule in the previous Chapter, we have guessed quite easily how to proceed, but what is the general method?

The answer is rather unusual: in the general formalism of Hilbert space, the structures that will play a key role and allow us to perform calculations are not the special form of such or such an observable, but rather the algebraic relations between the observables, in particular their commutation relations.

We know that in general two observables $\hat{A}$ and $\hat{B}$ do not commute. Their commutator $[\hat{A}, \hat{B}]$ is defined by

$$[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A} \quad . \tag{8.1}$$

8.1.1 Fundamental commutation relation

We have already seen that between position and momentum observables, we have, in wave mechanics

$$[\hat{x}, \hat{p}_x] = i\hbar\hat{I} \quad , \tag{8.2}$$

where $\hat{I}$ is the identity operator.

This relation is independent of the representation we use. It is the same in Dirac's representation.

This fundamental commutation relation will now be the definition of position and momentum observables along the same axis. More precisely, in three-dimensional space, if we note $\hat{x}_j$ $(j = 1, 2, 3)$ the components of a position vector and similarly for momentum $\hat{p}_j$, the set

$$[\hat{x}_j, \hat{p}_k] = i\hbar\delta_{j,k}\hat{I}, \quad [\hat{x}_j, \hat{x}_k] = [\hat{p}_j, \hat{p}_k] = 0 \tag{8.3}$$

defines a couple of vector observables of position $\hat{\boldsymbol{x}}$ and momentum $\hat{\boldsymbol{p}}$. And that's what we use in order to do calculations.

8.1.2 Other commutation relations

For observables that have a classical analogue, we still use the correspondence principle. For instance, for the angular momentum

$$\hat{\boldsymbol{L}} = \hat{\boldsymbol{r}} \times \hat{\boldsymbol{p}}$$

one can easily obtain, using the fundamental relation (8.2), the commutation relations

$$[\hat{L}_x, \hat{L}_y] = i\hbar\hat{L}_z \ , \tag{8.4}$$

and two others related by cyclic permutations. The three relations can be put together in the form:

$$\hat{\boldsymbol{L}} \times \hat{\boldsymbol{L}} = i\hbar\hat{\boldsymbol{L}} \quad . \tag{8.5}$$

We show in the next Chapter that what defines in general an angular momentum observable $\hat{\boldsymbol{J}}$ is not the particular form $\hat{\boldsymbol{L}} = \hat{\boldsymbol{r}} \times \hat{\boldsymbol{p}}$ (valid for a particle in space), but the algebraic relations

$$\hat{\boldsymbol{J}} \times \hat{\boldsymbol{J}} = i\hbar\hat{\boldsymbol{J}} \quad . \tag{8.6}$$

We show how this algebra allows us to calculate numbers, and furthermore that it is more general than $\hat{\boldsymbol{L}} = \hat{\boldsymbol{r}} \times \hat{\boldsymbol{p}}$ and leads to the existence of purely quantum angular momenta, which have no classical analogues, as the spin 1/2 of the electron.

When there is no classical analogue, there is no real substitute to guessing or postulating the corresponding algebra of observables. If it works, it's OK, if it doesn't we must postulate something else.

Here, we illustrate this on some simple and important results. First, we give the general proof of uncertainty relations for any couple of variables. Then we consider the time evolution of the expectation value of a physical quantity and the Ehrenfest theorem, which enables us to see how the classical limit appears in quantum mechanics. Next, we give an algebraic solution of the harmonic oscillator problem due to Dirac, which plays a crucial role in quantum field theory. Next, we say a few words about observables that commute, in order to have the concept of a complete set of commuting observables, which is essential in order to treat problems in several variables, in particular, three-dimensional problems. Finally, we briefly speak about Dirac's first major contribution to quantum mechanics.

8.1.3 Dirac in the summer of 1925

Before we start, let us say a few words about how Dirac entered quantum mechanics in 1925. Dirac was a student in Cambridge. He was born on August 8, 1902, so he was barely 23. One can talk for hours about him. He was completely unusual. Very discreet, very polite, and careful, he thought a lot. He had a great culture, in particular in mathematics. He knew about noncommutative algebras, but let's not anticipate.

At that time, Cambridge had an impressive intellectual richness and was similar to Göttingen in that respect. The professors were Keynes in economics, J.J. Thomson, E. Rutherford, A.S. Eddington, and R.H. Fowler in physics, as well as others.

And there, by accident, everything got started. On July 28, 1925, Heisenberg was invited to give a talk in Cambridge at the very fashionable Kapitza club.

That talk had no influence on Dirac (who thought instead of listening); neither of them remembered seeing the other on that day. But Heisenberg gave his article to Fowler who gave it to Dirac. He read the paper with some difficulty: it was full of philosophical ideas, but the mathematical formalism was clumsy. Dirac had nothing against philosophy, but he didn't particularly care for it in that context; what he wanted was good mathematics. Very rapidly he said, "This will lead us nowhere."

Two weeks later, he walked into Fowler's office and said, "It's remarkable; it contains the key of quantum mechanics!"

What had happened is that *Heisenberg had cracked up*!

Since 1924, he had elaborated a system of calculational rules and a theory that worked not too badly. Then Born had gotten mixed up with the whole business, and he had found that there were matrices. They had seen Hilbert. And Heisenberg, a positivist, had learned something tragic about his theory: some physical quantities did not commute! The product of a and b was not the same as the product of b and a! And that was contrary to any physical sense. The product of two quantities had never depended on the order. So Heisenberg was terrified, "My theory is not beautiful!" He was so terrified that he had put all the dust under the carpet and he wrote his papers in such a way to hide this noncommutativity as well as he could.

Dirac, very meticulous, had redone the calculations step by step, and he realized that everything boiled down to this noncommutativity. In particular, he proved independently from Born and Jordan the fundamental relation $[\hat{x}, \hat{p}_x] = i\hbar\hat{I}$ in August 1925 without mentioning matrices.

He knew the existence of noncommutative algebras. So he made an attempt to modify the classical equations in order to take into account this noncommutativity. After all nothing dictates that physical quantities should commute.

Some time after, Dirac said, "You know, Heisenberg was scared. He was scared by these foreign mathematics. I can understand that he was scared:

all his theory was at stake. I had an enormous advantage over him: I had no risk."

So, Dirac constructed his own version of quantum mechanics, based on noncommutative algebras, that is, on commutators.

8.2 Uncertainty relations

In 1927, Heisenberg stated his uncertainty principle. When Dirac saw it, he simply said, "Oh yes, indeed, I proved that in 1925."

The general proof of uncertainty relations for any couple of physical quantities is the following.

Consider two quantities A and B, and the corresponding observables $\hat{A}$ and $\hat{B}$. Let $|\psi\rangle$ be the state of the system. The measurement of A and of B gives the expectation values $\langle a\rangle$ and $\langle b\rangle$, and the dispersions Δa and Δb. We want to relate Δa and Δb for two given observables $\hat{A}$ and $\hat{B}$.

A simple calculation based on the Schwarz inequality and the manipulation of Dirac's formalism shows that, if $|\psi\rangle$ is the state of the system, then

$$\Delta a \ \Delta b \geq \frac{1}{2}|\langle\psi|[\hat{A},\hat{B}]|\psi\rangle| \quad . \tag{8.7}$$

Therefore,

1. If $\hat{A}$ and $\hat{B}$ do not commute, $[\hat{A},\hat{B}] \neq 0$, then, in general the two dispersions on A and B cannot be made simultaneously as small as possible except in special cases where the state $|\psi\rangle$ is such that $\langle\psi|[\hat{A},\hat{B}]|\psi\rangle = 0$.
2. For x and p_x, using (8.2), we have $\Delta x \ \Delta p_x \geq \hbar/2$. We remark that we no longer refer to the Fourier transformation. In this formalism, we can forget about it.
3. The uncertainty relations are therefore generalized to any couple of observables. They come from the noncommutativity of the corresponding observables.

Proof. We first center the variables, that is, we set $\hat{A}' = \hat{A} - \langle a\rangle$ and $\hat{B}' = \hat{B} - \langle b\rangle$, so that $\langle\hat{A}'\rangle = \langle\hat{B}'\rangle = 0$. We then have

$$(\Delta a)^2 = \langle\psi|\hat{A}'^2|\psi\rangle\,, \quad \text{and} \quad (\Delta b)^2 = \langle\psi|B'^2|\psi\rangle \quad .$$

Consider any state $|\psi\rangle$ and the vector $(\hat{A}' + i\lambda\hat{B}')|\psi\rangle$, with λ real. The square of the norm of this vector is

$$\begin{aligned}\|(\hat{A}' + i\lambda\hat{B}')|\psi\rangle\|^2 &= \langle\psi|(\hat{A}' - i\lambda\hat{B}')(\hat{A}' + i\lambda\hat{B}')|\psi\rangle \\ &= \langle\psi|\hat{A}'^2|\psi\rangle + \lambda^2\langle\psi|\hat{B}'^2|\psi\rangle + i\lambda\langle\psi|[\hat{A}',\hat{B}']|\psi\rangle \\ &= \Delta a^2 + \lambda^2\Delta b^2 + i\lambda\langle\psi|[\hat{A}',\hat{B}']|\psi\rangle \quad .\end{aligned}$$

Because $\hat{A}'$ and $\hat{B}'$ are Hermitian, the operator $i[\hat{A}',\hat{B}']$ is also Hermitian and the last term is real. The above expression is the square of the norm

of a vector, thus it must be positive whatever the value of λ. Because it is positive for $\lambda \to \infty$ it must not change sign. Therefore the discriminant of the trinomial in λ must be negative, and

$$\Delta a \; \Delta b \geq \frac{1}{2} |\langle \psi | [\hat{A}, \hat{B}] | \psi \rangle| \tag{8.8}$$

because $[\hat{A}', \hat{B}'] = [\hat{A}, \hat{B}]$.

8.3 Evolution of physical quantities

We now address the following problem: what is the time variation of the expectation value of a physical quantity?

8.3.1 Evolution of an expectation value

Consider the expectation value $\langle a \rangle$ of a physical quantity

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

We take the time derivative:

$$\frac{d}{dt}\langle a \rangle = \left(\frac{d}{dt} \langle \psi | \right) \hat{A} | \psi \rangle + \langle \psi | \left(\frac{\partial}{\partial t} \hat{A} \right) | \psi \rangle + \langle \psi | \hat{A} \left(\frac{d}{dt} | \psi \rangle \right) \quad .$$

Using the Schrödinger equation and its Hermitian conjugate expression

$$i\hbar \frac{d|\psi\rangle}{dt} = \hat{H} |\psi\rangle \,, \quad \text{and} \quad -i\hbar \frac{d\langle\psi|}{dt} = \langle \psi | \hat{H} \quad ,$$

we obtain

$$\frac{d}{dt}\langle a \rangle = \frac{1}{i\hbar} \langle \psi | [\hat{A}, \hat{H}] | \psi \rangle + \langle \psi | \frac{\partial \hat{A}}{\partial t} | \psi \rangle \quad . \tag{8.9}$$

This formula is due to Ehrenfest (1927) (but Dirac had found it in 1925).

We remark that, in a similar way to the Schrödinger equation, the Hamiltonian governs the time evolution of a physical quantity. Here it does so through its commutator with the observable.

If the operator $\hat{A}$ does not explicitly depend on time, we have

$$\frac{d}{dt}\langle a \rangle = \frac{1}{i\hbar} \langle \psi | [\hat{A}, \hat{H}] | \psi \rangle \quad . \tag{8.10}$$

8.3.2 Particle in a potential, classical limit

We call q_i $(i = 1, 2, 3)$ the three position variables x, y, z and p_i $(i = 1, 2, 3)$ the coordinates of the momentum p_x, p_y, p_z. The operators $\hat{q}_i$ and $\hat{p}_i$ obey the canonical commutation relations:

$$[\hat{q}_i, \hat{q}_j] = 0\,, \quad [\hat{p}_i, \hat{p}_j] = 0\,, \quad [\hat{q}_j, \hat{p}_k] = i\hbar\delta_{j,k} \quad . \tag{8.11}$$

From these relations, one can prove the commutation relations

$$[\hat{q}_j, \hat{p}_j^m] = m(i\hbar)\hat{p}_j^{m-1}\,, \quad [\hat{p}_j, \hat{q}_j^n] = -n(i\hbar)\hat{q}_j^{n-1}\,, \tag{8.12}$$

which we can generalize to any function $\hat{F} = F(\hat{q}_i, \hat{p}_i)$ of the operators $\hat{q}_i$ and $\hat{p}_i$ which can be expanded in a power series

$$[\hat{q}_j, \hat{F}] = i\hbar\frac{\partial \hat{F}}{\partial \hat{p}_j}\,, \quad \text{and} \quad [\hat{p}_j, \hat{F}] = -i\hbar\frac{\partial \hat{F}}{\partial \hat{q}_j} \quad . \tag{8.13}$$

If the Hamiltonian does not depend explicitly on time, choosing $\hat{F} = \hat{H}$, we obtain the evolution equations:

$$\frac{d}{dt}\langle q_i\rangle = \left\langle \frac{\partial \hat{H}}{\partial \hat{p}_j} \right\rangle\,, \quad \frac{d}{dt}\langle p_j\rangle = -\left\langle \frac{\partial \hat{H}}{\partial \hat{q}_j} \right\rangle \quad . \tag{8.14}$$

These forms have a great similarity to the equations of Hamilton's analytical mechanics.

The Hamiltonian of a particle in a potential is

$$\hat{H} = \frac{\hat{\boldsymbol{p}}^2}{2m} + V(\boldsymbol{r}) \quad . \tag{8.15}$$

Substituting in (8.14), we obtain

$$\frac{d\langle \boldsymbol{r}\rangle}{dt} = \frac{\langle \boldsymbol{p}\rangle}{m}\,, \tag{8.16}$$

$$\frac{d\langle \boldsymbol{p}\rangle}{dt} = -\langle \nabla V(\boldsymbol{r})\rangle\,. \tag{8.17}$$

Equation (8.16) is the correct definition of the group velocity. It is the same for expectation values as the classical relation between the position and the velocity. Actually, equation (8.17) differs slightly from the classical equation, which would be for expectation values

$$\frac{d\langle \boldsymbol{p}\rangle}{dt} = -\nabla V(\boldsymbol{r})\Big|_{\mathbf{r}=\langle \mathbf{r}\rangle} \quad ,$$

because, in general, $f(\langle \boldsymbol{r}\rangle) \neq \langle f(\boldsymbol{r})\rangle$. However, if the distribution in $\boldsymbol{r}$ is peaked around some value $\boldsymbol{r}_0$, then $\langle \boldsymbol{r}\rangle \sim \boldsymbol{r}_0$ and $f(\langle \boldsymbol{r}\rangle) \sim \langle f(\boldsymbol{r})\rangle$, and in this case equations (8.16) and (8.17) are, for the expectation values, essentially the same as the classical equations of the motion. This observation is the Ehrenfest theorem (1927).

One-dimensional classical limit

In one dimension, we have

$$\frac{d}{dt}\langle p\rangle = \left\langle -\frac{dV}{dx}\right\rangle \neq -\frac{dV}{dx}\Big|_{x=\langle x\rangle} .$$

We expand the function $f(x) = -\partial V/\partial x$ in the vicinity of $x = \langle x\rangle$, and we obtain

$$f(x) = f(\langle x\rangle) + (x - \langle x\rangle)f'(\langle x\rangle) + \frac{1}{2}(x - \langle x\rangle)^2 f''(\langle x\rangle) + \cdots ,$$

that is, taking the expectation value,

$$\langle f\rangle = f(\langle x\rangle) + \frac{\Delta x^2}{2} f''(\langle x\rangle) + \cdots ,$$

where $\Delta x^2 = \langle (x - \langle x\rangle)^2\rangle$.

The nonclassical term in the evolution of the expectation value will be negligible if

$$|\Delta x^2 f''(\langle x\rangle)/f(\langle x\rangle)| \ll 1 \quad ,$$

or, if we come back to the potential V,

$$\left|\Delta x^2 \frac{\partial^3 V}{\partial x^3}\right| \ll \left|\frac{\partial V}{\partial x}\right| \quad ,$$

that is, if the potential varies slowly on the extension of the wave packet (or on an interval of the order of the de Broglie wavelength). That is the actual content of the Ehrenfest theorem.

We recall that a macroscopic system (i.e., a system whose orders of magnitude are such that $\hbar$ seems negligible) is not necessarily a classical system. We can perfectly well consider a wave packet having two peaks located at a large macroscopic distance from each other. As we said in Chapter 4 such a state, macroscopic but quantum mechanical, is very unstable.

One can check that for a harmonic oscillator, $V = m\omega_0^2 x^2/2$, or, more generally if the potential is a second degree polynomial, the Ehrenfest theorem gives identically the classical equation of motion $d^2\langle x\rangle/dt^2 = -\omega_0^2\langle x\rangle$.

8.3.3 Conservation laws

There are two cases where $d\langle a\rangle/dt = 0$:

- Either the *observable* commutes with the Hamiltonian, $[\hat{A}, \hat{H}] = 0$. The quantity A is always conserved, it is a constant of the motion.
- Or the *state* $|\psi\rangle$ is such that $\langle\psi|[\hat{A}, \hat{H}]|\psi\rangle = 0$. This is in particular the case for stationary states, that is eigenstates of $\hat{H}$ for which no quantity evolves in time.

Here are some examples of applications of these results:

Conservation of the norm

Consider the identity operator $\hat{A} = \hat{I}$; this gives the conservation of the norm

$$\frac{d}{dt}\langle\psi|\psi\rangle = 0 \quad ,$$

that is to say, conservation of probability.

Conservation of energy for an isolated system

Consider a time-independent problem; the choice $\hat{A} = \hat{H}$ gives

$$\frac{d}{dt}\langle E\rangle = 0 \quad .$$

Conservation of momentum

Consider the motion of a free particle of Hamiltonian $\hat{H} = \hat{\boldsymbol{p}}^2/2m$. The observables $\hat{p}_x, \hat{p}_y, \hat{p}_z$ commute with $\hat{H}$ and we obtain conservation of the momentum:

$$\frac{d}{dt}\langle p_i\rangle = 0\,, \quad i = x, y, z \quad .$$

This is generalized to an N particle system whose Hamiltonian is translation invariant (pairwise interactions $V(x_i - x_j)$). The total momentum $\langle \boldsymbol{P}\rangle = \langle \sum \boldsymbol{p}_i\rangle$ is conserved.

8.4 Algebraic resolution of the harmonic oscillator

In order to see how the principles work, and to see the importance of the commutation relations, we show how the one-dimensional harmonic oscillator problem can be solved using the algebra of observables. This calculation is also due to Dirac.

In order to simplify the proof, we make use of qualitative results we already know. Consider the Hamiltonian

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega^2\hat{x}^2 \quad .$$

With the change of observables

$$\hat{X} = \hat{x}\sqrt{\frac{m\omega}{\hbar}}\,, \quad \hat{P} = \frac{\hat{p}}{\sqrt{m\hbar\omega}} \quad , \tag{8.18}$$

we obtain

$$\hat{H} = \hbar\omega\hat{\mathcal{H}} \quad , \tag{8.19}$$

with

$$\hat{\mathcal{H}} = \frac{1}{2}\left(\hat{X}^2 + \hat{P}^2\right) \quad . \tag{8.20}$$

We must solve the eigenvalue problem

$$\hat{\mathcal{H}}|\nu\rangle = \varepsilon_\nu|\nu\rangle \quad ,$$

where we have assumed that the eigenvalues ε_ν are not degenerate. We know this result from Chapter 5, but this can be proven easily in the present context.

The commutation relation of $\hat{X}$ and $\hat{P}$ can be deduced from (8.2),

$$[\hat{X}, \hat{P}] = i \quad . \tag{8.21}$$

8.4.1 Operators $\hat{a}$, $\hat{a}^\dagger$, and $\hat{N}$

Solving the eigenvalue problem is performed by introducing the following operators

$$\hat{a} = \frac{1}{\sqrt{2}}(\hat{X} + i\hat{P}) \,, \quad \hat{a}^\dagger = \frac{1}{\sqrt{2}}(\hat{X} - i\hat{P}) \tag{8.22}$$

whose commutator is

$$[\hat{a}, \hat{a}^\dagger] = 1 \quad . \tag{8.23}$$

Consider now the operator

$$\hat{N} = \hat{a}^\dagger\hat{a} = \frac{1}{2}(\hat{X}^2 + \hat{P}^2 - 1) \tag{8.24}$$

that satisfies the commutation relations

$$[\hat{N}, \hat{a}] = -\hat{a} \,, \quad [\hat{N}, \hat{a}^\dagger] = \hat{a}^\dagger \quad . \tag{8.25}$$

We have

$$\hat{\mathcal{H}} = \hat{N} + \frac{1}{2}\hat{I} \quad ;$$

therefore, $\hat{\mathcal{H}}$ and $\hat{N}$ have the same eigenvectors. Let ν be the eigenvalues of $\hat{N}$ and $|\nu\rangle$ its eigenvectors; we have, coming back to the initial Hamiltonian

$$\hat{H}|\nu\rangle = \left(\nu + \frac{1}{2}\right)\hbar\omega|\nu\rangle \quad . \tag{8.26}$$

8.4.2 Determination of the eigenvalues

The determination of the eigenvalues comes from the following lemmas:

Lemma 1. *The eigenvalues ν of the operator $\hat{N}$ are positive or zero.*

In fact, consider a vector $|\nu\rangle$ and the norm of the vector $\hat{a}|\nu\rangle$:

$$\|\hat{a}|\nu\rangle\|^2 = \langle\nu|\hat{a}^\dagger\hat{a}|\nu\rangle = \langle\nu|\hat{N}|\nu\rangle = \nu\langle\nu|\nu\rangle = \nu\||\nu\rangle\|^2 \quad . \tag{8.27}$$

Therefore, $\nu \geq 0$ and

$$\hat{a}|\nu\rangle = 0\,, \quad \text{if and only if} \quad \nu = 0\,. \tag{8.28}$$

Lemma 2. *The vector $\hat{a}|\nu\rangle$ is either an eigenvector of $\hat{N}$, corresponding to the eigenvalue $\nu - 1$, or the null vector.*

In fact, consider the vector $\hat{N}\hat{a}|\nu\rangle$. If we use the commutation relation of $\hat{N}$ and $\hat{a}$, we obtain

$$\hat{N}(\hat{a}|\nu\rangle) = \hat{a}\hat{N}|\nu\rangle - \hat{a}|\nu\rangle = \nu\hat{a}|\nu\rangle - \hat{a}|\nu\rangle = (\nu - 1)(\hat{a}|\nu\rangle) \quad .$$

Therefore,

- Either $\nu - 1$ is an eigenvalue of $\hat{N}$ and $\hat{a}|\nu\rangle$ is an associated eigenvector;
- Or $\nu - 1$ is not an eigenvalue of $\hat{N}$ and $\hat{a}|\nu\rangle$ is the null vector.

By an analogous argument one can show that:

Lemma 3. *If $\nu + 1$ is an eigenvalue of $\hat{N}$, $\hat{a}^\dagger|\nu\rangle$ is an eigenvector associated with the eigenvalue $\nu + 1$ and*

$$\|\hat{a}^\dagger|\nu\rangle\|^2 = (\nu + 1)\||\nu\rangle\|^2 \quad . \tag{8.29}$$

It is now simple to prove that:

Theorem. *The eigenvalues of $\hat{N}$ are the nonnegative integers.*

In fact, because the eigenvalues are nonnegative, one of them must be smaller than the others. Let us call it $\nu_{\min}$. Since $\nu_{\min}$ is the smallest eigenvalue, $\nu_{\min} - 1$ is not an eigenvalue. Therefore, $\hat{a}|\nu_{\min}\rangle$ is the null vector and its norm is zero. Inasmuch as we have shown that $\|\hat{a}|\nu\rangle\|^2 = \nu\||\nu\rangle\|^2$ for all values of ν, this means that $\nu_{\min} = 0$.

Starting with this eigenvalue and the corresponding eigenvector $|\nu_{\min} = 0\rangle$, we can generate all other eigenvalues and corresponding eigenvectors by repeatedly applying the operator $\hat{a}^\dagger$. We recover the energy levels of the harmonic oscillator.

8.4.3 Eigenstates

Ground state

The ground state $|0\rangle$ satisfies (8.28)

$$\hat{a}|0\rangle = 0\,, \quad \text{or} \quad \left(\hat{X} + i\hat{P}\right)|0\rangle = 0 \quad . \tag{8.30}$$

In the language of wave functions, this amounts to

$$\left(\frac{m\omega}{\hbar}x + \frac{d}{dx}\right)\varphi_0(x) = 0 \quad , \tag{8.31}$$

where $\varphi_0(x)$ is the ground state wave function. The solution is

$$\varphi_0(x) = C_0\ e^{-(m\omega/\hbar)x^2/2} \quad , \tag{8.32}$$

where C_0 is a normalization constant. We recover the result of Chapter 5.

Excited states

We assume the eigenstates are normalized: $\langle n|n\rangle = 1$. Owing to the above Lemmas 1 and 2 and to equations (8.27) and (8.29),

$$\hat{a}|n\rangle = \sqrt{n}|n-1\rangle\,, \quad \hat{a}^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle \quad . \tag{8.33}$$

Hence, the name of annihilation operator (for $\hat{a}$) and of creation operator (for $\hat{a}^\dagger$) because they allow us to get from an energy state $(n+1/2)\hbar\omega$ to energy states $(n+1/2\mp 1)\hbar\omega$, and they respectively annihilate or create an energy quantum $\hbar\omega$. Similarly, the operator $\hat{N}$ corresponds to the number of quanta in the state $|n\rangle$. The sequence of states $|n\rangle$ is generated from the ground state $|0\rangle$ by applying repeatedly the operator $\hat{a}^\dagger$,

$$|n\rangle = \frac{1}{\sqrt{n!}}(\hat{a}^\dagger)^n|0\rangle \quad . \tag{8.34}$$

This allows us to find the wave function $\varphi_n(x)$ of the energy state $(n+1/2)\hbar\omega$ in terms of the ground state wave function:

$$\varphi_n(x) = \frac{1}{\sqrt{n!}}\frac{1}{\sqrt{2^n}}\left[x\sqrt{\frac{m\omega}{\hbar}} - \sqrt{\frac{\hbar}{m\omega}}\frac{d}{dx}\right]^n \varphi_0(x) \quad . \tag{8.35}$$

This is an explicit and compact formula for the Hermite functions.

Similarly, one can show with (8.33) and the definition of $\hat{a}$ and $\hat{a}^\dagger$, that we have

$$\hat{x}|n\rangle = \sqrt{\frac{\hbar}{2m\omega}}\left(\sqrt{n+1}|n+1\rangle + \sqrt{n}|n-1\rangle\right) \,, \tag{8.36}$$

$$\hat{p}|n\rangle = i\sqrt{\frac{m\hbar\omega}{2}}\left(\sqrt{n+1}|n+1\rangle - \sqrt{n}|n-1\rangle\right) \quad . \tag{8.37}$$

We see, in this example, the elegance and the power of Dirac's algebraic method. This treatment of the harmonic oscillator, and the operators $\hat{a}, \hat{a}^\dagger$, and $\hat{N}$, are fundamental tools in many branches of physics such as quantum field theory, statistical mechanics, and the many-body problem.

8.5 Commuting observables

When two observables commute, there is no constraint such as the uncertainty relations. This case is, however, very interesting in practice.

8.5.1 Theorem

We know that if two matrices commute, one can diagonalize them simultaneously.

This remains true in the infinite-dimensional case. If two observables $\hat{A}$ and $\hat{B}$ commute, there exists a common eigenbasis of these two observables.

This theorem is generalized immediately to the case of several observables $\hat{A}$, $\hat{B}$, $\hat{C}$, which all commute.

Proof. Let $\{|\alpha, r_\alpha\rangle\}$ be the eigenvectors of $\hat{A}$, where the index r_α means that an eigenvector associated with an eigenvalue a_α belongs to an eigensubspace of dimension $d_\alpha \geq 1$,

$$\hat{A}|\alpha, r_\alpha\rangle = a_\alpha |\alpha, r_\alpha\rangle \,, \quad r = 1, \ldots, d_\alpha \quad . \tag{8.38}$$

By assumption, we have $[\hat{A}, \hat{B}] = 0$, that is,

$$\hat{A}\hat{B}|\alpha, r_\alpha\rangle = \hat{B}\hat{A}|\alpha, r_\alpha\rangle = a_\alpha \hat{B}|\alpha, r_\alpha\rangle \,, \quad r = 1, \ldots, d_\alpha \quad . \tag{8.39}$$

Therefore, the vector $\hat{B}|\alpha, r_\alpha\rangle$ is an eigenvector of $\hat{A}$ with the eigenvalues a_α. It therefore belongs to the corresponding eigensubspace. We call this vector $|\alpha, \beta, k_{\alpha\beta}\rangle$; the index $k_{\alpha\beta}$ means that again this vector may be nonunique. Therefore, this vector is a linear combination of the vectors $\{|\alpha, r_\alpha\rangle\}$, that is,

$$\hat{B}|\alpha, r_\alpha\rangle = \sum_{r_\alpha} b_{r_\alpha} |\alpha, r_\alpha\rangle \quad ,$$

which can be diagonalized with no difficulty. In other words, if $\hat{A}$ and $\hat{B}$ commute, they possess a common eigenbasis.

The reciprocal is simple. The Riesz theorem says that the othonormal eigenvectors of an observable form a Hilbert basis. Suppose $\hat{A}$ and $\hat{B}$ have in common the basis $\{|\psi_n\rangle\}$ with eigenvalues a_n and b_n:

$$\hat{A}|\psi_n\rangle = a_n|\psi_n\rangle \quad \text{and} \quad \hat{B}|\psi_n\rangle = b_n|\psi_n\rangle \quad . \tag{8.40}$$

If we apply $\hat{B}$ to the first expression and $\hat{A}$ to the second, and subtract, we obtain

$$(\hat{A}\hat{B} - \hat{B}\hat{A})|\psi_n\rangle = (a_n b_n - b_n a_n)|\psi_n\rangle = 0 \quad .$$

Because $\{|\psi_n\rangle\}$ is a Hilbert basis, we therefore have

$$[\hat{A}, \hat{B}]|\psi\rangle = 0 \,, \quad \text{whatever } |\psi\rangle \quad ,$$

which means $[\hat{A}, \hat{B}] = 0$.

8.5.2 Example

Actually, we have not yet seen examples of this because we have considered only one-dimensional problems. But the statement is nothing very complicated.

Consider, for instance, an isotropic two-dimensional harmonic oscillator. The eigenvalue problem of the Hamiltonian is a priori a difficult problem because it seems to be a partial differential equation in two variables. But the Hamiltonian can be written as the sum of two independent Hamiltonians acting on different variables:

$$\hat{H} = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + \frac{1}{2}m\omega^2 x^2 - \frac{\hbar^2}{2m}\frac{\partial^2}{\partial y^2} + \frac{1}{2}m\omega^2 y^2 = \hat{H}_x + \hat{H}_y \quad . \tag{8.41}$$

The two operators $\hat{H}_x$ and $\hat{H}_y$, which are both operators in one variable and which act on different variables commute obviously. One can solve the eigenvalue problems of $\hat{H}_x$ and $\hat{H}_y$ separately:

$$\hat{H}_x\varphi_n(x) = E_n\varphi_n(x)\ ; \quad \hat{H}_y\varphi_n(y) = E_n\varphi_n(y)\ .$$

The eigenvalues of $\hat{H}$ are the sums of eigenvalues of $\hat{H}_x$ and $\hat{H}_y$ with eigenfunctions that are the products of corresponding eigenfunctions:

$$E_n = E_{n_1} + E_{n_2} = (n_1 + n_2 + 1)\hbar\omega\,, \quad \psi_n(x,y) = \varphi_{n_1}(x)\varphi_{n_2}(y) \quad .$$

In other words, a sum of Hamiltonians that commute has for eigenvalues the sum of eigenvalues of each of them, and for eigenfunctions the product of corresponding eigenfunctions.

8.5.3 Tensor structure of quantum mechanics

This example has another interest. It shows that in the case of systems with several degrees of freedom, the Hilbert space is the tensor product of the Hilbert spaces in which each individual degree of freedom is described.

In the above example, it is the fashionable way of saying that the products of eigenfunctions of the one-dimensional harmonic oscillator

$$\psi_{n,m}(x,y) = \varphi_n(x)\varphi_m(y) \tag{8.42}$$

form a basis of square integrable functions in two variables (x, y), or that the space of square integrable functions in two variables $\mathcal{L}^2(\mathcal{R}^2)$ is the tensor product of two spaces $\mathcal{L}^2(\mathcal{R})$ of square integrable functions in one variable.

The tensor structure of quantum mechanics is important and useful in complex systems. One can find it at various degrees of sophistication in the literature.[1] In Dirac's notations the elements of the basis $\varphi_{n_1}(x)\varphi_{m_2}(y)$ are written as

[1] See J.-L. Basdevant and J. Dalibard, *Quantum Mechanics*, Chapter 5, Section 6.

$$|\psi_{n,m}\rangle = |1:\varphi_n\rangle \otimes |2:\varphi_m\rangle \quad , \tag{8.43}$$

where (1) and (2) stand for the two degrees of freedom (or subsystems) and n and m are the corresponding eigenstates. The symbol $\otimes$ stands for "tensor product," which is just an ordinary product in equation (8.42). Any state of the global system $|\psi\rangle$ can be written as

$$|\psi\rangle = \sum_{n,m} c_{n,m} |1:\varphi_n\rangle \otimes |2:\varphi_m\rangle \quad . \tag{8.44}$$

An important property is that the Hermitian scalar product of two factorizable vectors $|u\rangle \otimes |v\rangle$ and $|u'\rangle \otimes |v'\rangle$ factorizes as the product

$$(\langle u'| \otimes \langle v'|)(|u\rangle \otimes |v\rangle = \langle u'|u\rangle\langle v'|v\rangle \quad . \tag{8.45}$$

We show in Chapter 11, with spin, an example just as simple but not as trivial of a tensor product of Hilbert spaces.

8.5.4 Complete set of commuting observables (CSCO)

This brings us to a notion that is useful both conceptually and technically.

A set of operators $\hat{A}, \hat{B}, \hat{C} \ldots$, is said to form a complete set of commuting observables (CSCO) if the common eigenbasis of this set of operators is unique. In other words, to each set of eigenvalues $a_\alpha, b_\beta, c_\gamma, \ldots$ there corresponds a single eigenvector $|\alpha\beta\gamma \ldots\rangle$ (up to a phase factor). If an operator $\hat{O}$ commutes with all of the operators of the CSCO, then it is a function of these operators. For a given system, there exists an infinite number of CSCOs. We show in the following Chapters that one chooses a CSCO according to criteria of convenience. Neither the nature nor the number of observables that form a CSCO are fixed a priori.

For a one-dimensional harmonic oscillator, the Hamiltonian

$$\hat{H}_x = \frac{\hat{p}_x^2}{2m} + \frac{1}{2} m\omega^2 \hat{x}^2$$

is a CSCO by itself. There is only one eigenbasis of $\hat{H}_x$ formed by the Hermite functions $\psi_n(x)$.

For the two-dimensional isotropic oscillator above (8.41), this is not the case. A possible basis is formed by the set of functions $\{\psi_{n_1}(x)\psi_{n_2}(y)\}$, where $\psi_{n_1}(x)$ and $\psi_{n_2}(y)$ are, respectively, the eigenfunctions of $\hat{H}_x$ and $\hat{H}_y$. The eigenvalue corresponding to $\psi_{n_1}(x)\psi_{n_2}(y)$ is

$$E_{n_1,n_2} = \hbar\omega(n_1 + n_2 + 1) \quad ,$$

which is degenerate, except for $n_1 = n_2 = 0$. This implies that there are actually several eigenbases of $\hat{H}$ (actually an infinite number). For instance, in the subspace associated with $2\hbar\omega$, the possible base elements are

$$\{\cos\theta\psi_1(x)\psi_2(y) + \sin\theta\psi_2(x)\psi_1(y), -\sin\theta\psi_1(x)\psi_2(y) + \cos\theta\psi_2(x)\psi_1(y)\} \quad .$$

Therefore, $\hat{H}$ is not a CSCO by itself. A possible CSCO is the set of two Hamiltonians $\{\hat{H}_x, \hat{H}_y\}$. In fact, the two eigenvalues $\{E_{n_x} = (n_x + 1/2)\hbar\omega, E_{n_y} = (n_y + 1/2)\hbar\omega\}$ uniquely specify an eigenvector.

This does not imply that in this problem a CSCO is necessarily composed of two observables. A rigorous theoretician may object that the operator $\hat{H}_\pi = \hat{H}_x + \pi\hat{H}_y$ forms a CSCO by itself. Indeed, its eigenvalues are $n_\pi = (n_x + 1/2) + \pi(n_y + 1/2)$, and since π is transcendental there is a unique couple of integers corresponding to a given eigenvalue. An experimentalist will reply that it is simpler to measure two numbers (n_x, n_y) rather than investing in a measuring apparatus that directly gives the answer n_π.

8.5.5 Completely prepared quantum state

Why is the notion of a completely prepared quantum state important physically? If we want to specify as accurately as possible the initial conditions of an experiment, we must know whether we start from a specific quantum state or with some ill-defined situation. In the above case of the isotropic harmonic oscillator, if we know the initial energy $n\hbar\omega$, we only know that the initial state belongs to a subspace of dimension n, generated by the n functions $\psi_{n_1}(x)\psi_{n_2}(y)$ with $n_1 + n_2 + 1 = n$. A measurement of the energy is not sufficient to specify unambiguously the initial state. If we measure the vibrational energies along both the x and y axes, we know the state. One says that one deals with a *completely prepared quantum state*.

More generally, consider two observables $\hat{A}$ and $\hat{B}$ that commute. There exists a common eigenbasis $\{|\alpha, \beta, \gamma\rangle\}$ of $\hat{A}$ and $\hat{B}$, where the index γ indicates that $\hat{A}$ and $\hat{B}$ are not necessarily a CSCO by themselves.

Consider a state $|\psi\rangle$, and suppose that by measuring A on $|\psi\rangle$, we find the value a_0. After the measurement, the state of the system has changed. It is

$$|\psi\rangle \to |\psi_0\rangle = \lambda \left(\sum_{\beta,\gamma'} |\alpha_0; \beta, \gamma'\rangle\langle\alpha_0; \beta, \gamma'| \right) |\psi\rangle \quad ,$$

where λ is a normalization factor such that $\langle\psi_0|\psi_0\rangle = 1$.

On this new state $|\psi_0\rangle$, we measure B. Suppose the result is b_1. Similarly, we obtain another state $|\psi_1\rangle$ after the measurement:

$$|\psi_0\rangle \to |\psi_1\rangle = \lambda' \left(\sum_{\alpha\gamma} |\alpha, \beta_1; \gamma\rangle\langle\alpha, \beta_1; \gamma| \right) |\psi_0\rangle \quad .$$

If we insert the expression of $|\psi_0\rangle$ in this formula, we obtain

$$|\psi_1\rangle = \lambda\lambda' \sum_{\alpha\beta\gamma'} |\alpha, \beta_1; \gamma\rangle\langle\alpha, \beta_1; \gamma|\alpha_0, \beta; \gamma'\rangle\langle\alpha_0, \beta; \gamma'|\psi\rangle \; .$$

But, by assumption, $\langle \alpha, \beta; \gamma | \alpha', \beta'; \gamma' \rangle = \delta_{\alpha\alpha'} \delta_{\beta\beta'} \delta_{\gamma\gamma'}$. Therefore, the expansion of $|\psi_1\rangle$ reduces to

$$|\psi_1\rangle = \lambda\lambda' \sum_{\gamma} |\alpha_0, \beta_1; \gamma\rangle \langle \alpha_0, \beta_1, \gamma | \psi \rangle \quad .$$

In other words, $|\psi_1\rangle$ is still an eigenvector of $\hat{A}$ with the same eigenvalue a_0.

This is an important result. If two observables $\hat{A}$ and $\hat{B}$ commute, if we successively measure A, with a result a_0, and then B with a result b_1, this second measurement does not affect the value found previously for A. If we redo a measurement of A, we find the same result a_0 with probability one.

This result can be extended to any number of commuting observables. If one measures all the physical quantities of a CSCO, any ("immediate") new measurement of one of these quantities will always give the same result. The state vector defined by this series of measurements is defined uniquely. One says that with this series of measurements one obtains a state that is completely prepared or completely known.

8.6 Sunday, September 20, 1925

Let us get back to Dirac, in the summer of 1925. His problem was to find how to incorporate noncommutativity in classical mechanics.

Dirac knew that in analytical mechanics, which had been developed one century before, Hamilton had found a quite fruitful formulation.[2]

In this version of mechanics, the state of a particle is described at any time by its position x and its momentum p.

The system is characterized by a Hamilton function or Hamiltonian H which, for a particle of mass m in a potential $V(x)$ is

$$H = \frac{p^2}{2m} + V(x) \quad . \tag{8.46}$$

The evolution equations of the state variables x and p are then given by the canonical equations

$$\frac{dx}{dt} = \frac{\partial H}{\partial p} \quad \text{and} \quad \frac{dp}{dt} = -\frac{\partial H}{\partial x} \quad . \tag{8.47}$$

Dirac wanted to construct his quantum mechanics with noncommutative algebras, that is, with commutators.

On Sunday, September 20, 1925, Dirac committed a crime! He was very well organized and he had the habit of working hard all week and of relaxing on Sundays by talking a walk. But on that Sunday he committed a crime.

[2] See, for instance, J.-L. Basdevant, *Variational Principles in Physics*, Chapter 4. New York: Springer (2006).

In fact, during his walk he started thinking. He had given himself the right to walk when he was thinking, but never to think on his Sunday walk! He thought about commutators, and, suddenly, he thought about something he had been told of, and about which he had read. He thought about the works of a former student of Polytechnique, in Paris, Siméon-Denis Poisson! And Dirac rushed to the library, but there was some suspense because "On Sundays it's closed." He had to wait. On Monday morning, he rushed in and read about what Carl Gustav Jacob Jacobi had called the greatest discovery of Poisson: the Poisson brackets.

Consider two physical quantities f and g, which are functions of the state variables (x, p); the Poisson bracket of f and g is the quantity

$$\{f, g\} = \left(\frac{\partial f}{\partial x}\frac{\partial g}{\partial p} - \frac{\partial f}{\partial p}\frac{\partial g}{\partial x}\right) \quad . \tag{8.48}$$

For the state variables (x, p), we find the relation

$$\{x, p\} = 1 \quad , \tag{8.49}$$

and

$$\{x, f\} = \frac{\partial f}{\partial p}\,, \quad \{p, f\} = -\frac{\partial f}{\partial x} \quad . \tag{8.50}$$

In three dimensions, (x_i, p_i) $(i = 1, 2, 3)$, (8.49) generalizes as

$$\{x_i, x_j\} = 0\,, \quad \{p_i, p_j\} = 0\,, \quad \{x_i, p_j\} = \delta_{ij} \quad . \tag{8.51}$$

The time evolution of a physical quantity $f(x, p)$ is

$$\dot{f} = \frac{df}{dt} = (\frac{\partial f}{\partial x}\dot{x} + \frac{\partial f}{\partial p}\dot{p}) \quad . \tag{8.52}$$

But, using Hamilton's equations (8.47), one obtains

$$\dot{f} = \{f, H\} \quad . \tag{8.53}$$

In particular, the equations (8.47) can be written in the symmetric way

$$\dot{x} = \{x, H\}\,, \quad \dot{p} = \{p, H\} \quad . \tag{8.54}$$

Dirac was fascinated. The quantum commutators, divided by $i\hbar$, play a completely similar role as the Poisson brackets in analytical mechanics. Just compare the fundamental commutation relation (8.2) and the relation (8.49), and, similarly, the Ehrenfest theorem (8.10) and equation (8.53).

That is how Dirac understood the actual form of the correspondence principle. One must, in the classical equations, replace the Poisson brackets by the quantum commutators divided by $i\hbar$. Dirac called the quantum physical quantities "q-numbers" which are noncommutative contrary to the classical commutative "c-numbers".

He finished his work on November 7, 1925, and he published his article "The Fundamental Equations of Quantum Mechanics" in December 1925. In November, he wrote to Heisenberg who replied that his work was "most beautiful and remarkable." He was apparently insensitive to the publication of the works of Born and Jordan in November 1925 and of Born, Heisenberg, and Jordan in January 1926, where one can find a number of his results (proven independently).

9

Angular momentum

Angular momentum and, more generally, the rotations of systems play a fundamental role in physics. The conservation laws of angular momentum play a central role, in the same way as energy conservation. We show this when we study the hydrogen atom.

The quantization of angular momentum will display a special feature in that it is universal. Contrary to energy levels, which depend on the system under consideration, the possible discrete values of the angular momentum must be chosen in a given universal catalogue which we establish. They depend on the system under consideration, but they are "ready to wear" and not "made-to-measure." The reason is that Planck's constant has the dimension of an angular momentum, and that $\hbar$ is the natural unit of angular momenta.

The field of applications of our results is very broad:

- From the technical point of view, we use these results in order to study the hydrogen atom or the spin of the electron and other particles.
- The rotation spectra of molecules play an important role, both in chemistry and in astrophysics.
- The angular distributions of final particles in collisions or decays in nuclear and elementary particle physics allow us to determine the structure of particles and the nature of fundamental interactions.
- Last, and not least, all of magnetism, on which we base the experimental analysis of our results, comes from rotating charges. We discover that there exist angular momenta that do not have any classical analogues, in particular spin 1/2 to which we devote chapter 11 and which is one of the most revolutionary discoveries of the 1920s.

9.1 Fundamental commutation relation

9.1.1 Classical angular momentum

Our starting point is the correspondence principle. The angular momentum $\boldsymbol{L}$ of a particle of momentum $\boldsymbol{p}$ and position $\boldsymbol{r}$ with respect to the origin is classically the vector product

$$\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p} \quad .$$

Quantum mechanically we therefore postulate that the vector observable $\hat{\boldsymbol{L}}$ (i.e., the set of three observables) corresponding to this angular momentum is $\hat{\boldsymbol{L}} = \hat{\boldsymbol{r}} \times \hat{\boldsymbol{p}}$, or, in the wave function formalism,

$$\hat{\boldsymbol{L}} = \frac{\hbar}{i} \hat{\boldsymbol{r}} \times \hat{\boldsymbol{\nabla}} \quad .$$

However, as said previously, more than this particular representation, it is the commutation relations of these three observables $\hat{L}_x$, $\hat{L}_y$, and $\hat{L}_z$, that are of interest because they are independent of the representation.

These commutation relations are, as pointed out in (8.4),

$$[\hat{L}_x, \hat{L}_y] = i\hbar \hat{L}_z \ , \tag{9.1}$$

and two other relations obtained by cyclic permutations. These can be put together in the compact form

$$\hat{\boldsymbol{L}} \times \hat{\boldsymbol{L}} = i\hbar \hat{\boldsymbol{L}} \quad . \tag{9.2}$$

It is these algebraic relations that we use in the present discussion.

9.1.2 Definition of an angular momentum observable

We take as a general definition of any angular momentum observable $\hat{\boldsymbol{J}}$ the algebraic relation between its coordinates

$$\hat{\boldsymbol{J}} \times \hat{\boldsymbol{J}} = i\hbar \hat{\boldsymbol{J}} \ , \quad \text{or} \quad [\hat{J}_x, \hat{J}_y] = i\hbar \hat{J}_z \quad . \tag{9.3}$$

By definition, any triplet of observables corresponding to the components of a vector quantity that satisfy this relation is an angular momentum observable. We keep the letter L in the specific case of orbital angular momenta, which have classical analogues.

We first show how one can obtain numbers (i.e., eigenvalues) by manipulating this algebra. This was done by Heisenberg and Jordan in 1925 (but they missed the most interesting part of their results). This method is quite similar to Dirac's method for the harmonic oscillator (chapter 8, Section 8.4). The results are obtained in a much simpler way than with wave functions.

9.1.3 Results of the quantization

The results we obtain are as follows.

Consider a vector observable $\hat{\boldsymbol{J}}$ that has the algebraic properties (9.3) and the corresponding physical quantity. It is an experimental result that

1. Whatever the system under consideration, a measurement of the square of the angular momentum

$$\hat{J}^2 = \hat{J}_x^2 + \hat{J}_y^2 + \hat{J}_z^2 \tag{9.4}$$

gives the values

$$j(j+1)\hbar^2\,, \quad \text{where} \quad 2j\text{is an integer} \quad .$$

2. After finding this value for the square of the angular momentum, the measurement of one of the components of $\boldsymbol{J}$ along any axis gives one of the $(2j+1)$ values

$$m\hbar\,, \quad \text{where} \quad m \in \{j, j-1, \cdots, -j\} \quad .$$

3. In the case of orbital angular momenta (which have a classical analogue) the numbers j and m are integers.

Actually, the algebra (9.3) is very important in mathematics. It is called the Lie algebra of the group of rotations in three dimensions. And these results were proven as soon as 1913 by Elie Cartan in his classification of Lie groups. Why the physicists of the 1920s didn't know it is a mystery! Elie Cartan did not forget to point out their ignorance in his book *Lectures on the Theory of Spinors* in 1935.

How did Cartan proceed, inasmuch as he did not know the Planck constant $\hbar$? Because $\hbar$ plays no role in this game. $\hbar$ is the natural unit of angular momenta. In (9.3), if we divide both sides by $\hbar^2$ we obtain an algebraic relation between dimensionless operators.

9.2 Proof of the quantization

9.2.1 Statement of the problem

Elie Cartan's proof is crystal clear. The problem is to find all matrices that satisfy the relation (9.3) and to calculate their eigenvalues and eigenvectors.

1. $\hat{J}_x$, $\hat{J}_y$, and $\hat{J}_z$ do not commute. We can only diagonalize one of them. However, they obviously have the same eigenvalues.

2. Looking at the physics of the problem enables us to understand a methodology that the mathematicians Sophus Lie and Elie Cartan knew at the end of the 19th century.

The square of the angular momentum $\hat{J}^2$ commutes with the three components

$$[\hat{J}^2, \hat{J}_x] = [\hat{J}^2, \hat{J}_y] = [\hat{J}^2, \hat{J}_z] = 0 \quad .$$

This is a simple and direct calculation. It is easy to understand because J^2 does not depend on the system of axes. It is rotation invariant and it make no discrimination among the components.

Therefore, we can diagonalize simultaneously $\hat{J}^2$ and any one of the components of $\hat{\boldsymbol{J}}$, for instance, $\hat{J}_z$.

Note that $\hat{J}^2$ and one of the components is the maximal set of operators that are functions of only $\hat{J}_x$, $\hat{J}_y$, and $\hat{J}_z$, which can be diagonalized simultaneously. $\hat{J}^2$ and $\hat{J}_z$ form a CSCO if one considers only angular momentum.

Our problem is to find the eigenvectors common to $\hat{J}^2$ and $\hat{J}_z$ and the corresponding eigenvalues.

Analytic solution

At this point, we could use wave functions. Because $\boldsymbol{L}$ is invariant under dilatation, it acts only on angular variables. The calculation of eigenfunctions and eigenvalues had been done in the 19th century by Legendre and Fourier.

If we work in spherical variables with Oz as the polar axis, r is the radial coordinate, θ the colatitude, and ϕ the azimuth. The expression of $\hat{L}^2$ and $\hat{L}_z$ is not too complicated

$$\hat{L}_z = \frac{\hbar}{i}\frac{\partial}{\partial\varphi} \quad , \tag{9.5}$$

$$\hat{L}^2 = -\hbar^2\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}\right) \quad . \tag{9.6}$$

This is what Legendre and Fourier did. We look for eigenfunctions $Y_\ell^m(\theta,\varphi)$ common to $\hat{L}^2$ and $\hat{L}_z$. These are called the spherical harmonics:

$$\hat{L}^2 Y_\ell^m = \lambda^2\hbar^2 Y_\ell^m \quad \text{and} \quad \hat{L}_z Y_\ell^m = \mu\hbar Y_\ell^m \quad ,$$

where λ^2 and μ are the eigenvalues.

However, this resolution of differential equations is tedious and, more important, we would miss half of the results.

Cartan's algebraic relation is much more elegant.

9.2.2 Vectors $|j, m>$ and eigenvalues j and m

We look for the eigenvectors common to $\hat{J}^2$ and $\hat{J}_z$, and the corresponding eigenvalues. We define the vectors $|j, m>$ and the numbers j and m as

$$\hat{J}^2|j,m\rangle = j(j+1)\hbar^2|j,m\rangle \ , \tag{9.7}$$

$$\hat{J}_z|j,m\rangle = m\hbar|j,m\rangle \tag{9.8}$$

with $j \geq 0$ (the eigenvalues of $\hat{J}^2$ are positive because $\langle\psi|\hat{J}^2|\psi\rangle \geq 0$). For the moment, no other constraint exists on the possible values of j and m. We assume these vectors are orthonormal

$$\langle j, m|j', m'\rangle = \delta_{j,j'}\,\delta_{m,m'} \quad .$$

It is a theorem that this eigenbasis is unique, because $\hat{J}^2$ and $\hat{J}_z$ form a CSCO.

There are two indices because two operators are diagonalized simultaneously.

The method consists of first finding the eigenvalues of $\hat{J}^2$, then for a given value of $\hat{J}^2$, we search the eigenvalues of $\hat{J}_z$ in the corresponding eigensubspace.

This is geometrically intuitive. We fix the norm $|j|$ of a vector, and we seek the values of its components. Classically, all values between j and $-j$ are allowed; quantum mechanically, there is only a finite discrete set of values for the projection.

9.2.3 Operators $\hat{J}_\pm = \hat{J}_x \pm i\hat{J}_y$

We now follow Elie Cartan. The technique is similar to Dirac's creation and annihilation operators.

i. Consider the two operators $\hat{J}_+$ and $\hat{J}_-$:

$$\hat{J}_+ = \hat{J}_x + i\hat{J}_y\,, \quad \text{and} \quad \hat{J}_- = \hat{J}_x - i\hat{J}_y \quad . \tag{9.9}$$

$\hat{J}_+$ and $\hat{J}_-$ are Hermitian conjugates of each other $\hat{J}_+^\dagger = \hat{J}_-$, $\hat{J}_-^\dagger = \hat{J}_+$. We show that these operators enable us to move from one vector to another in the eigensubspace of $\hat{J}^2$ by increasing by one unit of $\hbar$ the eigenvalues of $\hat{J}_z$.

ii. Commutation relation of $\hat{J}_\pm$ with $\hat{J}^2$ and $\hat{J}_z$

Because $\hat{J}_\pm$ are linear combinations of $\hat{J}_x$ and $\hat{J}_y$, which commute with $\hat{J}^2$, they commute with $\hat{J}^2$

$$[\hat{J}^2, \hat{J}_\pm] = 0 \quad . \tag{9.10}$$

However, $\hat{J}_\pm$ do not commute with $\hat{J}_z$. In fact, using the relations (9.3),

$$[\hat{J}_z, \hat{J}_\pm] = [\hat{J}_z, \hat{J}_x] \pm i[\hat{J}_z, \hat{J}_y] = i\hbar\hat{J}_y \pm i(-i\hbar\hat{J}_x)\;;$$

therefore,

$$[\hat{J}_z, \hat{J}_\pm] = \pm\hbar\hat{J}_\pm \quad . \tag{9.11}$$

iii. The states $\hat{J}_\pm|j, m\rangle$ and their normalization

We apply these commutation relations to a vector $|j, m>$. We see that

$$\hat{J}^2\hat{J}_\pm|j,m\rangle = \hat{J}_\pm\hat{J}^2|j,m\rangle = j(j+1)\hbar^2\hat{J}_\pm|j,m\rangle\ , \tag{9.12}$$

$$\hat{J}_z\hat{J}_\pm|j,m\rangle = (\hat{J}_\pm\hat{J}_z \pm \hbar\hat{J}_\pm)|j,m\rangle = (m\pm 1)\hbar\hat{J}_\pm|j,m\rangle\ . \tag{9.13}$$

Therefore, the vectors $\hat{J}_\pm|j,m\rangle$ are either eigenvectors of $\hat{J}^2$ and $\hat{J}_z$ with eigenvalues $j(j+1)\hbar^2$ and $(m\pm1)\hbar$, or equal to the null vector ($\hat{J}_\pm|j,m\rangle = 0$) if $j(j+1)\hbar^2$ and $(m\pm1)\hbar$ are not eigenvalues of $\hat{J}^2$ and $\hat{J}_z$.

Therefore, the operators $\hat{J}_\pm$ can be repeatedly applied to any eigenvector and to increase or decrease the value of m by any integer in the eigensubspace of $\hat{J}^2$. They allow us to move around in this subspace.

However, as our physical intuition tells us, to a given value of j, the projection m is bounded on both sides: $-j \le m \le j$.

In fact, the square of the norm of the vector $\hat{J}_\pm|j,m\rangle$ is

$$\|\hat{J}_\pm|j,m\rangle\|^2 = \langle j,m|\hat{J}_\pm^\dagger\hat{J}_\pm|j,m\rangle = \langle j,m|\hat{J}_\mp\hat{J}_\pm|j,m\rangle\ .$$

However,

$$\hat{J}_\mp\hat{J}_\pm = (\hat{J}_x \mp i\hat{J}_y)(\hat{J}_x \pm i\hat{J}_y) = \hat{J}_x^2 + \hat{J}_y^2 \pm i[\hat{J}_x,\hat{J}_y] = \hat{J}^2 - \hat{J}_z^2 \mp \hbar\hat{J}_z \quad .$$

Therefore, owing to (9.7), we have:

$$\|\hat{J}_\pm|j,m\rangle\|^2 = j(j+1)\hbar^2 - m^2\hbar^2 \mp m\hbar^2 = (j(j+1) - m(m\pm1))\hbar^2 \ge 0\ . \tag{9.14}$$

We therefore conclude that:

1. Because the square of the norm of a vector is positive, we necessarily have the inequalities:
$$-j \le m \le j \quad . \tag{9.15}$$
2. If $m+1 \le j$ (or $m-1 \ge -j$), the vector $\hat{J}_\pm|j,m\rangle$ is nonzero and it is proportional to the vector $|j,m\pm1\rangle$. The proportionality coefficient is deduced from the norm calculated above:
$$\hat{J}_\pm|j,m\rangle = \sqrt{j(j+1) - m(m\pm1)}\,\hbar\,|j,m\pm1\rangle \tag{9.16}$$
(we make an implicit choice for the phase).
3. If the eigenvalues $m=j$ and $m=-j$ exist, we have
$$\hat{J}_+|j,j\rangle = 0\,, \quad \hat{J}_-|j,-j\rangle = 0 \quad .$$

9.2.4 Quantization

Consider the maximum value of m, $m_{\max}$. This means that $m_{\max}+1$ is not an eigenvalue. Therefore the vector $\hat{J}_+|j,m_{\max}\rangle$ is the null vector, and its norm is zero. Therefore, according to (9.14),

$$m_{\max} = j$$

and the vector $|j, j\rangle$ exists.

If we apply repeatedly the operator $\hat{J}_-$ to this vector $|j, j\rangle$, we generate a whole series of eigenvectors of $\hat{J}_z$ corresponding to the eigenvalues $\hbar(j-1), \hbar(j-2)$, and so on. However, there exists a minimum value $m_{\min}$ such that $m_{\min} - 1$ is not an eigenvalue and the vector $\hat{J}_-|j, m_{\min}\rangle$ is the null vector. Therefore, because of (9.14),

$$m_{\min} = -j \quad .$$

Consequently, in the repeated application of $\hat{J}_-$ to the vector $|j, j\rangle$, there exists an integer N such that

$$j - N = -j \quad .$$

In other words, the eigenvalues of the square of the angular momentum (9.7) are such that $2j$ is an integer

$$j = N/2 \quad . \tag{9.17}$$

For a given value of $j = N/2$, the corresponding eigensubspace is of dimension $2j + 1 = N + 1$. The eigenvalues of J_z corresponding to the set of the $N + 1$ values

$$m \in \{-j, -j+1, \ldots, j-1, j\} \tag{9.18}$$

are either integers or half integers according to the value of j.

There we are! We have found the eigenvalues, that is, the catalogue for which we were looking. If $\hat{\boldsymbol{J}}$ is an observable such that $\hat{\boldsymbol{J}} \times \hat{\boldsymbol{J}} = i\hbar\hat{\boldsymbol{J}}$, the eigenvalues of the observable $\hat{J}^2 = \hat{J}_x^2 + \hat{J}_y^2 + \hat{J}_z^2$ are of the form $j(j+1)\hbar^2$, where j is either an integer or a half integer, positive or zero. The eigenvalues of the observable $\hat{J}_z$ are of the form $m\hbar$, where m is an integer or a half-integer.

For a system in an eigenstate of $\hat{J}^2$ corresponding to the value j, the only possible values of m are the $2j + 1$ numbers $\{-j, -j+1, \ldots, j-1, j\}$.

Construction of the states $|j, m\rangle$

This algebraic analysis allows us to construct the eigenstates $|j, m\rangle$ for a given j, starting from (9.16). At this stage, we are on a formal level. Concrete applications come later.

The state $|j, j\rangle$ satisfies

$$\hat{J}_+|j, j\rangle = 0 \quad ,$$

which defines it, as we show. The states $|j, m\rangle, m = j - n$, are obtained by using (9.16) and by repeatedly applying the operator $\hat{J}_-$:

$$|j, j-n\rangle = \gamma_n(\hat{J}_-)^n|j, j\rangle \quad ,$$

where the γ_n are calculated with (9.16) and the normalization of the $|j, m\rangle$.

9.3 Orbital angular momenta

Consider now the orbital angular momentum of a particle with respect to the origin $\hat{\boldsymbol{L}} = \hat{\boldsymbol{r}} \times \hat{\boldsymbol{p}}$, which we mentioned in the beginning. After the results of the previous section, if $\ell(\ell+1)\hbar^2$ are the eigenvalues of $\hat{L}^2 (\ell \geq 0)$ and $m\hbar$ the eigenvalues of $\hat{L}_z$, then 2ℓ and $2m$ are integers.

However, in this case ℓ and m are integers.

9.3.1 Formulae in spherical coordinates

In addition to the formulae (9.5) and (9.6) above, a useful formula for our purpose in spherical coordinates is the following. The operators $\hat{L}_\pm$ have the form

$$\hat{L}_\pm = \hat{L}_x \pm i\hat{L}_y = \hbar e^{\pm i\varphi} \left(\pm \frac{\partial}{\partial \theta} + i \cot\theta \frac{\partial}{\partial \varphi} \right) \quad . \tag{9.19}$$

9.3.2 Integer values of m and ℓ

The state of a particle in space can be described by a wave function $\psi(\boldsymbol{r}) = \psi(x, y, z)$. In spherical coordinates, this wave function becomes a function $\Psi(r, \theta, \varphi)$. The operator $\hat{L}_z$ has the very simple form

$$\hat{L}_z = \frac{\hbar}{i} \frac{\partial}{\partial \varphi} \quad . \tag{9.20}$$

Consider an eigenstate of the projection on the z-axis of the angular momentum of the particle with the eigenvalues $m\hbar$. The corresponding wave function $\psi_m(\boldsymbol{r})$ satisfies

$$\hat{L}_z \psi_m(\boldsymbol{r}) = m\hbar \psi_m(\boldsymbol{r}) \quad .$$

The form (9.20) of $\hat{L}_z$ gives us the simple φ dependence of the wave function

$$\psi_m(\boldsymbol{r}) = \Phi_m(r, \theta) e^{im\varphi} \quad ,$$

where $\Phi_m(r, \theta)$ is arbitrary at this point. The function $\psi_m(\boldsymbol{r})$ is a particular case of a wave function $\psi(x, y, z)$. In the change $\varphi \to \varphi + 2\pi$, x, y and z do not change and the function ψ_m is unchanged. It must therefore be a periodic function of φ with period 2π. Therefore,

$$e^{im\varphi} = e^{im(\varphi + 2\pi)} \Rightarrow e^{i2\pi m} = 1 \quad .$$

Therefore, in the case of an orbital angular momentum, m must be an integer.

In the above analysis, we have seen that m and j differ by an integer. Therefore, for an orbital angular momentum the value of ℓ is a nonnegative integer.

9.3.3 Spherical harmonics

The eigenfunctions common to $\hat{L}^2$ and $\hat{L}_z$, denoted $Y_\ell^m(\theta,\varphi)$, are called the spherical harmonics. The spherical harmonics associated with the eigenvalues $\ell(\ell+1)\hbar^2$ and $m\hbar$ satisfy

$$\hat{L}^2 Y_\ell^m(\theta,\varphi) = \ell(\ell+1)\hbar^2\, Y_\ell^m(\theta,\varphi)\ , \tag{9.21}$$

$$\hat{L}_z Y_\ell^m(\theta,\varphi) = m\hbar\, Y_\ell^m(\theta,\varphi) \quad . \tag{9.22}$$

We have just seen that their φ dependence is simply $(e^{im\varphi})$, so that they factorize as

$$Y_\ell^m(\theta,\varphi) = F_{\ell,m}(\theta)\, e^{im\varphi}\ .$$

The spherical harmonics form a Hilbert basis of square integrable functions on the sphere of radius one. They are completely defined in the following way:

1. They are normalized:

$$\int\int (Y_\ell^m(\theta,\varphi))^*\ Y_{\ell'}^{m'}(\theta,\varphi)\, \sin\theta\, d\theta\, d\varphi = \delta_{\ell,\ell'}\, \delta_{m,m'} \quad .$$

2. Their phases are such that the recursion relation (9.16) which we repeat below, is satisfied and that $Y_\ell^0(0,0)$ is real and positive:

$$\hat{L}_\pm Y_\ell^m(\theta,\varphi) = \sqrt{\ell(\ell+1) - m(m\pm 1)}\ \hbar\ Y_\ell^{m\pm 1}(\theta,\varphi) \quad . \tag{9.23}$$

3. Starting from the relation

$$\hat{L}_+ Y_\ell^\ell(\theta,\varphi) = 0\ , \tag{9.24}$$

we obtain with (9.19)

$$Y_\ell^\ell(\theta,\varphi) = C(\sin\theta)^\ell e^{i\ell\varphi}\ ,$$

where the normalization constant C is determined by the above constraints.

Examples of spherical harmonics

The spherical harmonics play an important role in atomic and molecular physics. They form with their linear combinations the atomic orbitals of one-external electron atoms, in particular atomic hydrogen which we describe in the next chapter. The first are

$$\ell = 0 \qquad Y_0^0(\theta,\varphi) = \frac{1}{\sqrt{4\pi}} \tag{9.25}$$

$$\ell = 1 \qquad Y_1^1(\theta,\varphi) = -\sqrt{\frac{3}{8\pi}}\sin\theta\, e^{i\varphi} \tag{9.26}$$

$$Y_1^0(\theta,\varphi) = \sqrt{\frac{3}{4\pi}}\cos\theta \tag{9.27}$$

$$Y_1^{-1}(\theta,\varphi) = \sqrt{\frac{3}{8\pi}}\sin\theta\, e^{-i\varphi} \quad . \tag{9.28}$$

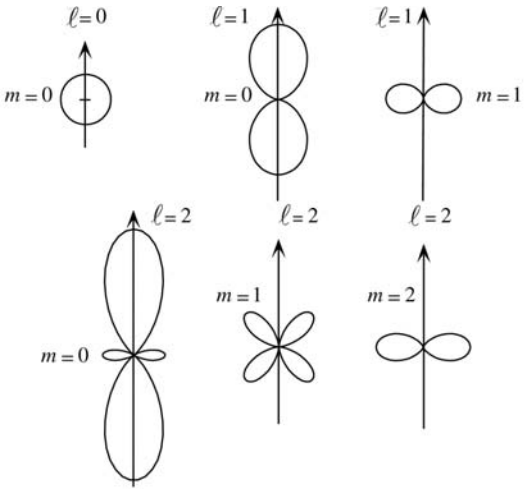


Fig. 9.1. Graph as a function of the polar angle θ of $|Y_\ell^m(\theta,\varphi)|^2 = |F_{\ell,m}(\theta)|^2$ for $\ell = 0, 1, 2$ and $|m| \leq l$.

In Figure 9.1, the squares $|Y_\ell^m(\theta,\varphi)|^2 = |F_{\ell,m}(\theta)|^2$ of spherical harmonics corresponding to the first values of ℓ and m are represented in polar coordinates in terms of θ.

Wave function of a particle in an eigenstate of the orbital angular momentum

The wave functions $\psi_{\ell,m}(\boldsymbol{r})$ of particles in an eigenstate of the orbital angular momentum are therefore of the form

$$\psi_{\ell,m}(\boldsymbol{r}) = R_{\ell,m}(r) Y_\ell^m(\theta,\varphi) \quad .$$

The radial dependence of these functions is contained in the radial wave function $R_{\ell,m}(r)$, which can have any form a priori.

9.4 Rotation energy of a diatomic molecule

It seems that the first one who understood empirically the quantization of angular momentum was Ehrenfest (in an article published on June 15, 1913) just just before Bohr's article on the hydrogen atom (published in July 1913). Ehrenfest noticed that $\hbar$ had the dimension of an angular momentum, and he postulated the quantization, without giving any proof, in order to explain the variation with temperature of the specific heats of diatomic molecular gases (he found an improvement on the theory of Einstein and Stern).

9.4.1 Diatomic molecule

A simple illustration of the quantization of the values of $\hat{L}^2$ is obtained through the rotational energy spectrum of a molecule. Such a spectrum is presented in Figure 9.2 for the diatomic cesium molecule Cs_2. It has been obtained[1] by measuring the frequency of the photons needed to ionize the Cs_2 molecules that are formed in a very cold atomic vapor of cesium atoms (temperature $\sim 100\ \mu$K). The data of Figure 9.2, which represent only a small fraction of the total spectrum, exhibit a series of peaks characteristic of a quantized rotational energy.

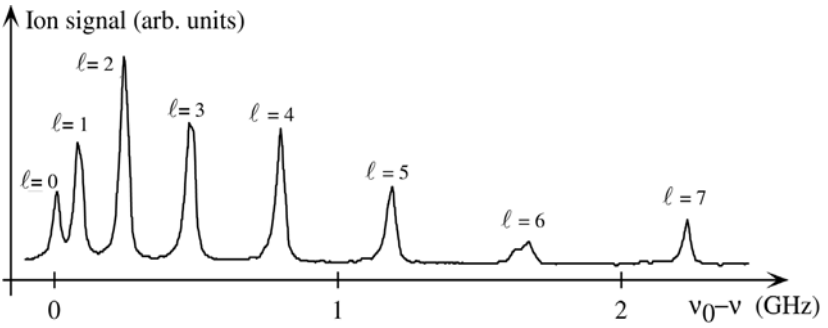


Fig. 9.2. Rotational spectrum of cold Cs_2 molecules, showing the quantization of $\hat{L}^2$. This spectrum is obtained by measuring the number of molecular ions produced by a laser beam crossing the assembly of cold molecules, as a function of the laser frequency ν. The height of each peak is proportional to the population of the corresponding rotational level ℓ. (Courtesy of Pierre Pillet.)

One can visualize a diatomic molecule formed by two atoms of mass M separated by a distance R as a two-body system bound by a harmonic potential. Classically, if the interatomic distance R is in its equilibrium position, the molecule has a rotational energy

$$E_{\text{rot}} = \frac{L^2}{2I} \ , \tag{9.29}$$

where $I = MR^2/2$ is the moment of inertia of the system and L is its angular momentum with respect to its center of gravity. In quantum mechanics, this result transposes into

$$E_{\text{rot}}(\ell) = \frac{\hbar^2 \ell(\ell+1)}{2I} \ , \tag{9.30}$$

where the rotational energy is quantized. Formula 9.30 gives a very good account for the series of peaks of fig. 9.2. The distance between two consecutive

[1] These data, corresponding to the seventeenth excited vibrational state, are extracted from A. Fioretti et al., *Eur. Phys. J.* D **5**, 389 (1999).

peaks increases linearly with the peak index, as expected from

$$E_{\rm rot}(\ell) - E_{\rm rot}(\ell - 1) = \frac{\hbar^2}{I}\ell \ .$$

The moment of inertia deduced from this spectrum corresponds to a distance $R = 1.3$ nm between the two cesium atoms. This distance, much larger than the usual interatomic spacing in diatomic molecules, indicates that the Cs_2 dimer is actually prepared in a long-range molecular state.

If one investigates the absorption spectrum of the cold molecular gas on a much wider range, one finds several series of lines such as the one of Figure 9.2. Each series corresponds to a given vibrational state of the molecule. The moments of inertia associated with these series differ slightly from one another: this is a consequence of the variation of the average distance between the two atoms in the various vibrational states of the molecule.

Remark. The study of rotational excitations of molecules is an important field of research in physics, chemistry, and astrophysics. If we choose x, y, and z along the principal axes of inertia of a rigid rotator and denote I_x, I_y, and I_z the corresponding moments of inertia, the rotational energy spectrum arises from the Hamiltonian:

$$\hat{H}_R = \frac{\hat{L}_x^2}{2I_x} + \frac{\hat{L}_y^2}{2I_y} + \frac{\hat{L}_z^2}{2I_z} \quad ,$$

where we neglect the vibrational energies (there exist subtleties in this problem because $\hat{L}_x$, $\hat{L}_y$, $\hat{L}_z$ refer to a body-fixed reference system and not to a space-fixed one). If the three moments of inertia are all different, the diagonalization of such an operator cannot be written in closed form, except for low values of $\hat{\boldsymbol{L}}^2$. If two of the moments of inertia are equal, for instance, $I_x = I_y \equiv I$, the spectrum is simple because $\hat{H}_R = \left(\hat{L}^2 - \hat{L}_z^2\right)/(2I) + \hat{L}_z^2/(2I_z)$ whose eigenvalues are

$$E_{l,m} = \hbar^2 \left(\frac{l(l+1) - m^2}{2I} + \frac{m^2}{2I_z} \right) \ .$$

From this point of view, a diatomic molecule can be considered as a rigid rotator for which $I_x = I_y = MR^2/2$ and $I_z \simeq 0$. Therefore the excitation energies of the z term are very large, and we can restrict to the ground state $m = 0$, thus ending up with the form (9.30).

9.4.2 The CO molecule

The CO molecule plays an important role in astrophysics, as we show in the last chapter. It emits in the millimeter range.

The symmetric molecules such as H_2 or O_2 have the defect that they do not have an electric dipole moment (such as the molecule NH_3) and therefore they do not have an electric dipole emission.

The nonsymmetric molecule CO does possess a permanent electric dipole moment and it emits intensively in these transitions. This molecule has a size $R = 0.1128$ nm and the masses $M_{\mathrm{C}} = 12$ amu, $M_{\mathrm{O}} = 15.99492$ amu, which leads to a frequency of the transition $\ell = 2 \to \ell = 1$ of 230.54 GHz, or equivalently a series of transitions of frequencies $\nu_\ell = 115.27\,\ell$ with $\ell = 1, 2, \ldots$.

Carbon and oxygen are comparatively abundant in the interstellar medium because these elements are synthesized in most stars. The CO molecule has a relative abundance of the order of 10^{-5} compared to the hydrogen atom. Its distribution is in general more dense around galactic nuclei rather than on the edges. The CO molecule is a very useful indicator because it is one of the most visible molecules in millimetric radioastronomy. We show examples in chapter 14.

9.5 Angular momentum and magnetic moment

How can we directly compare our results with experiment? Molecular spectra are interesting, but they do not give us access to nonclassical angular momenta.

In fact, we have discovered possible half-integer values of the angular momentum, and we know that angular momenta with classical analogues correspond to integer values of (j, m). Are such half-integer values mathematical artifacts or do they exist in nature?

It is quite possible to measure angular momenta directly, but at our level the description of the corresponding experiments would be complicated.

We rely on a phenomenon that is intuitively related to angular momentum, that is, magnetism, which has numerous applications from nuclear magnetic resonance to superconductivity.

The experimental evidence for the quantization of angular momenta relies to a large extent on the fact that when charged particles rotate, they possess magnetic moments.

9.5.1 Classical model

Classically, to a rotating charge distribution there corresponds a magnetic moment that is proportional to the angular momentum.

We can make a classical model of an atom (let's say hydrogen for simplicity) by considering a particle with mass m_e and charge $-q$ (the electron), rotating with a uniform velocity v along a circle of radius r centered on a charge $+q$. This positive fixed charge represents the nucleus and it is supposed to be much heavier than the electron. The angular momentum of this system is

$$\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p} = m_e r v\, \boldsymbol{u}\ , \tag{9.31}$$

where $\boldsymbol{u}$ is the unit vector orthogonal to the orbital plane of the electron. The magnetic moment of this elementary current loop is

$$\boldsymbol{\mu} = IS\,\boldsymbol{u}\ , \tag{9.32}$$

where $I = -qv/(2\pi r)$ is the intensity in the loop, and $S = \pi r^2$ is the loop area. We then find a remarkably simple relation between the angular momentum and the magnetic moment of this classical system:

$$\boldsymbol{\mu} = \gamma_0 \boldsymbol{L}\,, \quad \text{with} \quad \gamma_0 = \frac{-q}{2m_e} \quad . \tag{9.33}$$

Note that the proportionality coefficient, called the gyromagnetic ratio, does not depend on the radius r of the trajectory of the electron, nor on its velocity v. Strictly speaking, the presence of an external magnetic field perturbs the electronic motion and modifies this very simple relation, but one can show that this perturbation is very weak for realistic fields, and we neglect it here.

If we place this magnetic moment in a magnetic field $\boldsymbol{B}$ the system has a magnetic energy

$$W_M = -\boldsymbol{\mu} \cdot \boldsymbol{B}\ , \tag{9.34}$$

and a torque

$$\boldsymbol{\Gamma} = \boldsymbol{\mu} \times \boldsymbol{B} \tag{9.35}$$

is exerted on the magnetic moment.

From (9.34), one could naively expect that the magnetic moment of the atom would get aligned with the local magnetic field, as does the needle of a compass. However, the proportionality between the magnetic moment and the angular momentum gives rise to a radically different phenomenon, analogous to the gyroscopic effect. The evolution of the angular momentum is given by $d\boldsymbol{L}/dt = \boldsymbol{\Gamma}$. The proportionality between $\boldsymbol{L}$ and $\boldsymbol{\mu}$ then implies

$$\frac{d\boldsymbol{\mu}}{dt} = -\gamma_0\ \boldsymbol{B} \times \boldsymbol{\mu} \quad . \tag{9.36}$$

Consequently, for an atom at $\boldsymbol{r}$, the magnetic moment does not align with the axis of the local magnetic field $\boldsymbol{B}(\boldsymbol{r})$, but it precesses around this axis with the angular frequency

$$\omega_0 = -\gamma_0 B(\boldsymbol{r})\ , \tag{9.37}$$

as shown in Figure 9.3. The quantity ω_0 is called the Larmor frequency.

This precession phenomenon is very important in practice. It is a particular case of a general theorem of electrodynamics[2] proven by Larmor in 1897. This problem was considered independently the same year by H.A. Lorentz.

[2] See, for example, J.D. Jackson, *Classical Electrodynamics*. New York: Wiley, (1975).

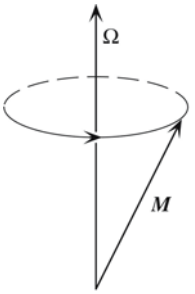


Fig. 9.3. Time evolution of the components of a magnetic moment placed in a field $\boldsymbol{B}(\boldsymbol{r})$ along the z-axis.

9.5.2 Quantum transposition

The quantum transposition of this result consists of assuming that the same proportionality relation remains true in quantum mechanics. Any system in an eigenstate of the square of the angular momentum $\hat{J}^2$ possesses a magnetic moment $\hat{\boldsymbol{\mu}}$ that is proportional to $\hat{\boldsymbol{J}}$,

$$\hat{\boldsymbol{\mu}} = \gamma \hat{\boldsymbol{J}} \quad . \tag{9.38}$$

This is an hypothesis. It is verified experimentally.

For an orbital electron, the gyromagnetic factor is the same as in classical physics

$$\gamma_0 = \frac{-q}{2m_e} \quad . \tag{9.39}$$

In general, the gyromagnetic factor γ of (9.38) depends on the value of j in a complex system, such as a nucleus.

9.5.3 Experimental consequences

In a magnetic field $\boldsymbol{B}$, the system has a magnetic potential energy $\hat{H}_M$ which we deduce from (9.34),

$$\hat{H}_M = -\hat{\boldsymbol{\mu}} \cdot \boldsymbol{B} \quad . \tag{9.40}$$

The quantization of angular momenta therefore leads to a perfectly analogous quantization of magnetic moments, up to a coefficient. But measuring magnetic moments is easier and more intuitive than measuring angular momenta.

Orbital angular momenta

From the properties of $\hat{\boldsymbol{L}}$, we deduce the following results, anticipating the conclusions of the next chapter:

1. Consider an electron moving in a central potential. We suppose that the electron is in a given energy level E_n and in an eigenstate of the orbital angular momentum with eigenvalue $\ell(\ell+1)\hbar^2$. As a consequence of rotation invariance, in the absence of an external magnetic field, the $2\ell+1$ states corresponding to $m=-\ell,\ldots,+\ell$ have the same energy E_n. Let us call these states $|n,\ell,m\rangle$; they are eigenstates of $\hat{L}_z$ with eigenvalues $m\hbar$. Using the assumption (9.40), the state $|n,\ell,m\rangle$ is also an eigenstate of $\hat{\mu}_z$. The corresponding eigenvalue is $\mu_z=\gamma_0 m\hbar$. The negative quantity

$$\mu_B=\gamma_0\hbar=\frac{-q\hbar}{2m_e}\sim -9.27\ 10^{-24}\ \mathrm{J\ T^{-1}} \tag{9.41}$$

is called the Bohr magneton. From the properties of orbital angular momenta $\hat{\boldsymbol{L}}$, we can deduce the following.

2. If we place the system in a magnetic field $\boldsymbol{B}$ parallel to z, the degeneracy is lifted. The state $|n,\ell,m\rangle$ is an eigenstate of the observable $\hat{H}_M$, with eigenvalue:

$$W_m=-m\mu_B B \quad .$$

We therefore expect to observe a splitting of the atomic energy level E_n into $2\ell+1$ sublevels, equally spaced by the interval $\Delta E=-\mu_B B$. This is called the Zeeman effect. It can be observed in a transition $E_n\to E_{n'}$. In the absence of a magnetic field, this transition occurs at a single frequency $(E_n-E_{n'})/2\pi\hbar$. If we apply a field B, several lines appear. The number of such lines is directly related to the angular momenta ℓ and ℓ' of the initial and final levels.

Notice that if all angular momenta were orbital, then $2j+1$ would always be odd. This simple qualitative prediction existed in classical physics. It was made by Lorentz and by Larmor in 1897. (They both had in mind a three-dimensional harmonic oscillator motion of electrons.)

9.5.4 Larmor precession

Another consequence of this proportionality relation between $\hat{\boldsymbol{J}}$ and $\hat{\boldsymbol{\mu}}$ is the Larmor precession phenomenon, which takes place at the quantum level for the expectation values $\langle\boldsymbol{\mu}\rangle$. The Ehrenfest theorem yields

$$\frac{d}{dt}\langle\boldsymbol{\mu}\rangle=\frac{1}{i\hbar}\langle[\hat{\boldsymbol{\mu}},\hat{H}_M]\rangle \quad ,$$

where we assume that only the term $\hat{H}_M$ in the Hamiltonian does not commute with $\hat{\boldsymbol{\mu}}$. Indeed the other terms of the Hamiltonian are supposed to be rotation invariant. Hence, they commute with $\hat{\boldsymbol{J}}$ and with $\hat{\boldsymbol{\mu}}$. Owing to (9.3), the commutation relations of $\hat{\boldsymbol{\mu}}$ are

$$\hat{\boldsymbol{\mu}}\times\hat{\boldsymbol{\mu}}=i\hbar\gamma\hat{\boldsymbol{\mu}} \quad .$$

Therefore, a simple calculation yields

$$\frac{d}{dt}\langle \boldsymbol{\mu} \rangle = -\gamma \boldsymbol{B} \times \langle \boldsymbol{\mu} \rangle \quad .$$

The expectation value $\langle \boldsymbol{\mu} \rangle$ satisfies the same equations of motion as those found above for the classical quantity (9.36). This comes from the fact that the Hamiltonian is linear in $\hat{\boldsymbol{\mu}}$. The measurement of the Larmor precession frequency provides a direct determination of the gyromagnetic ratio γ and a consistency test of the results. We therefore have an experimental means to measure angular momenta via the measurements of magnetic moments.

9.5.5 What about half-integer values of j and m?

To conclude this chapter, we come back to the half-integer values of j and m that we found in the general derivation of the eigenvalues of angular momenta. Concerning an orbital angular momentum, we did not accept such values. Nevertheless, one may wonder whether these values appear in Nature, or whether they are simply a mathematical artifact.

If all angular momenta are orbital angular momenta, j is an integer and $2j + 1$ is odd. Initially, Zeeman performed his experiments with cadmium and zinc (atoms with two external electrons), and he saw, as expected, an odd number of lines. But then there was a drama. In fact, Zeeman continued his experiments on alkali atoms, sodium, potassium, and he found an even number of lines. This was called the *anomalous Zeeman effect*. It is astonishing that, for 25 years, nobody was able to explain those even numbers. The anomalous Zeeman effect seemed to be the greatest challenge given to the physics community. Of course we come back to this when we study spin 1/2.

10

The Hydrogen Atom

The explanation of spectroscopic data was one of the first great victories of quantum theory. In modern science and technology, the mastery of atomic physics is responsible for decisive progress ranging from laser technology to the exploration of the cosmos.

The particular case of the hydrogen atom is perhaps the most striking. Its particularly simple spectrum delivered the first clues of quantum laws. It has been used as a testbed for the development of quantum theory. Its hyperfine structure is responsible both for the hydrogen maser and for a revolution in astrophysics, because the corresponding 21-cm line has been extensively studied in radioastronomy to probe the structure of the interstellar and intergalactic media as we show in chapter 14. Furthermore, the hydrogen atom is probably the physical system that is known with the greatest accuracy. It can be calculated "completely" in the sense that the accuracy of present experimental results is the same as the accuracy of theoretical computer calculations, of the order of 10^{-12} to 10^{-13} relative accuracy (the only competitor is celestial mechanics).

Here, we first consider technical points, that is, how a two-body problem, where the potential depends only on the distance of the particles, reduces to a one-particle problem. In the case of a central potential, we use the invariance properties of the problem in order to choose the CSCO made up with the Hamiltonian and the angular momentum, $\hat{H}, \hat{L}^2$ and $\hat{L}_z$, and we show the traditional quantum numbers $\{n, l, m\}$ used in atomic physics show up. We study the Coulomb potential and we calculate the bound state energies of hydrogen in the nonrelativistic approximation and recover the $E_n = -E_I/n^2$ formula obtained by Bohr in 1913. We end with some considerations on similar atoms, muonic atoms, where the electron is replaced by a heavier sibling, the muon whose mass is 207 times larger.

Taking into account relativistic kinematics and spin effects requires a formalism not covered in this book: the Dirac equation. These corrections are small compared to the leading terms. Up to that point, the problem can be

solved analytically. Other fine structure effects, such as the Lamb shift, require the more elaborate formalism of quantum field theory.

The theoretical treatment of complex atoms (i.e., atoms with more than one electron) involves serious computational problems, even at the nonrelativistic stage. The helium atom with its two electrons can only be calculated numerically. Actually this calculation was considered as the first true test of quantum mechanics, because the much simpler case of hydrogen could be treated successfully by several other approaches issued from the "old" Bohr–Sommerfeld quantum theory. Owing to the accuracy of present numerical calculations, the helium atom is considered to be known exactly.[1]

10.1 Two-body problem; relative motion

Consider a system of two particles, of masses M_1 and M_2, and of positions $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$, whose mutual interaction is given by a potential $V(\boldsymbol{r}_1 - \boldsymbol{r}_2)$. The potential depends only on the relative position of the particles. The Hamiltonian is

$$\hat{H} = \frac{\hat{\boldsymbol{p}}_1{}^2}{2M_1} + \frac{\hat{\boldsymbol{p}}_2{}^2}{2M_2} + V(\hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_2) \ , \tag{10.1}$$

and the system is described by wave functions $\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2)$.

We can separate the global motion of the center of mass of the system and the relative motion of the two particles. We introduce the position and momentum operators of the center of mass

$$\hat{\boldsymbol{R}} = \frac{M_1\hat{\boldsymbol{r}}_1 + M_2\hat{\boldsymbol{r}}_2}{M_1 + M_2} \ , \quad \hat{\boldsymbol{P}} = \hat{\boldsymbol{p}}_1 + \hat{\boldsymbol{p}}_2 \ , \tag{10.2}$$

and the relative position and momentum operators:

$$\hat{\boldsymbol{r}} = \hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_2 \ , \quad \hat{\boldsymbol{p}} = \frac{M_2\hat{\boldsymbol{p}}_1 - M_1\hat{\boldsymbol{p}}_2}{M_1 + M_2} \ . \tag{10.3}$$

We can rewrite the Hamiltonian as

$$\hat{H} = \hat{H}_{\text{c.m.}} + \hat{H}_{\text{rel}} \ , \tag{10.4}$$

with

$$\hat{H}_{\text{c.m.}} = \frac{\hat{\boldsymbol{P}}^2}{2M} \ , \qquad \hat{H}_{\text{rel}} = \frac{\hat{\boldsymbol{p}}^2}{2\mu} + V(\hat{\boldsymbol{r}}) \ . \tag{10.5}$$

We have introduced the total mass M and the reduced mass μ:

$$M = M_1 + M_2 \ , \quad \mu = \frac{M_1 M_2}{M_1 + M_2} \ . \tag{10.6}$$

[1] T. Kinoshita, "Ground state of the helium atom," *Phys. Rev.*, **105**, 1490 (1957).

Just as in classical mechanics, the Hamiltonian $\hat{H}$ separates in the sum of (i) the Hamiltonian $\hat{H}_{\text{c.m.}}$ describing the free motion of the center of mass (momentum $\boldsymbol{P}$, total mass M) and (ii) the Hamiltonian $\hat{H}_{\text{rel}}$ which describes the relative motion of the two particles in the potential $V(\boldsymbol{r})$ (momentum $\boldsymbol{p}$, reduced mass μ).

Let $\{\hat{X}_i\}$ and $\{\hat{P}_i\}$ be the coordinates of $\hat{\boldsymbol{R}}$ and $\hat{\boldsymbol{P}}$, and $\{\hat{x}_i\}$ and $\{\hat{p}_i\}$ those of $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{p}}$. The commutation relations are

$$[\hat{X}_j, \hat{P}_k] = i\hbar\delta_{jk}\,, \quad [\hat{x}_j, \hat{p}_k] = i\hbar\delta_{jk} \tag{10.7}$$

and

$$[\hat{X}_j, \hat{p}_k] = 0\,, \quad [\hat{x}_j, \hat{P}_k] = 0 \quad . \tag{10.8}$$

In other words, the position and momentum operators of the center-of-mass and of the relative variables obey the canonical commutation relations (10.7), and any variable associated with the center-of-mass motion commutes with any other variable associated with the relative motion (10.8).

These commutation relations imply

$$[\hat{\boldsymbol{P}}, \hat{H}_{\text{rel}}] = 0\,, \quad [\hat{\boldsymbol{P}}, \hat{H}] = 0\,, \quad [\hat{H}, \hat{H}_{\text{rel}}] = 0\ . \tag{10.9}$$

Consequently, there exists a basis of eigenfunctions of $\hat{H}$ that are simultaneously eigenfunctions of $\hat{\boldsymbol{P}}$ and $\hat{H}_{\text{rel}}$. The eigenfunctions of $\hat{\boldsymbol{P}}$ are the plane waves $e^{i\boldsymbol{K}\cdot\boldsymbol{R}}$, where $\boldsymbol{K}$ is an arbitrary wavevector. Therefore, the desired basis of eigenfunctions of $\hat{H}$ has the form

$$\Psi(\boldsymbol{R}, \boldsymbol{r}) = e^{i\boldsymbol{K}\cdot\boldsymbol{R}}\, \psi(\boldsymbol{r})\ ,$$

where $\psi(\boldsymbol{r})$ is an eigenfunction of $\hat{H}_{\text{rel}}$:

$$\hat{H}_{\text{rel}}\, \psi(\boldsymbol{r}) = E\, \psi(\boldsymbol{r})\ . \tag{10.10}$$

The eigenvalues E_{tot} of $\hat{H}$ are

$$E_{\text{tot}} = \frac{\hbar^2 K^2}{2M} + E\ , \tag{10.11}$$

the sum of the kinetic energy of the global system ($\hat{H}_{\text{c.m.}}$) and the internal energy ($\hat{H}_{\text{rel}}$).

One consequence of this is the conservation of the total momentum $d\langle\boldsymbol{P}\rangle/dt = 0$. This is due to the fact that the potential depends only on the relative variable $\boldsymbol{r} = \boldsymbol{r}_1 - \boldsymbol{r}_2$; in other words, the Hamiltonian of the system is translation invariant.

Because $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{p}}$ have canonical commutation relations, the problem reduces to the quantum motion of a particle of mass μ in the potential $V(\boldsymbol{r})$. For an atomic system made of an electron ($M_1 = m_e$) and the rest of the atom (M_2), we have $M_2 \gg m_e$. Therefore, we can neglect the small difference between the reduced mass μ and the electron mass m_e, remembering that it is easy to correct for reduced mass effects if necessary. We are interested in the eigenvalue problem of H_{rel}.

10.2 Motion in a central potential

The Coulomb potential is central, that is, it only depends on the distance $r = |\boldsymbol{r}|$ of the two particles. The problem is therefore rotation invariant.

10.2.1 Spherical coordinates, CSCO

Owing to the symmetry of the problem, it is appropriate to work in spherical coordinates.

The Laplacian Δ has the following expression (which is easily obtained by writing the square of the angular momentum $\hat{\boldsymbol{L}} = -i\hbar \boldsymbol{r} \times \boldsymbol{\nabla}$ and expanding)

$$\Delta = \frac{1}{r}\frac{\partial^2}{\partial r^2} r - \frac{1}{r^2\hbar^2}\hat{L}^2 \quad . \tag{10.12}$$

Equation (10.10) is then written as

$$\left(-\frac{\hbar^2}{2m_e}\ \frac{1}{r}\ \frac{\partial^2}{\partial r^2} r + \frac{\hat{L}^2}{2m_e r^2} + V(r) \right) \psi(\boldsymbol{r}) = E\ \psi(\boldsymbol{r}) \quad . \tag{10.13}$$

The Hamiltonian $\hat{H}_{\rm rel}$ commutes with the three angular momentum operators $\hat{L}_i$, $i = x, y, z$. Each $\hat{L}_i$ commutes with $\hat{L}^2$. In addition $\hat{L}_i$ only acts on the variables θ and φ, and it commutes with $r, \partial/\partial r,\ V(r)$. In other words, the Hamiltonian $\hat{H}_{\rm rel}$, which from now on is denoted $\hat{H}$ for simplicity, commutes with the angular momentum:

$$[\hat{H}, \hat{\boldsymbol{L}}] = 0 \quad .$$

Consequently, $\hat{H}, \hat{L}^2$ and a given component of $\hat{\boldsymbol{L}}$ (e.g., $\hat{L}_z$) form a set of commuting observables. We verify a posteriori that this set is complete by checking that the basis corresponding to the common eigenfunction is unique.

The relation $[\hat{H}, \hat{\boldsymbol{L}}] = 0$ implies the conservation of the angular momentum: $d\langle \boldsymbol{L} \rangle / dt = 0$. This is due to the fact that the Hamiltonian of the system is rotation invariant.

10.2.2 Eigenfunctions common to $\hat{\boldsymbol{H}}$, $\hat{\boldsymbol{L}}^2$, and $\hat{\boldsymbol{L}}_z$

Separation of the angular variables

Part of the eigenvalue problem (10.13) is already solved because we know the form of the eigenfunctions common to $\hat{L}^2$ and $\hat{L}_z$. These are the spherical harmonics. We separate the variables in the following way:

$$\psi_{\ell,m}(\boldsymbol{r}) = R_\ell(r)\, Y_{\ell,m}(\theta, \varphi) \ , \tag{10.14}$$

$$\hat{L}^2 \psi_{\ell,m}(\boldsymbol{r}) = \ell(\ell+1)\hbar^2\, \psi_{\ell,m}(\boldsymbol{r}) \ , \tag{10.15}$$

$$\hat{L}_z \psi_{\ell,m}(\boldsymbol{r}) = m\hbar\, \psi_{\ell,m}(\boldsymbol{r}) \ , \tag{10.16}$$

where ℓ and m are integers, with $|m| \leq \ell$. Substituting in (10.13), the eigenvalue equation becomes

$$\left(-\frac{\hbar^2}{2m_e}\frac{1}{r}\frac{d^2}{dr^2}r + \frac{\ell(\ell+1)\hbar^2}{2m_e r^2} + V(r)\right) R_\ell(r) = E\ R_\ell(r)\ . \tag{10.17}$$

This equation is independent of the quantum number m. This is why we have not put an index m for the unknown function $R_\ell(r)$ in equation (10.14). This differential equation is the radial equation and $R_\ell(r)$ is called the radial wave function.

The normalization of the wave function, which we must impose on finding bound states, is $\int |\psi(\boldsymbol{r})|^2\, d^3r = 1$; in spherical coordinates,

$$\int d^2\Omega \int_0^\infty dr\ r^2\ |\psi(r,\theta,\varphi)|^2 = 1\ .$$

Here, Ω is the solid angle with $d^2\Omega = \sin\theta\, d\theta\, d\varphi$. Because the spherical harmonics are normalized, we obtain, for the radial wave function $R_\ell(r)$, the condition

$$\int_0^\infty dr\ r^2\ |R_\ell(r)|^2 = 1 \quad . \tag{10.18}$$

Introducing the reduced wave function $u_\ell(r) = r\, R_\ell(r)$, the Schrödinger equation becomes

$$\left(-\frac{\hbar^2}{2m_e}\frac{d^2}{dr^2} + \frac{\ell(\ell+1)\hbar^2}{2m_e r^2} + V(r)\right) u_\ell(r) = E\ u_\ell(r)\ , \tag{10.19}$$

with $\int_0^\infty |u_\ell(r)|^2\, dr = 1$. One can prove that any normalizable solution $R_\ell(r)$ is bounded at the origin, and therefore $u_\ell(0) = 0$.

This equation has the structure of the Schrödinger equation describing the one dimensional motion of a particle of mass m_e in the potential:

$$V_{\text{eff}}(r) = V(r) + \frac{\ell(\ell+1)\hbar^2}{2m_e r^2}\ . \tag{10.20}$$

This effective potential is the superposition of the interaction potential between the two particles 1 and 2, and a centrifugal barrier term which is repulsive and increases as the angular momentum increases (Figure 10.1).

10.2.3 Quantum numbers

The energy levels depend on the parameter ℓ, but they do not depend on the projection m. For each value of ℓ, corresponding to a given value of the square of the angular momentum, each level has a degeneracy of degree $(2\ell+1)$. For a given ℓ, we deal with a one-dimensional problem of the same type as studied in chapter 5. The bound state energy levels ($E < 0$) correspond to solutions $R_\ell(r)$ that satisfy (10.18).

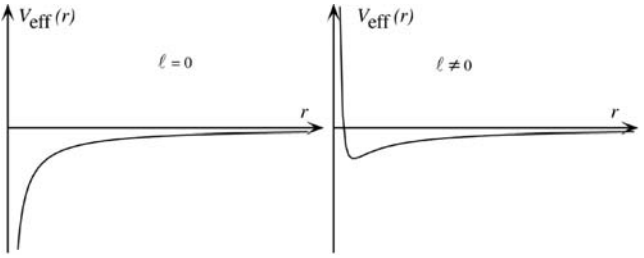


Fig. 10.1. Effective potential that enters the one-dimensional Schrödinger equation for the reduced radial wave function $u_\ell(r)$. For $\ell = 0$ (left) the motion occurs in the "bare" potential $V(r)$; for $\ell \neq 0$ (right) the effective potential is the superposition of $V(r)$ and of the centrifugal barrier $\ell(\ell+1)\hbar^2/(2m_e r^2)$. The figure is drawn for a Coulomb potential $V(r) \propto 1/r$.

The radial quantum number n'

For a given ℓ, we can arrange the possible values of the bound state energies in an increasing sequence, which we label by an integer n' ($n' = 0, 1, 2 \ldots$), the state $n' = 0$ being the most strongly bound. Depending on the potential, this sequence may be finite (as for a square well potential) or infinite (as for the Coulomb potential).

The general mathematical properties of the differential equation (10.17), together with the conditions that $R_\ell(0)$ is finite and that $R_\ell(r)$ can be normalized (10.18), show that this number n' corresponds to the number of nodes of the radial wave function; the number of times it vanishes between $r = 0$ and $r = \infty$. This is independent of the form of the potential $V(r)$ (provided it is not too pathological).

The quantum number n' is called the radial quantum number. A radial wave function, defined by the two quantum numbers ℓ and n' and normalized to unity, is unique (up to a phase factor). The eigenvalues of the Hamiltonian are therefore labeled in general by the two quantum numbers n' and ℓ. They do not depend on the quantum number m as a consequence of the rotation invariance of the system. This means that the $2\ell + 1$ states corresponding to given values of n' and ℓ and to different values of m, have the same energy and are degenerate.

These general considerations apply to any two-body system with a central potential: the hydrogen atom and also to a certain extent alkali atoms, diatomic molecules, the deuteron, and quark systems.

The principal quantum number n

In the following section, we solve equation (10.19) in the case of a Coulomb potential $V(r) = -q^2/4\pi\varepsilon_0 r$. In this particular case, the energy levels only depend on the quantity $n' + \ell + 1$. It is therefore customary to label atomic

levels with the three quantum numbers ℓ, m, and the positive integer n, called principal quantum number, defined by the relation:

$$n = n' + \ell + 1 \quad .$$

The energy eigenstates are then classified by increasing values of n, ($n = 1, 2, 3, \ldots$). The classification of atomic states by the three integers (n, ℓ, m) is just a redefinition of a catalogue with respect to the classification in terms of (n', ℓ, m). For a given value of n, there are only n possible values of ℓ: $\ell = 0, 1, \ldots, n-1$. For each value of ℓ, there are $2\ell+1$ possible values of m. The wave function of an energy eigenstate is labeled with the three corresponding quantum numbers ($\psi_{n,\ell,m}(\boldsymbol{r})$) and the corresponding energy is denoted $E_{n,\ell}$.

Spectroscopic notation (states $s, p, d, f, \ldots$)

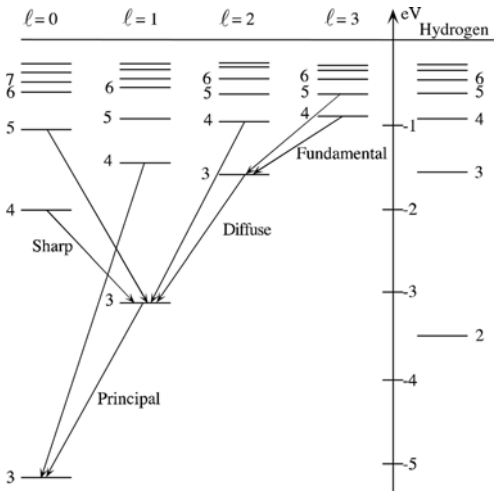


Fig. 10.2. Energy levels of the external electron of sodium (left) and energy levels of hydrogen (right).

The measurement of the energy levels of an atom often comes from the observation of the wavelengths of its spectral lines. We show in Figure 10.2 the energies $E_{n,\ell}$ of the valence electron of sodium and some of the observed transitions. Each horizontal line represents a state; the number on the left is the value of the principal quantum number n. Each column corresponds to a given value of ℓ. The energy of the state is given on the vertical axis (for instance, $E_{3,0} = -5.13$ eV). On the right, we give the energy levels E_n of hydrogen, which, as we show, only depend on n.

The quantum theory of the emission of a photon by an excited atom imposes selection rules. In the transition from a state (n, ℓ) to a state (n_0, ℓ_0)

by emission of a photon of energy $\hbar\omega = E_{n,\ell} - E_{n_0,\ell_0}$, all transitions are not allowed. Only the transitions for which $\ell = \ell_0 \pm 1$ are intense.

Experimental observations in the 19th century showed that one can group the lines in series which were given names according to their aspect. In the case of sodium, after the theory had been understood, it turned out that these series correspond to the following transitions,

$$\begin{aligned} \text{the sharp series} \quad & \hbar\omega = E_{n,\ell=0} - E_{3,1} \\ \text{the principal series} \quad & \hbar\omega = E_{n,\ell=1} - E_{3,0} \\ \text{the diffuse series} \quad & \hbar\omega = E_{n,\ell=2} - E_{3,1} \\ \text{the fundamental series} \quad & \hbar\omega = E_{n,\ell=3} - E_{3,2} \quad . \end{aligned}$$

Each of these four series corresponds to transitions from a state of given ℓ (and various values of n) to a well-defined state. Consequently, the tradition consists of attributing to a given value of ℓ the initial of the corresponding series (spectroscopic notation):

Symbolic letter :	s	p	d	f	g	h
Corresponding value of ℓ :	0	1	2	3	4	5

.

A state of well-defined energy is then denoted by a number (the value of n) followed by a letter (corresponding to the value of ℓ):

$$n = 1\,, \ell = 0 \ : \text{ state } 1s\,; \quad n = 3\,, \ell = 2 \ : \text{ state } 3d \quad .$$

10.3 The hydrogen atom

We now consider the specific case of the hydrogen atom. Here, we consider the problem in its first approximation, where we neglect spin effects. We consider the problem of a particle of mass m_e in the Coulomb field of the proton, which is considered infinitely massive (the reduced mass correction is straightforward):

$$V(r) = -\frac{q^2}{4\pi\varepsilon_0 r} = -\frac{e^2}{r} \ .$$

q is the elementary charge and we set $e^2 = q^2/4\pi\varepsilon_0$. The radial equation is

$$\left(-\frac{\hbar^2}{2m_e}\frac{1}{r}\frac{d^2}{dr^2}r + \frac{\ell(\ell+1)\hbar^2}{2m_e r^2} - \frac{e^2}{r} \right) R_\ell(r) = E\ R_\ell(r) \quad . \tag{10.21}$$

10.3.1 Atomic units; fine structure constant

The above equation involves three constants: $\hbar$ (action), m_e (mass), and e^2 (product of an energy by a length). Using these three constants, it is useful to form a length unit and an energy unit relevant for our problem.

Fine structure constant

We notice that $e^2/\hbar$ has the dimension of a velocity. Unless the differential equation had pathologies (which is not the case), it must represent the typical velocity v of the electron in the lowest energy levels of the hydrogen atom. This velocity must be compared with the velocity of light c, which is the absolute velocity standard in physics. The ratio between these two velocities is a dimensionless constant α, which is a combination of the fundamental constants $q, \hbar$, and c

$$\alpha = \frac{e^2}{\hbar c} = \frac{q^2}{4\pi\varepsilon_0 \hbar c} \sim \frac{1}{137} \ .$$

The smallness of this constant α guarantees that the nonrelativistic approximation is acceptable up to effects of the order of $v^2/c^2 \sim 10^{-4}$. The constant α is called, for (unfortunate) historical reasons, the fine structure constant. A more appropriate terminology would have been: fundamental constant of electromagnetic interactions.

Any charge Q is an integer multiple of the elementary charge $Q = Zq$ (or an integer multiple of $q/3$ if one incorporates quarks). Therefore the fundamental form of Coulomb's law between two charges $Q = Zq$ and $Q' = Z'q$ is $V(r) = \alpha Z Z'(\hbar c/r)$, with Z and Z' integers, which only involves mechanical quantities. The introduction of electric units and of ε_0 is only a convenient manner to describe macroscopic cases where Z and Z' are very large. The experimental determination of the fundamental constant α is a key point in physics: $1/\alpha = 137.0359779\,(32)$.

The fact that this is a dimensionless number stirred minds at the beginning. One cannot change the value of α by changing units.

For a long time, after this discovery (i.e., the discovery of the universality of Planck's constant) people have tried to obtain it starting from transcendental numbers e, π, the Euler constant γ, and so on. For instance $e^{(-\pi^2/2)} \simeq 1/139$, or better $(1/2)\pi^{(-e^2/2)} \simeq 137.3$.

The great astronomer Eddington, at the end of his life, had made an arithmetic theory of α. This constant is encountered in the hydrogen atom, made of a proton and and electron. We live in a four-dimensional space–time, and the proton and electron both have 4 degrees of freedom. Now $4 \times 4 = 16$, and if we consider the symmetric real 16×16 matrices, they have 136 different elements, to which we must add spin, hence the number 137.[2]

[2] The "Eddington numbers" were $E_n = n^2(n^2+1)/2$, that is, the sequence $E_1 = 1$, $E_2 = 10$, $E_3 = 45$, $E_4 = 136$, $E_5 = 325$, $E_6 = 666$, and so on. This provoked a remark of Max Born, who pointed out that Eddington's theory had a much broader scope than physics. Indeed, one could find the numbers of The Apocalypse Of Saint John (Rev. 13: 1–18) "I saw a beast coming out of the sea, having ... *ten* horns, ... and power was given to him to do *forty two* months... He that hath understanding, let him count the number of the beast. For it is the number of a man: and the number of him is *six hundred sixty six*." Obviously, pointed out

Hans Bethe had replied, "But the theory of Mr. Eddington is very interesting because it explains features of ideal gases. Indeed, the simplest ideal gas is molecular hydrogen. If each of the two atoms in the molecule H_2 has 137 degrees of freedom, then $2 \times 137 = 274$, to which we must subtract 1 since the atoms are bound, which gives 273, namely the temperature of the absolute zero."

The truth came with the great revolution of quantum field theory introduced by Ken Wilson in the 1970's. The understanding of what is called the renormalization group enabled us to understand that the value of the fine structure constant, as well as other dimensionless constants in elementary interactions, depends on the energy, or on the distance. In fact, experiments in the LEP collider at CERN have shown that around 100 GeV center of mass energy, the value of the fine structure constant is larger, $\alpha \simeq 1/129$. There is some hope to calculate it in unified field or superstring theories.

Atomic units

The length unit of the problem is the Bohr radius:

$$a_1 = \frac{\hbar^2}{m_e e^2} = \frac{1}{\alpha}\,\frac{\hbar}{m_e c} \sim 0.53 \text{ Å} \quad ,$$

where $\hbar/m_e c$ is the Compton wavelength of the electron. The Bohr radius is the typical size of an atom.

The energy unit relevant for the hydrogen atom is

$$E_I = \frac{m_e e^4}{2\hbar^2} = \frac{1}{2} m_e c^2 \alpha^2 \sim 13.6 \text{ eV} \ ,$$

which, as we show, is the ionization energy of the atom. The electron-volt is a typical energy for external atomic electrons.

The atomic time scale is $2\pi\hbar^3/(m_e e^4) \sim 1.5\ 10^{-16}$ s. It represents the period of the classical circular motion of the electron around the proton for the energy $-E_I$.

10.3.2 The dimensionless radial equation

Having identified the relevant length and energy scales of the problem, we introduce the dimensionless quantities $\rho = r/a_1$ and $\varepsilon = -E/E_I$. We define ε with a minus sign so that this quantity is positive if we are dealing with a bound state, whose energy is negative. We obtain the following dimensionless equation,

$$\left(\frac{1}{\rho}\frac{d^2}{d\rho^2}\rho - \frac{\ell(\ell+1)}{\rho^2} + \frac{2}{\rho} - \varepsilon \right) R_\ell(\rho) = 0 \quad . \tag{10.22}$$

Born, $10 = E_2$ and $666 = E_6$, but there remains a fundamental unsolved problem: is 42 equal to $E_3 - 3E_1$ or to $(E_4 - E_2)/3$?

This equation is well known to mathematicians. Laguerre gave the solutions in 1860. The following properties can be proven:

1. For each value of ℓ, we obtain a infinite set of normalizable solutions labeled by an integer $n' = 0, 1, \ldots$:

$$R(\rho) = e^{-\sqrt{\varepsilon}\rho}\, \rho^{\ell}\, Q_{n',\ell}(\rho) \quad , \tag{10.23}$$

   where $Q_{n',\ell}(\rho) = C_0 + C_1\rho + \cdots + C_{n'}\rho^{n'}$ is called a Laguerre polynomial of degree n'. It has n' real zeros between $\rho = 0$ and $\rho = +\infty$.
2. These normalizable solutions correspond to the eigenvalues

$$\varepsilon = \frac{1}{(n' + \ell + 1)^2} \quad . \tag{10.24}$$

As already mentioned above, the integer n' corresponds to the number of nodes of the radial wave function and is called the radial quantum number. The principal quantum number is the integer $n = n' + \ell + 1$. The first radial wave functions $R_{n,\ell}(\rho)$ are given in Table 10.1. We remark that, owing to (10.24), $\varepsilon_n = 1/n^2$ is an eigenvalue of all radial equations corresponding to values of ℓ smaller than $n : \ell = 0, 1, \ldots, n-1$.

$n = 1$	$\ell = 0$	$2\, e^{-\rho}$
$n = 2$	$\ell = 0$	$\frac{1}{\sqrt{2}} \left(1 - \frac{\rho}{2}\right) e^{-\rho/2}$
	$\ell = 1$	$\frac{1}{2\sqrt{6}}\, \rho\, e^{-\rho/2}$
$n = 3$	$\ell = 0$	$\frac{2}{3^{3/2}} \left(1 - \frac{2}{3}\rho + \frac{2}{27}\rho^2\right) e^{-\rho/3}$
	$\ell = 1$	$\frac{2^{5/2}}{3^{7/2}}\, \rho \left(1 - \frac{\rho}{6}\right) e^{-\rho/3}$
	$\ell = 2$	$\frac{2^{3/2}}{3^{9/2}\sqrt{5}}\, \rho^2\, e^{-\rho/3}$

Table 10.1. Radial wave functions $R_{n,\ell}(\rho)$ for the Coulomb problem, for $n = 1, 2, 3$.

We do not give a rigorous proof here that the normalizable solutions of (10.22) can indeed be cast in the form (10.23), but we can check that these solutions make sense both around the origin and at infinity.

Around $\rho = 0$

The Coulomb term $1/\rho$ and the constant term ϵ are negligible compared with the centrifugal term $\ell(\ell + 1)/\rho^2$ (for $\ell \neq 0$). Assuming a power law depen-

dence $R_{n,\ell}(\rho) \propto \rho^s$ around $\rho = 0$, one finds that the only possible exponent s compatible with a normalizable solution is $s = \ell$ ($s = -\ell - 1$ is not square-integrable for $\ell \geq 1$). This corresponds to the expansion of (10.23) around $\rho = 0$. Notice that in an s-wave ($\ell = 0$), a function behaving as $1/r$ is square-integrable; however, it does not satisfy the Schrödinger equation because $\Delta(1/r) = -4\pi\delta(\boldsymbol{r})$.

At infinity

Keeping only the leading terms in the expansion in $R_{n,\ell}$, we have

$$R_{n,\ell}(\rho) \sim e^{-\sqrt{\varepsilon}\rho} \left(C_{n'}\rho^{n-1} + C_{n'-1}\rho^{n-2} + \cdots\right)$$

If one injects this expansion into the differential equation (10.22), it is immediate to check that the term in $e^{-\sqrt{\varepsilon}\rho}\ \rho^{n-1}$ always cancels out of the equation, and the coefficient of the next term $e^{-\sqrt{\varepsilon}\rho}\ \rho^{n-2}$ is proportional to $C_{n'}(1 - n\sqrt{\varepsilon})$. Therefore this term also cancels out for the particular choice $\varepsilon = 1/n^2$. The subsequent terms of the expansion, which depend on the centrifugal barrier, allow the determination of the coefficients $C_{n'}, C_{n'-1}, \ldots, C_0$.

Coming back to the initial variables for length and energy, we can summarize the above results:

Each solution of the Schrödinger equation (10.13) corresponding to a bound state for the Coulomb problem is labeled by three integers (or quantum numbers):

$$n = 0, 1, 2, \ldots\ , \quad \ell = 0, 1, \ldots, n-1\ , \quad m = -\ell, \ldots, \ell \quad .$$

The energy of a solution depends only on the principal quantum number n:

$$E_n = -\frac{E_I}{n^2}\ , \quad \textit{with} \quad E_I = \frac{m_e e^4}{2\hbar^2} \sim 13.6\ eV \quad .$$

To each energy level, there correspond several possible values of the angular momentum. The total degeneracy (in ℓ and m) of a level with given n is

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

The wave function corresponding to a given set n, ℓ, m is unique (up to a phase factor) and it reads:

$$\psi_{n,\ell,m}(\boldsymbol{r}) = Y_{\ell,m}(\theta, \varphi)\ e^{-r/(n\,a_1)} \left(\frac{r}{a_1}\right)^{\ell} \times \left(C_0 + C_1 \frac{r}{a_1} + \cdots + C_{n-\ell-1} \left(\frac{r}{a_1}\right)^{n-\ell-1}\right) \ . \quad (10.25)$$

where the C_k's ($k = 0, \cdots, n - \ell - 1$) are the coefficients of the Laguerre polynomials and where $a_1 = \hbar^2/(m_e e^2) \sim 0.53$ Å.

10.3.3 Spectrum of hydrogen

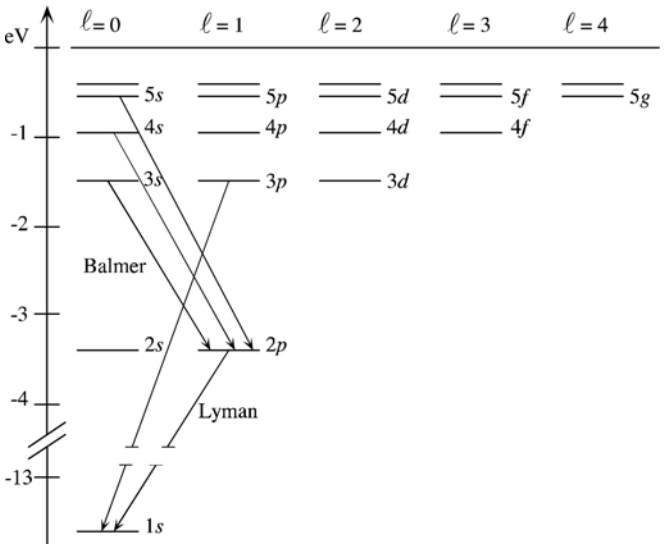


Fig. 10.3. Energy levels of hydrogen.

In Figure 10.3, we represent the energies E_n of the hydrogen atom. Each line represents an energy level, the number on the right is the value of n, the column corresponds to a given value of ℓ, and we give the value of the energy on the vertical axis.

The selection rule given for the observable spectral lines of the sodium atom $\ell = \ell_0 \pm 1$ still holds. The most famous series is the Balmer series. It corresponds to transitions from states ns to the state $2p$:

$$\hbar\omega = E_n - E_2 = 13.6\,\frac{n^2-4}{4n^2}\ \text{eV}\ .$$

The first lines of the Balmer series are in the visible part of the spectrum ($\hbar\omega \sim 2$ to 3 eV; $\lambda \sim 0.5\ \mu$m). The Lyman series, corresponding to transitions to the ground state, lies in the ultraviolet ($\lambda \leq 121.5$ nm).

10.3.4 Stationary states of the hydrogen atom

The ground state ($1s$)

The ground state corresponds to $n = 1$, therefore $\ell = 0$ and $m = 0$ ($1s$ state in the spectroscopic language). Because the spherical harmonic $Y_{0,0}(\theta, \varphi)$ is a constant equal to $1/\sqrt{4\pi}$, the normalized wave function of this state is

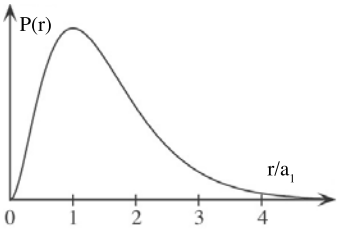


Fig. 10.4. Radial probability density $P(r)$, giving the probability of finding the electron between r and $r + dr$ in a hydrogen atom prepared in its ground state.

$$\psi_{1,0,0}(\boldsymbol{r}) = \frac{e^{-r/a_1}}{\sqrt{\pi a_1^3}} \quad .$$

The probability of finding the electron in a spherical shell of thickness dr, represented in Figure 10.4, is

$$P(r)dr = |\psi_{1,0,0}(\boldsymbol{r})|^2 \, 4\pi r^2 \, dr \ .$$

The probability density per unit volume is proportional to the exponential function e^{-2r/a_1}, and it is maximum for $r = 0$. The most probable distance between the electron and the proton is the Bohr radius $a_1 = 0.53$ Å.

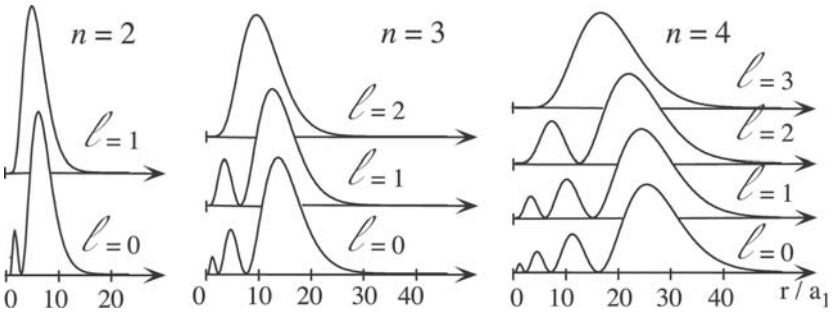


Fig. 10.5. Radial probability density $r^2 \, |R_{n,\ell}(r)|^2$ of the states $n = 2, 3, 4$ of hydrogen.

Other states

Figure 10.5 represents the radial probability density $P_{n,\ell}(r) = r^2 \, |R_{n,\ell}(r)|^2$ for various states n, ℓ. We note the reduction of the number of nodes of the radial wave function as ℓ increases for a given n. For a level n, l, the function $P_{n,\ell}(r)$ has $n' = n - \ell - 1$ zeroes, where n' is the degree of the corresponding Laguerre polynomial. In particular, for $\ell = n - 1$, we remark that $P(r)$ has a single maximum, located at a distance $r = n^2 a_1$ (Equation (10.25)).

Figure 10.6 represents some spatial probability densities $|\psi_{n,\ell,m}(\boldsymbol{r})|^2$ in a plane $y = 0$ (these are axial symmetric functions around the z-axis). For large quantum numbers $n \gg 1$, we notice that one gets closer to "classical" situations, corresponding to a localized particle.

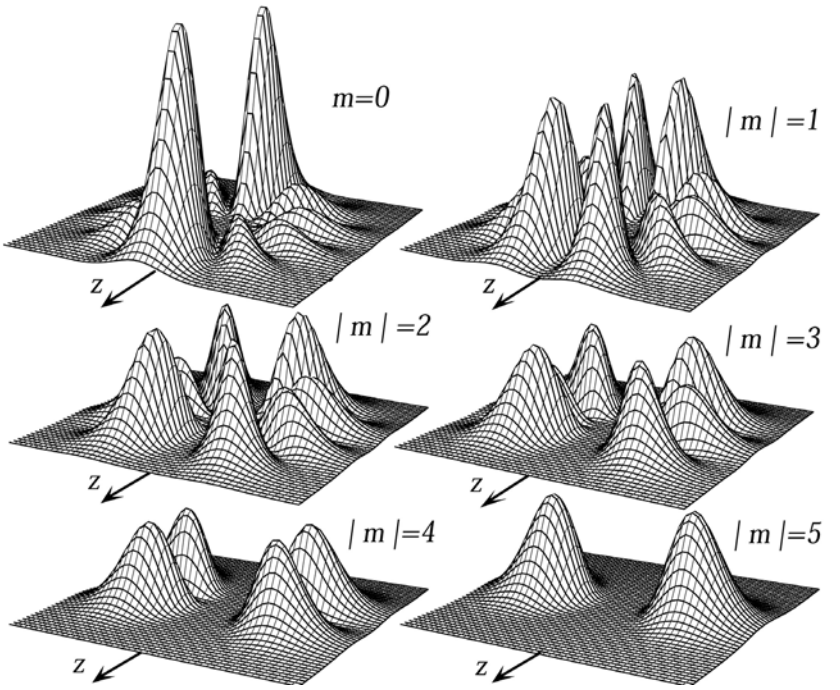


Fig. 10.6. Probability density $|\psi_{n,\ell,m}(\boldsymbol{r})|^2$ in the $y = 0$ plane for $n = 6$, $\ell = 5$ (mesh size: $60\,a_1 \times 60\,a_1$). For $m = 0$, the particle is localized in the vicinity of the z axis. For large $|m|$ (in particular $m = \pm 5$), the particle is localized in the plane $z = 0$, in the vicinity of a circle centered at the origin, of radius $r = 30\,a_1$ (circular state). The vertical scale of the surface $m = 0$ has been reduced by a factor of 2 with respect to the five other surfaces in order to improve the visibility.

10.3.5 Dimensions and orders of magnitude

Consider a hydrogen atom prepared in a stationary state $|n, \ell, m\rangle$. Using the virial theorem, one can show that the classical relation between the kinetic energy and the potential energy still holds for the expectation values of these quantities:

$$E_n^{(\mathrm{kin})} = \langle \frac{p^2}{2m_e} \rangle = -E_n = \frac{E_I}{n^2} \ , \tag{10.26}$$

$$E_n^{(\mathrm{pot})} = \langle \frac{-e^2}{r} \rangle = 2\,E_n = -\frac{2E_I}{n^2}\ . \tag{10.27}$$

Using the properties of Laguerre polynomials, one finds that the mean radius has the following variation with n and ℓ,

$$\langle r \rangle = \frac{a_1}{2}\left(3n^2 - \ell(\ell+1)\right)\ , \tag{10.28}$$

as well as

$$\langle \frac{1}{r} \rangle = \frac{1}{n^2 a_1},\ \langle \frac{1}{r^2} \rangle = \frac{2}{n^3\,(2\ell+1)\,a_1^2},\ \langle r^2 \rangle = \frac{n^2\,a_1^2}{2}\ \left(5n^2 + 1 - 3\ell(\ell+1)\right)\ .$$

Also, setting $\rho = r/a_1$, one gets for $p > -2\ell - 1$:

$$\frac{p+1}{n^2}\langle \rho^p \rangle\ -\ (2p+1)\langle \rho^{p-1} \rangle\ +\ \frac{p}{4}\left((2\ell+1)^2 - p^2\right)\langle \rho^{p-2} \rangle = 0\ .$$

Remark. Because of the n^2 variation of the mean atomic radius, the maximal probability density for a given radial wave function decreases like $1/n^4$. This is why a readjustment of scales is necessary in order to visualize Figure 10.5 properly.

10.3.6 Historical landmarks

The importance of this calculation done by Schrödinger in 1926, is not so much in the result, which was known, but in the fact that it was obtained by a systematic method, which could be generalized to other cases.

Spectroscopy

The law involving integers was a crowning achievement for spectroscopy.

After Newton's discovery of the decomposition of white light in 1680, Wollaston, in 1802, and mainly Fraunhofer, after 1814, had discovered dark and shining lines in the solar spectrum and in stars.

People had realized that spectral lines observed in the light emitted by a body were characteristic of the elements in this body. The observation of regularities allowed the attribution of a series of lines to a given element (there was a catalogue of 5000 lines in 1890). Very soon, people had had the idea that spectral lines were analogous to harmonics in acoustics and that there should exist some simple relations between them.

But that didn't lead very far. Fundamental physicists considered spectroscopy to be too complicated and outside the scope of physics, such as the harmonics of a piano, which depend on the shape of the instrument as a whole. Why? Because the spectra were too complicated.

The spectrum of hydrogen

One had to start with the spectrum of the simplest atom, hydrogen. But people didn't know that, because the spectrum of atomic hydrogen was found late. It was difficult to obtain the spectrum of atomic hydrogen. One had to operate with a discharge tube, and disentangle the spectra of atomic and molecular hydrogen, and of other species such as nitrogen and water.

The first line of hydrogen was discovered in 1853 by the Swedish physicist Anders Jonas Ångstrom. Then in 20 years, he found three others that were called, including the first, α, β, γ, δ.

In 1881 Huggins discovered in the star α of the Lyre, ten other lines "whose sequence seemed to be following the visible lines of hydrogen." Hydrogen is abundant in stars, but one had to guess it at that time!

Balmer

Integers played an important role in science in the 19th century. Examples can be found in chemical reactions, atomic theory, classification and evolution of species in zoology and in botany, and so on. It is by chance that, in 1885, Balmer, who was a high school teacher in Basel, and who was fascinated by numerology, learned the positions of the first four lines of hydrogen. He realized that the wavelengths of the lines could be represented with an accuracy of 10^{-3} by a formula involving integers

$$\lambda = A\frac{n^2}{n^2-4}\,, \quad \text{with} \quad A = 0.3646\ \mu\text{m} \quad .$$

And this law was good to one part in a thousand when applied to the ten lines of Huggins! Balmer predicted the next ones and the limit.

Although he was not a physicist, Balmer found the simplicity of the formula quite striking. He called that constant A the "fundamental number" of hydrogen. In his 1885 paper, he wrote, "It appears to me that hydrogen ... more than any other substance is destined to open new paths to the knowledge of the structure of matter and its properties."

In 1912, Niels Bohr, who was 27, was working with Rutherford on an atomic model. He was not aware of Balmer's formula and of analogous results obtained by Rydberg for alkali atoms. One day, by chance, he learned the existence of Balmer's formula; it only took him a few weeks to construct his celebrated model of the hydrogen atom, which is one of the turning points of quantum physics.

10.4 Muonic atoms

To end this chapter, we describe an application of what we have done in nuclear and elementary particle physics.

The μ lepton, or muon, discovered in 1937, has physical properties of a heavy electron. As is the electron, it is elementary or pointlike; it has the same electric charge, the same spin, but it is 200 times more massive: $m_\mu = 206.8\ m_e$. It is unstable and it decays into an electron and two neutrinos: $\mu \to e + \overline{\nu}_e + \nu_\mu$ with a lifetime of $2\ 10^{-6}$ s. Therefore, physically it is a true heavy electron except that it is unstable.

In particle accelerators, one can produce muons, slow them down, and have them captured by atoms where they form hydrogenlike systems. In a complex atom, the muon is not constrained with respect to electrons by the Pauli principle. The muon expels electrons, cascades from one level to another, and eventually falls in the vicinity of the nucleus at a distance $a_\mu = \hbar^2/Zm_\mu e^2$, which is 200 times smaller than the distance of the internal electrons. Therefore, it forms a hydrogenlike atom, because it does not feel the screened electron electric field. The lifetime of the muon is much longer than the total time of the cascades ($\sim 10^{-14}$ s). It is also much longer than the typical atomic time $\hbar^3/m_\mu e^4 \sim 10^{-19}$ s. The muon can therefore be considered as stable on these time scales.

It then forms a *muonic atom*. The Bohr radius of a muonic atom is of the same order as a nuclear radius. The effect of external electrons is screened. Consider lead ($Z = 82$), whose nuclear radius is $R \approx 8.5$ fm. One finds $a_\mu \approx 3.1$ fm, which means that the μ penetrates the nucleus noticeably. In fact, in the ground state, it has a 90% probability of being inside the nucleus. The description of the nucleus as a point particle is inadequate. Consequently, spectra of muonic atoms provide information on the structure of nuclei, in particular concerning their charge distributions.

For a spherical nucleus, the potential is harmonic inside the nucleus (assuming a constant charge density), and Coulomblike outside the nucleus. If the nucleus is deformed, flattened, or cigar-shaped, spherical symmetry is broken, and the levels will no longer be degenerate in the magnetic quantum number m. This results in a splitting of the spectral lines, which enables us to determine the charge distribution, that is the proton distribution inside the nucleus.

Figure 10.7, obtained at CERN, shows the spectra of muonic atoms in the cases of gold ($Z = 79$), which is a spherical nucleus, and of uranium ($Z = 92$), which is a deformed nucleus. We notice the more complicated structure of the higher energy line for uranium. This is a very accurate method to determine the deformations of nuclei.

The existence of the muon has been a mystery for more than 40 years. When it was discovered, Rabi said, "Who ordered that?" Why a heavy electron? All matter we know about around us can be built starting with protons, neutrons, electrons, and neutrinos, or, in terms of fundamental constituents, with the family of quarks and leptons $\{u, d, e, \nu\}$, using the Schrödinger equation. (Of course, to construct a zebra or a humming bird, there are simpler methods, and we are interested in simple features such as 2 ears, 4 feet, 1 tail for a rabbit, 2 humps for a camel. But with a powerful enough computer, it's

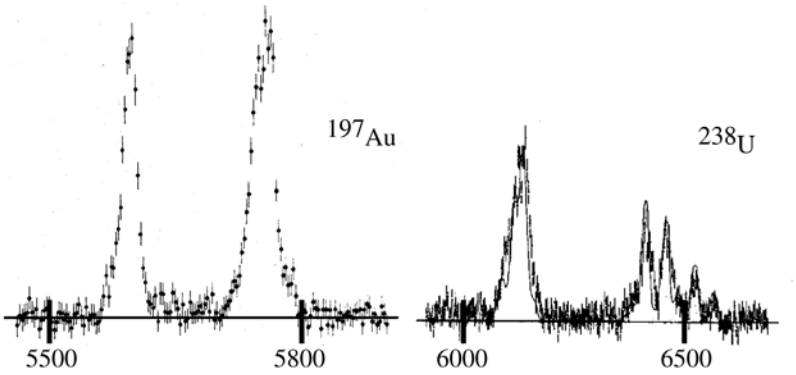


Fig. 10.7. Transition line from the $2p$ level (actually split into two sublevels $2p_{1/2}$ and $2p_{3/2}$, to the $2s$ level in muonic atoms of gold ($Z = 79, A = 197$) and uranium ($Z = 92, A = 238$) (scale in keV). Gold is spherical, and its spectrum has a simple shape; uranium is deformed, and the upper peak is split in four lines. (CERN document.)

in principle feasible). So, why a heavy electron, with which one can imagine a Gulliver universe: atoms, molecules, a chemistry, a biology, 200 times smaller, but 200 times more energetic than the matter we know?

There are many applications of muons, such as probing nuclei, probing crystals, and probing pyramids, but why are they there? What role is the muon supposed to play?

In 1974, with the discovery of a new quark, the charm quark c, it was realized that the muon forms, with its neutrino, the charm quark, and the strange quark s, discovered in the 1940's, a new family of quarks and leptons (c, s, μ, ν_μ). This family generates at higher scales, a new atomic and nuclear physics, but its members are unstable. In 1975–1976, with the discovery of a new lepton τ, another new quark was discovered, the b quark (beautiful or bottom), and, in 1995, the top quark t, hence a third family (t, b, τ, ν_τ).[3]

In 1989, at the LEP electron collider, it was proven that the (light) fundamental constituents of matter consist of these three families only. Present ideas are that a more fundamental theory could not exist, and that the big bang could not have happened so nicely, if these two extra (useless) families of quarks and leptons did not exist. They are of importance in order to create the world. But, after that, they are merely toys for physicists.

However, we still do not understand the masses of these quarks and leptons (therefore their stability). The origin of mass is one of the great problems of modern physics.

[3] M. Perl, "The leptons after 100 years," *Phys. Today*, October (1997), p. 34.

11

Spin 1/2

Spin 1/2 is the first truly revolutionary discovery of quantum mechanics. The properties of this physical quantity in itself, the importance of its existence, and the universality of its physical effects were totally unexpected.

The physical phenomenon is the following. In order to describe completely the physics of an electron, one cannot use only its degrees of freedom corresponding to translations in space. One must take into account the existence of an internal degree of freedom that corresponds to an intrinsic angular momentum. In other words, the electron, which is a pointlike particle, "spins" on itself. We use quotation marks for the word "spins". One must be cautious with words, because this intrinsic angular momentum is purely a quantum phenomenon. It has no classical analogue, except that it is an angular momentum.

One can use analogies and imagine that the electron is a sort of quantum top. But we must keep in mind the word "quantum". The electron is a pointlike object down to distances of 10^{-18} m. One must admit that a pointlike object can possess an intrinsic angular momentum. (As we have already pointed out, in this respect, the photon, which is pointlike, has an intrinsic angular momentum, and is a *zero mass* particle, is from this point of view even more strange.)

11.1 Experimental results

Experimentally, this intrinsic angular momentum, called spin, has the following manifestations (we do not enter in any technical detail):

1. If we measure the projection of the spin along any axis, whatever the state of the electron, we find either of two possibilities:

$$+\hbar/2 \quad \text{or} \quad -\hbar/2 \quad .$$

There are two and only two possible results for this measurement.

2. Consequently, if one measures the square of any component of the spin, the result is $\hbar^2/4$ with probability one.
3. Therefore, a measurement of the square of the spin $S^2 = S_x^2 + S_y^2 + S_z^2$ gives the result
$$S^2 = \frac{3\hbar^2}{4} \quad .$$
4. A system that has a classical analogue, such as a rotating molecule, can rotate more or less rapidly on itself. Its intrinsic angular momentum can take various values. However, for the electron, as well as for many other particles, it is an amazing fact that the square of its spin S^2 is always the same. It is fixed: all electrons in the universe have the same values of the square of their spins $S^2 = 3\hbar^2/4$. The electron "spins" on itself, but it is not possible to make it spin faster.

One can imagine that people did not come to that conclusion immediately. The discovery of the spin 1/2 of the electron is perhaps the most breathtaking story of quantum mechanics.

The elaboration of the concept of spin was certainly the most difficult step of all quantum theory during the first quarter of the 20th century. It is a real suspense that could be called the various appearances of the number 2 in physics. There are many numbers in physics; it is difficult to find a simpler one than that.

And that number 2 appeared in a variety of phenomena and enigmas that seemed to have nothing to do a priori with one another, or to have a common explanation. The explanation was simple, but it was revolutionary. For the first time people were facing a purely quantum effect, with no classical analogue. Nearly all the physical world depends on this quantity, the spin 1/2.

The challenge existed for a quarter of a century (since 1897). Perhaps, there was never such a long collective effort to understand a physical structure. It is almost impossible to say who discovered spin 1/2, even though one personality dominates, Pauli, who put all his energy into finding the solution.

We show that in order to manipulate spin 1/2, and understand technicalities we essentially know everything already. We have done it more or less on two-state systems.

But for anybody to have a really intuitive representation is a complicated matter. It is really a personal matter, as can be seen in Figure 11.1. These gentlemen are simply discussing spin effects.

11.2 Spin 1/2 formalism

The measurement results fit perfectly with the general framework of the theory of angular momenta. The electron spin is a half-integer angular momentum that corresponds to the quantum numbers $j = 1/2$, $m = 1/2$. The corresponding Hilbert space is two-dimensional.

Fig. 11.1. Three physicists discussing the optimal way to measure spin effects in proton collisions at the Argonne ZGS accelerator (CERN document).

Any spin state is a linear superposition of two basis states and the degree of freedom corresponding to spin is described in a two-dimensional Hilbert space: $\mathcal{E}_{\text{spin}}$.

11.2.1 Representation in a particular basis

We choose a basis of states in which both $\hat{S}^2$ and $\hat{S}_z$ are diagonal, which we denote $\{|+\rangle, |-\rangle\}$:

$$\hat{S}_z|+\rangle = \frac{\hbar}{2}|+\rangle\,, \quad \hat{S}_z|-\rangle = -\frac{\hbar}{2}|-\rangle\,, \quad \hat{S}^2|\pm\rangle = \frac{3\hbar^2}{4}|\pm\rangle \quad . \tag{11.1}$$

Using the notation of chapter 9, the states $|\pm\rangle$ would be $|j = 1/2, m = \pm 1/2\rangle$. The action of $\hat{S}_x$ and $\hat{S}_y$ on the elements of the basis is written as (see Equation (9.16))

$$\hat{S}_x|+\rangle = \hbar/2|-\rangle\,, \qquad \hat{S}_x|-\rangle = \hbar/2|+\rangle\,, \tag{11.2}$$

$$\hat{S}_y|+\rangle = i\hbar/2|-\rangle\,, \qquad \hat{S}_y|-\rangle = -i\hbar/2|+\rangle \quad . \tag{11.3}$$

An arbitrary spin state $|\Sigma\rangle$ can be written as

$$|\Sigma\rangle = \alpha\,|+\rangle \,+\, \beta\,|-\rangle\,, \quad |\alpha|^2 + |\beta|^2 = 1 \quad . \tag{11.4}$$

The probabilities of finding $+\hbar/2$ and $-\hbar/2$ in a measurement of S_z on this state are $P(+\hbar/2) = |\alpha|^2,\ P(-\hbar/2) = |\beta|^2$.

11.2.2 Matrix representation

It is convenient to use matrix representations for the states and the operators:

$$|+\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |-\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad |\Sigma\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \quad . \tag{11.5}$$

We can use the Pauli matrices $\hat{\boldsymbol{\sigma}} \equiv \{\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z\}$

$$\hat{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{\sigma}_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \hat{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{11.6}$$

which satisfy the commutation relations

$$\hat{\boldsymbol{\sigma}} \times \hat{\boldsymbol{\sigma}} = 2i\, \hat{\boldsymbol{\sigma}} \quad . \tag{11.7}$$

The spin observables are represented as

$$\hat{\boldsymbol{S}} = \frac{\hbar}{2}\, \hat{\boldsymbol{\sigma}} \quad . \tag{11.8}$$

In this basis, the eigenstates $|\pm\rangle_x$ of $\hat{S}_x$ and $|\pm\rangle_y$ of $\hat{S}_y$ are

$$|\pm\rangle_x = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm 1 \end{pmatrix}, \quad |\pm\rangle_y = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm i \end{pmatrix} \quad . \tag{11.9}$$

The major complication, compared to the ammonia molecule of chapter 7 is to deal with a vector degree of freedom, and a vector observable. But by studying angular momenta, we understood how to deal with that situation.

We remark that, as we already know in different systems, the state $|+\rangle$ for which we are sure that $S_z = \hbar/2$ can also be considered as a linear superposition with equal weights of $S_x = \pm\hbar/2$ or $S_y = \pm\hbar/2$.

If the spin "points" upwards, it points both on the left and on the right with equal probabilities. This is again the superposition principle.

Notice that we just said that the spin points upwards, and we have spoken classically of a purely quantum property. That is not too severe, because we have a mathematical dictionary to translate it properly. But remember that Galileo just escaped being burned to death, and Giordano Bruno was burned to death, because they said that it is the earth that turns around itself and that the sun is fixed. Nevertheless, this does not prevent people nowadays from saying that the sun rises and sets everyday. Once things have been understood, it's much simpler, in everyday life, to come back to traditional ways of visualizing things

11.3 Complete description of a spin 1/2 particle

We now turn to the complete description of the state of an electron taking into account both the space degrees of freedom and the spin.

Here, we follow a more direct path than the traditional way of showing a nontrivial tensor product of Hilbert spaces.

The result is quite intuitive. We deal with two random variables, position and spin along the z-axis. We need the probability law of this couple of variables.

The state of the particle in space is described by a square-integrable function that belongs to what we call the "external" Hilbert space $\mathcal{E}_{\text{external}}$. At any point in space, the spin state is described by a two-dimensional vector of $\mathcal{E}_{\text{spin}}$, the "internal" degree of freedom.

We therefore need two wave functions to describe the state of the electron in space with its spin. We must double the number of wave functions (or the "dimension" of the Hilbert space). A tensor product of space is as simple as that in this particular case.

There are several possible representations of the states, together with corresponding representations of observables. Choosing a particular representation is a matter of convenience.

Mixed representation

The state is represented by a vector of $\mathcal{E}_{\text{spin}}$ whose components are square-integrable functions

$$\psi_+(\boldsymbol{r},t)\,|+\rangle + \psi_-(\boldsymbol{r},t)\,|-\rangle \quad . \tag{11.10}$$

The physical interpretation of this representation is as follows. $|\psi_+(\boldsymbol{r},t)|^2\,d^3r$ (resp., $|\psi_-(\boldsymbol{r},t)|^2\,d^3r$) is the probability of finding the particle in a volume d^3r around the point $\boldsymbol{r}$, with a spin component $+\hbar/2$ (resp., $-\hbar/2$) along the z-axis.

Two-component wave function

The state vector is represented in the form:

$$\begin{pmatrix} \psi_+(\boldsymbol{r},t) \\ \psi_-(\boldsymbol{r},t) \end{pmatrix} \quad . \tag{11.11}$$

The physical interpretation of ψ_+ and ψ_- as probability amplitudes for the couple of random variables $(\boldsymbol{r}, S_z)$ is the same as above.

11.3.1 Observables

The space observables $(x, id/dx)$ only act on wave functions $\psi_\pm(\boldsymbol{r},t)$. They do not make any difference between spin states. In the matrix representation the space observables are diagonal 2×2 matrices whose elements are operators and that do not act on spin variables.

Spin observables are 2×2 matrices with numerical coefficients seen above. They do not act on space variables.

In the general case, there exist observables that can act on both sets of variables. These are simply the products of the 2×2 matrices by the operators; for instance,

$$\hat{\boldsymbol{r}} \cdot \hat{\boldsymbol{S}} = x\hat{S}_x + y\hat{S}_y + z\hat{S}_z \quad .$$

More generally, the products of operators $\hat{A}_{\rm ext}\hat{B}_{\rm sp}$ follow the rule

$$\left(\hat{A}_{\rm ext}\hat{B}_{\rm sp}\right)\,(\psi_\sigma(\boldsymbol{r},t)|\sigma\rangle) = \left(\hat{A}_{\rm ext}\psi_\sigma(\boldsymbol{r},t)\right)\left(\hat{B}_{\rm sp}|\sigma\rangle\right)\,, \quad \sigma = \pm \quad . \tag{11.12}$$

Functions of a two-valued variable

In some problems, it is useful to use a single wave function depending on four variables (plus time) $\psi(\boldsymbol{r},\sigma;t)$, where the fourth variable σ can take the values ±1. One has obviously

$$\psi(\boldsymbol{r},\sigma;t) = \psi_\sigma(\boldsymbol{r},t)\,, \quad \sigma = \pm1.$$

Atomic states

In many problems of atomic physics, it is useful to use the quantum numbers n,ℓ,m to classify the states $|n,\ell,m\rangle$ that form a basis of $\mathcal{E}_{\rm external}$. The introduction of spin is done in the space spanned by the family $\{|n,\ell,m\rangle\otimes|\sigma\rangle\}$ where the spin quantum number can take the two values ±1. It is convenient to use the compact notation

$$|n,\ell,m,\sigma\rangle\ , \tag{11.13}$$

where the states of an electron are described by four quantum numbers. The action of space operators on the states $|n,\ell,m\rangle$ is known (see chapter 10), therefore the action of general operators on the states $|n,\ell,m,\sigma\rangle$ can readily be inferred from the considerations developed above.

11.4 Physical spin effects

Physical spin effects belong to the following main categories:

1. There are angular momentum effects, in particular in nuclear and particle physics. It is compulsory to take into account spin in order to observe conservation of the total angular momentum. The structure of fundamental interactions relies on the spin 1/2 of the electron, quarks, and neutrinos.

This is too difficult to examine here.

2. There are surprising and revolutionary effects compared to classical physics that are due to the Pauli principle. Such effects are fundamental in order to understand the structure of matter, atoms, molecules, solids, and

liquids. We show some of them in the next chapter. The Pauli principle plays a central role in statistical physics and in chemistry.

3. There are magnetic effects that actually are at the origin of the discovery of spin 1/2. We mainly use those effects in order to develop our experimental discussion.

11.5 Spin magnetic moment

Indeed, all what has been said above is for the moment simple matrix calculus. But it will acquire a much greater physical flavor because to the spin of the electron (or any spin 1/2 particle) there corresponds an intrinsic magnetic moment, a spin magnetic moment, that is proportional to the spin.

$$\hat{\boldsymbol{\mu}} = \gamma\, \hat{\boldsymbol{S}} = \mu_0\, \hat{\boldsymbol{\sigma}} \ . \tag{11.14}$$

The gyromagnetic ratio γ, the value of which we come back to, is a characteristic of the particle under consideration.

The action of $\hat{\boldsymbol{\mu}}$, which is proportional to $\hat{\boldsymbol{S}}$, on the state vectors is obvious.

It is through measurements of this magnetic moment that we can find the easiest access to spin measurements.

For instance, if we place this spin magnetic moment in a field $\boldsymbol{B}$, there is an interaction Hamiltonian

$$\hat{W} = -\hat{\boldsymbol{\mu}} \cdot \boldsymbol{B} \quad , \tag{11.15}$$

which leads to a number of experimental observations.

11.5.1 Hamiltonian of a one-electron atom

In what follows we consider an atom with one external electron in its ground state. Globally, the atom is electrically neutral. However, it carries the spin 1/2 of the external electron, and the corresponding magnetic moment.

Suppose the atom is moving in space and that it is placed in a potential $V(\boldsymbol{r})$, and that furthermore it is placed in a magnetic field $\boldsymbol{B}$. The magnetic potential energy is given by (11.15).

The Hamiltonian is the sum of two terms

$$\hat{H} = \hat{H}_{\text{ext}} + \hat{W} \ , \tag{11.16}$$

where

$$\hat{H}_{\text{ext}} = \frac{\hat{p}^2}{2m} + V(\hat{\boldsymbol{r}})$$

is of the type seen in previous chapters. In particular, $\hat{H}_{\text{ext}}$ does not act on the spin variables. The operator $\hat{W}$ is given by (11.15). It acts in $\mathcal{E}_{\text{spin}}$ through

the three operators $\hat{\mu}_x, \hat{\mu}_y, \hat{\mu}_z$. If the field is inhomogeneous, it also acts on space variables through the three functions $B_x(\hat{\boldsymbol{r}})$, $B_y(\hat{\boldsymbol{r}})$, $B_z(\hat{\boldsymbol{r}})$.

The Schrödinger equation is

$$i\hbar \frac{d}{dt}|\psi\rangle = \hat{H}|\psi\rangle \ . \tag{11.17}$$

If we choose the representation of states (11.11) and if we decompose on the orthonormal basis $\{|+\rangle, |-\rangle\}$, we obtain the coupled differential system

$$\begin{aligned} i\hbar \frac{\partial}{\partial t}\psi_+(\boldsymbol{r},t) &= \left(-\frac{\hbar^2}{2m}\Delta + V(\boldsymbol{r}\,)\right)\psi_+(\boldsymbol{r},t) \\ &\quad + \langle +|\hat{W}|+\rangle\ \psi_+(\boldsymbol{r},t) + \langle +|\hat{W}|-\rangle\ \psi_-(\boldsymbol{r},t)\ , \\ i\hbar \frac{\partial}{\partial t}\psi_-(\boldsymbol{r},t) &= \left(-\frac{\hbar^2}{2m}\Delta + V(\boldsymbol{r}\,)\right)\psi_-(\boldsymbol{r},t) \\ &\quad + \langle -|\hat{W}|+\rangle\ \psi_+(\boldsymbol{r},t) + \langle -|\hat{W}|-\rangle\ \psi_-(\boldsymbol{r},t)\ . \end{aligned}$$

The matrix elements of $\hat{W}$ in the basis $\{|+\rangle, |-\rangle\}$ are functions of the external space variables. They add to the usual potential energy terms, which are diagonal, and, in general, they couple the evolution equations of the components ψ_+ and ψ_-.

We now possess all tools to be able to do physics and to discover the suspense of the discovery of spin 1/2.

11.6 The Stern–Gerlach experiment

The first measurement of the intrinsic magnetic moment of the electron, and the first undeniable appearance of the number 2 came in 1921–1922 with the experiment of Stern and Gerlach.

11.6.1 Principle of the experiment

A collimated beam of silver atoms is sent in a region where an inhomogeneous magnetic field is applied along the z-direction, perpendicular to the initial velocity of the atoms (Figure 11.2a). The possible deflection of the beam by the field gradient is then measured by observing the impacts of the atoms on a detection plate perpendicular to the initial direction of the beam. The silver atom has one external electron in an orbital angular momentum state $\ell = 0$. Therefore, the atom's magnetic moment is equal to the intrinsic magnetic moment of the valence electron.

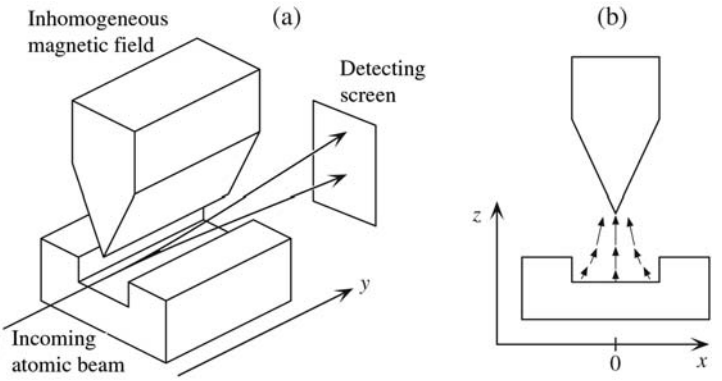


Fig. 11.2. (a) The Stern–Gerlach experiment: atoms from a collimated beam are deflected as they cross a region where an inhomogeneous magnetic field is applied. This experiment can be interpreted as a measurement of the component of the atomic magnetic moment along the direction of the field (z in the Figure). (b) Magnetic gradient between the polar pieces of the magnet.

11.6.2 Semi-classical analysis

We first analyze this experiment within classical mechanics. The atoms are neutral and they are not subject to a magnetic Lorentz force. However, because they have a nonvanishing magnetic moment $\boldsymbol{\mu}$, the force

$$F_z = \mu_z \frac{\partial B_z}{\partial z} \ , \tag{11.18}$$

parallel to the z-direction, acts on them and deflects their trajectory.

We recall that when a magnetic moment $\boldsymbol{\mu}$ is placed in a magnetic field $\boldsymbol{B}$ the magnetic interaction energy is

$$W = -\boldsymbol{\mu} \cdot \boldsymbol{B} \ , \tag{11.19}$$

and the torque

$$\boldsymbol{\Gamma} = \boldsymbol{\mu} \times \boldsymbol{B} \tag{11.20}$$

is exerted on the magnetic moment. In addition, if the magnetic field is inhomogeneous, the force (11.18), or more generally

$$\boldsymbol{F} = \boldsymbol{\nabla}(\boldsymbol{\mu} \cdot \boldsymbol{B}) = \sum_{i=x,y,z} \mu_i(t)\, \boldsymbol{\nabla} B_i \ , \tag{11.21}$$

acts on the dipole.

As we have seen in (9.36), the magnetic moment of an atom at $\boldsymbol{r}$ does not align with the axis of the local magnetic field $\boldsymbol{B}(\boldsymbol{r})$, but it precesses along this axis with the angular frequency

$$\omega_0 = -\gamma_0 B(\boldsymbol{r}) \ , \tag{11.22}$$

called the Larmor frequency.

We assume that the classical trajectory of the atoms lies in the plane of symmetry $x = 0$ of the magnet (see Figure 11.2b). Along this trajectory the magnetic field is always parallel to the z-axis , so that the Larmor precession takes place around z. Also, owing to the symmetry of the device, the quantities $\partial B_z/\partial x$ and $\partial B_z/\partial y$ vanish along the atomic beam trajectory (we neglect possible edge effects). If the displacement of the magnetic moment during a single precession period $2\pi/\omega_0$ is small compared with the typical variation scale of the magnetic field, we can average the force (11.21) over the Larmor period. The contributions of μ_x and μ_y to (11.21) then vanish, and one is left only with the z-component of the force $F_z = \mu_z(t)\ \partial B_z/\partial z$. In addition we deduce from (9.36) that μ_z stays constant as the atom moves in the magnetic field gradient, which is helpful to justify the result (11.18).

11.6.3 Experimental results

In the absence of a magnetic field gradient one observes a single spot on the detecting plate, in the vicinity of $x = z = 0$ (Figure 11.3a). The magnetic field gradient provides a way to measure the z-component of the magnetic moment of the atoms as they enter the field zone. Let us assume that all the atoms carry the same magnetic moment of norm μ_0, and that this moment is oriented at random when an atom enters the field zone. Classically, this should produce some continuous distribution of μ_z between the two extreme values $-\mu_0$ and $+\mu_0$, so that one would expect that the impacts of the atoms on the screen form an extended line parallel to z (Figure 11.3b). The endpoints of the line correspond to atoms whose magnetic moments are oriented respectively such as $\mu_z = +\mu_0$ and $\mu_z = -\mu_0$.

The experiment is difficult. But the observation differs radically from this classical prediction. The set of impacts never forms a continuous line on the screen. For atoms such as silver, the impacts are grouped in two spots corresponding to $\mu_z = +\mu_0$ and $\mu_z = -\mu_0$, with $\mu_0 = 9.27\ 10^{-24}\ \mathrm{J\,T^{-1}}$ (Figure 11.3c).

The result μ_0 of the Stern–Gerlach experiment is consistent with

$$\mu_0 = \hbar\,|\gamma_0| = \frac{\hbar q}{2m_e} \ , \tag{11.23}$$

which amounts to taking $L = \hbar$ in (9.33). The quantity (11.23) is the absolute value of the Bohr magneton.

11.6.4 Explanation of the Stern–Gerlach experiment

Our theory explains the observed spatial separation of the states $|\pm\rangle_z$. We consider an incident atomic beam propagating along y; each atom possesses

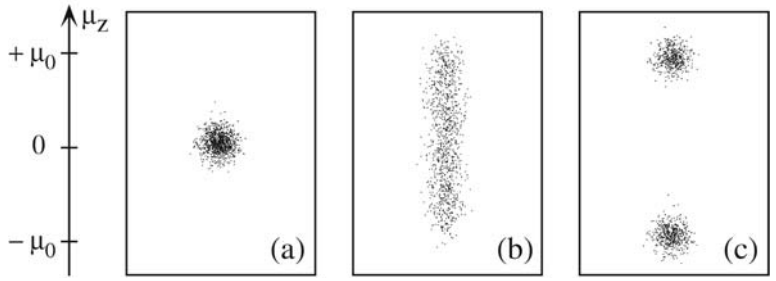


Fig. 11.3. Possible results of a Stern–Gerlach experiment. (a) In the absence of a magnetic gradient no deflection of the atomic trajectories occurs, and the atoms form a single spot around the point $x = z = 0$; each dot represents the impact of an atom on the detection screen. (b) Simulation of the result expected from classical mechanics, assuming that all atoms carry the same magnetic moment $\boldsymbol{\mu}_0$ with a random orientation; the distribution of the z-component of the magnetic moment is uniform between $-\mu_0$ and $+\mu_0$. (c) Simulation of the result found experimentally with silver atoms: the experiment, which can be considered as a measurement of the z-component of the magnetic moment, yields only the two results $+\mu_0$ and $-\mu_0$.

a magnetic moment. In a region of length L, a magnetic field $\boldsymbol{B}$ parallel to z is applied with a gradient along z:

$$\boldsymbol{B}(\boldsymbol{r}) = B_z(\boldsymbol{r})\boldsymbol{u}_z\,, \quad \text{with} \quad B_z(\boldsymbol{r}) = B_0 + b'z\ . \tag{11.24}$$

In full rigor, the equation (11.24) is incorrect because the field $\boldsymbol{B}(\boldsymbol{r})$ does not satisfy $\boldsymbol{\nabla}.\boldsymbol{B} = 0$. A more realistic calculation can be done with a field $\boldsymbol{B} = B_0\boldsymbol{u}_z + b'(z\boldsymbol{u}_z - x\boldsymbol{u}_x)$ which satisfies Maxwell equations. If the dominant part of the field $B_0\boldsymbol{u}_z$ is much larger than the transverse field $-b'x\boldsymbol{u}_x$ on the transverse extension Δx of the atomic wave packet (i.e., $B_0 \gg b'\Delta x$), the eigenstates of $-\hat{\boldsymbol{\mu}} \cdot \boldsymbol{B}$ remain practically equal to $|\pm\rangle_z$ and the present approach is valid.

Under these conditions, the Schrödinger equation (11.17) can be decoupled into two equations:

$$i\hbar\frac{\partial}{\partial t}\psi_+(\boldsymbol{r},t) = \left(\frac{\hat{p}^2}{2m} - \mu_0 B\right)\psi_+(\boldsymbol{r},t)\ , \tag{11.25}$$

$$i\hbar\frac{\partial}{\partial t}\psi_-(\boldsymbol{r},t) = \left(\frac{\hat{p}^2}{2m} + \mu_0 B\right)\psi_-(\boldsymbol{r},t)\ . \tag{11.26}$$

These two equations are both of the same type as the Schrödinger equation seen in chapter 5, but the potential is not the same for ψ_+ and ψ_-. In order to proceed further, we set

$$\pi_\pm = \int |\psi_\pm(\boldsymbol{r},t)|^2\, d^3r\,, \quad \pi_+ + \pi_- = 1\ , \tag{11.27}$$

where π_+ and π_- are the probabilities of finding $\mu_z = +\mu_0$ and $\mu_z = -\mu_0$. We deduce from (11.25) and (11.26) that

$$\frac{d\pi_+}{dt} = \frac{d\pi_-}{dt} = 0 \quad . \tag{11.28}$$

We define the functions

$$\phi_\pm(\boldsymbol{r}, t) = \psi_\pm(\boldsymbol{r}, t)/\sqrt{\pi_\pm} \ , \tag{11.29}$$

which are the conditional probability amplitudes of particles for which $\mu_z = \pm\mu_0$. These normalized functions also satisfy the Schrödinger-type equations (11.25) and (11.26).

We now define

$$\langle \boldsymbol{r}_\pm \rangle = \int \boldsymbol{r} \, |\phi_\pm(\boldsymbol{r}, t)|^2 \, d^3r \ , \tag{11.30}$$

$$\langle \boldsymbol{p}_\pm \rangle = \int \phi_\pm^*(\boldsymbol{r}, t) \, \frac{\hbar}{i} \, \boldsymbol{\nabla} \phi_\pm(\boldsymbol{r}, t) \ d^3r \ , \tag{11.31}$$

where $\langle \boldsymbol{r}_+ \rangle$ (resp., $\langle \boldsymbol{r}_- \rangle$) is the average position of particles for which $\mu_z = +\mu_0$ (resp., $\mu_z = -\mu_0$), and $\langle \boldsymbol{p}_\pm \rangle$ are their average momenta. A simple application of the Ehrenfest theorem gives

$$(d/dt)\langle \boldsymbol{r}_\pm \rangle = \langle \boldsymbol{p}_\pm \rangle / m \ , \tag{11.32}$$

$$(d/dt)\langle p_{x_\pm} \rangle = (d/dt)\langle p_{y_\pm} \rangle = 0 \ , \tag{11.33}$$

$$(d/dt)\langle p_{z_\pm} \rangle = \pm\mu_0 b' \quad . \tag{11.34}$$

At $t = 0$, we assume that

$$\langle \boldsymbol{r}_\pm \rangle = 0 \, , \quad \langle p_{x_\pm} \rangle = \langle p_{z_\pm} \rangle = 0 \, , \quad \langle p_{y_\pm} \rangle = mv \quad .$$

We obtain at time t

$$\langle x_\pm \rangle = 0 \, , \quad \langle y_\pm \rangle = vt \, , \quad \langle z_\pm \rangle = \pm\mu_0 b' t^2 / 2m \quad . \tag{11.35}$$

Therefore, there is a spatial separation along z of the initial beam into two beams. One beam corresponds to $\mu_z = +\mu_0$, and the other to $\mu_z = -\mu_0$. When the beams leave the magnet of length L, their separation is

$$\delta z = \langle z_+ \rangle - \langle z_- \rangle = \frac{\mu_0 b'}{m} \, \frac{L^2}{v^2} \quad . \tag{11.36}$$

If the field gradient is sufficiently strong so that $\delta z > \Delta z$ (separation larger than the spatial extension of each wave packet), we obtain two beams: one in the internal state $|+\rangle$, the other in the state $|-\rangle$. Therefore, the formalism we have developed in this section explains completely the Stern–Gerlach experiment and its results. As was announced in the case of the population

inversion for the ammonia molecule in chapter 7, this experiment elicits two fundamental aspects of a measurement process in quantum mechanics:

- A measurement requires a finite spatial extension ($\delta z = 0$ if $L = 0$).
- A measurement is never instantaneous ($\delta z = 0$ if $T = L/v = 0$).

These two aspects were absent in the formulation of the principles of quantum mechanics presented in chapter 6.

Finally, a simple inspection of the evolution of the expectation value of the separation of the two spots, and of their dispersion at the exit of the magnet, leads to the following result. Let $T = L/v$ be the time the atoms spend in the inhomogeneous magnetic field, and let us note $E_\perp = \langle p_z^2 \rangle / 2m$ the transverse energy communicated to the atom by the field gradient. In order to observe the splitting, the following condition must be satisfied,

$$T\, E_\perp \geq \hbar/2 \quad .$$

This condition, where the value of the field gradient has disappeared, is one important aspect of the so-called *time-energy uncertainty relation* that appears in any quantum measurement.

11.6.5 Successive Stern–Gerlach setups

Consider now the experimental situation shown in Figure 11.4. We place two consecutive magnets. The first has a field gradient directed along z and it splits the incident beam into two beams corresponding to the two internal states $|+\rangle_z$ and $|-\rangle_z$. When the beams leave the field zone, we stop the beam corresponding to the state $|-\rangle_z$ and we keep only the beam in the state $|+\rangle_z$. This latter beam is then sent into another Stern–Gerlach device whose axis is along the x-axis, orthogonal to z. We therefore perform a measurement of the x-component of the atomic magnetic moment, whose corresponding observable is $\hat{\mu}_x$. The result observed experimentally is that the beam is again split in two beams of equal intensities corresponding to values of the magnetic moment along x equal to $+\mu_0$ and $-\mu_0$, respectively.

This, naturally, is the manifestation of the superposition principle. The eigenstates of $\hat{\mu}_x$ and of $\hat{\mu}_y$ are

$$|\pm\rangle_x = \frac{1}{\sqrt{2}} \left(|+\rangle_z \pm |-\rangle_z \right) , \quad |\pm\rangle_y = \frac{1}{\sqrt{2}} \left(|+\rangle_z \pm i|-\rangle_z \right) \ . \tag{11.37}$$

There is nothing surprising.

11.6.6 Measurement along an arbitrary axis

Suppose we are interested in measuring the component of the magnetic moment along an arbitrary axis. This is shown in Figure 11.5. We place a Stern–Gerlach apparatus along an arbitrary direction defined by the unit vector $\boldsymbol{u}_\theta$ such that

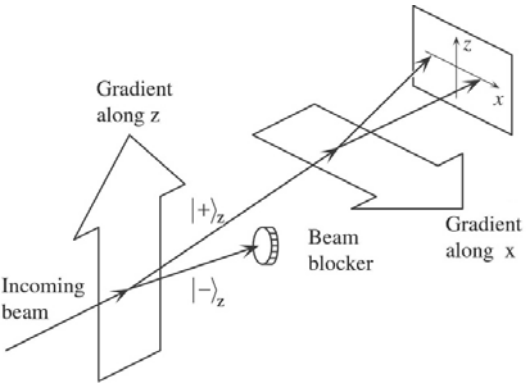


Fig. 11.4. A beam of silver atoms crosses two magnetic field zones. The first creates a field gradient along z, the second a field gradient along x. After the first magnet, a shutter only lets the atoms in the internal state $|+\rangle_z$ pursue. The second magnet allows us to perform a measurement of the x-component of the magnetic moment. One finds the two results $+\mu_0$ and $-\mu_0$ with equal probabilities.

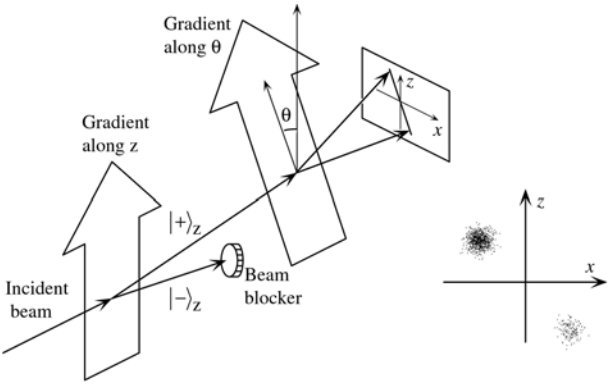


Fig. 11.5. A beam of silver atoms is prepared in the state $|+\rangle_z$. It then crosses a field gradient directed along $\boldsymbol{u}_\theta$. In this measurement of the component of the magnetic moment along $\boldsymbol{u}_\theta$, the two possible results are $+\mu_0$ and $-\mu_0$ with respective probabilities $\cos^2\theta/2$ and $\sin^2\theta/2$. The graph in the lower right corner shows a typical result for $\theta = \pi/4$.

$$\boldsymbol{u}_\theta = \boldsymbol{u}_x \sin\theta + \boldsymbol{u}_z \cos\theta \ . \tag{11.38}$$

This corresponds to a measurement of the component μ_θ of the magnetic moment along $\boldsymbol{u}_\theta$ (i.e., $\mu_\theta = \mu_x \sin\theta + \mu_z \cos\theta$). Using the correspondence principle, we assume that the corresponding observable is

$$\hat{\mu}_\theta = \hat{\mu}_x \sin\theta + \hat{\mu}_z \cos\theta = \mu_0 \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} \ . \tag{11.39}$$

This choice guarantees that the expectation values $\langle\mu_x\rangle, \langle\mu_y\rangle$, and $\langle\mu_z\rangle$ of the components of the magnetic moment transform as the components of a usual three-vector under rotations, which is essential.

Just as $\hat{\mu}_x, \hat{\mu}_y, \hat{\mu}_z$, the operator $\hat{\mu}_\theta$ has the eigenvalues $+\mu_0$ and $-\mu_0$. Its eigenvectors are

$$|+\rangle_\theta = \quad |+\rangle_z \cos(\theta/2) + |-\rangle_z \sin(\theta/2) = \begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix} , \tag{11.40}$$

$$|-\rangle_\theta = -|+\rangle_z \sin(\theta/2) + |-\rangle_z \cos(\theta/2) = \begin{pmatrix} -\sin(\theta/2) \\ \cos(\theta/2) \end{pmatrix} . \tag{11.41}$$

The experimental observations are the following. If a beam prepared in the state $|+\rangle_z$ is sent in the field gradient directed along $\boldsymbol{u}_\theta$, one finds that this beam is split into two beams corresponding to a magnetic moment along $\boldsymbol{u}_\theta$ equal to $+\mu_0$ and $-\mu_0$, with relative intensities $I_+(\theta) = I_+(0)\cos^2(\theta/2)$ and $I_-(\theta) = I_+(0)\sin^2(\theta/2)$.

In order to account for this result, we apply the principles of chapter 6. A measurement of $\hat{\mu}_\theta$ can give two possible values, the eigenvalues $+\mu_0$ and $-\mu_0$; if the initial system is in the state $|+\rangle_z$, the respective probabilities for these two issues are

$$p_+ = |_\theta\langle +|+\rangle_z|^2 = \cos^2(\theta/2) , \tag{11.42}$$

$$p_- = |_\theta\langle -|+\rangle_z|^2 = \sin^2(\theta/2) . \tag{11.43}$$

Therefore, equation (11.39) explains why the experimental measurement, which involves a large number of atoms, gives two spots with relative intensities $\cos^2(\theta/2)$ and $\sin^2(\theta/2)$. The measurement only gives a result with probability 1 when θ is 0 or π, that is, when the preparation axis $\boldsymbol{u}_z$ and the measurement axis $\boldsymbol{u}_\theta$ are parallel or antiparallel.

11.7 The discovery of spin

11.7.1 The hidden sides of the Stern–Gerlach experiment

Coming back to Stern and Gerlach, one naturally thinks that they have discovered spin 1/2! They have found the quantization of the magnetic moment, the superposition principle, and the explanation of the electron's magnetic moment if not its spin.

It is with that type of reasoning and observations that Fresnel had founded the wave theory of light, in particular, the laws of polarization, which was a similar and difficult problem in the 19th century.

Absolutely not! The result was considered

1. as perfectly natural, and
2. as a brilliant confirmation of the old quantum theory of Bohr and Sommerfeld (remember this happened in 1922).

If people had thought about it, they might have found the answer. But at that time, physicists were concerned with quite a different problem: they wanted to prove the old quantum theory and the quantization of trajectories that Bohr had used in his model of the hydrogen atom.

Actually, the Stern–Gerlach experiment had been induced by theorists, Born, Bohr, Sommerfeld, and Pauli, for months if not for years, and they had predicted the result.

In physics, there is some sort of credibility principle for an experimental result. One believes in a result only if a theory has predicted it. An example is the discovery of the 3 K cosmic background radiation by Penzias and Wilson. It was considered as a background noise until they were told about the prediction of Gamow of the existence of background radiation in the big bang theory.

The quantization of L_z had been guessed very early.

Sommerfeld, in the old quantum theory, had predicted the spatial quantization of trajectories and the directional quantization in a magnetic field $\boldsymbol{B}$. He knew that $\boldsymbol{\mu} = \gamma \boldsymbol{L}$, and that the orbital gyromagnetic ratio is $q/2m$. Sommerfeld understood the principle of the experiment as soon as 1918. And he expected a lot from it because it would have been the first proof of quantization in a nonradiative process.

But one could argue that there should be three spots and not two. Imagine an electron in a circular uniform motion around a proton. The quantization of angular momentum is an integer multiple of $\hbar$. In a magnetic field, the plane of the trajectory could have three directions corresponding, respectively, to an angular momentum parallel, antiparallel, or perpendicular to the field $\boldsymbol{B}$ with $L_z = \hbar$, $L_z = -\hbar$, or $L_z = 0$.

Not at all! As soon as 1918, Bohr proved that the trajectory $L_z = 0$ was unstable. One must therefore observe only two spots $L_z = \pm 1$.[1]

Born insisted in 1920 (he was 30), "This experiment must absolutely be done." At that time, Born was a professor in Frankfurt, where there was an artist of atomic and molecular beams, Otto Stern (32 at that time), but Stern wasn't interested.

So, Born, who was a mathematician, decided to do experiments. And he managed to do so thanks to a talented assistant Fraulein Elizabeth Bormann. This new activity of Born was a surprise to all physicists. (One day, Rutherford asked him if he had a relative doing experiments. Born answered, "No, but I have a good assistant.")

But Born had to face the facts; he suffered from the Pauli effect: the better you are as a theorist, the more you are a disastrous experimentalist. Whenever Pauli entered a laboratory, everything went wrong. One day, in Göttingen, an experimental setup of Franck exploded. Everyone looked for Pauli, but there was no trace of him. Some time later, someone learned that at the precise

[1] It is one of the few times that Bohr predicted correctly the result of an experiment. He did it with a wrong argument.

time of the explosion, Pauli was on a train, which had stopped in Göttingen, on the way from Munich to Hanover. The Pauli effect acted at a distance!

Born eventually convinced Stern. Actually, Stern did not know what to think. At first he proposed the experiment, but some time later he was skeptical, "Quantum restrictions on trajectories are simply calculational rules. I'm going to show once for all that what theorists say is nonsense." However, Stern suffered somewhat from the Pauli effect. All his experimental setups were constructed by his technician. He knew remarkably how to conceive them, but he wasn't very skillful.

And, then, it was too difficult: neither the technician nor Fraulein Bormann succeeded. Fortunately, Gerlach, who was a very talented 21-year-old experimentalist, had just arrived in Frankfurt, after graduating in Tübingen. Born said "Thank God, now we have at last someone who knows how to do experiments!" Gerlach took care of everything – the technician, Fraulein Bormann, and Stern's ideas – and he did the experiment. He was successful, and found the two spots. It seemed to be a triumph for Sommerfeld.

Pauli (22 at that time) congratulated Gerlach and said to: "Let us hope now that the old unbeliever Stern will now be convinced of directional quantization!"

The triumph was even greater because, by measuring μ_0 they found to a few percent

$$\mu_0 = |\gamma_0^{orb}|\hbar = \frac{q\hbar}{2m_e} \quad , \tag{11.44}$$

exactly the prediction of Bohr and Sommerfeld!

At that time, nobody could suspect that Nature had played a bad trick. Equation (11.44) must be read as

$$\mu^{spin} = (\frac{q}{m_e})(\frac{\hbar}{2}), \quad \text{and not} \quad (\frac{q}{2m_e})\hbar \quad !$$

In other words, the spin gyromagnetic ratio is twice the orbital gyromagnetic ratio, and the angular momentum is $\hbar/2$.

Dirac proved that in 1927, for any charged pointlike spin 1/2 particle, in his theory of a relativistic electron.

Einstein used to say that the Lord is not mean, but he is subtle. On that point, the Lord had really been nasty!

11.7.2 Einstein and Ehrenfest's objections

However, there were some skeptical people who thought about the physics. Because the experiment is amazing, it is completely opposed to classical conceptions. Einstein and Ehrenfest performed a critical analysis of the experiment in semi-classical terms as in Sommerfeld's theory.

They calculated the time it would take a loop of current to orient itself in the magnetic field, and they found a value of $t \simeq 10^9$ s, that is, 30 years! Yet the atoms stay in the inhomogeneous field for 10^{-4} s.

Einstein and Ehrenfest concluded: "We must make a complete revisal of our classical ideas ... this experimental shows a conceptual dead end ... we must find a new quantum idea."

Indeed, for us, in quantum mechanics, this is not a problem. In the initial beam there is a probability amplitude for the spin to point upwards or downwards, and the orientation of the magnetic moment is "already there," to speak in classical terms.

Now, one can think that with Einstein getting involved in that business, and with that conclusion of his, people started wondering and found spin 1/2. The discovery of the number 2 must have occurred at that time. Not at all; nobody paid any attention.

11.7.3 Anomalous Zeeman effect

At that time, physicists were all concerned with a true intellectual challenge that seems to have nothing to do with all that, the anomalous Zeeman effect.

We place an atom, prepared in a state of energy E and angular momentum j, in a magnetic field $\boldsymbol{B}$ parallel to z. The magnetic energy is

$$\hat{W} = -\hat{\boldsymbol{\mu}} \cdot \boldsymbol{B} \quad . \tag{11.45}$$

The corresponding level is split in $2j + 1$ sublevels of respective energies:

$$E - \gamma \hbar B_0 m \, , \quad m = -j, \ldots, j \; .$$

A corresponding splitting of each line is observed in the spectrum. If all angular momenta are orbital angular momenta (i.e., they have a classical interpretation), j must be an integer. In that case $2j + 1$ is odd and we expect a splitting in an odd number of levels. The splitting of spectral lines in a magnetic field, first observed by Zeeman in the period 1896–1903, showed that in many cases, in particular for alkali atoms, this is not true. There is a splitting into an even number of levels.

Faraday had been convinced as early as 1845 that there was a deep connection between optical and magnetic phenomena. In one of the last experiments of his life, in 1862, he attempted to find the influence of magnetic fields on radiation. Many technical problems prevented him from obtaining a positive answer. It was only in 1896 that these experiments were redone successfully by Zeeman. Already at that time, theorists, in particular H.A. Lorentz, predicted one should observe a splitting in an odd number of lines (1 – no splitting – or 3). Zeeman first confirmed this result on the spectra of cadmium and zinc. The discovery, in the particular case of sodium, of what was to be called the "anomalous Zeeman effect", namely an even number of lines, remained for more than 25 years a real challenge for the scientific community, which was totally confused by this phenomenon. It was only after many struggles, in the years 1925–1926, with the ideas of Pauli, and of Uhlenbeck and Goudsmit, that

the introduction of the notion of spin completely clarified the problem. The "anomalous Zeeman effect" then appeared as a perfectly normal phenomenon.

It was a real antique tragedy. One must realize that if we have the theory in front of us (i.e. that there exist half-integer angular momenta for which $2j+1$ is even) physicists did not at that time, despite the works of Elie Cartan in 1913.

11.7.4 Bohr's challenge to Pauli

The anomalous Zeeman effect seemed to be an impenetrable wall. Bohr thought there was only one person capable of solving that problem: Pauli, whom Bohr invited to Copenhagen in 1922. Pauli was 22, he had been Sommerfeld's student in Munich, and he was known for his remarkable book on relativity. He had a difficult temper, as we said; for him most things were either obvious or stupid. One day, Ehrenfest had said to him, "You know, Herr Pauli, I find your paper on relativity much nicer than you are." Pauli replied, "That's funny! For me it's the opposite, I find you much nicer than all the rubbish you write."

Of course, Pauli's colleagues were terrified. Franck told him; "Please behave when you get to Copenhagen." "I'll have to learn Danish," said Pauli.

Niels Bohr asked Pauli two questions:

- Why are there shells in complex atoms, containing 2, 8, 18, and so on; that is $2n^2$ electrons, $n = 1, 2, 3 \ldots$? This remark had been made by Rydberg in 1903. Why aren't the electrons all in the ground state?
- From where does the Zeeman effect come?

And Pauli started thinking. He thought very hard. So hard that he said, in his Nobel prize lecture in 1945, that one day, as he was strolling in the beautiful streets of Copenhagen, a colleague had said to him in a friendly manner, "You look very unhappy;" whereupon Pauli replied fiercely, "How can one look happy when he is thinking about the anomalous Zeeman effect." Pauli understood that both the anomalous Zeeman effect and the effect of the inner core of electrons are a conceptual dead end in classical terms.

And, at the beginning of 1924, W. Pauli, Jr. wrote "The electron has a new, specifically quantum property, which corresponds to a two-valued physical quantity, which cannot be described classically. In an atom, an electron is characterized by four quantum numbers, (n, l, m, σ) where $\sigma = \pm 1$." And Pauli stated the principle of exclusion of equivalent electrons, called by Heisenberg the "Pauli verbot," which is very deep and also revolutionary.

At this point, one is tempted to say, "Fine, Pauli found the answer." Not at all.

11.7.5 The spin hypothesis

The hypothesis of the electron spin, such as presented above, that the electron has an intrinsic angular momentum $\boldsymbol{S}$ whose measurement along an axis gives

one of the two values $\pm\hbar/2$, and whose associated magnetic moment $\boldsymbol{\mu}_S = \gamma_S \boldsymbol{S}$ is such that $\gamma_S = q/m = 2\gamma_{orbital}$, was proposed by two young Dutch physicists in 1925, Uhlenbeck and Goudsmit. Uhlenbeck hesitated between a career in physics and in history, and Goudsmit did not have his degree yet. As soon as they understood that their hypothesis explained many experimental facts which were obscure up to then, they discussed it with their professor, P. Ehrenfest, who encouraged them to publish their work. Their idea was received with various reactions in the physics community. Bohr was very enthusiastic, whereas Pauli and Lorentz found serious objections. The objection of Lorentz comes from relativity. If one imagines that the electron is a sphere whose electrostatic energy is equal to its mass $m_e c^2$, one finds a radius of the order of $e^2/(m_e c^2)$ and the equatorial velocity of the sphere is much larger than the velocity of light, if the angular momentum is $\hbar/2$ (one obtains $v_{\text{eq}} \sim c/\alpha = 137c$). We know that this objection is not valid, because spin is a purely quantum concept.

11.7.6 The fine structure of atomic lines

In 1925, Pauli did not believe in spin. In January 1926 he called it a heresy, "Irrlehre."

Pauli's criticism concerned what is known as fine structure effects. It is one of these effects of fine splitting of atomic lines that constitutes a world record in the agreement between experiment and theory. This splitting comes from the interaction of the electron's spin magnetic moment with its orbital angular momentum. Consider the hydrogen atom and let us imagine we sit on the electron. It sees the proton circling around it with an angular momentum $\boldsymbol{L}$ and producing a magnetic field that interacts with the spin magnetic moment. The resulting interaction is $W = 2\alpha^2 E_I (a_0/r)^3 (\boldsymbol{L}\cdot\boldsymbol{S})/\hbar^2$. And, once all calculations are performed, this gives a splitting that is two times too large!

Fortunately, everything was saved by an English physicist. In March 1926, L.H. Thomas remarked that the rest frame of the electron is not an inertial frame, and that a correct relativistic calculation introduces a factor of 1/2 in the formula, called the Thomas precession.[2] (Remember Pauli was a great specialist of relativity!)

On March 12, 1926, Pauli wrote to Bohr: "I must surrender!"

So, there is the incredible story of the number two in physics. It appeared everywhere, but with traps. For a long time, nobody thought that all these numbers two had a common origin. There are lots of numbers in physics, but that one is quite simple. Spin was a two-valued quantum number. Its gyromagnetic ratio was twice the orbital gyromagnetic ratio. The "normal" Zeeman effect appeared in two-electron atoms. The anomalous Zeeman effect was a splitting in an even number of spectral lines. Complete electron shells in

[2] See, for example, J.D. Jackson, *Classical Electrodynamics*, Section 11.8. New York: Wiley, (1975).

atoms contained $2n^2$ electrons. The exclusion principle said that two electrons could not be in the same state. The fine structure of atoms consisted of a splitting in two lines. The Thomas precession introduced a factor of $1/2$.

Maybe the Creator thought those physicists were going too far and he was annoyed to see them discover the structure of the world He had created. So He introduced this number two with all possible traps, thinking that they would stop teasing him. But He had forgotten that He had created Pauli. One cannot think of every detail, even when one is a Creator.

11.8 Magnetism, magnetic resonance

In his Nobel Lecture, on December 12, 1946, Otto Stern spoke about his experiment, but he insisted much less on its revolutionary aspects than on the measurement of the parameter μ_0 which he called the elementary magneton, the elementary quantum of magnetism.

The discovery of magnetism goes very far back in time. It is probable that the first physicists of mankind knew this extraordinary phenomenon, which appeared as a force at a distance at a human scale. One can find examples of compasses in China in 2600 B.C., and with the Vikings in the 12th century.

The treatise "De Magnete" of William Gilbert, in 1600, classifies magnetic phenomena into three categories:

- The strong and permanent magnetism of substances such as iron, cobalt, and nickel
- The weaker paramagnetism, induced by strong fields in crystals and in fluids
- The diamagnetism, weak and repulsive, that appears in all substance and that appears as the square of the applied field

All these forms of magnetism come from electrons.

Paramagnetism is the response of a magnetic moment, be it orbital or spin, to an external field.

Ferromagnetism does not come from the orbital motion of electrons, but from a subtle collective consequence of Pauli's principle. In the transition metals, such as Fe, Co, Ni, the wave functions of external electrons are such that the electron spins point spontaneously in the same direction. This results in a permanent macroscopic magnetization.

Stern knew he had discovered the elementary quantum of magnetism proposed by Pierre Weiss in 1911. For a long time, physicists had understood that ferromagnetism cannot be explained by currents. Weiss introduced the idea of an "elementary magneton," analoguous to the elementary charge, as being the "Greatest common divisor of molecular magnetic moments."[3]

[3] The Weiss magneton was actually one fifth of the value found by Stern.

But, in the same type of experiment, Stern went much further. In 1933, he managed to measure the magnetic moment of the proton. This was much more difficult, because it is 1000 times smaller because of the mass factor (q/m) and one must suppress electronic effects. Stern used H_2 or HD molecules, where the effect of paired electrons cancel. Stern found a gyromagnetic ratio roughly 2.5 times what one expects from a pointlike particle (i.e., the nuclear magneton $\mu_N = q\hbar/2m_p = 3.1525\,10^{-8}\,\mathrm{eV\,T^{-1}}$).

This observation was the starting point of the discovery of a fourth type of magnetism, nuclear magnetism, which is a major discovery of the 20th century. There exists a nuclear ferromagnetism, a nuclear paramagnetism. They are very weak, but their consequences are huge.

This is when Rabi appears. He was born in 1899, he had worked on a thesis on magnetism at Cornell; then he had gone to Columbia where he learned some quantum mechanics. In the course of his work, he made a little calculation which had enormous consequences. In particular, between 1933 and 1939, he gained a factor of 1000 in accuracy compared to Stern's measurements.

11.8.1 Spin effects, Larmor precession

Uncorrelated space and spin variables

In most physical situations, such as the Stern–Gerlach experiment, the space and the spin variables are correlated. For instance, we only studied in chapter 10 a first approximation to the hydrogen atom where we neglected spin effects. If we include the spin degree of freedom in this approximation, we find that the states $|n,\ell,m,+\rangle$ and $|n,\ell,m,-\rangle$ are degenerate. Actually, there exist corrections to this approximation, such as the fine structure of the hydrogen atom, which we mentioned above, which is due to the interaction between the spin magnetic moment and the electromagnetic field created by the proton. The degeneracy is then partially lifted and the new states are combinations of the initial states $|n,\ell,m,\sigma\rangle$. In other words, in a given eigenstate of the total Hamiltonian, the spatial wave function of an electron depends on its spin state, and the two random variables $\boldsymbol{r}$ and S_z are correlated.

This correlation is sometimes extremely weak. In such cases, the two variables $\boldsymbol{r}$ and S_z can be considered as independent and their probability law is factorized. Such a physical situation is represented by a factorized state vector

$$\Phi(\boldsymbol{r},t)\begin{pmatrix}\alpha_+(t)\\ \alpha_-(t)\end{pmatrix}\ . \tag{11.46}$$

If one performs spin measurements in this case, the results are independent of the position of the particle. The only relevant observables are 2×2 Hermitian matrices with numerical coefficient (which can depend on time).

Such cases happen in practice, in particular in magnetic resonance experiments. We then use the term *spin state of the proton* instead of *state of the proton* because the position of the proton in space does not play any role in the experiment under consideration.

11.8.2 Larmor precession in a fixed magnetic field

We choose the z-axis parallel to a magnetic field $\boldsymbol{B}_0$. Ignoring space variables, the Hamiltonian is

$$\hat{H} = -\hat{\boldsymbol{\mu}} \cdot \boldsymbol{B}_0 = -\mu_0 B_0 \, \hat{\sigma}_z \quad . \tag{11.47}$$

We set

$$-\mu_0 B_0/\hbar = \omega_0/2 \,, \quad \text{that is,} \quad \omega_0 = -\gamma B_0 \quad . \tag{11.48}$$

The eigenstates of $\hat{H}$ are the eigenstates $|+\rangle$ and $|-\rangle$ of $\hat{\sigma}_z$.

Consider an arbitrary state $|\psi(t)\rangle$ such that $|\psi(0)\rangle = \alpha|+\rangle + \beta|-\rangle$ with $|\alpha|^2 + |\beta|^2 = 1$. Its time evolution is

$$|\psi(t)\rangle = \alpha \, e^{-i\omega_0 t/2} \, |+\rangle + \beta \, e^{i\omega_0 t/2} \, |-\rangle \quad . \tag{11.49}$$

The expectation value $\langle \boldsymbol{\mu} \rangle$ reads

$$\langle \mu_x \rangle = 2\mu_0 \, \mathrm{Re} \left(\alpha^* \beta \, e^{i\omega_0 t} \right) = C \, \cos(\omega_0 t + \varphi) \,, \tag{11.50}$$

$$\langle \mu_y \rangle = 2\mu_0 \, \mathrm{Im} \left(\alpha^* \beta \, e^{i\omega_0 t} \right) = C \, \sin(\omega_0 t + \varphi) \,, \tag{11.51}$$

$$\langle \mu_z \rangle = \mu_0 \, \left(|\alpha|^2 - |\beta|^2 \right) \,, \tag{11.52}$$

where C and φ are, respectively, the modulus and phase of the complex number $\alpha^*\beta$. We recover the Larmor precession that we derived for an arbitrary angular momentum in chapter 9, Section 9.5.4. The projection $\langle \mu_z \rangle$ of the magnetic moment along the field is time-independent, and the component of $\langle \boldsymbol{\mu} \rangle$ perpendicular to $\boldsymbol{B}$ rotates with the angular velocity ω_0. The fact that μ_z is a constant of motion is a consequence of the commutation relation $[\hat{H}, \hat{\mu}_z] = 0$ and of the Ehrenfest theorem.

This provides a simple method to measure the angular frequency ω_0. We place a coil in a plane parallel to $\boldsymbol{B}_0$ and we prepare a macroscopic quantity of spins all in the same spin state $|\psi(0)\rangle$. The precession of $\langle \boldsymbol{\mu} \rangle$ at the frequency ω_0 causes a periodic variation of the magnetic flux in the coil, and this induces an electric current at the same frequency. This method is, however, not as accurate as the resonance experiment we present below.

11.8.3 Rabi's calculation and experiment

Superposition of a fixed field and a rotating field

A technique invented by Rabi in the 1930s, allows us to perform a very accurate measurement of ω_0 by a resonance phenomenon. With this technique, he was able to gain a factor of 1000 in the accuracy of nuclear magnetic moments.

We place the magnetic moment in a known field B_0, on which we superimpose a weak field B_1 which rotates at a variable angular velocity ω in the xy plane. Such a field can be obtained with two coils along the x- and y-axes, with a current at a frequency ω and phase-shifted by $\pi/2$ (one works

with radio frequencies). At the resonance, for $\omega = \omega_0$, the spin flips between the two possible states $|\pm\rangle$. Notice that this calculation, characteristic of a driven two-state system, is similar to that performed for the ammonia maser in chapter 7.

The form of the Hamiltonian is

$$\hat{H} = -\hat{\boldsymbol{\mu}} \cdot \boldsymbol{B} = -\mu_0 B_0 \hat{\sigma}_z - \mu_0 B_1 \cos\omega t\, \hat{\sigma}_x - \mu_0 B_1 \sin\omega t\, \hat{\sigma}_y \quad . \tag{11.53}$$

We set

$$|\psi(t)\rangle = a_+(t)|+\rangle + a_-(t)|-\rangle \quad . \tag{11.54}$$

The Schrödinger equation yields the differential system

$$i\,\dot{a}_+ = \frac{\omega_0}{2}\, a_+ + \frac{\omega_1}{2} e^{-i\omega t}\, a_- \ , \tag{11.55}$$

$$i\,\dot{a}_- = \frac{\omega_1}{2} e^{i\omega t}\, a_+ - \frac{\omega_0}{2}\, a_- \ , \tag{11.56}$$

where we have defined $\mu_0 B_0/\hbar = -\omega_0/2,\ \mu_0 B_1/\hbar = -\omega_1/2$.

The change of functions $b_\pm(t) = \exp(\pm i\omega t/2) a_\pm(t)$ leads to

$$i\,\dot{b}_+ = -\frac{\omega - \omega_0}{2}\, b_+ + \frac{\omega_1}{2}\, b_- \ , \tag{11.57}$$

$$i\,\dot{b}_- = \frac{\omega_1}{2}\, b_+ + \frac{\omega - \omega_0}{2}\, b_- \ . \tag{11.58}$$

The above transformation is the quantum form of a change of reference frame. It transforms from the laboratory frame to the frame rotating with the magnetic field at an angular velocity ω around the z-axis. With this change of reference frame, the basis of Hilbert space is time-dependent, whereas the Hamiltonian is time-independent:

$$\hat{\tilde{H}} = \frac{\hbar}{2} \begin{pmatrix} \omega_0 - \omega & \omega_1 \\ \omega_1 & \omega - \omega_0 \end{pmatrix} = -\frac{\hbar}{2}(\omega - \omega_0)\hat{\sigma}_z + \frac{\hbar}{2}\omega_1 \hat{\sigma}_x \quad .$$

In this rotating reference system, the problem becomes time-independent!

One can check that the equations (11.57), and (11.58) imply $\ddot{b}_\pm + (\Omega/2)^2\, b_\pm = 0$ with

$$\Omega^2 = (\omega - \omega_0)^2 + \omega_1^2 \ . \tag{11.59}$$

Suppose the spin is initially in the state $|+\rangle$; that is, $b_-(0) = 0$. One finds

$$b_-(t) = -\frac{i\omega_1}{\Omega} \sin\left(\frac{\Omega t}{2}\right) \ , \tag{11.60}$$

$$b_+(t) = \cos\left(\frac{\Omega t}{2}\right) + i\frac{\omega - \omega_0}{\Omega} \sin\left(\frac{\Omega t}{2}\right) \quad . \tag{11.61}$$

The probability that a measurement of S_z at time t gives the result $-\hbar/2$ is

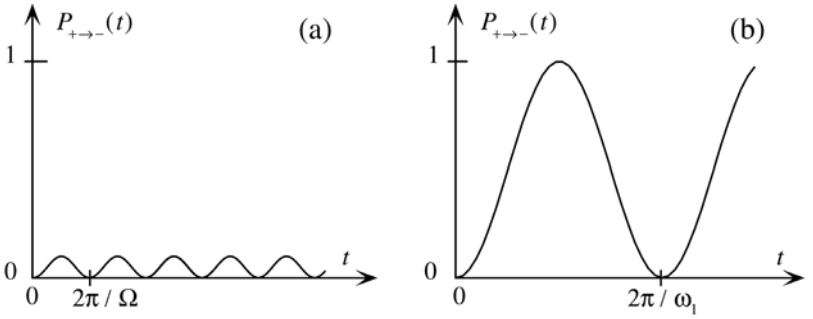


Fig. 11.6. Rabi oscillations (a) slightly off resonance $\omega - \omega_0 = 3\omega_1$; (b) at resonance $\omega = \omega_0$.

$$\mathcal{P}_{+\to -}(t) = |\langle -|\psi(t)\rangle|^2 = |a_-(t)|^2 = |b_-(t)|^2$$
$$= \left(\frac{\omega_1}{\Omega}\right)^2 \sin^2\left(\frac{\Omega t}{2}\right) \quad . \tag{11.62}$$

This formula, which is due to Rabi, exhibits the resonance phenomenon:

- If the frequency ω of the rotating field is noticeably different from the frequency ω_0 we want to measure, more precisely, if $|\omega - \omega_0| \gg \omega_1$, the probability that the spin flips, that is, that we measure $S_z = -\hbar/2$, is very small for all t.
- If we choose $\omega = \omega_0$, then the probability for a spin flip is equal to one at times $t_n = (2n+1)\pi/\omega_1$ (n integer) even if the amplitude of the rotating field $\boldsymbol{B}_1$ is very small.
- For $|\omega - \omega_0| \sim \omega_1$, the probability amplitude oscillates with an appreciable amplitude, smaller than one.

In Figure 11.6, we have drawn the time oscillation of the probability $\mathcal{P}_{+\to -}$ off resonance and at resonance. For a typical magnetic field of 1 Tesla, the resonance frequency is $\omega_e/2\pi \sim 28$ Ghz for an electron, and 2.79 $\omega_N/2\pi \sim$ 43 MHz for a proton. These frequencies correspond to decametric waves in the nuclear case, and centimetric waves in the electronic case.

At the resonance, the system periodically absorbs and emits the energy $2\mu_0 B$ by absorption and stimulated emission.

Notice that, here, we have closed the loop, concerning the starting point in chapter 2. We have proven in a particular case that absorption and stimulated emission of radiation do occur at the Bohr frequency of the system $\nu = \Delta E/h$.

Rabi's experiment

The resonance effect described above was understood by Rabi in 1939. It provides a very accurate measurement of a magnetic moment. The device

used by Rabi consists of a source, two Stern–Gerlach deflectors with magnetic fields in opposite directions, and a detector (Figure 11.7). Between the two Stern–Gerlach magnets, one places a zone with a superposition of a uniform field $\boldsymbol{B}_0$ and a rotating field $\boldsymbol{B}_1$, as described above.

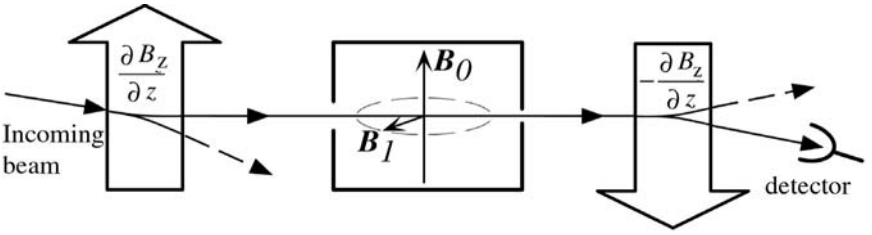


Fig. 11.7. Apparatus developed by Rabi for observing the magnetic resonance effect. In the absence of magnetic resonance, all particles emitted in the state $|+\rangle$ reach the detector. If resonance occurs, the spins of the particles flip between the two magnets, and the signal drops.

Consider first the effect of the two Stern–Gerlach magnets in the absence of the fields $\boldsymbol{B}_0$ and $\boldsymbol{B}_1$. A particle emitted by the source in the spin state $|+\rangle$ undergoes two successive deflections in two opposite directions and it reaches the detector. When the fields $\boldsymbol{B}_0$ and $\boldsymbol{B}_1$ are present, this is not true anymore. If the frequency ω of the rotating field is close to the Larmor frequency ω_0, the resonance phenomenon will change the component μ_z of the particle. When such a spin flip occurs between the two Stern–Gerlach magnets, the two deflections have the same direction (upward in the case of Figure 11.7), and the particle misses the detector. The signal registered on the detector as a function of the frequency of the rotating field $\boldsymbol{B}_1$ undergoes a sharp drop for $\omega = \omega_0$ (Figure 11.8). This leads to a measurement of the ratio $|\mu|/j = \hbar\omega_0/B_0$ for a particle of angular momentum j. Actually, this measurement is so precise that the main source of error comes from the determination of B_0. In practice, as shown in Figure 11.8, the frequency ω stays fixed and one varies the magnitude of the field $\boldsymbol{B}_0$, or equivalently the frequency ω_0.

In 1933, Stern, with his apparatus, had measured the proton magnetic moment with 10% accuracy. That was a very difficult experiment, because nuclear magnetic moments are 1000 times smaller than electronic ones. In 1939, with his resonance apparatus, Rabi gained a factor of 1000 in accuracy. The resonance is very selective in frequency and the presence of other magnetic moments causes no problem. Rabi's result struck the minds of people. It was greeted as a great achievement. Stern remarked that Rabi had attained the theoretical accuracy of the measurement, which is fixed by the uncertainty relations. When Hulthén announced that Rabi was awarded the Nobel prize on December 10, 1944, on the Stockholm radio, he said that "By this method

Rabi has literally established radio relations with the most subtle particles of matter, with the world of the electrons and of the atomic nucleus."

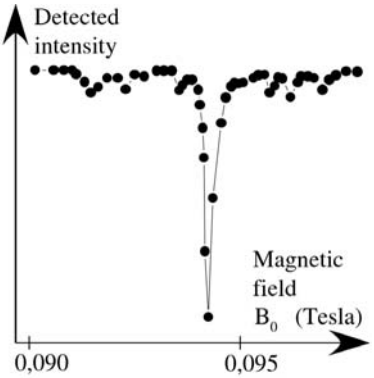


Fig. 11.8. Signal (obtained by Rabi) recorded on the detector of Figure 11.7 with a beam of HD molecules, as a function of the field B_0 ($B_1 = 10^{-4}$ T, $\omega/2\pi = 4$ MHz).

The Nobel prize for 1943 was awarded to Stern, that of 1944 to Rabi, and that of 1945 to Pauli. Rabi did not go to the reception. The world had undergone a terribly hard period.

11.8.4 Nuclear magnetic resonance

Stern and Rabi agreed in 1943 on the fact that there was an intrinsic limitation with atomic and molecular beams, both in time and in intensity. They knew that eventually one would have to operate on condensed matter.

The great breakthrough of the applications of nuclear magnetic resonance (N.M.R.) came with the works of Felix Bloch at Stanford and of Edward Purcell at MIT, in 1945. Owing to the development of radiowave technologies, Bloch and Purcell were able to operate on condensed matter, and not on molecular beams. One uses macroscopic numbers of spins, thereby obtaining much more intense signals and more manageable experiments. The resonance is observed, for instance, by measuring the absorption of the wave generating the rotating field $\boldsymbol{B}_1$. The imbalance between the populations of the two states $|+\rangle$ and $|-\rangle$, which is necessary in order to get a signal, results from the conditions of thermal equilibrium. In a field $B_0 = 1$ T, the magnetic energy for a proton is 2.79 $\mu_N B \sim 10^{-7}$ eV and the relative population difference between the two spin states due to the Boltzmann factor at room temperature is $\pi_+ - \pi_- \sim 4\ 10^{-6}$. This relative difference is small, but quite sufficient to observe a significant signal because one deals with samples containing a macroscopic number of spins (typically 10^{23}).

Bloch and Purcell discovered nuclear paramagnetism, a fourth type of magnetism.

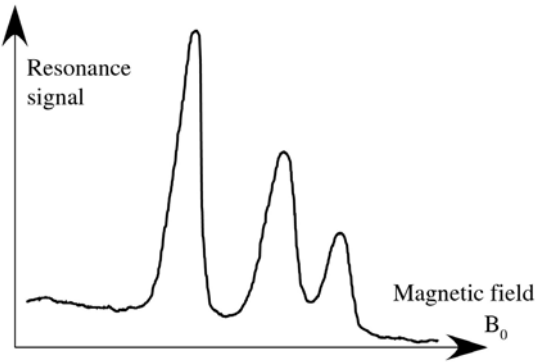


Fig. 11.9. One of the first examples of nuclear magnetic resonance applied to chemistry: the resonance signal obtained with the protons of the ethanol molecule CH_3CH_2OH consists of a three-peak structure. These peaks are associated respectively with the three protons of the CH_3 group, with the two protons of the CH_2 group, and with the unique proton of the OH group. The magnetic field B_0 is $\sim$0.8 T, and the total trace is 7.5 μT wide.

The applications of magnetic resonance are numerous in domains ranging from solid-state physics and low temperatures, chemistry, biology, or medicine. By its magnetic effects, spin can play the role of a local probe inside matter. NMR has transformed chemical analysis and the determination of the structure of molecules (see, for instance, Figure 11.9). It has become an invaluable tool in molecular biology. Since 1980, NMR has also caused a revolution in medical diagnosis and physiology. It allows us to measure and visualize in three dimensions and with a spatial precision better than a millimeter the concentration in water of "soft matter" (muscles, brain, and so on), which, in contrast with bones, is difficult to observe with X-rays. One studies in this way the structure and the metabolism of living tissues; one can detect internal injuries, tumors, and so on. The nuclear spin, which was a curiosity for some visionary physicists of the 1940s and 1950s, has become one of the great hopes of modern medicine.

One can, by now, visualize the activity of the brain in real-time. It is possible to localize and register the response of the visual cortex to some stimulation. The following step, after submitting a volunteer to a sequence of such excitations, is to ask her to think about the signal. The NMR response of the brain is the same as that obtained by an external stimulation. This may be considered a direct experimental proof that we think, which is somewhat comforting for the mind.

11.8.5 Magnetic moments of elementary particles

The electron, the proton, and the neutron are spin 1/2 particles. The corresponding spin magnetic moment is related to the spin $\boldsymbol{S}$ by the relation $\hat{\boldsymbol{\mu}} = \gamma \hat{\boldsymbol{S}}$. Experiments give the following values of the gyromagnetic ratios

$$\begin{aligned} \text{electron} \quad & \gamma \simeq 2\gamma_0 = -q/m_e \ , \\ \text{proton} \quad & \gamma \simeq +2.79 \ q/m_p \ , \\ \text{neutron} \quad & \gamma \simeq -1.91 \ q/m_p \ . \end{aligned}$$

The possible results of the measurement of the component of these magnetic moments along a given axis are therefore

$$\begin{aligned} \text{electron} \quad & \mu_z = \pm\mu_B = \mp q\hbar/2m_e \ , \\ \text{proton} \quad & \mu_z = \pm 2.79 \ q\hbar/2m_p \ , \\ \text{neutron} \quad & \mu_z = \mp 1.91 \ q\hbar/2m_p \ . \end{aligned}$$

The quantity $\mu_B = -9.274\,10^{-24}$ J T^{-1} is called the Bohr magneton. The quantity $\mu_N = q\hbar/2m_p = 5.051\,10^{-27}$ J T^{-1} is called the nuclear magneton.

Dirac's relativistic theory of the electron explains the value of the electron magnetic moment

$$\hat{\boldsymbol{\mu}} = g_e \left(\frac{q}{2m_e} \right) \hat{\boldsymbol{S}} , \quad \text{with} \quad g_e = 2 \quad .$$

The value measured experimentally for the gyromagnetic factor g_e nearly coincides with this prediction. One can account for the slight difference between the experimental result and Dirac's prediction by taking into account the coupling of the electron with the quantized electromagnetic field (quantum electrodynamics). This constitutes one of the most spectacular successes of fundamental physics. The experimental and theoretical values of the quantity g_e coincide within the accuracy limits of experiments and of computer calculations. At present, one has for the electron, setting $g_e = 2(1 + a)$,

$$a^{\text{theo.}} = 0.001\ 159\ 652\ 2\underline{00}\ (40) \ , \tag{11.63}$$

$$a^{\text{exp.}} = 0.001\ 159\ 652\ 1\underline{93}\ (10) \ ; \tag{11.64}$$

the errors in parentheses bear on the two last digits. To lowest order in the fine structure constant, quantum electrodynamics gives the result $a = \alpha/2\pi = 0.00116$.

The coefficients $+2.79$ and -1.91 for the proton and for the neutron are due to the internal structure of these particles. They can be measured with great accuracy by magnetic resonance experiments: $\mu_p/\mu_N = 2.792\ 847\ 3\underline{86}\ (63)$ and $\mu_n/\mu_N = -1.913\ 042\ \underline{75}\ (45)$; they can be calculated with 10% accuracy in the quark model.

11.9 Entertainment: Rotation by 2π of a spin 1/2

It seems obvious and of common sense geometrically that the rotation of 2π of a system around a fixed axis is equivalent to the identity. However, strictly speaking this is not true for a spin 1/2 particle.

Let us return to the calculation of Section 11.8.2 and suppose that, at time $t = 0$ the state of the system is $|+x\rangle$:

$$|\psi(t=0)\rangle = \frac{1}{\sqrt{2}}\left(|+\rangle + |-\rangle\right) \quad .$$

The mean value of the magnetic moment, given by (11.50), through (11.52), is $\langle \boldsymbol{\mu} \rangle = \mu_0 \, \boldsymbol{u}_x$. Equation (11.49) gives the evolution of this state. After a time $t = 2\pi/\omega_0$, classically, the system has precessed by an angle 2π around $\boldsymbol{B}$. In quantum mechanics one verifies that the expectation value is back to its initial value $\langle \boldsymbol{\mu} \rangle = \mu_0 \, \boldsymbol{u}_x$. What can we say, however, about the state vector? We can check that $|\psi(t)\rangle$ is still an eigenvector of $\hat{S}_x$ (or of $\hat{\mu}_x$) with an eigenvalue $+\hbar/2$. However, we notice that, quite surprisingly, the state vector has changed sign:

$$|\psi(t = 2\pi/\omega_0)\rangle = -\frac{1}{\sqrt{2}}\left(|+\rangle + |-\rangle\right) = -|\psi(0)\rangle \quad .$$

A 2π rotation is therefore not equivalent to the identity for a spin 1/2. Only rotations of $4n\pi$ give back the initial state identically. This property can also be guessed from the dependence $e^{im\varphi}$ for orbital angular momenta: using the same formula for $m = 1/2$ and $\varphi = 2\pi$ would give $e^{i\pi} = -1$.

This peculiarity was understood as soon as spin 1/2 was discovered in 1926. It remained a controversial point for more than 50 years. Does the phase of the state vector after a rotation of 2π have a physical meaning? The positive experimental answer was only given in the 1980s in a series of remarkable experiments.[4] The spin 1/2 particles are sent in a two-channel interferometer. In one of them a magnetic field rotates the spins by multiples of 2π. The change in sign of the wave function is observed by a displacement of the interference fringes and the experimental signal confirms that a rotation of 4π is needed to recover the fringe pattern that is measured in the absence of rotation.

This property reflects an important mathematical structure that relates the two Lie groups SO(3) and SU(2). Rotations in Euclidian space R^3 form the well-known group SO(3). Mathematically, one says that there is a local isomorphism between the Lie algebras of the two groups SO(3) and SU(2), but that these two groups are not globally isomorphic. This formalism, which was called *spinor theory*, was developed in the early 20th century by the mathematician Elie Cartan. The minus sign is completely equivalent to the fact that on a Moebius strip one must make an even number of rotations in order to get back to the starting point.

[4] A.W. Overhauser, A.R. Collela, and S.A. Werner, *Phys. Rev. Lett.* **33**, 1237 (1974); **34**, 1472 (1975); **35**, 1053 (1975).

12

The Pauli Principle

The origin of geometry (from Pythagoras to Euclid) lies in our environment, and in the observation that one can model the world in which we live by a space where each object is described by a point or a set of points. The concept of space itself came after the simpler concept of "place" of an object. The idea of space arose with the question as to whether the place exists independently of the fact that some object occupies it. In this context, by definition, two objects cannot have the same position at the same time.

In this chapter, we address the quantum transposition of this problem. In the probabilistic quantum description, there is no reason a priori why the density probability for two particles to be at the same point in space should vanish, contrary to the classical observation. It is therefore legitimate to elevate the above question to state vectors (or wave functions) rather than positions. Can two particles be in the same state at the same time?

Naturally, two particles of different kinds, such as an electron or a proton, will never be in the same state: even if their wave functions coincide, their mass difference implies differences in the values of various physical quantities and one can always tell them from each other. However, there exist in Nature *identical particles*: all electrons in the universe have the same mass, the same charge, and so on. Can such particles, whose intrinsic properties are all the same, be in the same state? The answer lies in one of the simplest, but most profound, principles of physics, whose consequences on the structure of matter are numerous and fundamental: the *Pauli principle*.

The depth, the intellectual upheaval, and the philosophical implications of the Pauli principle are considerable. If this principle did not provoke the same interest as relativity among philosophers, and even among physicists, it is probably because it explained so many experimental facts (many more than relativity) that Fermi and Dirac had incorporated it immediately in the general theory of quantum mechanics.

The result is the following: The state of a state of N identical particles is either totally symmetric or totally antisymmetric if one exchanges any two of these particles. It is symmetric if the spin of these particles is integer; it is

antisymmetric if the spin is half-integer. This relation between the symmetry of states and the spin of particles is an experimental fact. However it is both a mystery and a triumph of modern theoretical physics that this property can be proven but cannot be explained in a simple way.

We show in Section 12.1 that there is a genuine physical problem: some predictions are ambiguous. The principles of chapter 6 do not suffice when one deals with systems containing identical particles. A new fundamental principle must be added in order to get rid of this ambiguity.

In Section 12.2, we introduce the notion of the total spin of a set of two particles, identical or not, both of which have spin 1/2. This belongs to the general algebraic problem of the addition of angular momenta in quantum mechanics but, in this book, we do not need the general results. We use these results on the total spin in other problems such as entangled states and the 21-cm line of hydrogen.

The essential point is that, by definition, two identical particles can be interchanged in a physical system without modifying any property of this system. The mathematical tool that corresponds to the interchange of two particles is the exchange operator which we introduce in Section 12.3. In Section 12.4 we express the Pauli principle as an additional axiom. Finally, in Section 12.5, we discuss some consequences of the Pauli principle.

12.1 Indistinguishability of two identical particles

12.1.1 Identical particles in classical physics

By definition, two particles are identical if all their intrinsic properties are the same. In classical mechanics, for a two-particle system, it is always possible to measure at a given time the position of each particle. At that instant, we can define which particle we call 1 and which one we call 2. It is also possible to follow the trajectory of particle 1 and that of particle 2. We can keep on distinguishing unambiguously each particle at any later time. For instance, in the collision of two billiard balls of the same color, we can unambiguously tell the difference between the two processes of Figure 12.1. Therefore, for any system that is described by classical physics, two particles are always distinguishable, whether or not they are identical (concerning macroscopic objects; the notion of identity is anyway an idealization).

12.1.2 The quantum problem

The situation is different in quantum mechanics. At a given time we can still measure the positions of the particles and label them with the indices 1 and 2. However, because the notion of a trajectory is not defined, it may be impossible to follow the two particles individually as time goes on. For instance, one cannot tell the difference between the two processes sketched

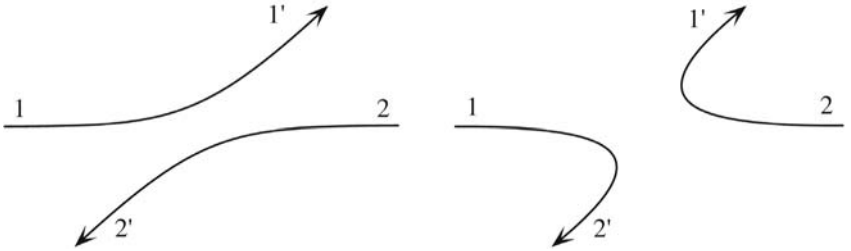


Fig. 12.1. Collision between two identical particles.

in Figure 12.1 if the two wave functions of particles 1 and 2 overlap. It is impossible to know whether particle 1 has become particle $1'$ or particle $2'$. In quantum mechanics two identical particles are indistinguishable.

Here, physics falsifies the famous "principle of the identity of indistinguishables," which is a basic principle in Leibniz's philosophy, where two real objects are never exactly similar. We show that there exist cases where N particles can be in the same state (Bose–Einstein condensate) although they are not a single entity. The number N of these particles is a measurable quantity, although they are indistinguishable from each other.

12.1.3 Example of ambiguities

In the framework of the principles of chapter 6, this indistinguishability leads to ambiguities in the predictions of physical measurements. Consider, for instance, two identical particles moving in a one-dimensional harmonic potential. We label the particles 1 and 2 and we assume that the Hamiltonian is

$$\hat{H} = \frac{\hat{p}_1^2}{2m} + \frac{1}{2}m\omega^2\hat{x}_1^2 + \frac{\hat{p}_2^2}{2m} + \frac{1}{2}m\omega^2\hat{x}_2^2 = \hat{h}^{(1)} + \hat{h}^{(2)} \ .$$

For simplicity we suppose that the particles have no mutual interaction. Let $(n+1/2)\hbar\omega$ and $\phi_n(x)$ $(n = 0, 1, \ldots)$ be the eigenvalues and eigenfunctions of the one-particle Hamiltonian $\hat{h} = \hat{p}^2/2m + m\omega^2\hat{x}^2/2$.

There is no problem in describing the physical situation where both particles are in the ground state of $\hat{h}$. The corresponding state is

$$\Phi_0(x_1, x_2) = \phi_0(x_1)\,\phi_0(x_2) \ ,$$

and its energy is $E_0 = \hbar\omega$.

On the contrary, the description of the first excited state of the system is ambiguous. This corresponds to one of the particles being in the first excited state of $\hat{h}$ and the other in the ground state. The total energy is $2\hbar\omega$. One possible state is $\phi_1(x_1)\,\phi_0(x_2)$; another possible state is $\phi_0(x_1)\,\phi_1(x_2)$. Because these two states are possible candidates, then, according to the superposition principle, any linear combination:

$$\Phi(x_1, x_2) = \lambda\, \phi_1(x_1)\, \phi_0(x_2) + \mu\, \phi_0(x_1)\, \phi_1(x_2)$$

also corresponds to an energy $2\hbar\omega$.

Therefore, there are several different states that appear to describe the same physical situation. This might not be a problem, provided no measurement could make the difference. Alas this is not true! These various states lead to different predictions concerning physically measurable quantities. Consider, for instance the product of the two positions, the observable $\hat{x}_1\hat{x}_2$, for which the labeling of the two particles is irrelevant. Its expectation value is:

$$\langle x_1\, x_2 \rangle = \frac{\hbar}{m\omega}\, \mathrm{Re}(\lambda^* \mu)\ .$$

This prediction depends on λ and μ. However, nothing in the theory that we have presented tells us the values of these parameters. Therefore there is a basic ambiguity in the predictions of our principles, and we need a prescription in order to fix the values of λ and μ.

It is a remarkable fact of Nature that the only allowed values are $\lambda = \pm\mu$, and that the sign only depends on the nature of the particles under consideration. The allowed states for a system of identical particles are therefore restrictions of the most general states that one could imagine if the particles were distinguishable.

12.2 Systems of two spin 1/2 particles, total spin

The case of two spin 1/2 particles is of particular interest in the following (Pauli principle, entangled states, 21-cm line of hydrogen.) We treat it in an elementary way, without considering the general problem of coupling two arbitrary angular momenta.

12.2.1 The Hilbert space of the problem

Consider a system of two spin 1/2 particles, for instance the electron and the proton in a hydrogen atom, or the two electrons of a helium atom. We call the particles 1 and 2. The Hilbert space of the system is such that it gives access to the probability laws of the two positions and of the two spins.[1] We perform a generalization of what we have already done in the case of a single spin 1/2 particle in three-dimensional space.

12.2.2 Hilbert space of spin variables

We are interested in the space $\mathcal{E}_s$ where we can describe the state of the two spins. $\mathcal{E}_s$ is a four-dimensional space generated by the family

[1] In a sophisticated language it is the tensor product
$\mathcal{E}_H = \mathcal{E}^1_{\text{external}} \otimes \mathcal{E}^1_{\text{spin}} \otimes \mathcal{E}^2_{\text{external}} \otimes \mathcal{E}^2_{\text{spin}}$.

$$\{|++\rangle, |+-\rangle, |-+\rangle, |--\rangle\} \quad . \tag{12.1}$$

A state $|\sigma_1, \sigma_2\rangle$ is by definition an eigenstate of the four observables (which obviously commute) $\hat{\boldsymbol{S}}_1^2$, $\hat{\boldsymbol{S}}_{\boldsymbol{2}}{}^2$, $\hat{S}_{1z}$ and $\hat{S}_{2z}$ with respective eigenvalues $(3/4)\hbar^2$, $(3/4)\hbar^2$, $\sigma_1\hbar/2$, and $\sigma_2\hbar/2$ (σ_1, $\sigma_2 = \pm 1$).

We can define the total spin operator $\hat{\boldsymbol{S}}$ of the system by

$$\hat{\boldsymbol{S}} = \hat{\boldsymbol{S}}_{\boldsymbol{1}} + \hat{\boldsymbol{S}}_{\boldsymbol{2}} \quad . \tag{12.2}$$

Because $\hat{\boldsymbol{S}}_1$ and $\hat{\boldsymbol{S}}_2$ commute, $\hat{\boldsymbol{S}}$ satisfies the commutation relations $\hat{\boldsymbol{S}} \times \hat{\boldsymbol{S}} = i\hbar\hat{\boldsymbol{S}}$, and it is an angular momentum.

The most general state (space + spin) $|\psi\rangle$ of this two spin 1/2 particle system can be written as

$$\begin{aligned} |\psi\rangle = {} & \psi_{++}(\boldsymbol{r}_1, \boldsymbol{r}_2)|++\rangle + \psi_{+-}(\boldsymbol{r}_1, \boldsymbol{r}_2)|+-\rangle \\ & + \psi_{-+}(\boldsymbol{r}_1, \boldsymbol{r}_2)|-+\rangle + \psi_{--}(\boldsymbol{r}_1, \boldsymbol{r}_2)|--\rangle \ . \end{aligned} \tag{12.3}$$

12.2.3 Matrix representation

We can use a matrix representation of the spin states and observables of this system. In the basis $\{|++\rangle, |+-\rangle, |-+\rangle, |--\rangle\}$, a state is represented by a four-component column vector whose components are square integrable functions of $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$.

The observables $\hat{\boldsymbol{S}}_1$ and $\hat{\boldsymbol{S}}_2$ can be written quite easily using the Pauli matrices, and a 2×2 block form for 4×4 matrices

$$\hat{S}_{1x} = \hbar/2 \begin{pmatrix} 0 & \vdots & 1 \\ \cdots & \cdots & \cdots \\ 1 & \vdots & 0 \end{pmatrix} \quad , \quad \hat{S}_{2x} = \hbar/2 \begin{pmatrix} \sigma_x & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & \sigma_x \end{pmatrix} \ ,$$

$$\hat{S}_{1y} = \hbar/2 \begin{pmatrix} 0 & \vdots & -i \\ \cdots & \cdots & \cdots \\ i & \vdots & 0 \end{pmatrix} \quad , \quad \hat{S}_{2y} = \hbar/2 \begin{pmatrix} \sigma_y & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & \sigma_y \end{pmatrix} \ ,$$

$$\hat{S}_{1z} = \hbar/2 \begin{pmatrix} 1 & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & -1 \end{pmatrix} \quad , \quad \hat{S}_{2z} = \hbar/2 \begin{pmatrix} \sigma_z & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & \sigma_z \end{pmatrix} \ .$$

12.2.4 Total spin states

We consider $\mathcal{E}_s$ and we denote $|S, M\rangle$ the eigenstates of $\hat{S}^2$ and $\hat{S}_z$ with respective eigenvalues $S(S+1)\hbar^2$ and $M\hbar$.

Because $\hat{S}_z = \hat{S}_{1z} + \hat{S}_{2z}$, the largest possible value of M is $1/2 + 1/2 = 1$. The corresponding state is unique; it is the state $|++\rangle$. Similarly, the smallest possible value of M is $-1/2 - 1/2 = -1$, and the corresponding eigenstate is $|--\rangle$.

Let us calculate the action of the square of the total spin on these two vectors:

$$\begin{aligned}\hat{S}^2|++\rangle &= \left(\hat{S}_1^2 + \hat{S}_2^2 + 2\hat{\boldsymbol{S}}_{\mathbf{1}} \cdot \hat{\boldsymbol{S}}_{\mathbf{2}}\right)|++\rangle \\ &= \left(3/4\hbar^2 + 3/4\hbar^2 + \hbar^2/2\left(\hat{\sigma}_{1x}\hat{\sigma}_{2x} + \hat{\sigma}_{1y}\hat{\sigma}_{2y} + \hat{\sigma}_{1z}\hat{\sigma}_{2z}\right)\right)|++\rangle \\ &= 2\hbar^2|++\rangle \quad .\end{aligned}$$

Similarly,

$$\hat{S}^2|--\rangle = 2\hbar^2|--\rangle \quad .$$

The two states $|++\rangle$ and $|--\rangle$ are therefore eigenstates of $\hat{S}^2$ with the eigenvalue $2\hbar^2$, which corresponds to an angular momentum equal to 1. With the notations of the first paragraph, we have therefore:

$$\begin{aligned}&|s_1 = 1/2, m_1 = 1/2; s_2 = 1/2, m_2 = 1/2\rangle \\ &= |s_1 = 1/2, s_2 = 1/2; S = 1, M = 1\rangle \ , \\ \text{and} \quad &|s_1 = 1/2,\ m_1 = -1/2; s_2 = 1/2,\ m_2 = -1/2\rangle \\ &= |s_1 = 1/2,\ s_2 = 1/2; S = 1,\ M = -1\rangle \ .\end{aligned}$$

Because we have recognized two states $|S = 1, M = \pm 1\rangle$ of angular momentum 1, we now look for the third one $|S = 1, M = 0\rangle$. In order to do this, we use the general relation found in chapter 9, and we obtain

$$\hat{S}_-|j, m\rangle \quad \propto \quad |j, m-1\rangle \ .$$

We therefore obtain

$$\hat{S}_-|S = 1, M = 1\rangle = \left(\hat{S}_{1-} + \hat{S}_{2-}\right)|++\rangle \quad \propto \quad |-+\rangle + |+-\rangle \ .$$

After normalization, we obtain

$$|S = 1, M = 0\rangle = \frac{1}{\sqrt{2}}\left(|+-\rangle + |-+\rangle\right) \ .$$

One can check that this state is indeed an eigenstate of $\hat{S}^2$ and $\hat{S}_z$ with respective eigenvalues $2\hbar^2$ and 0.

We have identified a three-dimensional subspace in $\mathcal{E}_s$ corresponding to a total angular momentum equal to 1. The orthogonal subspace, of dimension 1, is generated by the vector

$$\frac{1}{\sqrt{2}}\left(|+-\rangle - |-+\rangle\right) \quad .$$

One can readily verify that this vector is an eigenvector of $\hat{S}^2$ and of $\hat{S}_z$ with both eigenvalues equal to zero.

To summarize, the total spin in the particular case $j_1 = j_2 = 1/2$ corresponds to

$$S = 1\,, \quad \text{or} \quad S = 0\,,$$

and the four corresponding eigenstates, which form an eigenbasis of $\mathcal{E}_H$ common to $\hat{\boldsymbol{S}}^2$ and $\hat{S}_z$, are

$$|1, M\rangle : \begin{cases} |1,\ 1\rangle = |++\rangle \\ |1,\ 0\rangle = (|+-\rangle + |-+\rangle)/\sqrt{2} \\ |1,\ -1\rangle = |--\rangle \end{cases}\,, \tag{12.4}$$

$$|0, 0\rangle : \quad |0,\ 0\rangle = (|+-\rangle - |-+\rangle)/\sqrt{2}\,, \tag{12.5}$$

which are called, respectively, the triplet state and the singlet state.

Symmetry properties

The following symmetry properties are important when we consider identical particles and the Pauli principle.

The three states $|1, M\rangle$ are called collectively the *triplet state* of the two-spin system. They are symmetric with respect to the interchange of the z projections of the spins of the two particles, σ_1 and σ_2. The state $|0, 0\rangle$ is called the singlet state and it is antisymmetric in the same exchange. In mathematical terms, if we define a permutation operator $\hat{P}^s_{12}$ in $\mathcal{E}_s$ by the relation

$$\hat{P}^s_{12}|\sigma_1, \sigma_2\rangle = |\sigma_2, \sigma_1\rangle \quad , \tag{12.6}$$

the triplet and singlet states are eigenvectors of this operator:

$$\hat{P}^s_{12}|1, M\rangle = |1, M\rangle\,, \quad \hat{P}^s_{12}|0, 0\rangle = -|0, 0\rangle \quad . \tag{12.7}$$

12.3 Two-particle system; the exchange operator

12.3.1 The Hilbert space for the two-particle system

Within the framework that we have used up to now, we describe a two-particle system (distinguishable or not) by labeling these particles. A state of the system is therefore

$$|\psi\rangle = \sum_{k,n} C_{k,n}\ |1 : k\,;\, 2 : n\rangle. \tag{12.8}$$

Implicitly, the vectors $\{|1 : k\rangle\}$ form a basis of one-particle states. The states $\{|1 : k\,;\, 2 : n\rangle\}$ are factorized states where the first particle has the quantum numbers k and the second n. They form a basis of the two-particle Hilbert space.

12.3.2 The exchange operator between identical particles

The labeling of the particles that we used above has no absolute meaning if they are identical. Consequently, the predictions of experimental results must be independent of this labeling. In order to describe this property due to the exchange symmetry, we introduce the exchange operator $\hat{P}_{12}$ such that for any couple (k, n)

$$\hat{P}_{12}\,|1:k\,;\,2:n\rangle = |1:n\,;\,2:k\rangle\ . \tag{12.9}$$

One can verify that this operator is Hermitian and that it satisfies

$$\hat{P}_{12}^2 = \hat{I}\ . \tag{12.10}$$

Examples

1. For two spinless particles, we have

$$\hat{P}_{12} \equiv \hat{P}_{12}^{(\text{external})}\,, \quad \text{that is,} \quad \hat{P}_{12}\,\Psi(\boldsymbol{r}_1,\boldsymbol{r}_2) = \Psi(\boldsymbol{r}_2,\boldsymbol{r}_1) \quad .$$

2. For two spin 1/2 particles, then $\hat{P}_{12}$ exchanges both the orbital and the spin variables of the two particles:

$$\hat{P}_{12} = \hat{P}_{12}^{(\text{external})}\,\hat{P}_{12}^{(\text{spin})}\ .$$

3. **Permutation of two spin 1/2 particles.** In this case, one can write down explicitly the action of $\hat{P}_{12}$ by using the representation (12.3),

$$\hat{P}_{12}\sum_{\sigma_1,\sigma_2}\Psi_{\sigma_1,\sigma_2}(\boldsymbol{r}_1,\boldsymbol{r}_2)\,|1:\sigma_1;2:\sigma_2\rangle \;=\; \sum_{\sigma_1,\sigma_2}\Psi_{\sigma_1,\sigma_2}(\boldsymbol{r}_2,\boldsymbol{r}_1)\,|1:\sigma_2;2:\sigma_1\rangle$$

where $\sigma_i = \pm$ with $i = 1, 2$.
In order to discuss the properties of this permutation, it is convenient to work with the eigenbasis of the square of the total spin $\hat{\boldsymbol{S}} = \hat{\boldsymbol{S}}_{\boldsymbol{1}} + \hat{\boldsymbol{S}}_{\boldsymbol{2}}$ and $\hat{S}_z$

$$\begin{aligned}
&|S=1,\,m=1\rangle = |1:+\,;\,2:+\rangle\ ,\\
&|S=1,\,m=0\rangle = \frac{1}{\sqrt{2}}\left(|1:+\,;\,2:-\rangle + |1:-\,;\,2:+\rangle\right)\ ,\\
&|S=1,\,m=-1\rangle = |1:-\,;\,2:-\rangle\ ,\\
&|S=0,\,m=0\rangle = \frac{1}{\sqrt{2}}\left(|1:+\,;\,2:-\rangle - |1:-\,;\,2:+\rangle\right)\ .
\end{aligned}$$

We already noticed that

- the triplet states $(S = 1)$ are symmetric under the interchange of σ_1 and σ_2:

$$\hat{P}_{12}^{(\text{spin})}|S=1,m\rangle \;=\; |S=1,m\rangle \quad ,$$

- the singlet state $(S = 0)$ is antisymmetric in this interchange:

$$\hat{P}_{12}^{(\text{spin})}|S=0,m=0\rangle \;=\; -\,|S=0,m=0\rangle \quad .$$

12.3.3 Symmetry of the states

How can one fulfill the requirement that the experimental results must be unchanged as one goes from $|\Psi\rangle$ to $\hat{P}_{12}|\Psi\rangle$? These two vectors must represent the same physical state, therefore they can only differ by a phase factor, that is, $\hat{P}_{12}|\Psi\rangle = e^{i\delta}\ |\Psi\rangle$. Because $\hat{P}_{12}^2 = \hat{I}$, we have $e^{2i\delta} = 1$ and $e^{i\delta} = \pm 1$. Therefore,

$$\hat{P}_{12}|\Psi\rangle = \pm|\Psi\rangle\ . \tag{12.11}$$

We then reach the following conclusion:

The only physically acceptable state vectors for a system of two identical particles are either symmetric or antisymmetric under the permutation of the two particles.

Referring to (12.8), this implies $C_{k,n} = \pm C_{n,k}$. The only allowed states are either symmetric under the exchange of 1 and 2:

$$|\Psi_S\rangle \propto \sum_{k,n} C_{k,n}\left(|1:k\,;\,2:n\rangle + |1:n\,;\,2:k\rangle\right)\ ; \quad \hat{P}_{12}|\Psi_S\rangle = |\Psi_S\rangle \tag{12.12}$$

or antisymmetric:

$$|\Psi_A\rangle \propto \sum_{k,n} C_{k,n}\left(|1:k\,;\,2:n\rangle - |1:n\,;\,2:k\rangle\right)\ ; \quad \hat{P}_{12}|\Psi_A\rangle = -|\Psi_A\rangle\ , \tag{12.13}$$

where the coefficients $C_{k,n}$ are arbitrary.

This restriction to symmetric or antisymmetric state vectors is a considerable step forward in order to solve the ambiguity pointed out in the previous section. For instance, the expectation value $\langle x_1 x_2\rangle$ considered in Section 12.1.2 can now take only two values $\pm\hbar/(2m\omega)$, corresponding to the two choices $\lambda = \pm\mu = 1/\sqrt{2}$. However, it is not yet sufficient, and some essential questions are still open:

1. Can a given species, such as electrons, behave in some experimental situations with the plus sign in (12.11) and in other situations with the minus sign?
2. Assuming that the answer to the first question is negative, what decides which sign should be attributed to a given species?

The answers to these two questions lead us to one of the simplest and most fundamental laws of physics. It is called the Pauli principle, although the general formulation was derived from Pauli's ideas by Fermi and Dirac.

12.4 The Pauli principle

12.4.1 The case of two particles

All particles in Nature belong to one of the two following categories:

- Bosons for which the state vector of two identical particles is always symmetric under the operation $\hat{P}_{12}$.
- Fermions for which the state vector of two identical particles is always antisymmetric under the operation $\hat{P}_{12}$.
- All particles of integer spin (including 0) are bosons (photon, π meson, α particle, gluons, etc.).
- All particles of half-integer spin are fermions (electron, proton, neutron, neutrino, quarks, He^3 nucleus, etc.).

The state vectors of two bosons are of the form $|\Psi_S\rangle$ (Equation (12.12)), those of fermions are of the form $|\Psi_A\rangle$ (Equation (12.13)). The Pauli principle therefore consists of restricting the set of accessible states for systems of two identical particles. The space of physical states is no longer the tensor product of the basis states, but only the subspace formed by the symmetric or antisymmetric combinations.

The Pauli principle also applies to composite particles such as nuclei or atoms, provided experimental conditions are such that they are in the same internal state (be it the ground state or a metastable excited state). For instance, hydrogen atoms in their ground electronic state have a total spin $S = 0$ or $S = 1$ and they behave as bosons. Deuterium, the isotope of hydrogen is a fermion (nuclear spin 1, electron spin 1/2, total spin half-integer).

As we announced, this connection between the symmetry of states and the spin of the particles is an experimental fact. However, it is both a triumph and a mystery of contemporary physics that this property, called the "spin–statistics connection," can be proven starting from general axioms of relativistic quantum field theory. It is a mystery because it is probably the only example of a very simple physical law for which the proof exists but cannot be explained in an elementary way.

Examples

1. The wave function of two identical spin zero particles must be symmetric: $\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2) = \Psi(\boldsymbol{r}_2, \boldsymbol{r}_1)$.
2. The state of two spin 1/2 particles must be of the form

$$|\Psi\rangle = \Psi_{0,0}(\boldsymbol{r}_1, \boldsymbol{r}_2)\,|S = 0, m = 0\rangle + \sum_m \Psi_{1,m}(\boldsymbol{r}_1, \boldsymbol{r}_2)\,|S = 1, m\rangle \ ,$$

where $\Psi_{0,0}$ and $\Psi_{1,m}$ are, respectively, symmetric and antisymmetric:

$$\Psi_{0,0}(\boldsymbol{r}_1, \boldsymbol{r}_2) = \Psi_{0,0}(\boldsymbol{r}_2, \boldsymbol{r}_1)\,, \quad \Psi_{1,m}(\boldsymbol{r}_1, \boldsymbol{r}_2) = -\Psi_{1,m}(\boldsymbol{r}_2, \boldsymbol{r}_1)\ .$$

Therefore, the orbital state and the spin state of two identical fermions are correlated.

12.4.2 Independent fermions and exclusion principle

Consider a situation where two fermions, for instance, two electrons, are independent, that is, they do not interact with each other. The total Hamiltonian then reads $\hat{H} = \hat{h}(1) + \hat{h}(2)$. In such conditions, the eigenstates of $\hat{H}$ are products of eigenstates $|n\rangle$ of $\hat{h}$: $|1 : n\,;\, 2 : n'\rangle$. We remark that, if $n = n'$, that is, if the two particles are in the same quantum state, the state $|1 : n\,;\, 2 : n\rangle$ is necessarily symmetric. This is forbidden by the Pauli principle, which results in the (weaker) formulation:

Two independent fermions in the same system cannot be in the same state.

If $|n\rangle$ and $|n'\rangle$ are orthogonal, the only acceptable state is the antisymmetric combination

$$|\Psi_A\rangle = \frac{1}{\sqrt{2}}\left(|1 : n\,;\, 2 : n'\rangle - |1 : n'\,;\, 2 : n\rangle\right)\ .$$

In this simplified form, the Pauli principle appears as an exclusion principle. This point of view is an approximation because two particles are never completely independent.

12.4.3 The case of N identical particles

In a system of N identical particles, we proceed in a similar manner. We introduce the exchange operator $\hat{P}_{ij}$ of the two particles i and j. The indistinguishability imposes that $\hat{P}_{ij}|\Psi\rangle$ leads to the same physical results as $|\Psi\rangle$. The general form of the Pauli principle is as follows:

> The state vector of a system of N identical bosons is completely symmetric under the interchange of any two of these particles.
>
> The state vector of a system of N identical fermions is completely antisymmetric under the interchange of any two of these particles.

For instance, for $N = 3$, one has

$$\begin{aligned}\Psi_{\pm}(u_1, u_2, u_3) \propto\ & \left(f(u_1, u_2, u_3) + f(u_2, u_3, u_1) + f(u_3, u_1, u_2)\right) \\ & \pm \left(f(u_1, u_3, u_2) + f(u_2, u_1, u_3) + f(u_3, u_2, u_1)\right)\ ,\end{aligned}$$

where f is any function of the three sets of variables u_1, u_2, u_3. The plus sign (resp., minus sign) corresponds to a function Ψ that is completely symmetric (resp., completely antisymmetric).

More generally, let us consider an orthonormal basis $\{|n\rangle\}$ of the one-particle states, and the $N!$ permutations P of a set of N elements. We want to describe the following physical situation, "One particle in the state $|n_1\rangle$, one particle in the state $|n_2\rangle$, $\cdots$, one particle in the state $|n_N\rangle$." In order to do this, we number in an arbitrary way the N particles from 1 to N.

The case of bosons

The state is

$$|\Psi\rangle = \frac{C}{\sqrt{N!}} \sum_P |1 : n_{P(1)} \, ; \, 2 : n_{P(2)} \, ; \, \ldots \, ; \, N : n_{P(N)}\rangle \quad , \tag{12.14}$$

where $\sum_P$ denotes the sum over all permutations. Notice that two (or more) indices $n_i, n_j, \ldots$ labeling the occupied states may coincide. The normalization factor C is expressed in terms of the occupation numbers N_i of the states $|n_i\rangle$:

$$C = (N_1! \, N_2! \ldots)^{-1/2} \quad .$$

The case of fermions

In the case of fermions, the result is physically acceptable if and only if the N states $|n_i\rangle$ are pairwise orthogonal. The state $|\Psi\rangle$ is then

$$|\Psi\rangle = \frac{1}{\sqrt{N!}} \sum_P \varepsilon_P \, |1 : n_{P(1)} \, ; \, 2 : n_{P(2)} \, ; \, \ldots \, ; \, N : n_{P(N)}\rangle \, , \tag{12.15}$$

where ε_P is the signature of the permutation P: $\varepsilon_P = 1$ if P is an even permutation and $\varepsilon_P = -1$ if P is odd. This state vector is often written in the form of a determinant, called the Slater determinant:

$$|\Psi\rangle = \frac{1}{\sqrt{N!}} \begin{vmatrix} |1 : n_1\rangle & |1 : n_2\rangle & \ldots & |1 : n_N\rangle \\ |2 : n_1\rangle & |2 : n_2\rangle & \ldots & |2 : n_N\rangle \\ \vdots & \vdots & & \vdots \\ |N : n_1\rangle & |N : n_2\rangle & \ldots & |N : n_N\rangle \end{vmatrix} \, . \tag{12.16}$$

If two particles are in the same state, two columns are identical and this determinant vanishes.

The set of states that can be constructed using (12.14) (resp., (12.15)) forms a basis of the Hilbert space of an N boson (resp., N fermion) system.

12.5 Physical consequences of the Pauli principle

We give below a few of the many physical consequences of the Pauli principle that concern both few-body systems and the macroscopic properties of a large number of bosons or fermions.

12.5.1 Exchange force between two fermions

Consider the helium atom and neglect magnetic effects, as we did for hydrogen in chapter 10. We label the electrons 1 and 2, and the Hamiltonian is

$$\hat{H} = \frac{\hat{p}_1^2}{2m_e} + \frac{\hat{p}_2^2}{2m_e} - \frac{2e^2}{\hat{r}_1} - \frac{2e^2}{\hat{r}_2} + \frac{e^2}{\hat{r}_{12}}, \quad \text{with} \quad \hat{\boldsymbol{r}}_{12} = \hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_2 \quad .$$

The eigenvalue problem is technically complicated and can only be solved numerically, but the results of interest here are simple (Figure 12.2). The ground state ($E_0 = -78.9$ eV) corresponds to a symmetric spatial wave function, whereas the first two excited states $E_{1A} = -58.6$ eV and $E_{1S} = -57.8$ eV have, respectively, antisymmetric and symmetric spatial wave functions. The symmetry of the wave function implies a specific symmetry of the spin state: E_0 and E_{1S} are singlet spin states, and E_{1A} is a triplet spin state. In the ground state, the two spins are antiparallel. In order to flip one of them to make them parallel, one must spend a considerable amount of energy ($\sim$20 eV).

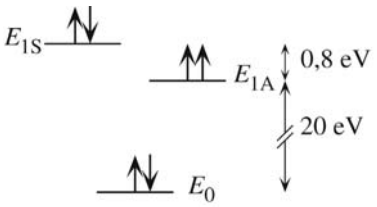


Fig. 12.2. The first three levels of the helium atom.

This corresponds to a "force" that maintains the spins in the antiparallel state. It is not a magnetic coupling between the spins: this magnetic interaction can be calculated, and it corresponds to an energy of the order of 10^{-2} eV. The "force" that we are facing here has an electrostatic origin, the Coulomb interaction, and it is transferred into a constraint on the spins via the Pauli principle. Such an effect is called an *exchange interaction*. The same effect is the basic cause of ferromagnetism.

12.5.2 The ground state of N identical independent particles

Consider N identical independent particles. The Hamiltonian is therefore the sum of N one-particle Hamiltonians:

$$\hat{H} = \sum_{i=1}^{N} \hat{h}^{(i)} \quad . \tag{12.17}$$

Let $\{\phi_n, \varepsilon_n\}$ be the eigenfunctions and corresponding eigenvalues of $\hat{h}$: $\hat{h}\phi_n = \varepsilon_n \phi_n$, where we assume that the ε_n are ordered: $\varepsilon_1 \leq \varepsilon_2 \cdots \leq \varepsilon_n \cdots$.

From the previous considerations, we see that the ground state energy of a system of N bosons is

$$E_0 = N\varepsilon_1 \ ,$$

whereas for a system of fermions, we have

$$E_0 = \sum_{i=1}^{N} \varepsilon_i \quad .$$

In this latter case, the highest occupied energy level is called the *Fermi energy* of the system and it is denoted ϵ_F. The occupation of the states ϕ_n is represented on Figure 12.3 both for a bosonic and a fermionic assembly.

Consider, for instance, N independent fermions of spin s confined in a cubic box of size L. We choose here a basis of states corresponding to periodic boundary conditions. Each eigenstate of the Hamiltonian $\hat{h}$ is a plane wave $\phi_{\boldsymbol{p}}(\boldsymbol{r}) = e^{i\boldsymbol{p}\cdot\boldsymbol{r}/\hbar}/\sqrt{L^3}$, associated with one of the $2s+1$ spin states corresponding to a well-defined component $m_s\hbar$ of the spin on a given axis ($m_s = -s, -s+1, \ldots, s$). The momentum $\boldsymbol{p}$ can be written $\boldsymbol{p} = (2\pi\hbar/L)\boldsymbol{n}$, where the vector $\boldsymbol{n} = (n_1, n_2, n_3)$ stands for a triplet of positive or negative integers. The ground state of the N fermion system is obtained by placing $2s+1$ fermions in each plane wave $\phi_{\boldsymbol{p}}$, as long as $|\boldsymbol{p}|$ is lower than the Fermi momentum p_F. This Fermi momentum p_F is determined using

$$N = \sum_{\boldsymbol{p}\ (p<p_F)} (2s+1) \ .$$

For a large number of particles, we can replace this discrete sum by an integral that yields

$$N \simeq (2s+1)\frac{L^3}{(2\pi\hbar)^3}\int_{p<p_F} d^3p = \frac{2s+1}{6\pi^2}\ \frac{L^3 p_F^3}{\hbar^3} \ . \tag{12.18}$$

This equation relates the density $\rho = N/L^3$ of the gas and the Fermi momentum, independently of the size of the box:

$$\rho = \frac{2s+1}{6\pi^2}\left(p_F/\hbar\right)^3 \ . \tag{12.19}$$

The average kinetic energy per particle can also be easily calculated:

$$\frac{\langle p^2\rangle}{2m} = \frac{1}{N}\sum_{\boldsymbol{p}\ (p<p_F)} (2s+1)\frac{p^2}{2m} \simeq \frac{2s+1}{N}\frac{L^3}{(2\pi\hbar)^3}\int_{p<p_F}\frac{p^2}{2m}d^3p \ ,$$

which leads to

$$\frac{\langle p^2 \rangle}{2m} = \frac{3}{5} \frac{p_F^2}{2m} \ . \tag{12.20}$$

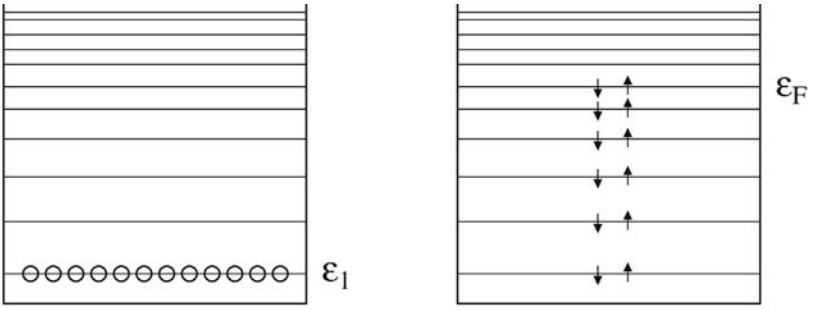


Fig. 12.3. Ground state of a system of N independent identical particles. Left: bosonic case, with all the particles in the ground state of the one-body Hamiltonian. Right: fermionic case, where the first $N/(2s+1)$ states of the one-body Hamiltonian are occupied (here $s = 1/2$).

12.5.3 Behavior of fermion and boson systems at low temperatures

The difference between the ground states of N-fermion or N-boson systems induces radically different behaviors of such systems at low temperature.

In a system of fermions at zero temperature and in the absence of interactions, we have just seen that all the energy levels of the one-body Hamiltonian are filled up to the Fermi energy ϵ_F. This simple model describes the conduction electrons in a metal remarkably well, and it accounts for many macroscopic properties of a solid, such as its thermal conductibility. Using the result (12.19), the Fermi energy $\epsilon_F = p_F^2/2m_e$ can be written in terms of the number density ρ_e of conduction electrons,

$$\epsilon_F = \frac{\hbar^2 \left(3\rho_e \pi^2\right)^{2/3}}{2m_e} \quad ,$$

where we used $2s + 1 = 2$ for electrons. This energy can reach large values ($\epsilon_F = 3$ eV for sodium). This is much larger than the thermal energy at room temperature ($k_B T \simeq 0.025$ eV). This explains the success of the zero-temperature fermion gas model for conduction electrons. At room temperature, very few electrons participate in thermal exchanges.

The application of the Pauli principle to fermionic systems has many consequences, ranging from solid-state physics to the stability of stars such as white dwarfs or neutron stars. In nuclear physics, the Pauli principle explains

why neutrons are stable under β decay inside nuclei. An isolated neutron is unstable and decays through the process $n \to p + e + \bar{\nu}$ with a lifetime of the order of 15 minutes. Inside a nucleus, a neutron can be stabilized if all the final states allowed by energy conservation for the final proton are already occupied. Therefore, the decay cannot occur because of energy conservation.

Concerning boson systems, a spectacular consequence of quantum statistics is the Bose–Einstein condensation. In the absence of interactions between the particles, if the number density $\rho = N/V$ is such that

$$\rho \Lambda_T^3 > 2.612\,, \text{ with } \quad \Lambda_T = \frac{h}{\sqrt{2\pi m k_B T}} \quad , \tag{12.21}$$

a macroscopic accumulation of particles occurs in a single quantum state, the ground state of the confining potential of the particles. This "gregarious" behavior of bosons contrasts with the "individualistic" character of fermions.

Until June 1995, the usual example of a Bose–Einstein condensation that was given in textbooks was the transition normal liquid $\to$ superfluid liquid of helium which happens at a temperature of $T = 2.17$ K. However, the complicated interactions inside the liquid make the quantitative treatment of the superfluid transition quite involved, and different from the simple theory of the Bose–Einstein condensation of an ideal gas.

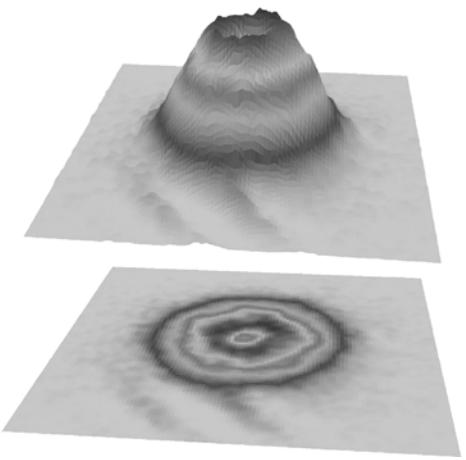


Fig. 12.4. Bose–Einstein condensate of a gas of rubidium 87 atoms. The atoms are confined in a harmonic isotropic trap in the xy-plane, and they are cooled by evaporation. This condensate (approximately 1 million atoms) has a vortex structure. It does not correspond to the ground state, but to a state immediately above of angular momentum $\ell = 1$. The two pictures are obtained by shining a laser beam on the condensate in the (xy)-plane and perpendicular to this plane. (Photos: F. Chevy, K. Madison, and J. Dalibard, ENS Paris.)

We now have at our disposal experiments[2] performed on gases of alkali atoms (lithium, sodium, rubidium) that are initially cooled by laser inside a vacuum chamber with a very low residual pressure (below 10^{-11} mBar). The atoms are then confined by an inhomogeneous magnetic field and they are cooled down further by evaporation until they reach Bose–Einstein condensation, at a temperature below 1 microKelvin. The evaporative cooling technique consists of eliminating the more energetic atoms, in order to keep only the slower ones. The collisions between the trapped atoms permanently maintain the thermal equilibrium. Starting from 10^9 atoms, one can obtain, after evaporation, a situation where the 10^6 remaining atoms are practically all in the ground state of the system. Figure 12.4 shows such a condensate of rubidium 87 atoms, confined in a magnetic trap and cooled down to the condensation point. The difference with the first Bose–Einstein condensates is that it has a vortex structure. It does not correspond to the ground state, but to a state immediately above, of angular momentum $\ell = 1$. These Bose–Einstein condensates possess remarkable coherence and superfluid properties, and this has been a very active field of research in recent years.

[2] See M.H. Anderson et al, *Science* **269**, 198 (1995), where the first Bose–Einstein condensate is shown in an experiment carried out with rubidium atoms.

13

Entangled states: The way of paradoxes

At the end of chapter 3, we mentioned Einstein's revolt against the probabilistic aspect of quantum mechanics and the uncertainty relations. As we said, Einstein was worried about two aspects. One is the notion of a complete description of reality. He thought that a complete description is possible in principle, and that the probabilistic description is simply easier to handle. The other aspect is the notion of determinism: same causes produce same effects. Einstein wrote the worldwide famous:

> The theory produces a good deal but hardly brings us closer to the secret of the Old One. I am at all events convinced that *He* does not play dice.[1]

which is usually contracted to "God does not play dice." Because this theory works, it must be an intermediate step toward an underlying theory, which would involve "hidden variables"[2] that are, at present, not accessible, and on which we average in the present version of quantum mechanics.

13.1 The EPR paradox

In 1935, Einstein, Podolsky, and Rosen pointed out, in a celebrated paper,[3] a paradox that has stirred the world of physics since then. Its starting point is a "gedanken experiment."

The original version is relatively simple. Consider a particle that decays into two particles. One of them comes in my direction, the other in yours. One can manage to make the momentum $\boldsymbol{P} = \boldsymbol{p}_1 + \boldsymbol{p}_2$ of the initial particle,

[1] "Die Theorie liefert viel, aber dem Geheimnis des Alten bringt sie uns kaum näder. Jedenfalls bin ich überzeugt, dass *Der* nicht würfelt."

[2] Again, Einstein never used that word.

[3] A. Einstein, B. Podolsky, and N. Rosen, "Can quantum-mechanical description of physical reality be considered complete?" *Phys. Rev.* **47**, 777 (1935).

that is, the total momentum, as well defined and as small as one wishes. For simplicity, we assume it vanishes. We therefore do not know exactly where the particle is, but this is not a problem; we assume we use large enough detectors. Similarly, we can measure with all wanted accuracy the relative position of the final particles $\boldsymbol{r} = \boldsymbol{r}_1 - \boldsymbol{r}_2$, inasmuch as $\boldsymbol{P}$ and $\boldsymbol{r}$ commute.

On the other hand, one can prove quite generally that in the decay, there is conservation of the total momentum. If $\boldsymbol{P} = 0$, the two final particles have exactly opposite momenta.

Let's check this experimentally. On a large number N of decays (all of which are such that both counters are activated), we place detectors sufficiently far from each other so that no information can be transmitted between both detectors when they measure their respective signals. This amounts to saying that the relative position is large enough. We proceed and perform, say, 1000 measurements. We notice afterwards, by comparing the set of time-ordered data, that systematically if you find a value $\boldsymbol{p}$ of the momentum of your particle, I always find $-\boldsymbol{p}$ for my particle. We have checked momentum conservation. We become convinced that we can repeat the operation as many times as we wish; we will always come to the same conclusion.

After a while, say event number 1001, I'm fed up and I let you measure the momentum of your particle while I measure the position of mine with as great an accuracy as I wish. After that measurement, I call you on the phone; you tell me the value $\boldsymbol{p}_1$ of the momentum that you have found. Therefore, I know exactly both the position $\boldsymbol{r}_1$, measured by me and the momentum $-\boldsymbol{p}_1$, measured by you, of my particle with as great an accuracy as I wish.

Notice, and it is important, that this scheme works because I have used a system such that I have information on the quantum state of the *set* of the two particles in order to obtain physical information from you on my particle.

One says that the state of the particles is entangled or correlated. This means that information on one of the particles is directly connected to information on the other, wherever they are located in space.

Actually, this is the central point in a key statement of the EPR article:

> If, without in any way disturbing a system, we can predict with certainty (i.e. with a probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity.

Notice the implicit and fundamental role of locality in the argument. Your measurement, which is done at a time and distance such that one cannot transmit any information to my particle, cannot by any means perturb the results of my measurements results. No reasonable definition of reality can allow the opposite claim Einstein, Podolsky, and Rosen.

Their conclusion is that the description of reality by quantum mechanics is not complete.

13.2 The version of David Bohm

What leads to measurable criteria is the version of the EPR argument given by David Bohm in 1952. This presentation is more convenient to explain and to treat mathematically than the initial version, although it is basically equivalent in its spirit.

Suppose we prepare two spin 1/2 particles a and b in the singlet spin state:

$$|\Psi_s\rangle = \frac{1}{\sqrt{2}} \left(|a : +z \,;\, b : -z\rangle - |a : -z \,;\, b : +z\rangle \right) \ . \tag{13.1}$$

Particle a is detected by Alice, who measures the component of its spin along an axis of unit vector $\boldsymbol{u}_a$ (Figure 13.1); similarly, particle b is detected by Bob who measures its spin component along an axis of unit vector $\boldsymbol{u}_b$.

If they both agree to make only one measurement of the projection of the spin of their particles along a given axis, each of them has a probability 1/2 to find $+\hbar/2$ or $-\hbar/2$.

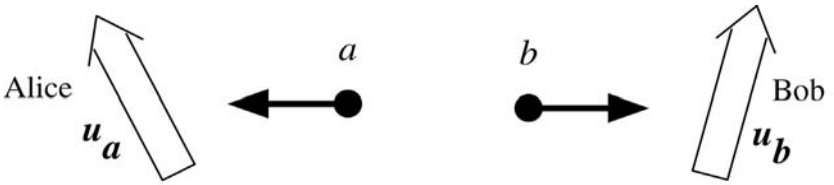


Fig. 13.1. Gedanken experiment corresponding to the EPR argument. Two spin 1/2 particles a and b are prepared in the singlet state. Alice measures the component of the spin of particle a along an axis $\boldsymbol{u}_a$. Bob measures the component of the spin of particle b along an axis $\boldsymbol{u}_b$.

Here, the law of interest is angular momentum conservation. The total angular momentum is zero in the singlet state. All its components are zero.

Let us assume first that Alice and Bob both choose the same axis z to do their measurements: $\boldsymbol{u}_a = \boldsymbol{u}_b = \boldsymbol{e}_z$. The argument presented in the introduction applies: with a probability 1/2, Alice will find $+\hbar/2$ and Bob will find $-\hbar/2$, and with the same probability 1/2, Alice will find $-\hbar/2$ and Bob will find $+\hbar/2$. Alice and Bob can never obtain the same result. There is a perfect correlation, or rather anticorrelation, of the two results.

Technically, this can be seen by writing the singlet state in another basis, for instance, the basis of spin eigenstates along the x-axis. In the eigenbasis of the observables $\hat{S}_{ax}$, $\hat{S}_{bx}$, using

$$|\pm z\rangle = \frac{1}{\sqrt{2}} \left(|+x\rangle \pm |-x\rangle \right) \ , \tag{13.2}$$

the singlet state (13.1) is written as

$$|\Psi_s\rangle = \frac{1}{\sqrt{2}} \left(|a : +x \,;\, b : -x\rangle - |a : -x \,;\, b : +x\rangle \right) \ . \tag{13.3}$$

The form of $|\Psi_s\rangle$ in this new basis is the same as previously. This remains true if one uses an arbitrary axis $\boldsymbol{\alpha}$,

$$|\Psi_s\rangle = \frac{1}{\sqrt{2}} \left(|a : +\alpha \,;\, b : -\alpha\rangle - |a : -\alpha \,;\, b : +\alpha\rangle \right) \ . \tag{13.4}$$

Such correlations appear frequently in daily life. Suppose we have two cards, one is red and the other one is yellow. We place each of them in a sealed envelope, we mix the envelopes in a closed box, and we give one of them to Alice and the other one to Bob. When Alice opens her envelope, she sees the color of her card (red with a probability 1/2, yellow with a probability 1/2). There is obviously a perfect anticorrelation with Bob's subsequent result. If Alice's card is red, Bob's card is yellow and vice versa. There is no paradox in these anticorrelations: the fact that Alice looks at the color of her card does not affect the color of Bob's card.

According to the EPR claim above, there is an element of physical reality associated with the color of Bob's card, because, without perturbing it in any manner, one can determine the color of this card by simply asking Alice what her result is. Similarly, there is an element of physical reality associated with the component S_{bz}, because without perturbing in any manner particle b, one can determine the value of S_{bz} that one would measure in an experiment: it is sufficient to ask Alice to measure the component S_{az} and to tell Bob the result. If Alice finds $+\hbar/2$, Bob is sure to find $-\hbar/2$ by measuring S_{bz}, and vice versa.

Actually, the EPR argument goes one step further. Because we can transpose the argument about the z-axis to the x-axis, there must also exist an element of physical reality associated with the component S_{bx} of particle b. If one prepares a two-particle system in the singlet state, Bob can determine the component S_{bx} without "touching" particle b. It is sufficient to ask Alice to measure the component S_{ax} and to tell him her result. Although the term "element of physical reality" is quite vague up to this point, we feel that we are reaching dangerous grounds. In fact, the observables $\hat{S}_{bx}$ and $\hat{S}_{bz}$ do not commute. How can they simultaneously possess this element of physical reality? (Notice that in this debate, no attention is paid to what one could do with that information experimentally in subsequent operations.)

Obviously, the above argument is contrary to the basic principles of quantum mechanics. When particles a and b are in an entangled state, such as the singlet state, it is risky to claim that one doesn't "act" on particle b when performing a measurement on a. Taken separately, particles a and b are not in well-defined states; only the global system $a + b$ is in a well defined quantum mechanical state. It is only for factorized states, of the type

$$|\Phi\rangle = |a : +\alpha \,;\, b : -\alpha\rangle \equiv |a : +\alpha\,\rangle |b : -\alpha\rangle \tag{13.5}$$

that the EPR argument can be applied safely. However, in that case there is no paradox: a measurement on a gives no information on a measurement that would be performed on b.

At this stage, we can have either of the following attitudes. We can stick to the quantum description that bears this paradoxical nonlocal character: the two particles a and b, as far as they may be from each other (a on Earth, b on the moon), do not have individual realities when their spin state is an entangled state. It is only after Alice (on the Earth) has measured S_{az} that the quantity S_{bz} (for the particle on the moon) acquires a well-defined value.[4] But if a measurement on Earth instantaneously affects a measurement on the moon, there is something we do not really understand in the theory.

On the contrary, we can adopt the point of view of Einstein, and hope that some day one will find a more "complete" theory than quantum mechanics. In that theory, the notion of locality will have the same meaning as it has in classical physics, and so will the notion of reality.

13.2.1 Bell's inequality

In 1964, John Bell, an Irish physicist working at CERN, made a decisive theoretical breakthrough. This allowed carrying this debate between two radically antagonistic conceptions of the physical world onto experimental grounds.

Bell's formulation is the following. Suppose the supertheory that Einstein was hoping for exists. It will involve for any pair (a, b) of the EPR problem described above, a parameter λ that determines completely the results of the measurements of Alice and Bob. For the moment, we know nothing about the parameter λ, which is absent in an orthodox quantum description.

We denote Λ the manifold in which the parameter λ evolves. In the supertheory framework, there must exist a function $A(\lambda, \boldsymbol{u}_a) = \pm\hbar/2$ for Alice and a function $B(\lambda, \boldsymbol{u}_b) = \pm\hbar/2$ for Bob, which give the results of their measurements. These results therefore depend on the value of λ: for instance, if λ pertains to some subset $\Lambda_+(\boldsymbol{u}_a)$, then $A(\lambda, \boldsymbol{u}_a) = \hbar/2$; if λ is in the complementary subset $\Lambda - \Lambda_+(\boldsymbol{u}_a)$, then $A(\lambda, \boldsymbol{u}_a) = -\hbar/2$. Locality plays a crucial role in the previous assumptions. In fact, we have assumed that the function A depends on the value of λ and on the direction of analysis $\boldsymbol{u}_a$ chosen by Alice, but not on the direction of analysis $\boldsymbol{u}_b$ chosen by Bob.

The parameter λ of the supertheory varies from one pair (a, b) to another, whereas in quantum mechanics, all pairs are prepared in the same state $|\Psi_s\rangle$ and nothing can make any difference between them. This parameter is therefore not accessible to a physicist who uses quantum mechanics: it is a hidden variable. All the beauty of Bell's argument is to prove that there exist strong

[4] One can check that this formulation does not allow the instantaneous transmission of information. In order to see the correlations with Alice's result, Bob must ask Alice what her result is, and the corresponding information travels (at most) at the velocity of light.

constraints on the theories with local hidden variables, and that these constraints can be established without any further assumptions than the ones given above. Notice that all correlations encountered in daily life can be described in terms of hidden variable theories. On the previous example of cards with different colors, the hidden variable comes from the shuffling of the cards. If a careful observer memorizes the motion of the cards in this shuffling, they can predict with probability 1 the result of Alice (red or yellow) and that of Bob (yellow or red).

In order to get to Bell's result, we introduce the correlation function $E(\boldsymbol{u}_a, \boldsymbol{u}_b)$.

Quite generally, if we consider two random variables x and y, corresponding to a probability density $p(x, y)$, one defines the linear correlation coefficient $r(x, y)$ by

$$r = \frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\Delta x \, \Delta y} \quad .$$

One can check that if the two variables are correlated, that is $x = ay + b$, then $r = \pm 1$ (according to the sign of a) and, if they are independent, that is, $p(x, y) = p_1(x)p_2(y)$, then $r = 0$.

This function $E(\boldsymbol{u}_a, \boldsymbol{u}_b)$ is equal to the expectation value of the product of the results of Alice and Bob, for given directions of analysis $\boldsymbol{u}_a$ and $\boldsymbol{u}_b$, divided by $\hbar^2/4$ in order to have a dimensionless quantity. Whatever the underlying theory, we have the following property:

$$|E(\boldsymbol{u}_a, \boldsymbol{u}_b)| \leq 1 \ . \tag{13.6}$$

Indeed, for each pair the product of Alice's and Bob's results is $\pm\hbar^2/4$.

For a hidden variable theory, the function $E(\boldsymbol{u}_a, \boldsymbol{u}_b)$ can be written as

$$E(\boldsymbol{u}_a, \boldsymbol{u}_b) = \frac{4}{\hbar^2} \int \mathcal{P}(\lambda) \, A(\lambda, \boldsymbol{u}_a) \, B(\lambda, \boldsymbol{u}_b) \, \mathrm{d}\lambda \ , \tag{13.7}$$

where the function $\mathcal{P}(\lambda)$ describes the (unknown) distribution law of the variable λ. The only constraints on $\mathcal{P}$ are

$$\text{for any } \lambda \, , \quad \mathcal{P}(\lambda) \geq 0 \, , \quad \text{and} \int \mathcal{P}(\lambda) \, \mathrm{d}\lambda = 1 \ . \tag{13.8}$$

Here, we assume that the function $\mathcal{P}(\lambda)$ does not depend on the directions of analysis $\boldsymbol{u}_a$ and $\boldsymbol{u}_b$. Indeed these directions can be chosen by Alice and by Bob after the pair with hidden parameter λ has been prepared.

In the framework of quantum mechanics, one can check that the value of the function $E(\boldsymbol{u}_a, \boldsymbol{u}_b)$ is

$$E(\boldsymbol{u}_a, \boldsymbol{u}_b) = \frac{4}{\hbar^2} \langle \Psi_s | (\hat{\boldsymbol{S}}_a . \boldsymbol{u}_a) \, (\hat{\boldsymbol{S}}_b . \boldsymbol{u}_b) | \Psi_s \rangle = -\boldsymbol{u}_a . \boldsymbol{u}_b \ . \tag{13.9}$$

In polar coordinates, in the (x, z)-plane where the y-axis is the axis between Alice and Bob, one has

$$E(\boldsymbol{u}_a, \boldsymbol{u}_b) = -\cos(\alpha - \beta) \ ; \tag{13.10}$$

α and β are the polar angles of $\boldsymbol{u}_a$ and of $\boldsymbol{u}_b$ in this plane. Bell's theorem can be stated in the following way:

Theorem 8. *(1) For a local hidden variable theory, the quantity*

$$S = E(\boldsymbol{u}_a, \boldsymbol{u}_b) + E(\boldsymbol{u}_a, \boldsymbol{u}'_b) + E(\boldsymbol{u}'_a, \boldsymbol{u}'_b) - E(\boldsymbol{u}'_a, \boldsymbol{u}_b) \tag{13.11}$$

always satisfies the inequality

$$|S| \leq 2 \ . \tag{13.12}$$

(2) This inequality can be violated by the predictions of quantum mechanics.

We first prove the inequality satisfied by hidden variable theories. We introduce the quantity

$$\begin{aligned} \mathcal{S}(\lambda) = {} & A(\lambda, \boldsymbol{u}_a)\ B(\lambda, \boldsymbol{u}_b) + A(\lambda, \boldsymbol{u}_a)\ B(\lambda, \boldsymbol{u}'_b) \\ & + A(\lambda, \boldsymbol{u}'_a)\ B(\lambda, \boldsymbol{u}'_b) - A(\lambda, \boldsymbol{u}'_a) B(\lambda, \boldsymbol{u}_b) \ , \end{aligned}$$

which enters into the definition of S:

$$S = \frac{4}{\hbar^2} \int \mathcal{P}(\lambda)\ \mathcal{S}(\lambda)\ d\lambda \ .$$

This quantity $\mathcal{S}(\lambda)$ can be rewritten as

$$\begin{aligned} \mathcal{S}(\lambda) = {} & A(\lambda, \boldsymbol{u}_a)\ \left(B(\lambda, \boldsymbol{u}_b) + B(\lambda, \boldsymbol{u}'_b)\right) \\ & + A(\lambda, \boldsymbol{u}'_a)\ \left(B(\lambda, \boldsymbol{u}'_b) - B(\lambda, \boldsymbol{u}_b)\right) \ , \end{aligned} \tag{13.13}$$

which is always equal to $\pm\hbar^2/2$. Indeed the quantities $B(\lambda, \boldsymbol{u}_b)$ and $B(\lambda, \boldsymbol{u}'_b)$ can only take the two values $\pm\hbar/2$. Therefore they are either equal or opposite. In the first case, the second line of (13.13) vanishes, and the first one is equal to $\pm\hbar^2/2$. In the second case, the first line of (13.13) vanishes, and the second term is $\pm\hbar^2/2$. We then multiply $\mathcal{S}(\lambda)$ by $\mathcal{P}(\lambda)$ and we integrate over λ in order to obtain the inequality for which we were looking.

Concerning the second point of Bell's theorem, it suffices to find an example for which the inequality (13.12) is explicitly violated. Consider the vectors $\boldsymbol{u}_a$, $\boldsymbol{u}'_a$, $\boldsymbol{u}_b$, and $\boldsymbol{u}'_b$ represented in Figure 13.2:

$$\boldsymbol{u}_b.\boldsymbol{u}_a = \boldsymbol{u}_a.\boldsymbol{u}'_b = \boldsymbol{u}'_b.\boldsymbol{u}'_a = -\boldsymbol{u}_b.\boldsymbol{u}'_a = \frac{1}{\sqrt{2}} \ . \tag{13.14}$$

Using (13.9), we find

$$S = -2\sqrt{2} \ , \tag{13.15}$$

which obviously violates the inequality (13.12).

After this remarkable step forward due to Bell, which transformed a philosophical discussion into an experimental problem, experimentalists had to find

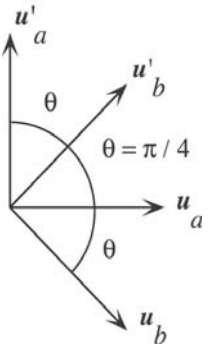


Fig. 13.2. Choice of directions of measurements of Alice and Bob that leads to a violation of Bell's inequality.

the answer. Is quantum mechanics always right, even for a choice of angles such as in Figure 13.2, which would eliminate any realistic and local supertheory, or, on the contrary, are there experimental situations where quantum mechanics can be falsified, which would allow for a more complete theory, as Einstein advocated?

13.2.2 Experimental tests

The first experimental attempts to find a violation of Bell's inequality started at the beginning of the 1970s. These experiments were performed on photon pairs rather than on spin 1/2 particles, because it is experimentally simpler to produce a two-photon entangled state of the type (13.1).

The previous argument can be transposed with no difficulty to photon pairs. The spin states $|+z\rangle$ and $|-z\rangle$ are replaced by the polarization states of the photon $|\uparrow\rangle$ and $|\rightarrow\rangle$, corresponding to vertical and horizontal polarizations. The states $|+x\rangle$ and $|-x\rangle$, which are symmetric and antisymmetric combinations of $|\pm : z\rangle$, are replaced by photon states linearly polarized at ± 45 degrees from the vertical direction:

$$|\nearrow\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow\rangle + |\rightarrow\rangle\right) , \quad |\nwarrow\rangle = \frac{1}{\sqrt{2}}\left(-|\uparrow\rangle + |\rightarrow\rangle\right) . \tag{13.16}$$

The first experimental tests, between 1970 and 1975 in the United States and in Italy, led to contradictory results concerning the violation of Bell's inequality. The experiments of Fry and Thomson in Texas in 1976, and particularly those of Aspect and his group in Orsay between 1980 and 1982, led to the first undeniable violation of Bell's inequality in a situation close to the gedanken experiment presented above.[5] The experiments of Aspect use pairs

[5] A. Aspect, P. Grangier, and G. Roger, *Phys. Rev. Lett.* **49**, 91 (1982); A. Aspect, J. Dalibard, and G. Roger, *Phys. Rev. Lett.* **49**, 1804 (1982).

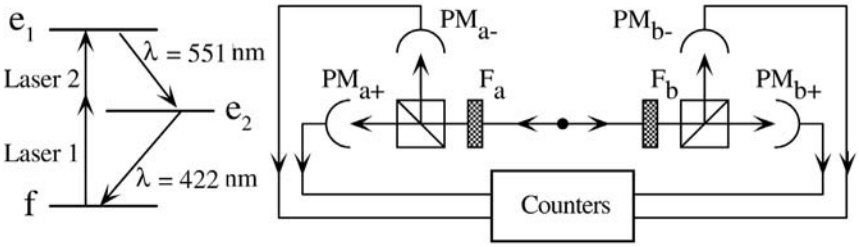


Fig. 13.3. Left: Levels of atomic calcium used in order to produce photon pairs with correlated polarizations. Right: The photons a and b are first filtered in frequency: F_a transmits photons a and stops photons b, and vice versa for F_b. They are then detected on photomutipliers PM_{a+}, PM_{a-}, PM_{b+}, and PM_{b-}. The polarizing cubes are analogues of Stern–Gerlach devices. They transmit photons with the polarization $|\uparrow\rangle$ toward the detectors PM_{a+}, PM_{b+}, and they deflect photons with polarization $|\rightarrow\rangle$ toward the detectors PM_{a-}, PM_{b-}. Because the photons are emitted nearly isotropically, only a small fraction ($\sim 10^{-5}$) of the emitted pairs is actually used.

of photons emitted in an atomic cascade of calcium atoms (Figure 13.3). These calcium atoms are prepared by lasers in an excited state e_1. This excited state has a lifetime of 15 ns, and decays toward an excited state e_2 by emitting a photon a, of wavelength $\lambda_a = 551$ nm. This latter level e_2 has a lifetime of 5 ns, and decays itself to the ground state f by emitting a second photon b, of wavelength $\lambda_b = 422$ nm. The initial level e_1 and the final level f have zero angular momentum, whereas the intermediate level e_2 has angular momentum 1. Under these conditions, one can show that the polarization state of the emitted photon pair is

$$|\Psi_p\rangle = \frac{1}{\sqrt{2}} \left(|a :\uparrow ; b :\uparrow\rangle + |a :\rightarrow ; b :\rightarrow\rangle \right) \ . \tag{13.17}$$

This entangled state leads to the same type of correlations as the singlet spin state considered above. The transposition of Bell's argument shows that some quantity S', involving correlation functions between the polarizations of detected photons, must verify $|S'| \leq 2$ for any local hidden variable theory. The Orsay result, $S' = 2.697 \pm 0.015$ violates this inequality, but it is in agreement with the quantum mechanical prediction $S' = 2.70$. Therefore the universal realistic local super-theory which was supposed to replace quantum mechanics, as Einstein believed, cannot exist, at least for this system. Physicists must learn to live with the genuine indeterminism of quantum mechanics.

The truth is that many physicists were disappointed with this results. They were of course hoping that the proof of the existence of a supertheory would give exciting research themes. But what is more disconcerting is that the result confirms the wave packet reduction. Something does happen instantaneously at a distance. We were discussing this with my great friend Albert Messiah on February 28, 2001, and, he told me

> Entangled states and Bell's inequality are what make quantum mechanics unbearable! When we observe long distance entanglement, we are not sure that we really understand what's going on. When Newton faced the problem of instantaneous action at a distance, he called upon God's finger. Quantum entanglement causes a similar reaction. Bell, with his inequalities, did hope to falsify quantum mechanics. The problem with that theory, is that it always manages to accommodate any situation, even with consequences that seem unbearable for us!

Contrary to what many people say, Einstein was quite right when he thought and said that the interpretation of quantum mechanics causes problems. It is the local hidden variables assumption that is contradicted by experiment.

13.3 Quantum cryptography; how to enjoy a nuisance

The ambition of cryptography is to transmit a message from an issuer (Alice) to a receiver (Bob) and to minimize the risk that a spy may intercept and decipher the message. In order to do that, classical cryptography uses sophisticated methods that cannot be "broken" in a reasonable amount of time, with the present capacities of computers. Quantum cryptography is based on a somewhat different principle. It allows Alice and Bob to make sure no spy has intercepted the message before actually sending it!

The principle of this technique is to profit from the fact that in quantum mechanics a measurement perturbs the state of the system, in particular with entangled states. One therefore devises a procedure that will prove the existence of a spy before sending the actual message!

13.3.1 The communication between Alice and Bob

A message can always be coded in binary language, that is, by a succession of 0s and 1s. Each number, 0 or 1, represents a piece of elementary information, or *bit*. In order to transmit her message, we assume that Alice sends Bob a beam of spin 1/2 particles in a well-controlled sequence, and that Bob detects these particles one after the other in a Stern–Gerlach type of apparatus. Each particle carries a bit coded through its spin state.

Suppose first that Alice sends each particle in the state $|+z\rangle$ or $|-z\rangle$. By convention, $|+z\rangle$ represents the value 1 and $|-z\rangle$ the value 0. Bob orients his Stern–Gerlach apparatus also along the z-axis, he measures the spin states of the particles that arrive, and he reconstructs Alice's message. Such a procedure has no quantum feature and it is simple to spy upon. The spy just sits between Alice and Bob and he places his own Stern–Gerlach apparatus in the z-direction. He measures the spin state of each particle, and

re-emits it toward Bob in the same spin state. He therefore reads the message and neither Alice nor Bob can detect his presence.

The situation changes radically if Alice chooses at random, for each of the particles she is sending, one of the four states $|+\,z\rangle$, $|-\,z\rangle$, $|+\,x\rangle$, or $|-\,x\rangle$, without telling anyone which axis she has chosen (x or z) for a given particle. Suppose Alice sends Bob a series of particles without trying, for the moment, to give it any intelligible form. There are 16 particles in the examples shown in Figures 13.4 and 13.5, but in practice one works with much larger numbers. It is only at the end of the procedure, as we show, that Alice will decide which particles should be taken into account in order to construct the message she wants to transmit.

What can Bob do in this situation? He can orient the axis of his Stern–Gerlach apparatus in an arbitrary way, x or z. On average, for half of the particles, his choice is the same as Alice's, in which case the bit he detects is significant. Indeed, if Alice sends a particle in the state $|+\,x\rangle$ and if Bob chooses the x-axis, he does measure $+$ with probability 1. For the other half of the particles, Alice and Bob choose different axes and Bob's results are useless: if Alice sends $|+\,x\rangle$ and if Bob chooses the z-axis, he will detect $+$ with probability 1/2 and $-$ with probability 1/2.

Number of the particle	1	2	3	4	5	6	7	8
Axis chosen by Alice (kept secret)	z	z	x	z	z	x	x	z
State chosen by Alice (kept secret)	$+$	$-$	$+$	$-$	$-$	$-$	$+$	$-$
Axis chosen by Bob (broadcast openly)	z	x	x	z	x	z	x	x
State measured by Bob (broadcast openly)	$+$	$-$	$+$	$-$	$-$	$+$	$+$	$+$
Useful measurement?	yes	no	yes	yes	no	no	yes	no

Fig. 13.4. Detection of a possible spy among Bob's results done along the same axis (particles 1,3,4, and 7), Alice looks for a possible difference that would mean a spy has operated. No anomaly appears here. In practice, in order to have a sufficient confidence level, one must use a number of events much larger than 8.

In order to make sure that no spy has intercepted the transmission, Bob announces openly the set of axes he has chosen, x or z, for each event. He also says what results he has obtained (i.e., $+$ or $-$) for a fraction of the particles. For instance, in the case of 16 particles shown in Figures 13.4 and 13.5, Bob announces publicly his 16 choices of axes, and his first 8 results. Alice examines the results, and she can detect whether a spy has operated. Her argument is the following. The spy does not know the directions x or z she has chosen for each particle. Suppose that the spy orients his Stern–Gerlach apparatus in a random way along x or z, and that he re-emits a particle whose

spin state is the same as what he has measured. If he chooses the x-axis and he gets the result $+$, he sends Bob a particle in the $|+\,x\rangle$ state. This operation is detectable, because it induces errors in Bob's observations.

Consider, for instance, the case where Alice has sent a particle in the state $|+\,z\rangle$, and Bob has also oriented his detector along the z-axis, but where the spy has oriented his own Stern–Gerlach apparatus along the x-axis. The spy can measure $+$ with a probability $1/2$ and $-$ with a probability $1/2$. According to his result, he re-emits to Bob a particle in the state $|+\,x\rangle$ or $|-\,x\rangle$. In both cases, because Bob's detector is oriented along z, Bob can measure $+$ with a probability $1/2$ and $-$ with a probability $1/2$. If the spy had not been present, Bob would have found $+$ with probability 1.

Therefore, among all the results announced by Bob, Alice looks at those where her own choice of axes is the same as Bob's (Figure 13.4). If no spy is acting, Bob's results must be identical to hers. If a spy is present, there must be differences in 25% of the cases. Therefore, if Bob announces publicly 1000 of his results, on average 500 will be useful for Alice (same axes), and the spy will have induced an error in 125 of them (on average). The probability that a spy is effectively present, but remains undetected by such a procedure is $(3/4)^{500} \sim 3 \times 10^{-63}$, which is negligible.

Number of the particle	9	10	11	12	13	14	15	16
Axis chosen by Alice (kept secret)	x	z	x	z	z	x	z	z
State chosen by Alice (kept secret)	$+$	$-$	$+$	$+$	$-$	$-$	$+$	$-$
Axis chosen by Bob (broadcast openly)	z	z	x	x	z	z	z	x
State measured by Bob (kept secret)	$-$	$-$	$+$	$+$	$-$	$+$	$+$	$+$
Useful measurement?	no	yes	yes	no	yes	no	yes	no

Fig. 13.5. After making sure that there is no spy, Alice chooses among the useful measurements those that allows communication of the message. For instance, to communicate the message "1,1", that is, "+, +", she openly asks Bob to look at the results of his measurements 11 and 15.

Once Alice has made sure that no spy has intercepted the communication, she tells Bob openly which measurements he must read in order to reconstruct the message she wants to send him. She simply chooses them among the sequence of bits for which Bob and she have made the same choice of axes, and for which Bob did not announce his result openly (Figure 13.5).

13.3.2 Present experimental setups

As for experimental tests of Bell's inequality, current physical setups use photons rather than spin 1/2 particles. Various methods can be used to code

information on photons. We only consider the coding of polarization, which is effectively used in practice. Alice uses four states that define two non-orthogonal bases, each of which can code the bits 0 and 1, for instance, in the form

$$|\uparrow\rangle : 1 \quad ; \quad |\rightarrow\rangle : 0 \quad ; \quad |\nearrow\rangle : 1 \quad ; \quad |\nwarrow\rangle : 0 \quad . \tag{13.18}$$

In present quantum cryptography devices, the challenge is to obtain sufficiently large distances of transmission. One currently reaches distances of the order of ten kilometers, by using optical telecommunication techniques, in particular photons in optical fibers.

An important point is the light source. A "noncloning" theorem, which is crucial in order to provide the safety of the procedure, only applies to individual photons. On the contrary, usual light pulses used in telecommunications contain very large numbers of photons, typically more than 10^6. If one uses such pulses for coding the polarization, the noncloning theorem no longer applies. Indeed it is sufficient for the spy to remove a small part of the light in each pulse and to let the remaining part propagate to Bob. The spy can measure in this way the polarization of the photons of the pulse without modifying the signal noticeably.

In order to guarantee the safety of the procedure, each pulse must contain a single photon. This is a difficult condition to satisfy in practice and one uses the following alternative as a compromise. Alice strongly attenuates the pulses so that the probability p of finding one photon in each pulse is much smaller than one. The probability of having two photons will be $p^2 \ll p$, which means that there will be very few two (or more) photon pulses. Obviously, most of the pulses will contain no photons, which is a serious drawback of the method because Alice must code the information redundantly. In practice, a value of p between 0.01 and 0.1 is considered to be an acceptable compromise.

Once this basic question is solved, the essential part of the system uses optical telecommunication technologies. The source is a strongly attenuated pulsed laser, and the coding polarization takes place directly in the optical fiber using integrated modulators. The attenuated pulses are detected with avalanche photodiodes, which transform a single photon into a macroscopic electrical signal by an electron multiplication process. In order to identify unambiguously the photons emitted by Alice and detected by Bob, electric pulses synchronized with the laser pulses are sent to Bob by conventional techniques, and they play the role of a clock. Finally, a computerized treatment of the data, involving a large number of pulses, fulfills the various stages of the procedure described above, in particular to test the absence of a possible spy on the line.

At present, the systems that have been built are more demonstration prototypes rather than operational systems. Several relevant parameters have been tested, such as the distance and the transmission rate, the error rate, and so on. Actually, developing these systems has for the moment a prospec-

tive character, because conventional (nonquantum) cryptographic systems are considered to be very reliable by civilian or military users.

13.4 Quantum teleportation

Quantum entanglement allows an amusing operation, "quantum teleportation." Suppose Alice has a spin 1/2 particle A in the spin state

$$\alpha|+\rangle + \beta|-\rangle\,, \quad \text{with} \quad |\alpha|^2 + |\beta|^2 = 1\ ,$$

that she wants to teleport, or fax, to Bob in a simple manner without measuring α and β.

13.4.1 Bell states

Consider a system of two particles of spin 1/2, whose spin state is written

$$\alpha|+;+\rangle + \beta|+;-\rangle + \gamma|-;+\rangle + \delta|-;-\rangle \tag{13.19}$$

with $|\alpha|^2 + |\beta|^2 + |\gamma|^2 + |\delta|^2 = 1$.

The probability of finding in a measurement $(+\hbar/2, +\hbar/2)$ is $|\alpha|^2$, the probability of finding $(+\hbar/2, -\hbar/2)$ is $|\beta|^2$, and so on. The four states of the Bell basis are defined as follows:

$$|\Psi_+\rangle = \frac{1}{\sqrt{2}}\left(|+;+\rangle + |-;-\rangle\right)\ , \quad |\Phi_+\rangle = \frac{1}{\sqrt{2}}\left(|+;-\rangle + |-;+\rangle\right)\ ,$$
$$|\Psi_-\rangle = \frac{1}{\sqrt{2}}\left(|+;+\rangle - |-;-\rangle\right)\ , \quad |\Phi_-\rangle = \frac{1}{\sqrt{2}}\left(|+;-\rangle - |-;+\rangle\right)\ .$$

The projectors $P_\Psi = |\Psi\rangle\langle\Psi|$ on one of these states $|\Psi\rangle$ can be called occupation operators of one of these states. In fact, the eigenvalues of these operators are 1 (occupied state) and 0 (vacant state).

A direct calculation gives the probabilities of finding (13.19) in each Bell state:

State $|\Psi_+\rangle$ occupied: probability $|\alpha + \delta|^2/2$,
State $|\Psi_-\rangle$ occupied: probability $|\alpha - \delta|^2/2$,
State $|\Phi_+\rangle$ occupied: probability $|\beta + \gamma|^2/2$,
State $|\Phi_-\rangle$ occupied: probability $|\beta - \gamma|^2/2$.

Because the basis of Bell states is orthonormal, the sum of these four probabilities is indeed 1.

13.4.2 Teleportation

In order for Alice to "teleport" the state $\alpha|+\rangle+\beta|-\rangle$ of particle A to Bernard, the procedure is as follows.

Alice and Bob also have in common a pair of spin 1/2 particles B and C, prepared in the singlet state (Figure 13.6):

$$\frac{1}{\sqrt{2}}\left(|+;-\rangle - |-;+\rangle\right) \ .$$

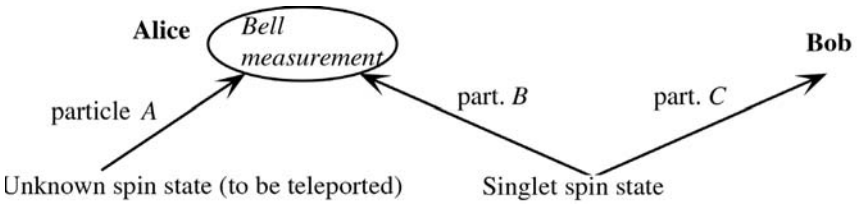


Fig. 13.6. Principle of the teleportation of the quantum state of a particle.

1. The state of the system of three spins (A, B, C) is

$$|\Psi\rangle = \frac{\alpha}{\sqrt{2}}|+;+;-;\rangle + \frac{\beta}{\sqrt{2}}|-;+;-;\rangle - \frac{\alpha}{\sqrt{2}}|+;-;+;\rangle - \frac{\beta}{\sqrt{2}}|-;-;+;\rangle \ .$$

   This state can be decomposed on the Bell basis of the spins A and B; the basis for spin C remains $|\pm\rangle$:

$$|\Psi\rangle = \frac{1}{2}|\Psi_+\rangle\left(\alpha|-\rangle - \beta|+\rangle\right) + \frac{1}{2}|\Psi_-\rangle\left(\alpha|-\rangle + \beta|+\rangle\right)$$
$$-\frac{1}{2}|\Phi_+\rangle\left(\alpha|+\rangle - \beta|-\rangle\right) - \frac{1}{2}|\Phi_-\rangle\left(\alpha|+\rangle + \beta|-\rangle\right) \ .$$

   If Alice measures the spin state of the couple AB that projects this state on one of the four vectors of the Bell basis of AB, the probability of finding the pair AB in each of the Bell states is the same $(|\alpha|^2 + |\beta|^2)/4 = 1/4$.
2. If Alice finds the pair AB in the state $|\Phi_-\rangle$, the state of the spin C is the original state $\alpha|+\rangle + \beta|-\rangle$.
3. In order to teleport the a priori unknown state $\alpha|+\rangle + \beta|-\rangle$ of particle A to particle C, Alice must not try to measure this state. She must simply make a Bell measurement on the pair AB and tell the result to Bob. When she finds that the state $|\Phi_-\rangle$ is occupied (in 25% of the cases), Bob need not do anything. The spin state of C after the measurement is the state of A before the measurement. In all other cases, Bob can reconstruct the initial state by a simple transformation. For instance, if Alice finds the pair AB in the Bell state $|\Phi_+\rangle$, the spin state of C is $\alpha|+\rangle - \beta|-\rangle$, which can be transformed into the initial state $\alpha|+\rangle + \beta|-\rangle$ by performing a rotation of π around the z-axis.

4. One cannot use this method to transmit information more rapidly than with usual means. As long as Alice doesn't tell Bob the result of her Bell measurement, Bob has no available information. It is only after he knows Alice's results, and he has rejected or reconstructed the fraction of experiments (75%) which do not give $|\Phi_-\rangle$ that he can make profit of the "teleportation" of the spin state of particle A.

Further reading

- Concerning the EPR problem: *Quantum Theory and Measurement*, edited by J.A. Wheeler and W.H. Zurek, Princeton University Press (1983).
- Concerning Bell's inequalities: J.S. Bell, *Physics* **1**, 195 (1964); see also J. Bell, Speakable and Unspeakable in Quantum Mechanics, Cambridge University Press (1993). The inequality proven here is due to J.F. Clauser, M.A. Horne, A. Shimony, and R.A. Holt, *Phys. Rev. Lett.* **23**, 880 (1969).
- Quantum cryptography is presented by C. Bennett, G. Brassard and A. Ekert, "Quantum Cryptography," *Scientific American*, October (1992).

14

Quantum mechanics in the Universe

The direct and intimate relationship of infinitely large-scale physics, stars, galaxies, clusters of galaxies, and cosmology, with infinitely small-scale physics, atoms, nuclei, elementary particles, and fundamental interactions was understood about 30 years ago.

The idea that on one hand the cosmos, and on the other hand the "atoms" of Democritus and Lucretius are interwoven, is a great endeavor of our time. This cross-fertilizing of the two infinities shows that our best laboratory of the future, if not the only one, is the Universe itself.

Our ideas about light, heat, motion, came first from the observation of the sun, the planets, and the stars. The key and the mastery of the understanding of radiation, atomic physics, nuclear physics, and particle physics was provided to a large extent by observation of the cosmos.

In return, it is fundamental physics that transformed astronomy into astrophysics because light carries information

- On matter through spectroscopy
- On motion through the Doppler effect
- On temperature through spectral distributions, and so on

Instead of simply observing the motion of celestial bodies, one has become able to answer questions about their chemical composition, their structure, and how they work. One must of course add to that all the technological progress in observational techniques since the second half of the 19th century.

Astronomy asks a series of questions:

- What is the world made of?
- How does it work?
- Where do we come from; where are we going?

However, these question seem hopeless at first, because the timescales of objects in the cosmos are huge compared to human timescales. The closest star, Proxima Centauri, is four light-years away. Our only means to probe the cosmos consists of collecting the radiation it emits such as light, infrared and ultraviolet radiation, radiowaves, X-rays, γ-rays, and neutrinos. All this radi-

ation propagates at the velocity of light, so that we see the sun as it was eight minutes ago, the nearest star as it was four years ago, the closest galaxy two million years ago, and far away objects such as quasars ten billion years ago.

In this chapter, we discover via a few examples the cross-fertilizing of quantum mechanics and astrophysics. This adventure has given a soul and a lifeline to the cosmos.

In the first part, we show how nuclear physics, together with human observations have allowed us to understand the mechanisms of stars, from their birth to their death. Next, we discover the progress made thanks to radioastronomy, which is our main subject of interest here. It is on one hand a technological tool that could not have emerged without quantum mechanics. In fact it would have been impossible to perform without the help of masers, which are selective noiseless amplifiers. Radioastronomy is also a very refined means to identify the actors in the cosmos owing to the accuracy of our knowledge in the domain of radiowaves. This leads us to the discovery of the cosmic background radiation by Penzias and Wilson in 1965. It is one of the strongest observational pieces of evidence in favor of the big bang and of the origin of the Universe. This discovery, and the ensuing results obtained with space astrophysics in the last decade have transformed cosmology. It used to be a series of intellectual speculations, and it has become a major observational field in astrophysics.

In Section 14.4, we will turn to a fundamental source of radiowave radiation, the emission at 21-cm or equivalently 1420 MHz of the hydrogen atom. This is due to the so-called "hyperfine" splitting of the ground state level of this atom. This physical phenomenon is important for many reasons. First, the hydrogen atom is the most abundant element in the Universe, and this emission is the only way to detect it in cold regions of the Universe. This brings us back to the interstellar medium, first of all in our galaxy, the Milky Way, which is difficult for us to observe because we sit in its plane. We discover how it has been possible to reconstruct the spiral structure of our galaxy. We show how the interstellar medium plays the role of a physical and chemical bath in which stars can regenerate permanently. In Section 14.6, we discover how large intergalactic clouds of hydrogen show the evolution of galaxies themselves. These objects are permanently evolving and interacting, sometimes violently.

With the discovery of more and more interstellar molecules, in particular organic molecules, we show at what point the debate on the location and the mechanism of the origin of life sits at present. There are fascinating observations on the origin of aminoacids and DNA, therefore on the simplest forms of life. We show what unexpected discoveries make us wonder about whether the origin of life is terrestrial or extraterrestrial, or perhaps a conjunction of both. This is an enigma at present. In fact, both components may be associated in prebiotic chemistry.

Finally, any good story must have a happy ending with a smile. We show how and why the quantum theory of the hydrogen atom provides a universal

cosmic language. It is the only one and the simplest in order to have, some day, an answer to the question: Are we alone in the Universe?

14.1 Quantum mechanics and astronomy

In astronomy, our theater is the world. It is a very special theater because one has no control over it. One cannot perform any experiment, not even applaud.

14.1.1 Life and death of stars

But we have an enormous advantage. My friend Alfred Vidal-Madjar, who is an astrophysicist, always says that, despite what people can say, astronomy is probably the oldest profession in the world (and of course the most beautiful). In fact we possess very ancient archives (probably the oldest existing ones) because humans attributed a special importance to what happened in the sky. They recorded the location and the date of events that they found of special importance, even though they were not able to say why they were important.

One example is the sudden and striking appearance of "new" stars, which were extremely bright because they shone in the daytime. We now know that these stellar events correspond to supernovae for which we possess very precious catalogues. Figure 14.1 shows the detail of a piece of bone found in China and which was carved during the Shang dynasty, 1500 B.C., which indicates the appearance of a new star near Antares.

Fig. 14.1. Left: Fragment of a Chinese carved bone of the Shang dynasty (1500 BC) indicating the arrival of a new star near Antares (ideogram indicated by an arrow). Right: Crab nebula, that is, the remains of the explosion of a supernova observed by Chinese astronomers in 1054. (Photo credit: FORS Team, 8.2-meter VLT, ESO.)

By now, we know the explanation of such exceptional events (about 20 of them per century occur in our galaxy; one can see more in other galaxies).

They correspond to catastrophic phases in the end of the life of massive stars (say ten solar masses). They are the most important nuclear phenomena since the big bang. The energies released are of the order of 10^{47} joules.[1]

It is during such events that elements heavier than carbon, nitrogen, and oxygen, such as iron, lead, or uranium are formed and dispersed in the interstellar medium. In this sense, we are made of stardust. Our hemoglobin contains iron. This shows we originated from the explosion of a supernova about which we know, by other means, that it exploded six billion years ago.

Stars synthesize heavy elements from the primordial hydrogen of the big bang via thermonuclear reactions.[2] In order for thermonuclear reactions to ignite, one must have high temperatures and high densities. The tunneling effect seen in chapter 5 plays a fundamental role in these reactions.

So, nowadays the sun quietly burns its hydrogen and transforms it into helium. As we have said in the first chapter, the fact that this burning is quiet and regular is amazing: we have on top of us a huge nuclear bomb that does not explode! The system is self-regulated and the thermal pressure of nuclear reactions (which tends to expand the system and make it explode) is exactly compensated by the inward gravitational pressure that tends to make it implode.

Nuclear physics teaches us that the sun will stop shining when its fuel is exhausted in five billion years. Nearly all its hydrogen will have been transformed into helium. The helium will then transform into carbon, oxygen and nitrogen, and it will contract under its own weight. At that point, further nuclear reactions will not be able to ignite because the mass of the sun is insufficient to fight against the pressure of the degenerate electron gas, therefore it will not reach high enough temperatures for further nuclear fusion. At that point, the sun will be a "white dwarf," a dead star that will simply quietly lose its heat.

For stars more massive than the sun, the next part of the story, which is essential for our existence, has a much more catastrophic scenario. When, in their core, massive stars have synthesized silicon and have reached iron, which is the most strongly bound nucleus (i.e., the ground state of nuclei), nuclear reaction can no longer produce heat. There is no more thermal pressure to balance gravitation. The weight is huge, and the core of these stars collapses onto itself until it reaches nuclear densities (one billion tons per cubic centimeter). At this point, one deals with compressed nuclear matter that bounces violently on itself. The rebound creates a shock wave that ejects the outer shells of the star. At the center, the compressed nuclear matter forms a neutron star, which manifests itself as a pulsar, whose mass is close to one

[1] This means $\sim 10^{28}$ one-hundred megaton nuclear weapons; some numbers are difficult to imagine.

[2] See, for instance, J.-L. Basdevant, J. Rich, and M. Spiro, *Fundamentals in Nuclear Physics; From Nuclear Structure to Cosmology*, chapter 8. New York: Springer (2005).

solar mass and whose size is a few kilometers. Pulsars emit typical periodic bursts of radiowaves. They are genuine astronomical radio beacons.

Thanks to the archives, it has been possible to spot the remains of such explosions. Figure 14.1 shows the Crab nebula, a supernova explosion observed by Chinese astronomers in 1054. One can see, 1000 years later, the ejected external shells. In 1967, the central pulsar was identified. One can therefore rebuild the film of this event over 1000 years thanks to the archives (and to quantum mechanics).

In modern times, we were lucky enough to see the explosion of a supernova, called SN1987A, near us (at 150,000 light-years) during the night of February 23 to 24, 1987. It occurred in the Tarantula nebula in the Large Magellanic Cloud (LMC), a satellite galaxy of the Milky Way.

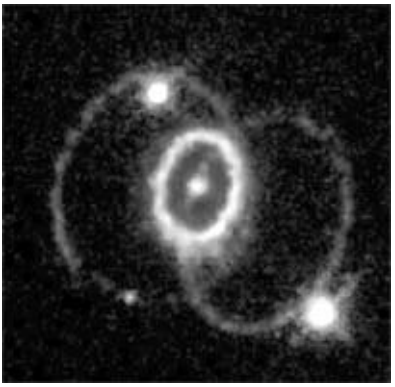


Fig. 14.2. Supernova SN1987a ten years after its explosion during the night of February 23 to 24, 1987. The rings observed in the matter ejected during the explosion are a surprising phenomenon for astrophysicists. (Photo credit: C. Burrows (ESA/ STSCI), HST, NASA) http://antwrp.gsfc.nasa.gov/apod/ap960705.html.)

Figure 14.2 shows what that supernova looks like a few years later. Such an observation has enormous interest. It enables us to verify our theories of these explosions, which, up to then, were only ideas.

In other words, thanks to quantum mechanics and to nuclear physics, we know better and better, since the middle of the 20th century, how stars work, how old they are, and what their history is.

Being an astrophysicist is beautiful. By combining fundamental physics and pictures, astrophysicists manage to tame the time in the Universe, to go and occupy the cosmos, and to write the scenario of the world's theater. The observation of far-away objects tells us our past history and predicts our future.

For instance, we observe that the place where we live is quite unexceptional, that galaxies all move away from us with a velocity proportional to

their distance. The Universe expands systematically as an inflated balloon. Then, astrophysicists imagine they run the film backwards. They understand that everything agrees with the idea that once upon a time, say 13 billion years ago, the Universe was infinitely small, infinitely dense, and hot, that there was a big bang, a "primordial singularity" that generated everything.

14.1.2 Spectroscopy

However, one must say that for a long time the oldest profession in the world was done for free. The only detectors people could make use of were human naked eyes with their limitations.

What triggered the transformation of astronomy into astrophysics is the development of spectroscopy which started in the early 1800s with Wollaston and more intensively with Fraunhofer in 1814. It was amazing to realize that stars gave us for free very valuable pieces of information in the light they send us!

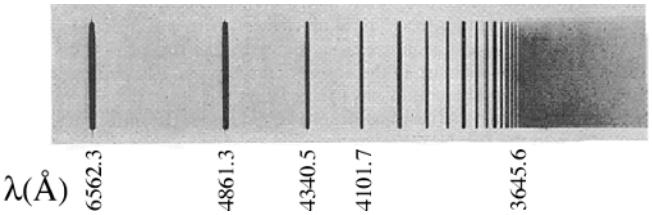


Fig. 14.3. Spectrum of atomic hydrogen (Balmer series) obtained by Hertzberg in 1927 at low temperature (27 K) with a prism. The first four lines on the left are in the visible part of the spectrum; the others are in the ultraviolet region.

It is needless to say how much atomic spectra of stars, which were accumulated during the 19th century, contributed to the birth of quantum mechanics. In return, the mastery of spectroscopy generated many steps forward in astrophysics.

Of course, the main stages of the progress of astronomy are primarily due to technological progress in the means of observation: Galileo's and Newton's telescopes, photography, space telescopes, X-ray and γ-ray astronomy, space probes, cosmic neutrino detectors, and so on. This happened in parallel with progress in fundamental research that enables a more and more refined analysis and interpretation of the data.

14.2 Radioastronomy, the interstellar medium

It seems easy to observe the cosmos. Our own galaxy, the Milky Way which is so beautiful to look at in the summer, is paradoxically one of the most

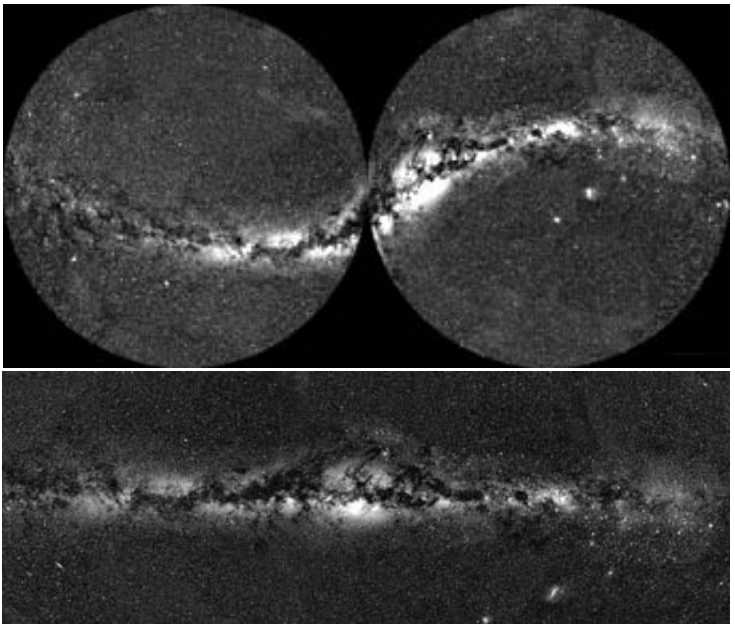


Fig. 14.4. The Milky Way. Top: Separate views of the northern and southern hemispheres. Bottom: The pictures have been put together in order to see the angular position of stars with respect to the plane of the galaxy, which is at a large angle from the plane of the solar system (this is what provokes the apparent asymmetries in the upper Figure). One can see on the bottom right the two Magellanic clouds which are only visible in the southern hemisphere. On both pictures, the contrast has been strongly amplified. (Photo credit Axel Mellinger; http://home.arcor-online.de/axel.mellinger/allsky.html.)

difficult galaxies to explore. There are so many objects that one only sees a few of them: 6000 with the naked eye, 100 million with telescopes, among the 200 billion stars which it contains. The nearby stars are very bright but they screen everything that lies behind them.

Here, we are interested in information obtained since the 1950s with the development of radioastronomy. Radiowaves are more penetrating than light waves and they carry a lot of complementary information.

14.2.1 The interstellar medium

In order to understand why radioastronomy has been a revolution in astrophysics, we must complete the list of actors in our theater, and come back to what galaxies contain. Because there are not only stars.

There are various forms of matter in galaxies:

- There are stars, which are dense and hot and emit in all the electromagnetic spectrum.
- There is the celebrated "dark matter," the amount of which is not fully evaluated at present; it could be $\simeq$10 times the observed usual or "baryonic" matter. That form of matter appears through its gravitational effects, such as the rotation velocities of stars in galaxies. Dark matter does not emit any radiation. Its nature is a puzzle of contemporary physics.
- There are very high energy particles and nuclei (up to 10^{21} eV), called cosmic rays, whose origin is not completely understood at present.
- There are "black holes" which, more and more, appear to play a central role in the life and structure of galaxies.
- Last, but not least for us, there is an interstellar medium, a diffuse medium in the galaxies, where the physical and chemical life of galaxies takes place. This medium has an overall large mass (10 to 50% of the total mass of a galaxy). It is very dilute (1 to 20 atoms per cm^3) and cold (50 to 100 K).

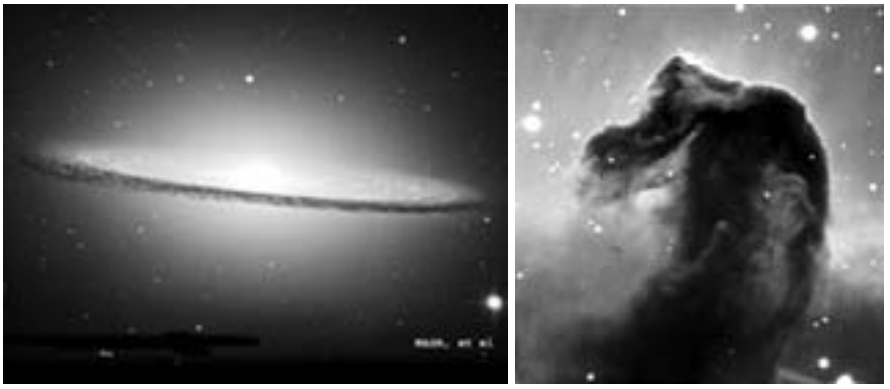


Fig. 14.5. Visible parts of the interstellar medium. Left: The galaxy Sombrero, very far away, viewed by the Hubble space telescope. It has a dense cloud of dust in its plane that screens the light of stars. Each little white point in this cottonlike object is a "sun." (Photo credit NASA and Hubble Heritage Team, http://heritage.stsci.edu/2003/28/big.html.) Right: The Horsehead Nebula, B33, which is a prominence of a molecular cloud in the Orion Nebula, close to us in the Milky Way. The cloud is surrounded by stars which heat it and bring light behind it. (Photo credit: instrument FORS2 at the VLT, http://www.eso.org/outreach/press-rel/pr-2002/phot-02-02.html.)

This medium consists primarily of atoms: 90% atomic hydrogen and 10% atomic helium (i.e., respectively, 75% and 25% in mass) that were formed just after the big bang, during the first three minutes of the Universe. There are other atoms, such as carbon, calcium, potassium, and others, but in smaller amounts. They have been formed in stellar nucleosynthesis and have been

ejected in the interstellar medium, for instance in supernovae explosions. The rest of the interstellar medium (from 1 to 50%) consists of molecules such as H_2O, CO_2, C_2H_5OH which we talk about, and interstellar dust.

The interstellar medium causes spectacular phenomena in astronomical observations. One can see in Figure 14.5 an example of a dust disk in the plane of the circular galaxy Sombrero, and, in the Milky Way, the famous Horsehead nebula, which is a prominence of a molecular cloud which is seen as a shadowgraph because of the stars behind it. It is inside such molecular clouds that stars form, as we show below.

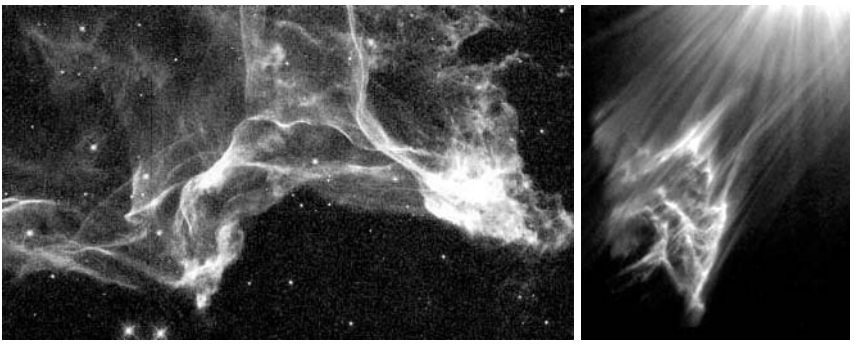


Fig. 14.6. Left: Veil of interstellar matter in the Cygnus (or Veil) nebula. (Photo credit: HST WFPC2 http://casa.colorado.edu/~maloney/CygnusLoop.gif.) Right: Veil of matter between two of the Pleiade stars. (Photo credit: NASA and Hubble Heritage Team http://heritage.stsci.edu/2000/36/big.html.)

Figure 14.6 shows veils of matter in the Cygnus nebula and between the Pleiade stars, which are young stars of our galaxy.

However, one can understand from these examples that this type of observation is quite limited. In order to see such objects or mass distributions, they must either emit light or be illuminated by light sources, and furthermore they must not be hidden by other objects (after all, we see a horsehead, but there may also be woodlice or lobsters).

The interstellar medium, which is cold, does not emit in the visible part of the spectrum, but it abundantly emits radiofrequency waves. The radiowaves emitted by the Milky Way, as well as by other galaxies, penetrate matter much more deeply than light waves, and they bring new information compared to what we can see directly.

However, this radiation comes from sources that are of extremely low densities and the interstellar medium has a small brilliance compared to stars. In order to observe such emissions, one needs high-performance selective noiseless amplifiers. This is where a major technical contribution of quantum mechanics to astrophysics was made. Masers, which we described in chapter 7, have precisely these features. The introduction of masers in astrophysics is due

to Townes himself, who turned to astrophysics in the mid 1960s and caused a true revolution. In fact, with masers one can observe the radiation coming from weak sources, that would be drowned in the background noise if one were to use traditional amplifiers. The masers used in astrophysics are solid-state masers, in particular, ruby masers that can deliver important amplification factors.

Fig. 14.7. Left: Very large array radiotelescope in Sorocco, New Mexico. (Image courtesy of NRAO/AUI http://www.nrao.edu/imagegallery/php/level1.php.) Right: Commercial hydrogen maser. (Photo credit ; NASA, Smithsonian Astrophysical Observatory, http://horology.jpl.nasa.gov/h_maser.html.)

The observation of the interstellar medium requires powerful radiotelescopes, and interferometric setups in order to reach an acceptable resolution (radio wavelengths are 100,000 times larger than optical ones), an example of which is shown in Figure 14.7. At present, a maximum resolution is obtained with the technique of very-large-baseline interferometry (VLBI). The data recorded with radiotelescopes located on different continents can be put together and synchronized with atomic clocks. The baseline is of the order of 10,000 km. Astrophysicists are contemplating putting space-radiotelescopes on large orbits in order to improve the resolution.

Radioastronomical observations represent a formidable task. The total energy received up to now in radioastronomy is of the order of the kinetic energy of the ashes of a cigar falling from 1 meter high.

One aims at various directions in the sky. By a sophisticated analysis of the observed signal, one can reconstruct the position, the density, the velocity (by Doppler effect), and the composition of interstellar clouds.

One example is shown in Figure 14.8 where one can see three hydrogen clouds of different masses on the line of sight. Two are coming toward us; the other is receding.

The pioneers of centimetric radioastronomy, which is fundamental in order to observe hydrogen, were Purcell and Van de Hulst in 1947. We come back to this below.

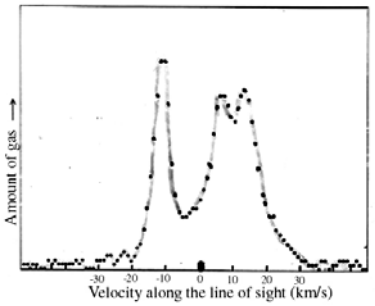


Fig. 14.8. Signal around the 21-cm line in a given direction. One can see three atomic hydrogen clouds of different masses on the line of sight. Two of them come toward us, the third one recedes. The abscissa is given in terms of the Doppler shift in km/s. (Credit: Hatt Creek Radiotelescope.)

14.3 Cosmic background radiation: Birth of the Universe

The founders of *millimetric* radioastronomy (nearly all molecules emit in the millimetric domain) were Penzias and Wilson, in 1965.

They were engineers at the Bell Telephone Company. They were constructing sophisticated antennas for radioastronomy. They used a ruby maser. They calibrated their antennas by pointing them at night toward the "vacuum." However, they were constantly disturbed by a persisting isotropic radiowave background noise, in the range 100 to 200 GHz, which was much larger than any of their expectations, and which scrambled their observations.

Penzias and Wilson actually made one of the greatest observational discoveries in cosmology. Again, this brings us back to the roots of quantum mechanics.

In fact, cosmologists realized rapidly that Penzias and Wilson had discovered the cosmic background radiation (CBR) of the big bang, which had been predicted by Gamow in his first works on the theory. It is the first, and probably the most important, experimental argument in favor of the big bang theory.

Roughly 400,000 years after the big bang, light uncoupled from matter, which was too dense before that, and was released. Since then, that radiation fills the Universe and cools down with its expansion.

At this point, we recover the cradle of quantum mechanics and Planck's formula for black-body radiation. The best way to verify Planck's prediction is to sit inside an oven at a given temperature, which is exactly our (cosmic) daily life. The frequency distribution is the best confirmation of Planck's formula, as we said in chapter 2. The measurement, shown in Figure 14.9, is in remarkable agreement with Planck's formula. We bathe in an isotropic cosmic background radiation at 2.728 K. The accuracy of the agreement between theory and experiment is within one part in a thousand. The measurement was made by the satellite COBE, Cosmic Background Explorer, launched in 1989.

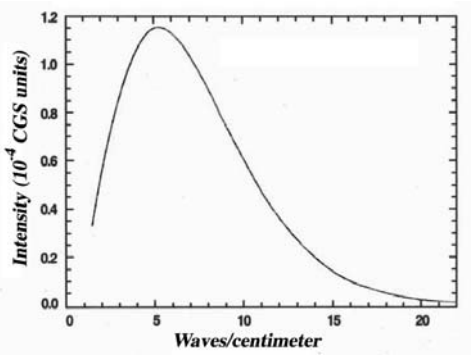


Fig. 14.9. Wave-number distribution of the cosmic background radiation measured in 1992 by the COBE satellite. The agreement between Planck's formula at a temperature $T = 2.728$ K lies within the line. (Photo credit: Mather et al., *Astrophys. J.*, **420**, 439 (1994) http://lambda.gsfc.nasa.gov/product/cobe/firas_image.cfm.)

But that is far from being the end of the story. The COBE data allowed us to do much better, thanks to the performance of masers.

The cosmic background radiation is not completely isotropic. If one performs a refined angular analysis, there is a relative anisotropy called the dipole. Figure 14.10 shows the angular dependence of the intensity of the radiation on the whole 4π steradians of the sky, with an accuracy of 10^{-3}.

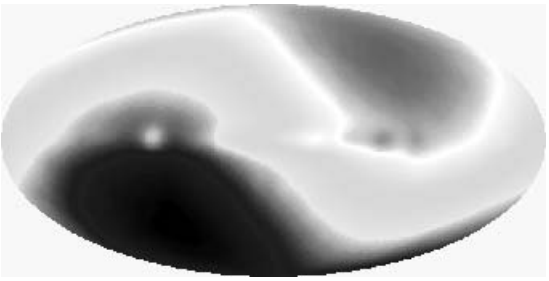


Fig. 14.10. Anisotropy of the CBR measured by COBE within 10^{-3}. Evidence for the "dipole." The left-bottom direction is blueshifted and the the right-top direction is redshifted. One can see the trace of the background noise of the Milky Way in its plane. (Photo credit: Wright et al., *Astrophys. J.*, 420, 450 (1994), "Interpretation of the COBE FIRAS CMBR Spectrum," and Fixsen et al., *Astrophys. J.*, 473, 576 (1996),"The Cosmic Microwave Background Spectrum from the Full COBE FIRAS Data Sets," http://lambda.gsfc.nasa.gov/product/cobe/firas_image.cfm.)

In the lower hemisphere, on the left of the picture, the temperature of the CBR is higher by 3.35 mK (blueshifted), and in the upper hemisphere on the right it is cooler by 3.35 mK (redshifted). This is simply the Doppler effect

due to the global relative motion of the solar system, and of all our galaxy as well as neighboring galaxies such as Andromeda, and the Magellanic clouds, with respect to the CBR. We all travel at a velocity of 600 km/s in the same direction toward the "Great attractor." Therefore we are moving with respect to an absolute space, the space where the CBR is isotropic.

It is quite amazing that by making a Doppler-shift correction of the data with a single velocity, both in norm and in direction, one can suppress completely, to that accuracy of 10^{-3}, the anisotropy (except for some little central part coming from the Milky Way).

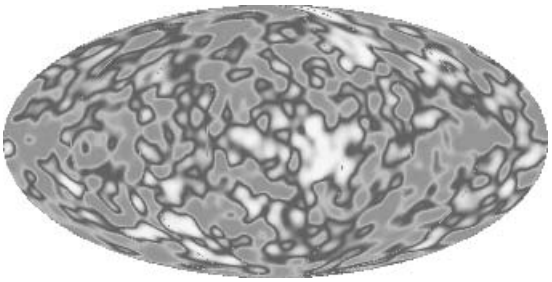


Fig. 14.11. Temperature fluctuations of the CBR as measured by COBE within 10^{-5}; appearance of matter distribution. (Photo credit, same as in Figure 14.10.)

Finally, after suppressing the dipole, if one pushes the accuracy to 10^{-5}, one can see in Figure 14.11 fluctuations of the order of 18 μK, which reflect the inhomogeneities, that is, the fluctuations of matter at the time when decoupling occurred. One can see protogalactic clusters forming 13 billion years ago. In other words, we see the matter distribution pouring out of the big bang, that is, *us*, who are lumps of condensed matter in a Universe that is nearly empty, with an average density of a few protons per cubic meters. This field of research has undergone spectacular progress in recent years with new experiments where the accuracies have been pushed toward 10^{-6} degrees. Figure 14.12 which was obtained by the experiment WMAP, Wilkinson Microwave Anisotropy Probe, proposed at NASA in 1995 and put in orbit in 2001, exhibits, in comparison with the previous one, the degree of accuracy that has been reached. Such results have generated spectacular progress in observational cosmology as well as in theoretical cosmology (one can now make a distinction between the two).

14.4 The 21-cm line of hydrogen

Let us come "back home," that is, to our galaxy and to its present neighborhood. One frequency dominates the radiowave emission and plays a key role:

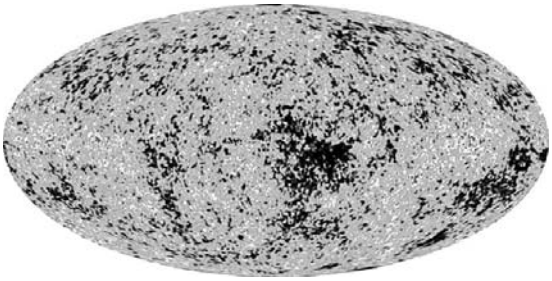


Fig. 14.12. Temperature fluctuations of the cosmic background radiation measured by the WMAP experiment with an accuracy of 10^{-6}. (Photo credit: NASA and Wilkinson Microwave Anisotropy Probe (WMAP) team http://map.gsfc.nasa.gov/m_mm.html.)

1420 MHz. It is on that frequency and only on that one, that hydrogen, the most abundant element in the Universe, emits radiation.

This effect is one of the best-known phenomena of atomic physics. It comes from the interaction of the electron and proton spin magnetic moments in the ground state of the hydrogen atom. There are only two relative orientations of the two spins, parallel and antiparallel. When the spins flip from one relative configuration to the other, there is emission or absorption on 1420 MHz, or at a wavelength of 21-cm. (This is in a domain close to television frequencies, and it is protected by an international agreement.)

14.4.1 Hyperfine structure of hydrogen

The theory of this effect was made by Fermi in 1930. Consider a system of two spin 1/2 particles as treated in chapter 12, Section 12.2.

The effect we are interested in (a splitting of the order of 6 10^{-6} eV) comes from the magnetic interaction between the two spin magnetic moments of the electron and the proton:

$$\hat{\boldsymbol{\mu}}_e = \gamma_e \hat{\boldsymbol{S}}_e \ , \quad \gamma_e = -q/m_e \ , \tag{14.1}$$

$$\hat{\boldsymbol{\mu}}_p = \gamma_p \hat{\boldsymbol{S}}_p \ , \quad \gamma_p \simeq 2.79 \ q/m_p \ . \tag{14.2}$$

This interaction is called the spin–spin, or hyperfine interaction. We only consider its effect in the ground state of hydrogen $n = 1, \ell = 0$.

We neglect here effects due to the internal structure of the proton and we treat it as a pointlike particle. The calculation of the magnetic field created at a point $\boldsymbol{r}$ by a magnetic dipole $\boldsymbol{\mu}_p$ located at the origin, is a well-known problem in magnetostatics.[3] The result can be written as

[3] See, for instance, J.D. Jackson, *Classical Electrodynamics*, Section 5.6. New York: Wiley (1975).

$$\boldsymbol{B}(\boldsymbol{r}) = -\frac{\mu_0}{4\pi r^3}\left(\boldsymbol{\mu}_p - \frac{3(\boldsymbol{\mu}_p\cdot\boldsymbol{r})\,\boldsymbol{r}}{r^2}\right) + \frac{2\mu_0}{3}\,\boldsymbol{\mu}_p\,\delta(\boldsymbol{r})\ . \tag{14.3}$$

The interaction Hamiltonian between the magnetic moment $\boldsymbol{\mu}_e$ of the electron and this magnetic field reads

$$\hat{W} = -\hat{\boldsymbol{\mu}}_e\cdot\hat{\boldsymbol{B}}\ . \tag{14.4}$$

For $\boldsymbol{r}\neq 0$, $\hat{W}$ reduces to the usual dipole–dipole interaction:

$$r\neq 0\,,\quad \hat{W}_{\rm dip} = \frac{\mu_0}{4\pi\hat{r}^3}\left(\hat{\boldsymbol{\mu}}_e\cdot\hat{\boldsymbol{\mu}}_p - \frac{3(\hat{\boldsymbol{\mu}}_e\cdot\hat{\boldsymbol{r}})(\hat{\boldsymbol{\mu}}_p\cdot\hat{\boldsymbol{r}})}{\hat{r}^2}\right)\ .$$

This interaction will not contribute to our calculation because of the following mathematical property. For any function $g(r)$, $(r=|\boldsymbol{r}|)$, regular at $r=0$, the angular integration vanishes:

$$\int g(r)\ W_{\rm dip}(\boldsymbol{r})\ d^3r = 0\ . \tag{14.5}$$

At $\boldsymbol{r}=0$, the field (14.3) is singular because of the contribution of the term proportional to $\delta(\boldsymbol{r})$. This leads to a contact interaction:

$$\hat{W}_{\rm cont} = -\frac{2\mu_0}{3}\hat{\boldsymbol{\mu}}_e\cdot\hat{\boldsymbol{\mu}}_p\,\delta(\hat{\boldsymbol{r}})\ .$$

The origin of the singularity in $\boldsymbol{r}=0$ is the pointlike nature of the proton that we have assumed in our analysis. It entails that all field lines converge to the same point. A calculation taking into account the finite size of the proton and the corresponding modification of the field leads to essentially the same result, because the size of the proton is very small compared with the variation of the probability distribution of the electron at the origin in the $1s$ state. Note that this pointlike model is strictly valid for positronium, which is an atom made of an electron and a positron; both of which are pointlike objects.

The observable $\hat{W}$ above is a perturbation that acts on both space and spin variables. The space variables cause no problem. Consider

$$\hat{H}_1 = \int \psi^*_{100}(\boldsymbol{r})\ \hat{W}\ \psi_{100}(\boldsymbol{r})\ d^3r\quad ,$$

where $\psi_{100}(\boldsymbol{r})$ is the ground state wave function found in chapter 10.

The probability density for the ground level $|\psi_{100}(\boldsymbol{r})|^2$ is isotropic. As a consequence of (14.5), $\hat{W}_{\rm dip}$ does not contribute to $\hat{H}_1$. The contact term is readily evaluated as

$$\hat{H}_1 = -\frac{2\mu_0}{3}\hat{\boldsymbol{\mu}}_e\cdot\hat{\boldsymbol{\mu}}_p\,|\psi_{100}(0)|^2\quad . \tag{14.6}$$

$\hat{H}_1$ is an operator that acts only on spin states. It can be cast in the form

$$\hat{H}_1 = \frac{A}{\hbar^2} \hat{\boldsymbol{S}}_e \cdot \hat{\boldsymbol{S}}_p \ , \tag{14.7}$$

where the constant A can be inferred from the values of γ_e, γ_p, and $\psi_{100}(0)$

$$A = -\frac{2}{3} \frac{\mu_0}{4\pi} \frac{4}{a_1^3} \gamma_e \gamma_p \hbar^2 = \frac{16}{3} \times 2.79 \ \frac{m_e}{m_p} \alpha^2 E_I \quad . \tag{14.8}$$

One obtains numerically

$$A \simeq 5.87 \ 10^{-6} \text{ eV} , \quad \nu = \frac{A}{h} \simeq 1417 \text{ MHz} , \quad \lambda = \frac{c}{\nu} \sim 21 \text{ cm}. \tag{14.9}$$

The diagonalization of $\hat{H}_1$ in the Hilbert space of spin states is simple. Considering the total spin $\hat{\boldsymbol{S}} = \hat{\boldsymbol{S}}_e + \hat{\boldsymbol{S}}_p$, one has

$$\hat{\boldsymbol{S}}_e \cdot \hat{\boldsymbol{S}}_p = \frac{1}{2} \left(\hat{S}^2 - \hat{S}_e^2 - \hat{S}_p^2 \right) \ ,$$

which is diagonal in the basis of the eigenstates $|S, M\rangle$ of the total spin, with eigenvalues:

$$\frac{\hbar^2}{2} \ (S(S+1) - 3/2) \ , \quad \text{with here} \quad S = 0 , \quad \text{or } S = 1 \quad .$$

The ground state $E_0 = -E_I$ of the hydrogen atom is therefore split by the hyperfine interaction in two sublevels corresponding to the triplet $|1, M\rangle$ and singlet $|0, 0\rangle$ states:

$$\begin{array}{lll} E_+ = E_0 + A/4 & \text{triplet state} & |1, M\rangle \ , \\ E_- = E_0 - 3A/4 & \text{singlet state} & |0, 0\rangle \quad . \end{array} \tag{14.10}$$

The difference of these two energies is equal to A, that is, 5.87 10^{-6} eV; it corresponds to the characteristic line of hydrogen at a wavelength $\lambda \sim 21$ cm.

14.4.2 Hydrogen maser

In its ground state, the hydrogen atom therefore constitutes a four-state system with two energy levels. By a method whose principle is similar to what we have discussed in chapter 7, it is possible (but technically more complicated) to devise a hydrogen maser.[4] Among other things, this allows the measurement of the constant A, or equivalently, the frequency $\nu = A/h$ with an impressive accuracy:

$$\nu = \underbrace{1\,4}_{\text{A}} \underbrace{2\,0}_{\text{B}} \underbrace{4\,0}_{\text{C}} \underbrace{5\,7\,5\,1}_{\text{D}} \ . \ \underbrace{7\,6\,8\,4}_{\text{E}} \pm 0.\ 0\ 0\ 1\ 7 \text{ Hz} \ .$$

[4] H.M. Goldenberg, D. Kleppner, and N.F. Ramsey, *Phys. Rev. Lett.* **8**, 361 (1960).

In this result, we have underlined several groups of digits. The first two (A) were obtained by Fermi in 1930; they correspond to the contact term considered above. The two following ones (B) are calculated using the Dirac equation, and the experimental value for the anomalous magnetic moment of the electron (deviation of the order of 10^{-3}). Other corrections account for the two following decimals (C): relativistic vacuum polarization corrections, finite size of the nucleus, and polarization of the nucleus, among others. The set (D,E) is out of range for theorists at present.

14.4.3 Importance of the 21-cm line

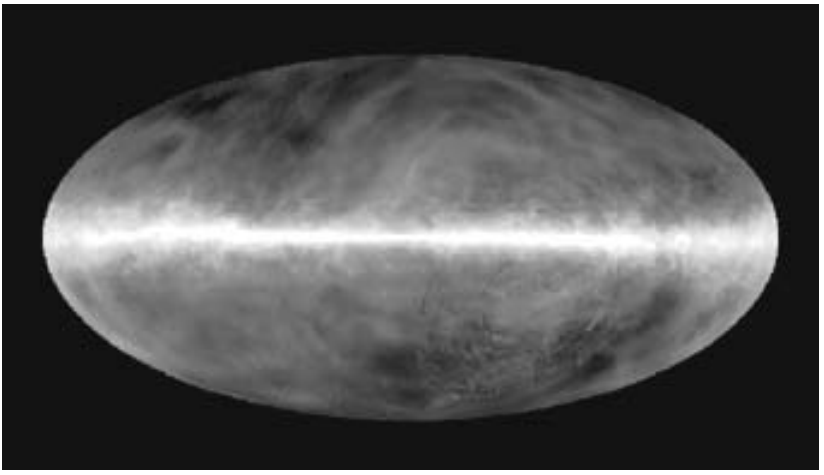


Fig. 14.13. Emission density of the 21-cm line in the sky. The emission is naturally brighter in the plane of the Milky Way where matter is closer and more dense on the average. (Photo credit: J. Dickey (UMn) and F. Lockman (NRAO), SkyView, http://antwrp.gsfc.nasa.gov/apod/ap010113.html.)

The importance of the hydrogen 21-cm line in astrophysics comes from the following facts:

- Hydrogen is very abundant.
- It absorbs any emission in the visible part of the spectrum (the sky is opaque for hydrogen lines in the visible and ultraviolet regions).
- On the other hand, cold interstellar atomic hydrogen emits this radiation abundantly. Furthermore, the interstellar medium and the entire galaxy are transparent for this radiowave. The temperature of interstellar clouds is typically 50 to 100 K. The corresponding thermal energy $kT \sim 10^{-2}$ eV is much smaller than E_I, and the atoms cannot be excited appreciably from the $1s$ state to other levels of the Lyman series. But thermal transitions

between the two hyperfine levels $S = 1$ and $S = 0$ are quite easy. The 21-cm emission corresponds to the spontaneous transition from the $S = 1$ state to the $S = 0$ state. This emission is very weak because the lifetime of the triplet state $S = 1$ is very long: $\tau \sim 3,5\ 10^{14}$ s $\sim 10^7$ years![5] Nevertheless, there are enormous amounts of hydrogen in the interstellar medium and the signals received on earth are appreciable.

- It is the only emission of hydrogen that can be collected, except for a few nearby stars. Its detection was predicted by Van de Hulst in 1947; the first observation is due to Purcell in 1951.

Its observation gives information on galactic and extragalactic structure and dynamics, and on the formation of galaxies.

The intensity distribution of the emission of the 21-cm line is shown in Figure 14.13. We notice immediately that the emission extends in directions far beyond those of visible light (Figure 14.4). The hydrogen clouds extend far beyond stars (i.e., condensed matter) in the sky. Their study is bound to bring new and abundant information.

The observation of this atomic hydrogen line (and that due to carbon monoxide, less abundant but very luminous) profoundly changed our understanding of the interstellar medium.

14.5 The Milky Way

There are countless results. By probing our own galaxy, the Milky Way, (which is difficult because the sun lies in its plane), one has been able to reconstruct it structure and to show that it has a spiral structure, as does its sister galaxy, Andromeda, M31. It has been possible to show that it has a radius of 50,000 light-years, and that the sun lies at some 35,000 light-years from the center (see Figure 14.14). Today, astronomers tend to think it has the structure of a barred spiral galaxy (see the right-hand side of Figure 14.14). The measured average density of the interstellar medium is 0.3 atoms per cm^3, the temperature varies from 20 K to 100 K, and there is on average one interstellar cloud every 1000 light-years along a line of sight. The thickness perpendicular to the galactic plane is roughly 1000 light-years.

There was a strange discovery in 2003. The Milky Way might be performing an act of cannibalism on its satellite galaxy, Canis Major. The detailed analysis of external hydrogen clouds, and infrared matter, reveals a closed winding filament-shaped structure that passes through a dwarf satellite galaxy of the Milky Way, Canis Major, as one can see in Figure 14.15. Canis Major

[5] This long lifetime is due to two effects: the energy difference between the two levels is small; and it is a magnetic dipole transition much weaker than the electric dipole transitions. The advantage is that, symmetrically, these radiowaves are absorbed weakly.

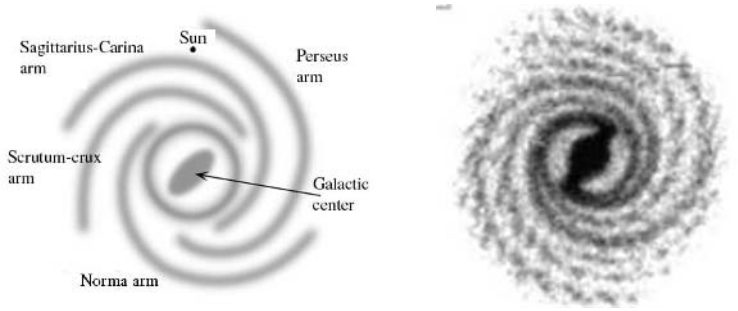


Fig. 14.14. Left: Spiral structure of the Milky Way as deduced from radioastronomical observations at a 21 cm wavelength. (Courtesy Frédéric Zantonio.) Right: Barred spiral structure of the Milky Way.

is at a distance of 25,000 light-years from the sun, it is the satellite galaxy closest to the center of the Milky Way.

Therefore, the Milky Way would be surrounded by a tidal stream of stars and matter. Such a feature is observed in other galaxies.

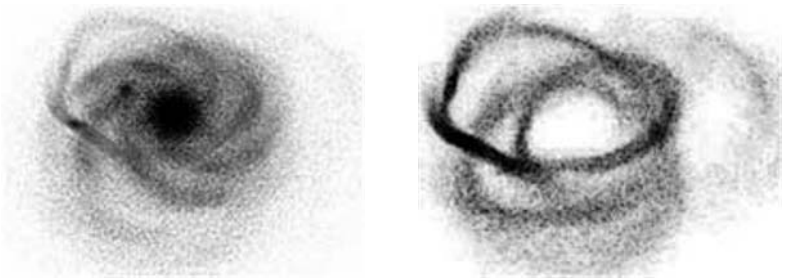


Fig. 14.15. Left: Global structure of atomic hydrogen and stars in and around the Milky Way. Right: Residual filament going through Canis Major, the darker region on the left, once the spiral arms nearest to the center of the Milky Way have been removed. (Photo credit: R. Ibata, Strasbourg Observatory, Two-Micron All Sky Survey or "2MASS" experiment, http://astro.u-strasbg.fr/images_ri/canm-e.html.)

14.6 The intergalactic medium; star wars

We now turn to a number of results obtained by observing outside our galaxy.

First, we can observe our sister galaxy, Andromeda, at two million light-years from us, twice as massive as the Milky Way. The pictures in the visible part of the spectrum are shown in Figure 14.16. One can see the spiral structure. In radiowave emission, one can see extended and powerful sources that

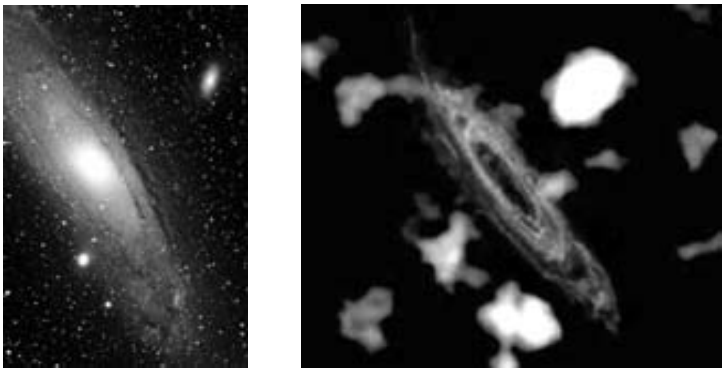


Fig. 14.16. Andromeda nebula, M31. Left: In visible light. (Photo credit: NASA and Robert Gendler, http:/antwrp.gsfc.nasa.gov/apod/ap021021.html.) Right: In 21-cm radioastronomy. (Photo credit: NRAO, Group of Astronomers at the Robert C. Byrd Green Bank Telescope (GBT) of the National Science Foundation, http://www.universetoday.com/am/publish/clouds_hydrogen_swarm_andromeda.html 422004.)

do not correspond to any visible matter. Astrophysicists tend to think that these clouds are materials which have not been used up to now in the construction of galaxies, but that will play (or are playing) a major role in galactic dynamics.

This reveals an interesting feature. In fact, atomic hydrogen extends much beyond stars in a galaxy. And that feature allows us to learn about the relationship of galaxies with respect to each other. This could not be guessed with usual telescopes on human time scales.

In the same way as stars live in an interstellar medium, there exists an intergalactic medium. Galaxies possess links. Contrary to the interstellar medium, the intergalactic medium contains practically only primordial hydrogen and helium from the big bang (there is no intergalactic nucleosynthesis).

The presence of intergalactic hydrogen clouds is obvious in the Cartwheel galaxy, which is 200 Mpc from us (1 pc = 3.26 light-years), Figure 14.17, where the contour lines of hydrogen clouds are superimposed on the optical photograph. This galaxy survived a head-on collision with a smaller galaxy 300 million years ago. This provoked the annular structure of the Cartwheel, with a rim of the size of the Milky Way and a nucleus. Numerical simulations confirm this idea. The contour lines of the important hydrogen cloud surrounding all these galaxies are marked. It leaves some doubt about which galaxy is responsible for this hit-and-run offense.

A famous example, which is very rich and was one of the first to be analyzed is the group of three galaxies M81, M82, NGC 3077 shown in Figure 14.18. This group is located in the constellation Ursa Major at 2.5 Mpc. These three galaxies (and other smaller ones) seem to orbit around each other quietly.

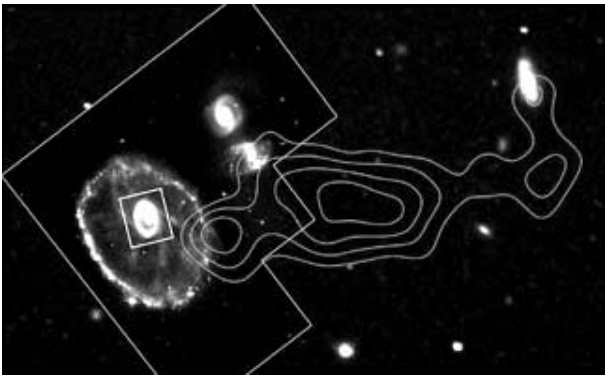


Fig. 14.17. The Cartwheel galaxy, A 0035. On the left, the ring shape, with a wheel of the size of the Milky Way and a nucleus. A hydrogen cloud joins this structure to the right hand side galaxy, and also to the middle galaxy, which is not a spiral galaxy. It is thought that this latter object, which is now in the axis of the Cartwheel at 100 kpc, had a head-on collision with it 300 million years ago, which gave rise to the structure. The extension of the hydrogen cloud leaves some doubt on the object responsible for the collision which could be the upper-right galaxy. (Photo credit: J. Higdon (NRAO), C. Struck, P. Appleton (ISU), K. Borne (Hughes STX), and R. Lucas (Stsci), NASA; http://antwrp.gsfc.nasa.gov/apod/ap970224.html.)

Fig. 14.18. Group of galaxies around M81. In the center, one can see M81, which is a large spiral galaxy, on the right, M82 which was thought to be very irregular, and on the left NGC 3077. (Photo credit: Robert Gendler, http://antwrp.gsfc.nasa.gov/apod/ap000209.html.)

M82 has an anomalous shape with a central prominence, perpendicular to its plane.

The 21-cm radioastronomical observation transforms these impressions, as one can see in Figure 14.19. One sees that the group bathes in hydrogen clouds. We notice that the spiral structure of M81 appears clearly, and that the emission at 21-cm, which is abundant in cold regions, is weak in the hot regions of galactic nuclei.

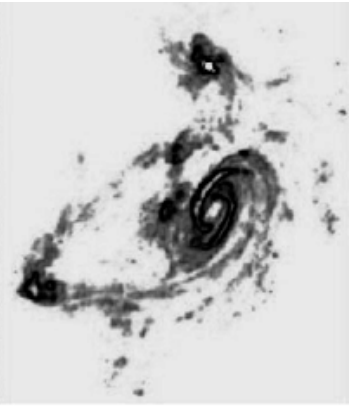


Fig. 14.19. The group of galaxies M81, M82, NGC3077 as seen on the 21-cm wavelength. The extension of the hydrogen cloud inside which the group is evolving is very large. One can see that M81 is eating up, on the left, a smaller galaxy, which is more difficult to see in visible light. We notice the absence of radio emissions in the vicinity of the hotter cores of galaxies. The orientation of this Figure is not the same as the optical picture of Figure 14.18; it is rotated by roughly 45 degrees. (Photo credit: Greydon Moore http://www.cosmicastronomy.com/bodes4.htm.)

These galaxies appear as floating islands in a common ocean of intergalactic hydrogen gas of a considerable extension. One observes the umbilical cords of hydrogen, in particular between M81 and NGC 3077, which show that these galaxies have a common history. One can see that the large galaxy M81 is in the process on performing cannibalism on a smaller galaxy, on its left, and of absorbing it. This phenomenon is very hard to see in visible light. Such observations are confirmed by numerical simulations.

For a long time, the galaxy M82 was considered as irregular, and that its nucleus was undergoing an eruption or an explosion. In fact, the visible picture, on the left-hand side of Figure 14.20, shows the unusual shape of a jellyfish. The radioastronomical observation of carbon monoxide CO, shown on the right, reveals that actually everything is perfectly normal. The central deformation of M82 results from important tidal effects produced on M82 by the presence of the nearby much more massive M81 galaxy.

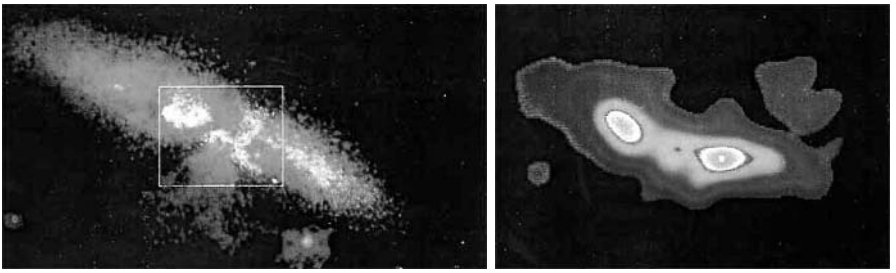


Fig. 14.20. M82 galaxy. Left: "Anomalous" aspect in the optical range with the shape of a jellyfish. Right: Radio emission at 230.54 GHz of carbon monoxide which shows a perfectly normal distribution in a ring around the galactic nucleus. (Photo credit: John Stawn Ward, Thesis www.jsward.com/publications/thesis.pdf.)

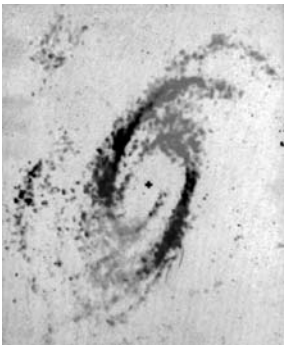


Fig. 14.21. Doppler effect on the 21-cm emission of M81 (the original picture was made with fake colors). The plane of the galaxy turns globally around an axis that is roughly at an angle of –30 degrees (black regions on this negative version of the picture) and that does not exhibit any Doppler effect in our direction. On this black and white picture, one cannot see that the upper part of the galaxy is redshifted and the lower part is blueshifted. (Photo credit: Robert Gendler, http://antwrp.gsfc.nasa.gov/apod/ap000209.html.)

Finally, the Doppler picture of the 21-cm emission of M81, shown in Figure 14.21, exhibits an unexpected feature. The plane of this galaxy rotates globally around an axis at roughly –30 degrees (black regions on the negative picture). The black and white picture does not allow us to see that the upper part of the galaxy is redshifted and the lower part is blueshifted. At this level of accuracy, no individual motion appears in the spiral arms themselves, whereas the overall rotation of the plane of the galaxy is manifest.

14.6.1 Spiral arms, birthplaces of stars

As we have just seen, the Doppler analysis of Figure 14.21 shows that if the plane of this galaxy rotates around its axis, at the scale of that rotation

the spiral arms do not move! Stars and gas clouds turn around the galactic nucleus, but the spiral arms themselves do not move! The spiral structure is a stationary situation of the gravitational field (the calculation of the field is a complicated nonlinear problem). Stars and interstellar matter turn around the center, and they slow down when they cross the spiral arms. This is why the spiral arms can be seen both in the visible part of the spectrum and in radiowaves. The spiral arms themselves do not move. They are brighter precisely because of the jamming phenomenon that happens in them, and matter is more abundant there.

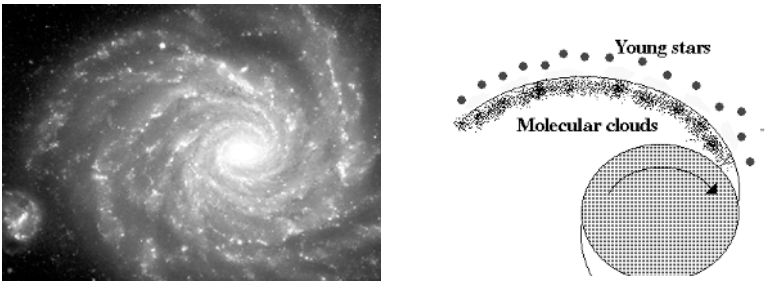


Fig. 14.22. Left: Spiral galaxy NGC 1232. (Photo credit: FORS1, 8.2-meter VLT Antu, ESO, http://www.star.ucl.ac.uk/ apod/apod/ap040125.html.) Right: Schematic star formation mechanism by compression of molecular clouds in spiral arms.

Spiral galaxies, an example of which is shown in Figure 14.22, resemble huge rotating fireworks. But, as we have just seen, the spiral arms do not move. They do not turn around!

Although it is not known at present how spiral arms are formed, one can show that if they exist, they are stable. It is a very difficult mathematical problem to explain how they arise.

Nevertheless, this observation shows a fundamental type of mechanism of star formation. Stars and the interstellar medium turn around the center of galaxies and slow down when they cross the spiral arms. During this slowing down, a "traffic jam" occurs and the molecular clouds are compressed as sketched in Figure 14.22. The compression results in an increase in temperature and in density. Locally, this led to a gravitational collapse, as described in chapter 1, followed by an ignition of thermonuclear reactions and the birth of stars in the spiral arms. This is how many stars form, in particular, massive stars.

Regions of star formation are among the most spectacular objects that one can see in the sky. In Figure 14.23 one can see one of these "star cradles," called the "Pillars of Creation" in the Eagle nebula, in the constellation Serpens. The compressed molecular cloud literally spits out bursts of young stars in the interstellar medium.

Fig. 14.23. The "Pillars of Creation" in the Eagle nebula, in the constellation Serpens. The two pictures of this star formation region are taken with different filters (and the prints do not have the same orientation). On the left, one can see the general shape of the molecular cloud in the nebula. The region of interest is a bit below the center. (Photo credit: ESA, ISO, ISOGAL Team, http://antwrp.gsfc.nasa.gov/apod/ap010914.html.) On the right, one can see a detailed picture of that region, where young bright stars are formed and ejected in the medium. (Photo credit: Jean-Charles Cuillandre (CFHT), Hawaiian Starlight, CFHT http://antwrp.gsfc.nasa.gov/apod/ap030213.html.)

14.7 Interstellar molecules, the origin of life

The observation of molecular clouds is of considerable interest because it has led in the last decades to the discovery of more and more interstellar molecules. At first, these were simply unusual astrophysical observations, but quite rapidly they raised the problem of the origin of life.

14.7.1 Rotation spectra of molecules

We have mentioned in chapter 9 the rotation spectra of molecules. Considering a molecule as a rigid rotator with principal axes x, y and z, and corresponding moments of inertia I_x, I_y, and I_z, the energy spectrum comes from the Hamiltonian:

$$\hat{H}_R = \frac{\hat{L}_x^2}{2I_x} + \frac{\hat{L}_y^2}{2I_y} + \frac{\hat{L}_z^2}{2I_z} \quad .$$

If two moments of inertia are equal, for instance, $I_x = I_y \equiv I$, and the third one is very much smaller, as is the case for a diatomic molecule, the spectrum is particularly simple. The energy levels are

$$E_{l,m} = \hbar^2 \left(\frac{l(l+1) - m^2}{2I} + \frac{m^2}{2I_z} \right) \quad .$$

The energy difference of consecutive levels increases linearly with the angular momentum ℓ,

$$E_{\rm rot}(\ell) - E_{\rm rot}(\ell - 1) = \frac{\hbar^2}{I}\ell \ . \tag{14.11}$$

Carbon monoxide CO has a length $R = 0.1128$ nm and masses $M_{\rm C} = 12$ amu, $M_{\rm O} = 15.99492$ amu, which corresponds to a transition frequency $\ell = 2 \to \ell = 1$ of 230.54 GHz.

Carbon and oxygen are comparatively abundant in the interstellar medium because these elements are synthesized in most stars. The relative abundance of the CO molecule is 10^{-5} compared to hydrogen. In general, its distribution is more dense near the center of galaxies than on the edge. This molecule radiates strongly because it has a permanent electric dipole moment, and it is easy to observe. We have already seen an example of this in Figure 14.20. This molecule is a very useful radioastronomical indicator.

14.7.2 Interstellar molecules

Many interstellar molecules have been identified by now in molecular clouds, which, in turn, generate stars. Figure (14.24) shows one of these natural sources of millimetric emission, a molecular interstellar cloud in the Orion nebula, the closest region of star formation, 1500 light-years from the sun, in the same spiral arm. At present, more than 200 molecules have been de-

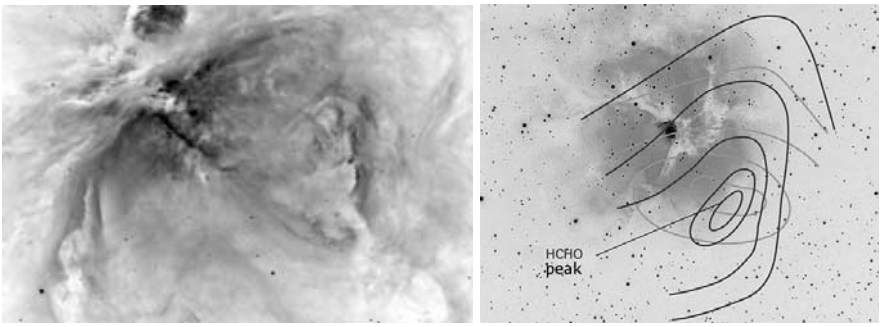


Fig. 14.24. Left: Orion nebula. (Photo credit: NASA, C.R. O'Dell and S.K. Wong, http://hubblesite.org/newscenter/newsdesk/archive/releases/1995/45/.) Right: Level curves of formaldehyde HCHO in the Trifida nebula, where stars also form. (Photo credit: James E. Brau, University of Oregon http://physics.uoregon.edu/ jimbrau/BrauImNew/Chap18/FG18_21.jpg.)

tected. Some of them are not observed in laboratories because they are too unstable at room temperature, such as acetylenic nitriles HC_3N, HC_5N, and so on, up to $HC_{11}N$. These are linear molecules whose moments of inertia can be calculated quite simply, therefore their spectra can be predicted with (14.11). (These molecules are identified by the quantum theory of angular momentum!). A well-known example is the case of fullerenes, spherical C_{60}

molecules that were identified in 1985 thanks to their spectrum, which is calculable. They were also found in meteorites, and they were synthesized in laboratories. Such molecules generated a major breakthrough in nanotechnologies. The 1996 Nobel prize in chemistry was awarded to Harold Kroto, Robert Curl, and Richard Smalley for the discovery of this new chapter of the chemistry of carbon.

Coming back to the Orion nebula, the great discovery lies in a large family of organic molecules. Figure (14.25) shows this diversity in the (small) frequency range 213 to 233 GHz. One observes:

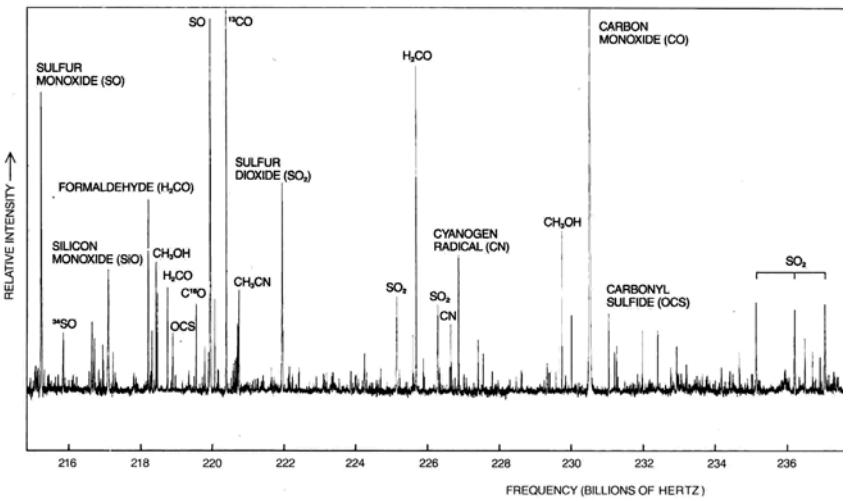


Fig. 14.25. Molecular spectrum from the Orion nebula in the frequency range 215 to 235 GHz. (Credit: Craig Kulesa, http://loke.as.arizona.edu/ ckulesa/research/overview.html.)

- The strong intensity of the ^{12}CO peak and its isotope ^{13}CO.
- Numerous organic molecules such as HCHO, CH_3OH, C_2H_5OH, and glycine, the simplest aminoacid (outside the spectral range of the figure).
- Molecules that are unknown in laboratories such as acetylenic nitriles mentioned above.

14.7.3 The origin of life

The observation of such molecules is interesting in many respects. It raises the question of the origin of life. In the spectrum of Orion, in Figure 14.25, two types of molecules are particularly interesting: HCN and HCHO, because these are very reactive molecules with which one can construct quite easily aminoacids, therefore biological molecules.

These organic molecules and the mechanism of their formation obviously raises the question of the place from where life originates. The presence of an aminoacid, glycine, whose formula is NH_2CH_2COOH is quite significant in this respect. It is quite possible that not only amino acids, but DNA can form in extrastellar regions.

However, this idea must be compared with the amazing observation made in San Diego, in 1953, by Stanley Miller, who was at that time a young 19-year-old chemistry student. He set up the initial conditions of the terrestrial atmosphere. In a mixture of NH_3, CH_4, H_2O, and H_2, he provoked artificial thunderstorms with electric discharges of 60,000 Volts. Within a week, he obtained 10 of the 20 amino acids that form living matter! This is a strong argument in favor of the terrestrial beginning of prebiotic chemistry.

This latter result has to be put together with measurements made on objects that belong to the solar system, in particular, comets and meteorites. Comets are difficult to observe. They were formed 4.5 billion years ago, at the same time as the sun, in the external regions of the original nebula. Since then, they have spent most of their time in the outer and cold regions of the solar system. Comet nuclei have evolved very little since they were formed. Their chemical composition gives an access to the chemical composition of the solar nebula when it was formed, 4.5 billion years ago. In 1997, the Hale–Bopp comet, which came close to the sun, allowed a considerable improvement of the catalogue of comet molecules. Observations at the Millimetric Radioastronomy Institute (IRAM), and at the Caltech Submillimeter Observatory (CSO) displayed the existence of seven new molecules: sulphur monoxide SO and dioxide SO_2, formic acid HCOOH, formamide NH_2CHO, cyanoacetylene HC_3N, methyl formiate $HCOOCH_3$, and ethanal CH_3CHO. These observations confirmed the presence of HNCO and OCS, which had been identified one year before in the comet Hyakutake. Two dozen interesting organic molecules have now been identified in comets.

The chemical composition of meteorites is easier to analyze. In 1969, 10 amino acids, out of the 20 which form DNA and RNA, were discovered in a meteorite that landed in Murchison, Australia. Most of them were the same as those found by Miller. They were all present in the racemic form (equal amounts of enantiomers) which excludes any terrestrial contamination. It is difficult to avoid relating this observation with the richness of interstellar molecules.

Therefore, the hypothesis that life can have an extraterrestrial or an interstellar origin cannot be dismissed. It may be that a conjunction of both processes is even more favorable. For instance, life could originate in the encounter of extraterrestrial amino acids and terrestrial nucleic acids. This field of research, quite close to fundamental research, opens fascinating perspectives.

14.8 Where are they? Quantum mechanics, the universal cosmic language

To end this story, let's come back home in order to discover an unexpected application of the 21-cm line of hydrogen.

Because a fundamental question remains: are we alone in the Universe?

14.8.1 Life, intelligence, and thought

This question is not science-fiction, it is no longer philosophical, but it is a genuine scientific question that may find an answer during this century or the next one. Life, whatever its complexity, as shown in Figure 14.26, does not mean intelligence, let alone thinking.

Fig. 14.26. Living and variously intelligent terrestrial beings. The South-African cheetah and ostrich are on both sides of Buridan's ass.

One of the quests of the space conquest adventures is to establish communication with possible extraterrestrial intelligence. In NASA, there exists a serious-minded committee in charge of studying this problem, SETI (Search for ExtraTerrestrial Intelligence), which was created in 1960, and where several Nobel prize winners collaborate.

The definition of words is subtle. Many people are convinced that one day or another, extraterrestrial life will be discovered in the simplest form of viruses. This is not what we are interested in here.

Intelligence exists among animals. There is an individual intelligence. Predators are capable of elaborating strategies, of memorizing them during REM sleep. When young birds learn to fly, they observe adults, they analyze gestures, and they memorize them. The "collective intelligence," often called "instinct" for convenience, which appears among social insects, in particular hymenoptera, is an endless source of amazement.

But there is also the problem of thinking. We can recall Descartes: "But what am I, really? Something that thinks. What is something that thinks? It is something that can doubt, that can conceive, that can claim, that can deny, that wants, that does not want, that imagines also, and that feels."

Mankind possesses two fundamental differences with all other animals (at least in the present context of the evolution of species), language and the ability of abstraction. We do not enter here into any discussion on which of these two properties causes the other. But these two properties demonstrate a radical difference between humans and other mammals many of which possess a brain which is not that different from ours. Language, be it spoken or written, is by itself an abstraction that transforms thought in signals that can be received, recorded, and understood. The various groups of human beings speak different languages, but all of them speak and some know how to translate a language into another one.

The ability of abstraction allows us, among other things, to construct other languages. Musical writing is an example. The most fascinating is probably mathematical language. Mathematics is a phenomenal tool that the human mind constructs and improves endlessly in order to understand the world.[6]

This capacity leads to acts that do not correspond to any logical necessity. Examples are the act of creation, be it artistic or scientific, the association of ideas, philosophy,[7] laughter, or dying for ideals.

Language allows a fundamental action for humans: to pass on one's thoughts and one's life-experience to other humans, and to learn theirs in turn. Mankind can preserve the memory of all that knowledge in an irreversible way, at the individual scale as well as at the scale of the society. (It is needless to say how much technical progress such as printing or computer science enhances these possibilities.)

Two hundred thousand years ago, Homo sapiens appeared. One hundred thousand years ago they invented the worship of the dead and art, two practices which have a priori no logical need. Twenty thousand years ago, mankind measured time and learned how to predict the state of the sky. Forty six hundred years ago, the pyramids were constructed. Twenty six hundred years ago, the Greeks simultaneously invented arithmetics, geometry, and democracy. A little less than one thousand years ago the Cathedrals and the Angkor temple were constructed. Eighty years ago, quantum mechanics was discovered.

The miracle of language (of successive languages) is that although it has taken 200,000 years for Homo sapiens to get to quantum mechanics (or cell phones), nowadays 20 years are sufficient for humans beings, after their birth, to understand it and use it.

One agrees on the fact that the essential part of all that happens inside the brain. But one also agrees that the brain cannot completely understand itself. For many people, this is a logical impossibility.

This is one of the many reasons for communicating with possible extraterrestrial "thinking" civilizations, or "intelligent" civilizations to be in agreement with the usual way of putting things. If we were to observe other in-

[6] I am grateful to Jean-Michel Bony for an enlightening discussion on this subject.
[7] "Making fun of philosophy, is truly philosophizing."Blaise Pascal, Pensées I,4.

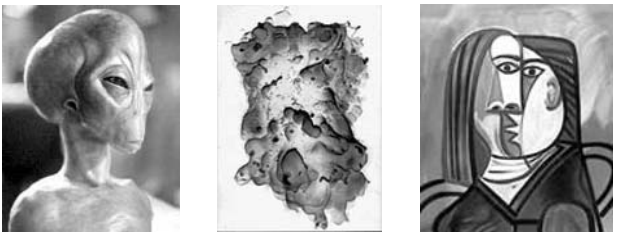


Fig. 14.27. Are extraterrestrials like this (left), like this (center), or like that (right)? (*Stargate SG-1*, Brad Wright, MGM; Blood cell fighting a virus; Picasso *Femme assise dans un fauteuil*, 1943.)

telligent beings, we might make progress in that direction. Perhaps we would discover unexpected mechanisms in the act of thinking.

14.8.2 Listening to extraterrestrials

Before all, we must know what we're talking about. Assume the beings that we are talking about exist; call them extraterrestrials. By definition they are very intelligent. We do not know anything about them, about their shape, or where they live; they might be like butterflies, or blood cells, or clouds, or something else. All possibilities exist. Some examples are shown in Figure 14.27.

Fig. 14.28. Radiotelescope of Arecibo in Puerto Rico. The antenna lies on the bottom of a 300-meter diameter volcano.(Photo credit: National Astronomy and Ionosphere Center, Cornell U., NSF http://antwrp.gsfc.nasa.gov/apod/ap981129.html.)

One thing seems quite likely. If they exist there must be many groups of them in the planetary systems around the 200 billion stars of the Milky Way. Our technological development started, say, 10,000 years ago, which is tiny

compared to the 4.5 billion year timescale of the solar system. Roughly half of the extraterrestrial civilizations must be ahead of us, therefore as advanced as we are, if not much more. And these beings must ask the same questions as we do. After all, we are not the center of the Universe; there are certainly some extraterrestrials who are more intelligent than we are and who have spotted us. Obviously, those guys must be trying to get in contact with us.

In order for them to send us messages, efficiency requires that they use a radiofrequency that reaches us with sufficient intensity. A radiotelescope is the ideal receiver. The greatest fixed radiotelescope presently installed is located at Arecibo, on the island of Puerto Rico which is shown in Figure 14.28. Its parabola is located at the bottom of an extinct volcano on which tropical vegetation grows. Its fixed antenna is 300 meters in diameter and it is capable of receiving signals from a civilization similar to ours at a distance of 15,000 light-years if the emissions come from a similar device.

As to which frequency one should tune in on, the initiators of the SETI program, Giuseppe Cocconi at Cornell University and Philip Morrison[8] at MIT, immediately pointed out that if extraterrestrial civilizations existed, they would most probably manifest themselves by emitting radio signals at the hydrogen frequency (1420 MHz). That is the universal radiofrequency, and by far the best-known frequency.

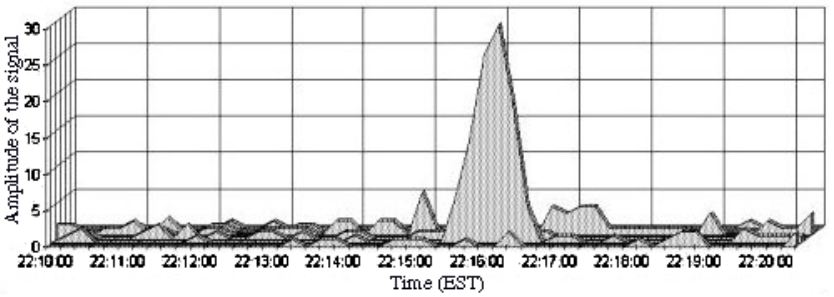


Fig. 14.29. WOW signal of August 15, 1977. (Photo credit: Astrosurf http://www.astrosurf.org/lombry/Images/seti-wow-histogramme.gif.)

In May 1960, the Green Bank radiotelescope registered a brief but strong signal coming from the direction of Epsilon Eridani on precisely 1420 MHz. Ten days later, the same signal happened again. One can understand that the scientific community was excited! Soon after, it was released that a secret military airplane had, by accident, flown above the site at high altitude when the recordings were made. (That was certainly cheaper for the taxpayer than to let the directors of SETI invest in too large equipment, with the support

[8] G. Cocconi and P. Morrison, "Searching for interstellar communications," *Nature*, **190**, 19, 1959.

of the general public as one can imagine.) The same type of phenomenon, named "WOW" which can be seen in Figure (14.29) was recorded on August 15, 1977 at Ohio State University exactly on 1420 MHz, with, most probably, the same explanation.

In July 1967, Antony Hewish, of the University of Cambridge, started recording more and more data coming from sources that emitted in a very precise manner radio signals whose regular periodicities were different from one source to another but all of the order of one second, as shown in Figure 14.30. The first observations were actually made by his student Jocelyn Bell Burtnell. These were human timescales! Many of his colleagues urged him to identify those radio signals with extraterrestrial signals. Hewish himself hesitated.[9] His caution was the right choice. He had made a decisive discovery in astrophysics: he had discovered pulsars, that is, extremely dense neutron stars rotating rapidly. He shared the 1974 Nobel prize with Martin Ryle for that discovery.

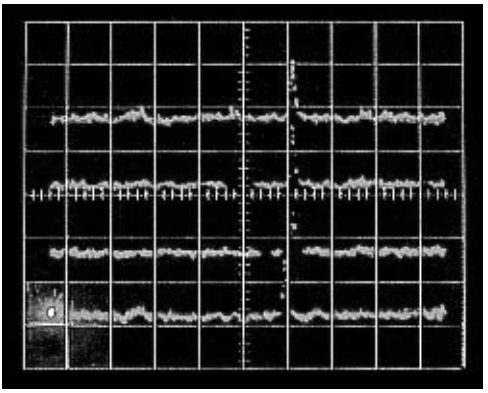


Fig. 14.30. Pulsar-blip signal such as those recorded by Jocelyn Bell Burtnell and A. Hewish in July 1967. (Photo credit: NASA/STSCI, http://www.astrosurf.org/lombry/seti4.htm.)

14.8.3 Quantum mechanics, the universal cosmic language

Now, we must face the facts. Up to now, not a single believable signal has ever been recorded in more than 40 years. It may be of course that, by chance, we are more advanced (taking into consideration the light-cone of relativity), or that other extraterrestrials are particularly stupid.

[9] The first name for the source was LGM-1 (for little green men); soon after it became CP 1919, now known as PSR 1919+21.

Therefore, in the absence of signals from space, the decision was made to send messages. Many radio messages have been sent in various promising directions, but the distance forbids our hoping to get an answer before a while.

However, written messages have also been sent on the space probes Voyager and Pioneer. This amounts to sending cosmic bottles into space with the hope that some day, in millions of years, some extraterrestrial will pick one up in his garden and will read it.

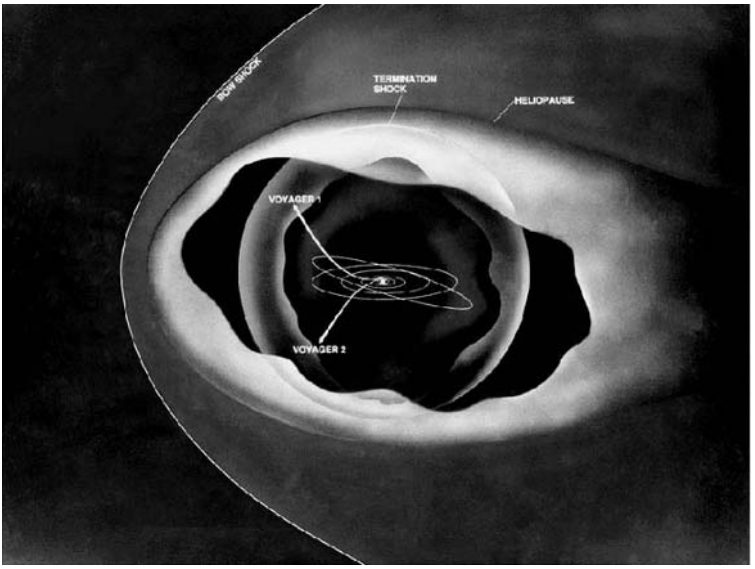


Fig. 14.31. Trajectories of the Voyager probes that have now left the solar system. (Photo credit: NASA Space Projects, http://spaceprojects.arc.nasa.gov/Space_Projects/pioneer/f23.gif.)

The trajectories of the Voyager 1 and 2 probes, launched at the end of the 1970s, are represented in Figure 14.31.

Now comes the big question: how will we communicate? This is a question of code, like the Rosetta stone. On that stone there is a text written by a group of priests to honor the Egyptian Pharaoh Ptolemy V. The message was meant to be understood by everybody. Therefore it was written in two languages, Egyptian and Greek, and three scripts, Hieroglyphic which was the honorable holy Egyptian script, Demotic, the local script, which has some similarities with Arabic, and Greek which was the universal language in the Mediterranean at that time.

Until the discovery of quantum mechanics, it was totally impossible to find the universal language of the cosmos. Here is how things work.

The simplest, most universal language of the cosmos is the structure of the hydrogen atom, which is known quasi-perfectly (provided one knows physics,

but again, by definition they are at least as advanced as we are, so they must know quantum mechanics). And, in that atom, the simplest structure is the hyperfine splitting and the 21-cm line (14.10).

That is the simplest fundamental physical structure. A splitting between two spin configurations, parallel spins ↗↗, or antiparallel spins ↗↙ generates the emission and absorption of a radiowave of 1420 MHz, that is, 21-cm.

$$\begin{array}{ll} \text{parallel spins} & \nearrow\nearrow \\ \text{antiparallel spins} & \nearrow\!\!\updownarrow\!\!\swarrow \end{array} \quad \begin{array}{l} \nu = \Delta E/h = 1420\,\text{MHz} \\ \lambda = c/\nu = 21\text{cm} \end{array} \tag{14.12}$$

This structure (14.12) provides us with a universal scientific language to start communicating in the cosmos. In fact it gives a length unit, 21-cm, and a frequency unit 1420 MHz, that is, a time unit. And all our observations of the physical world always boil down to space and time measurements.

Therefore, the decision was made to send messages on the spacecrafts Pioneer and Voyager.

But, what can we say to the extraterrestrials? Certainly not: How are things? One must, as in any correspondence, first say where and when we are, where the spacecraft comes from, and what our place looks like.

The result, which was put on the Pioneer 10 spacecraft is shown in Figure 14.32.

The code is simple. At the top left, in the circular symbol of the hydrogen atom, the parallel and antiparallel couples of arrows are the symbols of the hyperfine transition. This provides a length unit and a time unit. The line **I** between both symbols symbolizes the length unit (21-cm) which, at the scale of the drawing of the spacecraft, is in the correct proportion compared to the actual size of the spacecraft, so that they really understand what we mean. You have in front of you the simplest, most universal language to communicate with intelligent people. With other people, one can speak French, Morse, Esperanto, or Volapük.

We want to tell the date and the place. We want to locate where we are in space and time. The directions of the main pulsars around us are represented together with their frequencies expressed in the binary system in the unit **I**, that is, 1420 MHz. The variation of the frequency of pulsars with time is a universal law. If we tell the frequency of a given pulsar, then one can immediately tell when that frequency was measured. Pulsars are cosmic clocks. That space–time configuration has a large probability of being unique.

So, if we assume that they are not too stupid, and they have kept archives (as the Chinese astronomers did) they will quickly determine from which star the message came.

Next, one shows our daily environment, the solar system, and the spacecraft trajectory. They will understand that the spacecraft was launched from the third planet (distances are given in the binary system with the unit **I**) and that it followed the drawn trajectory.

The result is impressive. The message was handed over to 300 scientists and half of them understood it correctly. That's huge! But unanimously, they

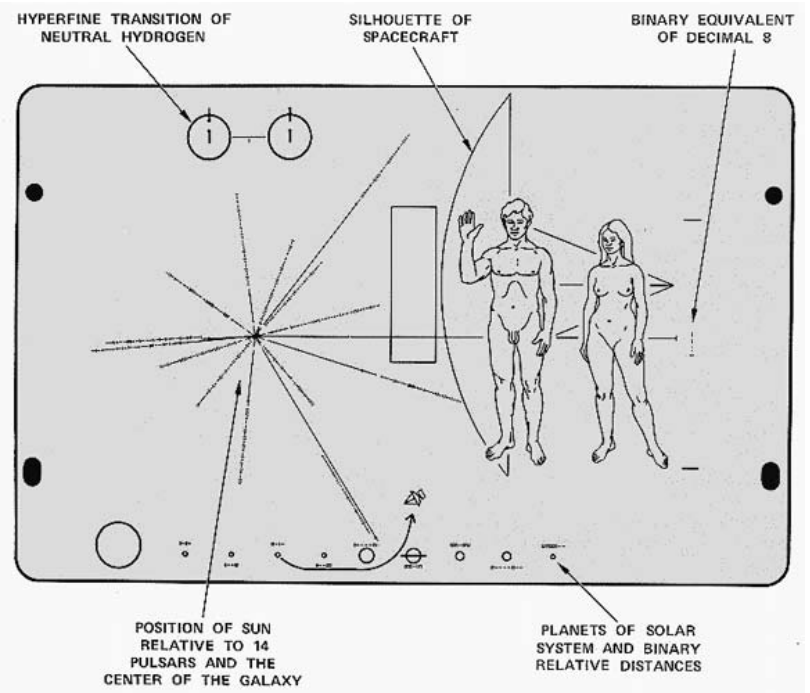


Fig. 14.32. Plaque on the Pioneer 10 spacecraft. (Photo credit: NASA Space Projects http://spaceprojects.arc.nasa.gov/Space_Projects/pioneer/PNimgs/Plaque.gif.)

all agreed that the drawings representing two human beings were totally incomprehensible.

An example of a possible misunderstanding is the following. Suppose the message falls in a world of spiders. When they look at the left-hand drawing, they will say, "Well, well! Here are fellow creatures!" Now if they count they will be horrified and say, "But these are monsters!" In fact, spiders have 10 legs, 4 pairs of walking legs plus the two chelicerae, which correspond to the antennas of insects. And the drawing shows 14 legs, that is, 4 too many!

Now, on the right, let's look more closely at those shapes. To first approximation, they are alike. The only real difference lies in the two appendices. On one of them, they are parallel; on the other they are antiparallel. These people have a really funny picture of the hydrogen atom and its hyperfine structure!

Therefore, on the other space probes, this part of the message was suppressed, and a more sophisticated message was sent with other information. In particular there is a CD recording of human voices and songs. (Most probably when they hear those atrocious sounds they will think that it comes directly from hell!).

One may wonder. The first time I saw this message and I got to know its meaning, I had two little boys (I still have them, but they aren't little boys

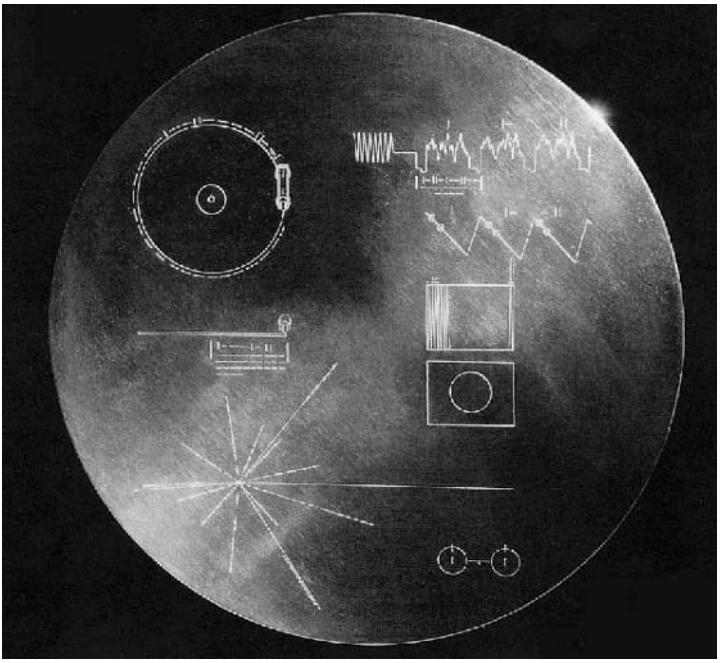


Fig. 14.33. Voyager Cover. (Photo credit: NASA Space Projects http://voyager.jpl.nasa.gov/spacecraft/images/VoyagerCover.jpg_2big.gif.)

any more) who found it fun. We used to go to the Louvre Museum on Sunday mornings.

They were interested by more or less disgusting things such as Egyptian mummies. And one day they suddenly stopped in front of an Egyptian papyrus, they called me and said "Look, we've found a message for extraterrestrials!"

It is a funeral text, a papyrus of the 21st dynasty (1000 BC), which was placed in the coffin of the pharaoh. It was supposed to protect him against various dangers during his trip to the sun, his final destination. The deceased pharaoh sits on his funeral boat and travels on a solar light ray toward his heavens. He is surrounded by familiar objects and symbols that are meant to protect him during this last trip.

What had stopped Nicolas and Olivier is that it is basically the same message as Figure 14.32! It has the same structure, the boat like the spacecraft, the light ray like the trajectory, and familiar objects like the sun and planets. And, furthermore, it contains the same mistake: the man and the woman; the pharaoh is between the sky (Nout) and the earth (Geb) whose arms are not oriented in the same way!

And, it is meant to be understood by people in the beyond, whom we don't know anything about!

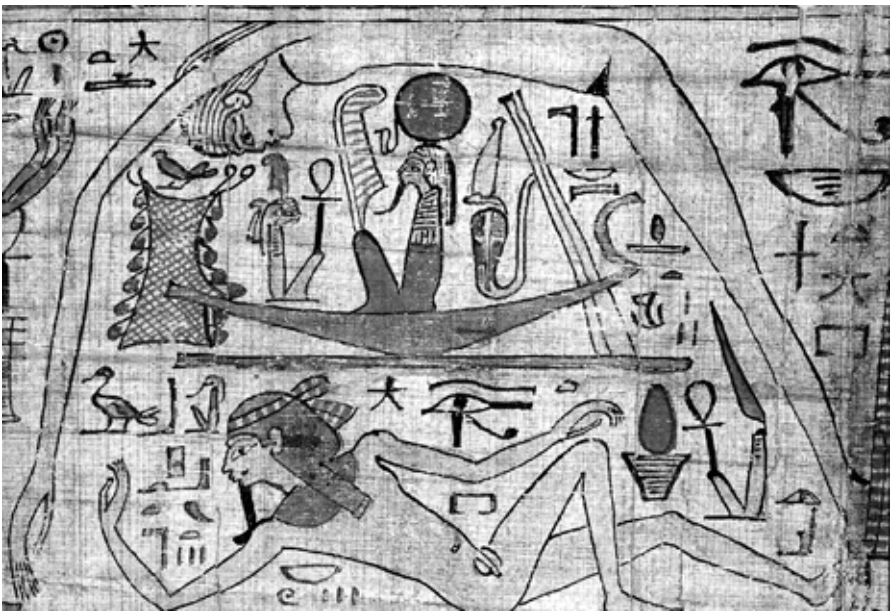


Fig. 14.34. "Man and the Cosmos," 21st dynasty (mythological papyrus of Neskapashouty, scribe accountant of the grains of the corn loft of Amon) Le Louvre Museum, Paris. This document, made 3000 years ago, was placed in the coffin of the pharaoh. The text contains a set of magic formulae that allow the king to protect himself from the dangers of the other world during his trip to his final destination toward the sun.

Here, as the above scientists, one can guess the possible misunderstandings. Just suppose we are extraterrestrials and we receive this message.

• The code is obvious. Undoubtedly there is a line that must represent the 21-cm line. Geb and Nout are representations of the hydrogen atom with the hyperfine structure represented by the relative orientations of their arms. We won't quarrel with these people because of any esthetical considerations.

• If one looks carefully, below the right of Nout's stomach, it is even written that it is 21. Obviously, the Egyptians had understood the hydrogen 21-cm line!

• The only ambiguous and unsolved question is the following. On the picture, these people must have drawn representations of themselves, of what they look like. But where are they? To what objects do they correspond?

This guy, with the sharp tip on his hat and the power symbol in his hands, who is speaking to a figurehead standing to attention, is most probably their chief.

But, for the others, it's really a difficult task to choose between all of these:

 ?  ?  ?

One thing, for sure, is that given the scale of the 21 centimeter line they are very small, just a few centimeters high.

Another thing is sure, they are not as beautiful as we are.

Fig. 14.35. Contemporary thinking extraterrestrials.

Index