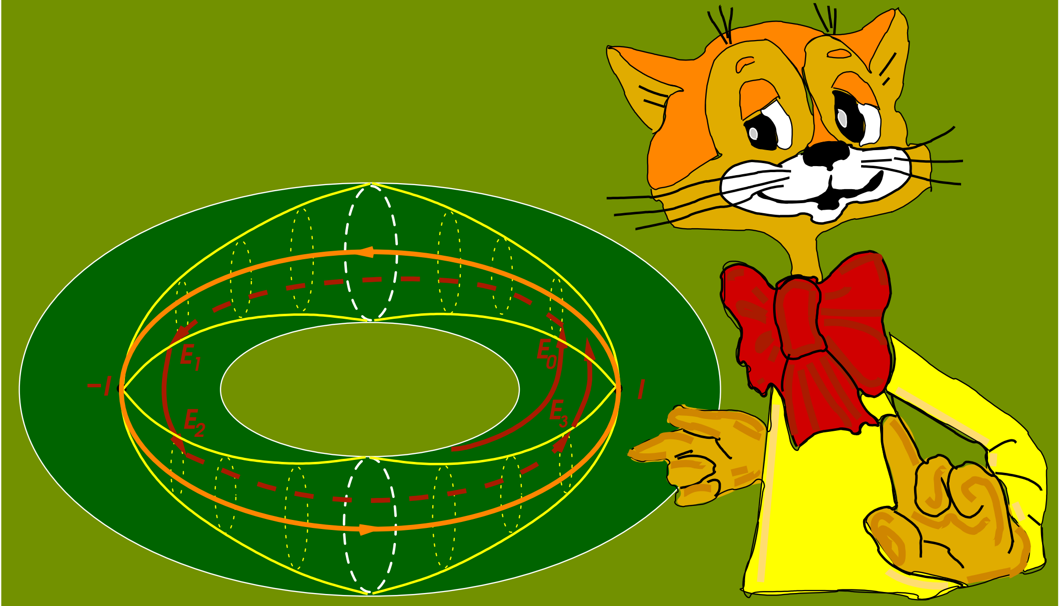
Introduction to Quantum Mechanics



Alexander
Givental

Introduction
to
Quantum Mechanics

by

Alexander Givental

Sumizdat

Published by Sumizdat
5426 Hillside Avenue, El Cerrito, California 94530, USA
http://www.sumizdat.org

ISBN 978-0-9779852-2-7

Contents

Preface

> *Israel M. Gelfand, at his 90th anniversary symposium, after the talk by Robbert Dijkgraaf on string theory:*
>
> We should all learn from Robbert. All these strings, schmings will go away, but the beautiful mathematics remains.

Creative urges often grow out of frustration. I encouraged my son to proceed with his college physics sequence by taking a first course in quantum mechanics. Browsing his required textbook, however, proved disappointing. So, by the spring break I was finishing my own.

The introductory quantum mechanics course I envisioned had to culminate at the hydrogen atom model, and show how the periodic system of elements can be explained by solving at the quantum level that same Kepler problem, solving which at the classical level allowed one to understand the solar system. This route from Newton to Schrödinger would also highlight the development of mathematics from multivariable calculus to the elements of representation theory.

Mathematicians are usually familiar with the machinery of quantum mechanics. In particular, quantum-mechanical notation is frequent in my own research papers. This doesn't mean at all that I understand quantum physics. So, I was compelled to read the very same textbook that mine was meant to augment or even replace.

Most undergraduate textbooks are "once you put it down, you simply can't pick it up" kinds of beasts. I suspect there is a close connection between the weight of a text in pounds and the number of digits used to label displayed formulas in it, or the factor by which it can be reduced without damaging the content. Some of my colleagues cynically admit that they cannot read the texts they teach from. But having put on the shoes of a student, I didn't have a choice.

Soon I encountered more fundamental difficulties. The subject begins with a description of some celebrated experiments which are meant to convince the student that quantum phenomena exist. Most experiments involve photons, which, naturally, move with the speed of light. Therefore they are relativistic creatures. The trouble was the quantum mechanics presented in the course was non-relativistic. Not only did I understand little at the beginning — there was little hope to understand anything by the end of it.

Consequently, to the mission of introducing quantum mechanics the way I myself wanted to be taught, there joined another one: to learn enough theory so that I could make sense of the early history of the subject including those experiments. My understanding of the latter is presented in the last, 14th section, which may be viewed as a supplement to the first thirteen. But it cannot be logically moved up in the pile: something from each of the previous sections is summoned there. Yet, *something* doesn't mean *everything*; in the contrary, heeding Gelfand's advice, in the rest of the book I did not hesitate to demonstrate the beauty and depth of mathematics somewhat beyond the bare practical necessity.

As a result, the mathematical temperature in the book gradually rises. If the starting level is the standard multivariable calculus and linear algebra, section 2 introduces Poisson brackets, and the idea of a group. Section 3 brings in asymptotics of oscillating integrals. Section 5 *de facto* applies the Fourier method to solving simple PDEs. By the way, the content here is fully plagiarized from that required textbook, though the computations are much improved. Section 6 teaches Hermitian linear algebra. In section 7 about bosons and fermions, tensor products make their appearance together with rudiments of super-geometry. A concise theory of spherical harmonics is developed in section 8 about the hydrogen atom, and the full-fledged classification of irreducible representations of the special unitary group SU_2 is built from scratch in section 9 in connection with spin. A minimalist's introduction to special relativity in section 10 reaches at least the formulation of the Dirac equation and links it to quaternions introduced earlier in section 9. Section 11 on quantum statistics makes use of partition functions, as well as tensor algebra developed in section 7. The band structure of solids is discussed in section 12 only after an exposition of the spectral theory of periodic one-dimensional potentials in terms of monodromy matrices and $SL_2(\mathbb{R})$, and is followed by an essay on the KdV equation. In section 13, meant to be both the epilogue to this and prelude to a higher level course, the calculus of variations (neglected up to that point) finally appears together with the least action principle, Feynman's path integrals, and what mathematicians call Wick's theorem: the graph summation formula for the asymptotical expansion of an oscillating integral.

None of this is assumed to be known beforehand. Rather, the reader is expected to be willing and ready, if not to fully absorb, then at least to make peace with new concepts after an informal explanation and a couple of examples.

Respectively, in this book, my concern for mathematical rigor does not exceed that of a true physicist — a feature that might displease a true mathematician. Yet, this is not a genuine physics text. To build a valid theory of a phenomenon, physicists must be mindful of the effects which were neglected by their model. This requires estimation of the orders of magnitudes, competence in measurements, materials, units. My effort in this direction was microscopic, so a true physicist won't be pleased either. Thus, my only hope is that a true student would find, that the two shortcomings counterbalance each other nicely, and enjoy approaching the subject the same way that I would find enjoyable.

Alexander Givental
Department of Mathematics
University of California Berkeley
March 2019

P.S. At the end of the book, the reader will find a collection of exercises, and a set of sample solutions, which were added in the process of teaching an upper-division course at UC Berkeley based on this text in the Fall semester of 2020.

To Emile
in a hope that one day he will read it

Introduction
to
Quantum Mechanics

1 Geometrical optics

Waves or particles? This debate about the nature of light has a long history that by 2-3 centuries precedes the discovery of quantum phenomena. The dilemma can be easily illustrated and even resolved within the realm of classical mechanics, and geometrical optics is a convenient place to start.

Everyone saw how a pebble dropped into water generates a circular *wavelet*. In 1678, Dutch mathematician and physicist Christiaan Huygens suggested that propagation of light can be explained as the superposition of similar locally generated waves. Namely, according to *Huygens' principle*, one assumes that the current state of a wave-like disturbance in a medium is described by a *wave front*, the locus of points where the disturbance has reached by the current moment, and then stipulates that each point of the wave front becomes the source of a spherically spreading wavelet. The position of the wave front at a later moment (say, after one second) is then described as the outer boundary of the region disturbed by all these spherical wavelets together. Each point of the new wave front, in its turn, becomes the source of a spherical wavelet (Figure 1).

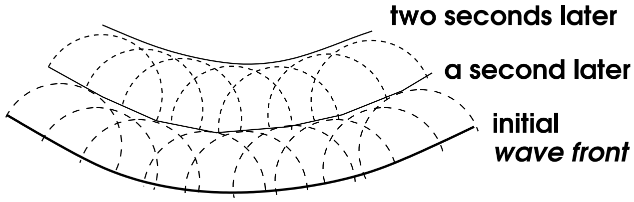


Figure 1: Huygens' principle

Now suppose that we position an observer at a point q in space, and want to find out how soon such propagation of the wave fronts will reach him. To answer this question, we should find the shortest path from q to the initial wave front, and divide the length of the path by the speed of wavelets' expansion (i.e. the speed of light if this is the phenomenon we are talking about). Of course, the shortest path will be straight, and moreover, it must be perpendicular to the initial wave front. Indeed (Figure 2), a straight segment connecting q with a point x on the initial wave front can be made shorter by moving the point x along the front *unless* the angle between the segment and the tangent to the front at x is 90°.

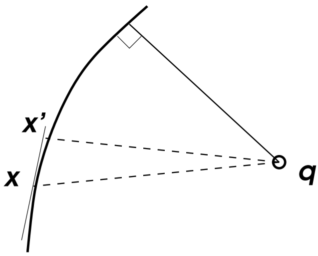


Figure 2: $|q - x'| < |q - x|$

In fact, since each wave front becomes the “initial” one at the appropriate moment, it follows from Huygens’ principle, that the shortest segment will be perpendicular to all intermediate wave fronts. Thus, given the initial wave front, one way to describe the position of the wave front t seconds later would be to draw all straight lines perpendicular to the initial one, and mark on each of them the same distance (equal to ct, where c is the speed of light). As time t varies, we obtain a family of wave fronts (in space, they will be the surfaces *equidistant* from the initial front) together with the family of lines (geometers say: *pencil of rays*) perpendicular to the family of wave fronts.

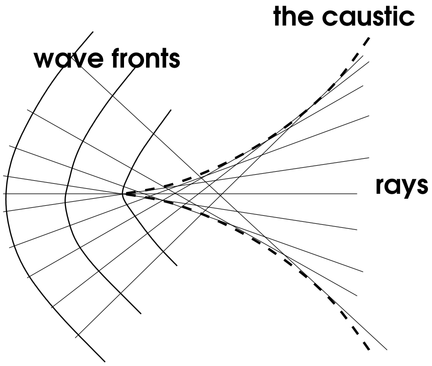


Figure 3: A pencil of rays

On Figure 3, an example of a pencil of rays emanating from a parabola and perpendicular to the family of wave fronts equidistant from the parabola are shown. The dashed bold line is *enveloping* the rays, and is called the *caustic*: the density of rays near it become infinite. In the region behind the caustic, through each point several rays are passing.

We see that the same instance of light propagation can be described in two related yet different ways: either in terms of a family of wave fronts, or as a pencil of light rays. The fronts convey the impression of a wave propagating through space, while the rays can be perceived by the observer as trajectories of particles emitted from the initial front as the source of light. These seemingly conflicting descriptions of the process are in fact logically consistent and even equivalent, as they represent two aspects of a single mathematical object: the *optical distance function.*

The eikonal equation. For the sake of simplicity, let us use the system of units in which the speed of wave propagation $c = 1$, and denote by $S(q)$ the "optical distance" $|q - x|$ from the position of the observer along a particular ray perpendicular to the initial wave front at the point x. In this system of units, $S(q)$ is the time the light particle travels from x to q along this ray. We consider S as a function in space, and in the example of Figure 3 (where the space is 2-dimensional), the graph of this function is shown on Figure 4.

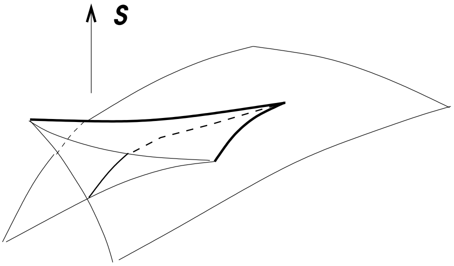


Figure 4: The swallow-tail

Comparing this picture (known as the *swallow-tail*) with Figure 3, one can notice that the function S, single valued in the region before the wave fronts reach the caustic, becomes triple-valued in the region beyond the caustic. This is because through every point q in the latter region, there pass three different rays, each bringing its own value of optical distance from the source to the point.

In terms of the optical distance function, the wave fronts are described as its *level sets* $S(q) = const.$ It might be challenging to imagine how these level sets look in the region where the function becomes triple-valued (try to meet this challenge using the name of the graph as a hint). Yet, the level sets everywhere have well-defined normal directions: these are the directions of the rays. Moreover, the optical distance grows along the rays with the speed of light, i.e. with speed 1 in our dimensionless units. Recalling the notion of the *gradient* of a function (which is the vector of length equal the maximal rate of growth of the function, and is always perpendicular to the level sets of it), we conclude that the gradient $|\nabla_q S| = 1$, or equivalently, $|\nabla_q S|^2 = 1$. This is known as *eikonal equation.*

In more realistic situation of wave propagation in space (with coordinates $q = (x, y, z)$), the optical distance function $S(x, y, z)$ satisfies the eikonal equation

$$\left(\frac{\partial S}{\partial x}\right)^2 + \left(\frac{\partial S}{\partial y}\right)^2 + \left(\frac{\partial S}{\partial z}\right)^2 = 1,$$

which is a nonlinear 1st order partial differential equation. Any solution to this equation describes a family of wave fronts, which are level surfaces of the solution, as well as a pencil of rays perpendicular to them. Thus, mathematically speaking, geometrical optics studies solutions to the eikonal equation. Any such a solution can be constructed from the initial wave front as the distance function to the front, which grows linearly with rate 1 along each ray perpendicular to the initial front.

In fact this particle-based approach is a special example of a general method (known as the *method of characteristics*) of reducing a single 1-st order partial differential equation to an ordinary differential equation (i.e. an equation whose solutions can be interpreted as trajectories of particles). In the next section, we will discuss this method in the context of Hamilton's approach to classical mechanics.

2 Hamiltonian mechanics

Hamilton equations. Newton's equation $F = ma$ can be interpreted as a second order ordinary differential equation

$$m\ddot{q} = F(q)$$

whose solutions represent trajectories $t \mapsto q(t)$ of a particle of mass m moving in any given force field F. The equation is *deterministic*, i.e. the past and future trajectory are uniquely determined by the initial conditions at the current moment t_0, and can be specified by the current position $q(t_0)$ and current velocity $\dot{q}(t_0)$.

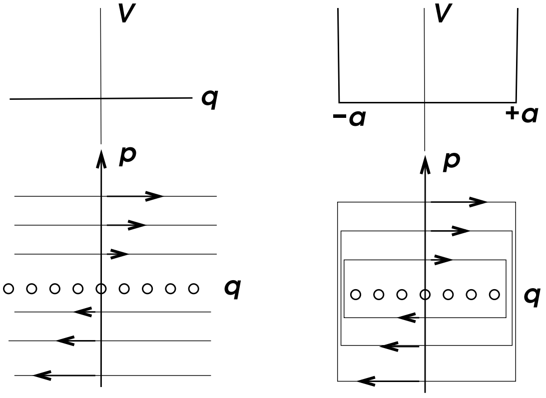
In fundamental physics, and in particular in celestial mechanics, one is primarily concerned with *conservative* force fields, i.e. the situation when the vector-valued function $F(q)$ is constructed from the gradient of a scalar-valued *potential energy* function: $F(q) = -\nabla_q V$. In this case, the equation of motion can be rewritten as the following system of *Hamilton equations*:

$$\left\{ \begin{array}{rcl} \dot{q} & = & \partial H/\partial p \\ \dot{p} & = - & \partial H/\partial q \end{array} \right. .$$

Here p is the *momentum* vector, $p = m\dot{q}$, and $H(p, q)$, called the *Hamilton function* or the *hamiltonian* of the system, is a scalar-valued function, namely the total (kinetic plus potential) energy:

$$H(p, q) = \frac{p^2}{2m} + V(q) = \frac{m\dot{q}^2}{2} + V(q).$$

The last formula is written as if there is only one q and one p (which would be true for a particle moving on a line), but also makes sense for a particle in space, if the vector notation is assumed. In components, we would have $q = (q_1, q_2, q_3)$, $p = (p_1, p_2, p_3)$, the notation $p^2/2m$ for the kinetic energy would hide the dot-product: $(p \cdot p)/2m = (p_1^2 + p_2^2 + p_3^2)/2m$, and the Hamilton system would consist of 6 first order ordinary differential equations: $\dot{q}_i = \partial H/\partial p_i$, $\dot{p}_i = -\partial H/\partial q_i$. Taking into account the explicit form of the Hamilton function, we find $\dot{q} = p/m$, $\dot{p} = -\nabla_q V$, which combine into Newton's 2nd order equation $m\ddot{q} = -\nabla_q V$.


(a) Free particle **(b) Infinite well**

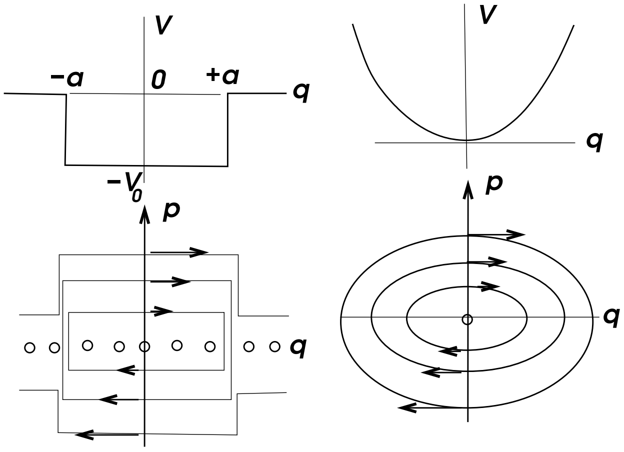

(c) Finite well **(d) Harmonic oscillator**

Figure 5: One degree of freedom

Examples in one degree of freedom.

A free particle on the line: $H(p, q) = \frac{p^2}{2m}$. There is no force field here, the equations of motion is $\dot{q} = p/m$, $\dot{p} = 0$, i.e. $\ddot{q} = 0$, and the solutions $q(t) = q(0) + vt$ describe the motion of the particle on the line with some constant speed v, i.e by inertia (Figure 5a).

One-dimensional billiard. Imagine the graph of the potential energy function as a "well" with the walls of infinite height (Figure 5b). The particle is still "free", but in the totalitarian meaning of this word, as it is forced to stay on the interval $(-a, a)$ of the q-axis. The physical phenomenon best described by this system is that of a billiard ball bouncing off the walls of the "well".

The finite square well (Figure 5c):

$$V(q) = \left\{ \begin{array}{cc} 0 & |q| > a \\ -V_0 & |x| < a \end{array} \right. .$$

When the particle is outside the well and has the initial kinetic energy greater than V_0 it will speed up while moving through $-a < q < a$, but past the well will slow down to the initial velocity. However, a particle with the kinetic energy below V_0 will bounce between the walls $q = \pm a$ like the billiard ball in the case of the infinitely deep potential well.

The harmonic oscillator:

$$H(p, q) = \frac{p^2}{2m} + K\frac{q^2}{2},$$

where K, in the interpretation of the oscillator as a mass-spring system, is the rigidity coefficient of the spring. The equations of motion are $\dot{q} = p/m$, $\dot{p} = -Kq$, i.e. $\ddot{q} = -(K/m)q$. The trajectories on the pq-plane are ellipses $H = const$ (Figure 5d), and all the motions are periodic with the same angular frequency $\omega = \sqrt{K/m}$.

The Kepler problem. Perhaps the most interesting mechanical system we will encounter is the motion of a particle in 3-space under the central force inverse proportional to the square distance to the center:

$$H(p, q) = \frac{p \cdot p}{2m} - \frac{G}{|q|}.$$

The Hamilton equations have the form $\dot{q} = p/m$, $\dot{p} = -Gq/|q|^3$, and show that the *angular momentum* vector $L := q \times p$ is conserved:

$$\dot{L} = \dot{q} \times p + q \times \dot{p} = (p/m) \times p - q \times Gq/|q|^3 = 0.$$

One consequence of this is that the direction of the cross-product $q \times p$ does not change in time, i.e. each trajectory remains in the plane determined by the initial position of the vectors q_0 and p_0. This reduces the problem from three to two dimensions.

The fact that not only the direction, but also the magnitude $|L|$ of the angular momentum 3-vector is conserved, translates into Kepler's 2nd law: the *sectorial velocity* is constant, i.e. the radius-vector sweeps equal areas in equal times. Using this, one can reduce the problem further to one degree of freedom, and obtain Kepler's 1st law, according to which the trajectories are conic sections (ellipses, hyperbolas, parabolas) with the origin $q = 0$ playing role of a focus.

Namely, let $r = (q \cdot q)^{1/2}$ denote the distance from the particle to the origin on the plane of motion. By differentiating twice in time, we find:

$$\dot{r} = \frac{(\dot{q} \cdot q)}{r}, \quad \ddot{r} = \frac{(\ddot{q} \cdot q)}{r} + \frac{(\dot{q} \cdot \dot{q})}{r} - \frac{(\dot{q} \cdot q)^2}{r^3}.$$

Note that

$$(\dot{q} \cdot \dot{q})(q \cdot q) - (\dot{q} \cdot q)^2 = |\dot{q}|^2 |q|^2 \sin^2 \theta,$$

where θ is the angle between the radius-vector q and the velocity $\dot{q}$. This expression is equal therefore to $|L|^2/m^2$, and is conserved along each trajectory. Combining this with the Newton equation $m\ddot{q} = -Gq/r^3$, we obtain

$$m\ddot{r} = -\frac{G}{r^2} + \frac{|L|^2}{mr^3}.$$

This can be interpreted as the Newton equation of a particle of mass m with one degree of freedom moving in the potential field with the *effective potential* $V(r) = -G/r + |L|^2/2mr^2$. The corresponding Hamilton function (depending on the value of $|L|$)

$$H(p, r) = \frac{p^2}{2m} + \frac{|L|^2}{2mr^2} - \frac{G}{r}$$

will reappear when we will study the quantum Kepler problem.

The effective equation can be solved by the usual technique of analyzing conservative systems with one degree of freedom. However, there is an elegant way, due to Joseph Louis Lagrange, to avoid computations and link the solutions directly to conic sections. Namely, consider the cone in space with coordinates (x, y, r) given by the equation $r^2 = x^2 + y^2$, where (x, y) are the components of the radius-vector q on the plane of the motion. Let a parametric curve $t \mapsto q(t) = (x(t), y(t))$ represent a solution to the Kepler non-linear equation $m\ddot{q} = -Gq/|q|^3$ with the sectorial velocity $|L|$, and respectively $r(t) = \sqrt{x(t)^2 + y(t)^2}$. Then, as it was noticed by Lagrange, the there functions $x(t)$, $y(t)$ and $r(t) - |L|^2/m^2$ also satisfy *the same linear 2nd order* differential equation (with time-dependent coefficients)

$$m\ddot{u} = -\frac{G}{r^3(t)} u.$$

But solutions of a linear 2nd order ordinary differential equation form a linear space of dimension two. Therefore the three solutions must be linearly dependent:

$$Ax(t) + By(t) + C\left(r(t) - \frac{|L|^2}{m^2}\right) = 0.$$

This yields the equation of a plane in the (x, y, r)-space. Its intersection with the cone $r^2 = x^2 + y^2$ is a conic section — typically an ellipse, or hyperbola. The projection of it to the (x, y)-plane (which is respectively an ellipse or hyperbola too) is the Keplerian trajectory.

We refer to our paper *Kepler's laws and conic sections* for further elementary treatment of the classical Kepler problem, and to an applet by Irina Boyadzhiev, providing a neat visualization of the results.

Symplectic geometry. At this stage one should begin to wonder what was the point of rewriting Newton's 2nd order differential equation in Hamilton's form of a system of two 1st order ones, and replacing at the same time the velocity $\dot{q}$ with the momentum $p = m\dot{q}$. What difference could this change make?

In fact the Hamilton form of the equations of motion has a number of important advantages over the old-fashioned, Newtonian approach. First the mechanical system does not have to consist of one particle, but could involve many, and not necessarily particles, but possibly more complicated objects (e.g. spinning tops). Generally speaking, the letter q represents a point in the *configuration space* of the mechanical system, i.e. the space of all possible positions of the system. (It could have any dimension, referred to as the number of *degrees of freedom* of the mechanical system). Respectively, p can represent the momentum of the mechanical system in some generalized sense, so that the pair (q, p) represents a point in the so-called *phase space* of the system (whose dimension is always twice the number of the degrees of freedom). For example, the configuration space of a soccer team has dimension 22 (two coordinates to specify the position of each player on the field), while the phase space of the team is 44-dimensional (to include the momentum vectors). Furthermore, in the Hamilton equations, the hamiltonian $H(p, q)$ does not have to be any particular expression resembling the sum of kinetic and potential energy, but could be any differentiable function on the phase space. Regardless of the expression, the "*energy conservation law*" easily follows from the chain rule and the form of the equations of motion:

$$\frac{d}{dt}H(p(t), q(t)) = \frac{\partial H}{\partial p}\dot{p} + \frac{\partial H}{\partial q}\dot{q} = -\frac{\partial H}{\partial p}\frac{\partial H}{\partial q} + \frac{\partial H}{\partial q}\frac{\partial H}{\partial p} = 0.$$

More fundamentally though (and here, risking to sound cryptic, we are going to stretch the reader's imagination to a dangerous degree), it turns out that this and many other meaningful properties which the Hamilton equations possess are explained not by their specific coordinate form, but are due to some remarkable geometric structure inherently present in the phase space of any conservative mechanical system. The structure is called *symplectic.* For those familiar with the language of differential forms we can say that it is a closed differential 2-form. E.g. in the above 44 phase coordinates (p, q) of the soccer team, it is given by the formula

$$dp_1 \wedge dq_1 + \cdots + dp_{22} \wedge dq_{22}.$$

For those unfamiliar, we can vaguely say that it is some kind of antisymmetric analog of the dot-product, and a device for computing signed areas of 2-dimensional surfaces in the phase space. This geometric approach brings into classical mechanics a lot of flexibility, which we are only going to illustrate here with the simplest nontrivial example.

Hamilton equations on S^2. On a 2-dimensional sphere, one cannot introduce a coordinate system (p, q) which would make the sphere look like the space of positions and velocities of a particle. Yet, regions of the sphere have areas, which make it a symplectic surface, and hence a legitimate phase space of a Hamiltonian mechanical system with one degree of freedom. For example, on the sphere $x^2 + y^2 + z^2 = 1$, consider the coordinate z as the Hamilton function and find the corresponding equations of motion.

Though unskilled in symplectic geometry, we can rely on the geometric fact known to the ancient Greeks: the axial projection of the sphere onto the cylinder of the same radius (see Figure 6) distorts shapes of regions, but preserves their areas. On the cylinder, introduce the coordinates (p, q), taking q to be the polar angle on the xy-plane, and $p = z$. Thus, in such coordinates, the cylinder is developed into the rectangle $0 \leq q \leq 2\pi$, $-1 \leq p \leq 1$. The point on the sphere, represented by (p, q), is found by the formulas $z = p$, $x = \sqrt{1 - p^2} \cos q$, $y = \sqrt{1 - p^2} \sin q$. In any case, the Hamilton function z in the cylindrical coordinates (p, q) coincides with $H(p, q) = p$, and the Hamilton equations assume the form: $\dot{q} = \partial H / \partial p = 1$, $\dot{p} = -\partial H / \partial q = 0$. The solutions are very simple: $q(t) = q_0 + t$, $p(t) = p_0$, and describe translation with speed 1 in the angular direction. The corresponding motion on the sphere is the steady rotation about the z axis with the angular velocity 1. The phase trajectories are shown on Figure 6 as parallel horizontal circles.

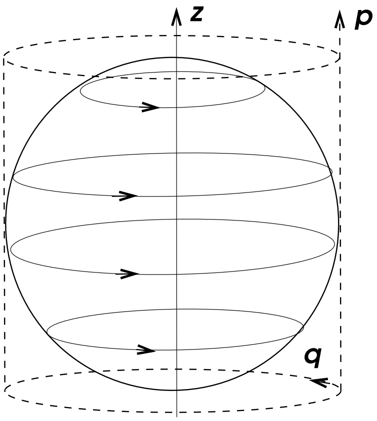


Figure 6: A Hamiltonian system on the sphere

One should ask here whether the equations of motion will remain the same if we choose to project the sphere onto another cylinder, e.g. the one around the x-axis. The answer is "yes", as the motion can be described geometrically without any reference to cylinders. Namely, by the energy conservation law, the trajectories are the level curves $z = const$ of the Hamilton function. The speed of motion along each trajectory is determined by the local density of the level curves: the phase flow on the sphere must preserves areas.

The latter fact is clear for the rotations on the sphere, but it is a manifestation of the general property of the phase flows of Hamiltonian mechanical systems: they preserve the geometric structure we called symplectic. In particular they preserve the *phase volume* of regions in the phase space. Indeed, the vector field with components $\dot{q}_i = \partial H/\partial p_i$, $\dot{p}_i = -\partial H/\partial q_i$ has divergence $\sum_i \partial^2 H/\partial p_i \partial q_i - \sum_i \partial^2 H/\partial q_i \partial p_i = 0$ by Clairaut's theorem. By Gauss' divergence theorem, the flow of such a vector field is volume-preserving. By the way this fact serves as the foundation for statistical mechanics.

Poisson brackets. There is, perhaps, a more accessible, algebraic way to describe the structure present in the phase spaces of Hamiltonian mechanics. Namely, let us think of infinitely differentiable functions $F(p, q)$ on such a phase space as physical quantities (they are called *classical observables*). For example, for a particle in $\mathbb{R}^3$, the components of the angular momentum vector $L = q \times p$ are such observables:

$$L_1 = q_2p_3 - q_3p_2, \quad L_2 = q_3p_1 - q_1p_3, \quad L_3 = q_1p_2 - q_2p_1.$$

When a point (p, q) of the phase space evolves in time according to the Hamilton equations $\dot{q}_i = \partial H/\partial p_i$, $\dot{p}_i = -\partial H/\partial q_i$, so does the value of an observable F at that point. The differential equation governing this change is found from the chain rule:

$$\frac{dF}{dt} = \sum_i \left(\frac{\partial F}{\partial q_i}\dot{q}_i + \frac{\partial F}{\partial p_i}\dot{p}_i \right) = \sum_i \left(\frac{\partial H}{\partial p_i}\frac{\partial F}{\partial q_i} - \frac{\partial H}{\partial q_i}\frac{\partial F}{\partial p_i} \right).$$

The last expression is called the *Poisson bracket* of H and F, and is denoted as $\{H, F\}$. Thus, $\dot{F} = \{H, F\}$ describes the evolution of all observables. The Hamilton equations $\dot{q}_i = \{H, q_i\}$, $\dot{p}_i = \{H, p_i\}$ are just special cases when the coordinates are taken for the observables.

Clearly, Poisson brackets are *anti*-commutative and distributive:

$$\{F, G\} = -\{G, F\}, \quad \text{and} \quad \{H, F + G\} = \{H, F\} + \{H, G\}.$$

It is not too hard to check that they also satisfies the *Jacobi identity*:

$$\{H, \{F, G\}\} + \{G, \{H, F\}\} + \{F, \{G, H\}\} = 0 \quad \text{for any } F, G, H.$$

In fact these properties amount to the definition of a *Lie algebra*. Beside this, the Poisson bracket is also a *bi-differentiation*, i.e. with respect to either operand, satisfies the product (Leibniz') rule:

$$\{H, FG\} = \{H, F\}G + F\{H, G\}.$$

This should not be surprising, since it simply says that $d(FG)/dt = (dF/dt)G + F(dG/dt)$ when the phase points move according to the Hamilton equations with the Hamilton function H.

Our point now is that an operation $\{\cdot, \cdot\}$ with the above properties given on the algebra of infinitely differentiable functions on some geometric space makes this geometric space a legitimate phase space of Hamiltonian mechanics (regardless of any specific coordinate formula for the bracket). Namely, given any hamiltonian H, the evolution of all observables is defined by the equation $\dot{F} = \{H, F\}$. Moreover, the following remarkable properties of Hamiltonian mechanics follow trivially:

- The energy conservation law: $\{H, H\} = 0$ (due to the anti-symmetry).

- *Noether's theorem* — symmetries generate conservation laws. *I.e.*, if the Hamiltonian flow defined on the phase space by a function F leaves our hamiltonian H invariant ($\{F, H\} = 0$), then F is a conservation law of our Hamiltonian system ($\{H, F\} = 0$).
- *Poisson's theorem:* If F and G are conservation laws of the Hamiltonian system with the Hamilton function H (i.e. $\{H, F\} = 0 = \{H, G\}$), then (as it follows from the Jacobi identity) their Poisson bracket $\{F, G\}$ is a conservation law too: $\{H, \{F, G\}\} = 0$.

For example, if two components of the angular momentum are conserved, then the third one is conserved as well. Indeed, the following "*commutation relations*" are straightforward to check:

$$\{L_1, L_2\} = -L_3, \quad \{L_2, L_3\} = -L_1, \quad \{L_3, L_1\} = -L_2.$$

By the way, they coincide with the cross-product rules: $e_1 \times e_2 = e_3$, etc. for $e_i = -L_i$.

Let us re-examine Hamiltonian mechanics on the sphere S^2 from this new point of view. In the 3-space with coordinates x, y, z, define the Poisson bracket by the cross-product scheme:

$$\{x, y\} = z = -\{y, x\}, \quad \{y, z\} = x = -\{z, y\}, \quad \{z, x\} = y = -\{x, z\}.$$

It is a general fact that once the (anti-symmetric) Poisson brackets between coordinates (say, x_i) are defined and obey the Jacobi identity, the operation can be extended to a Poisson bracket on all functions using Leibniz' rule:

$$\{F, G\} = \sum_{i,j} \frac{\partial F}{\partial x_i} \frac{\partial G}{\partial x_j} \{x_i, x_j\}.$$

In particular we can check that $r^2 = x^2 + y^2 + z^2$ has *zero* Poisson bracket with x, y, z:

$$\{x, r^2\} = 2x\{x, x\} + 2y\{x, y\} + 2z\{x, z\} = 2yz - 2zy = 0, \text{ etc.}$$

This means that r^2 is a *Casimir function*, i.e. it Poisson-commutes with all functions $H(x, y, z)$:

$$\{H, r^2\} = \frac{\partial H}{\partial x}\{x, r^2\} + \frac{\partial H}{\partial y}\{y, r^2\} + \frac{\partial H}{\partial z}\{z, r^2\} = 0.$$

Note that in our previous situation of the (p, q)-phase space, only constants had this property.

The consequence is that for any Hamilton function H the trajectories, found from the equations

$$\dot{x} = \{H, x\}, \ \dot{y} = \{H, y\}, \ \ \dot{z} = \{H, z\}$$

will lie on the spheres $x^2 + y^2 + z^2 = const$, i.e. r^2 is a universal conservation law. Thus, each sphere is a legitimate phase space of Hamiltonian mechanics.

To reconcile the discussion with what we have said earlier, consider the Hamilton function $H = z$. Then the equations of motion are

$$\dot{x} := \{z, x\} = y, \ \dot{y} := \{z, y\} = -x, \ \dot{z} := \{z, z\} = 0.$$

The first two equations describe on the (x, y)-plane clockwise rotations with constant angular velocity and period 2π (the same as in the case of the harmonic oscillator with the hamiltonian $(x^2+y^2)/2$), and the third one shows that z is conserved. Thus, the whole dynamics is described as the rotation about the z-axis, on each sphere $r^2 = const$ and in space as the whole.

Taking any linear function $\alpha x + \beta y + \gamma z$ on the role of the hamiltonian, we will obtain rotation in space with the angular velocity vector (α, β, γ). A fancy way to formulate this result is by saying that the space of linear functions equipped with this Poisson bracket (i.e. with the cross-product operation) forms *the Lie algebra of the Lie group SO_3 of Euclidean rotations in* $\mathbb{R}^3$.

Groups. This is not the last time we mention *groups*, and so it might be useful to say a few informal words about this notion.

The official definition is that a group is a set G equipped with an associative operation, the unit element, and the operation of inversion of all of its elements.

For example, all $n!$ permutations of the set $\{1, 2, \ldots, n\}$ of n integers form the *permutation group S_n* with respect to the operation of composition of permutations.

It is probably harder to imagine all permutations of an infinite set X. By definition, they are arbitrary *invertible mappings* from X to itself. Such mappings can be composed and inverted, and thus form a group, with the identity mapping playing the role of the unit element.

The group SO_3 of rotations of the Euclidean 3-space is a *subgroup* of the group of such permutations on the set $X = \mathbb{R}^3$. Namely, it consists of all those permutations of points in $\mathbb{R}^3$ which preserve several structures there: the structure of a linear space, the Euclidean dot-product, and the *orientation* (i.e. don't transform left gloves into right ones).

In fact this example just illustrates the universal way how groups arise. Whenever we have a set X equipped with some structure, all *symmetries* of the structure, i.e. all permutations on X which preserve this structure, form a group with respect to the operation of composition of mappings.

Indeed: composition of mappings is always associative; the identity mapping preserves whatever structure on X you have in mind; when a permutation preserves a structure, its inverse also preserves it; when each of two mappings preserves a structure, their composition also preserves it.

For example, even if we are not sure what a symplectic structure on the phase space of a Hamiltonian mechanical system is, we can be absolutely sure that symmetries of this structure form a group. It is called the group of *symplectomorphisms* of the phase space.

One of the claims we made earlier can be rephrased by saying that the transformations defined by solutions of Hamiltonian mechanical systems are symplectomorphisms. (In fact the converse is true locally on the phase space.)

The last example looks rather fancy. The groups we will really encounter in this book are: S_n, SO_3 and several of its close relatives, as well as *one-parametric groups* $t \mapsto U(t)$ of linear transformations in real and complex vector spaces (finite or infinite dimensional). Here "one-parametric" means that the transformations $U(t)$ are exponential functions of real numbers $t \in \mathbb{R}$: $U(t_1 + t_2) = U(t_1)U(t_2)$.

Most importantly, whatever a group G one ever encounters in "real life", the operation in the group is always the composition of mapping, and the group usually consists of all symmetries of a certain interesting structure.

The Hamilton-Jacobi equation. Let us return now from the skies to the ground and explain, as we have promised, in what way the wave–particle reciprocity we have seen in geometrical optics can be extended to general classical mechanical systems.

To an arbitrary Hamilton function $H(p, q)$, one can associate a 1st order partial differential equation, called *Hamilton-Jacobi equation* which generalizes that of eikonal:

$$H(\nabla_q S, q) = const, \quad \text{or} \quad H\left(\frac{\partial S}{\partial q_1}, \ldots, \frac{\partial S}{\partial q_n}, q_1, \ldots, q_n\right) = const,$$

where $S = S(q)$ is the unknown function on the configuration space.

Here is how the method of characteristics is generalized. Suppose that a function S satisfies the Hamilton-Jacobi equation, that is, the partial derivatives $p_i = \partial S/\partial q_i$ satisfy the algebraic equation $H(p(q), q) = const$. Differentiating in q_i and applying the chain rule, we find:

$$0 = \frac{\partial H}{\partial q_i} + \sum_{j=1}^{n} \frac{\partial H}{\partial p_j} \frac{\partial^2 S}{\partial q_j \partial q_i} = \frac{d}{dt}\left(-p_i + \frac{\partial S}{\partial q_i}\right).$$

The second equality is obtained by using the Hamilton equations $\partial H/\partial q_i = -\dot{p}_i$, $\partial H/\partial p_j = \dot{q}_j$ for the trajectories of the Hamilton system in the phase space, and the chain rule again. This result means that the trajectories are everywhere tangent to the n-dimensional surface given by the n equations $\partial S/\partial q_i - p_i = 0$. Therefore they lie inside this surface (Figure 7).

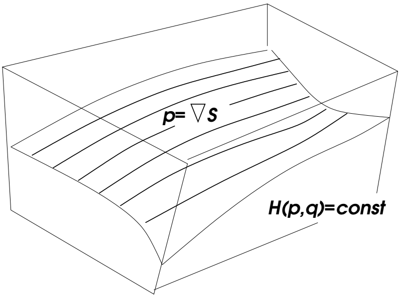


Figure 7: A solution to the Hamilton-Jacobi equation

This gives a transparent geometric interpretation to Hamilton-Jacobi equations and their solutions. The equation expresses the fact that in the phase space, the n-dimensional graph of the gradient map $p = \nabla_q S$ (in symplectic geometry, such graphs are called *Lagrangian submanifolds* of the phase space) lies inside the level set of the Hamilton function. Our computation shows that such a Lagrangian submanifold represents an $n-1$-parametric *pencil of trajectories* of the Hamiltonian mechanical system on a fixed level $H = const$ of the total energy. The function S is analogous to the optical distance function of geometrical optics, its level sets are analogous to the wave fronts, while the trajectories of the Hamiltonian system comprising the graph of ∇S are analogous to the rays, the main difference here being that the mechanical system can be arbitrary.

3 Short-wave optics

Newton's rings. While the idea that light is propagated by rays appears self-obvious, the wave nature of light seems harder to stomach. To see why, let us examine one bit of empirical evidence for it known as *Newton's rings*.

The phenomenon is schematically illustrated by Figure 8: A round lens is placed on top of a glass illuminated from below by monochromatic light, and is observed from above. The through-beam (shown in blue) is superimposed with the beam (green) reflected twice: from the surface of the lens, and then from the glass. The distances traveled by the "blue" and "green" beams differ by $2a$, which leads to the phase shift between the beams equal to $2a/\lambda \mod 2\pi$, where λ is the wavelength. When the shift is close to 0, the beams reinforce each other's intensity, and when it is close to π, they cancel. From geometry, $2a$ approximately equals d^2/R, where R is the curvature radius of the lens, and d is the shown distance of the reflection points from the optical axis. As the distance d changes, so does a, resulting in consecutive alternation between bright and dark rings around the axis. Namely, the radius d_k of the kth bright ring can be found from $d^2/R = 2\pi k$ as $d_k = \sqrt{2\pi k R\lambda}$, with the spacings $d_{k+1} - d_k$ between the rings getting smaller with k.

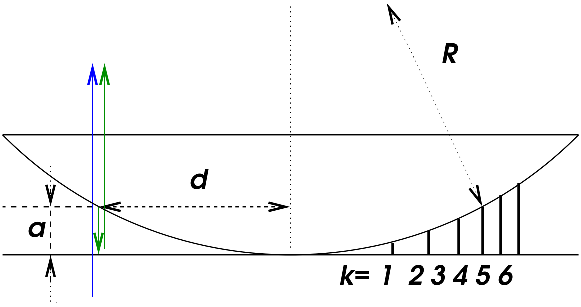


Figure 8: Newton's rings

That's the theory; now let's get down to numbers. The wavelength λ of visible light ranges between 390 and 700 nanometers (nm). For a lens of curvature radius $R = 10\,\text{cm}$, the radius d_1 of the central ring will be about 0.5 mm. Thus, the rings should better be observed under a microscope.

The same applies to other manifestations of the wave properties of light, such as e.g. *diffraction* i.e. the ability of waves to reach the region shadowed from the straight rays by an obstacle. As long as the wavelength λ remains small compared to the size of the objects, the wave effects remain hardly noticeable.

Oscillating integrals. We now revisit the geometrical optics setup with the aim of assessing the phenomenon of interference of light waves, remaining however within the assumption that the wavelength is small, or equivalently, the sizes of objects big relative to λ.

Suppose that a point-like source of monochromatic light of angular frequency ω, positioned at a point x, radiates a spherical wave of wavelength λ. When it reaches an observer positioned at q, it creates the oscillation which in complex notation can be described by the formula

$$\frac{A(x)}{|x-q|}e^{i(\omega t - 2\pi|x-q|/\lambda)}.$$

Here the phase shift $|x-q|/\lambda$ measures the optical distance between x and q, the amplitude $A(x)$ is proportional to the intensity of the source, while the denominator $|x-q|$ reflects the decay of the amplitude with distance. (Namely, the energy of the wave, proportional to the square of the amplitude, is distributed over the surface of the sphere proportional to the square of the radius $|x-q|$.)

Heeding Huygens' principle, we assume that each point x of the wave front at the moment $t = 0$ radiates such a spherical wave of possibly variable amplitude density $A(x)dx$, distributed somehow over the initial front. Superimposing all the spherical waves arriving at the point q from all points x of the initial front, and dropping the time factor $e^{i\omega t}$ common to all of them (as the initial source generating the current wave front could be considered monochromatic), we obtain the following general expression for the wave field, called an *oscillating integral*:

$$I(q) = \int a(x,q)\ e^{2\pi i f(x,q)/\lambda}\ dx.$$

Here the *phase* function $f(x,q)$ combines the optical distance from x to q with the optical distance from the initial source of light to the point x on the wave front at $t = 0$, while the *amplitude* factor $a(x,q)$ combines, in a single function of x depending on q as a parameter, the amplitude density on the front with the distance-caused decay. The whole integral represents the amplitude of the wave field at a point q.

Fresnel's integrals. Our current goal is to understand the behavior of an oscillating integral when λ is small.

One expects that the main contributions to the wave field at q come from *critical points* of the phase function $f(x,q)$ as a function of x, i.e. x_{cr} such that $\partial f(x_{cr},q)/\partial x = 0$. An heuristic explanation of this is that as the amplitude a varies slowly, i.e. remains practically constant at the wavelength scale: $a(x,q) \approx a(x+\lambda,q)$, and the phase function near a non-critical point varies at this scale roughly speaking linearly, the spherical wave issued from a point x would practically cancel with the wave coming from a nearby point where $\Delta f = \lambda/2$. To see why this explanation doesn't apply near a critical point, let us examine the following example, where the phase function is quadratic:

$$\int_{-\infty}^{\infty} e^{2\pi i x^2/\lambda} dx = \sqrt{\frac{\lambda}{2\pi}} \int_{-\infty}^{\infty} e^{iy^2} dy.$$

On Figure 9, the real part of the integrand is shown (and the reader is recommended to sketch the imaginary part $\sin y^2$). We see that for large y, $\cos y^2$ begins to oscillate between -1 and 1 faster and faster, and it seems plausible indeed that the areas below and above the horizontal axis will almost cancel each other. However, near the critical point $y = 0$ there is an interval above the axis which seems to have no partner to cancel with.

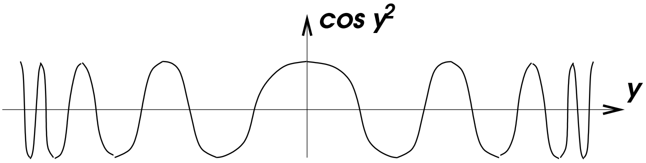


Figure 9: Fresnel's integral

By the way, this integral has to be manipulated with care, since when the integrand is replaced with its absolute value, the integral diverges to infinity. One way to find out what it is equal to would be to slightly perturb the integrand so that it would behave similar to the Gaussian distribution (assuming that $\epsilon > 0$:

$$\int_{-\infty}^{\infty} e^{(i-\epsilon)y^2} dy = \sqrt{\frac{\pi}{\epsilon - i}}.$$

When $\epsilon \to 0$, the limit value $(1+i)\sqrt{\pi/2}$ gives the correct result:

$$\lim_{\alpha\to+\infty}\int_{-\alpha}^{\alpha}\cos y^2\ dy=\sqrt{\frac{\pi}{2}}=\lim_{\alpha\to+\infty}\int_{-\alpha}^{\alpha}\sin y^2\ dy.$$

Short-wave asymptotics. The above heuristic expectations find their affirmation in the following mathematically rigorous statements about the asymptotical behavior of oscillating integrals as the wavelengths $\lambda \to 0$. We will see that:

(a) an oscillating integral tends to 0 as $\lambda \to 0$,

(b) it tends to 0 faster than any power of λ as long as the region of integration does not contain critical points x_{cr} of the phase function $f(\cdot, q)$, and

(c) each non-degenerate critical point x_{cr} of the phase function contributes to the wave field a summand of the order $\lambda^{d/2}$, polynomial in λ, where d is the number of the variables of integration (i.e. the dimension of the wave front).

These statements form the short-wave improvement of our observation in geometrical optics that the only relevant routes from x to q are those perpendicular to the initial wave front.

In order to keep notations simple, we (assuming q fixed) examine the model case of an oscillating integral in one variable:

$$\int_a^b A(x)\ e^{2\pi i f(x)/\lambda}\ dx.$$

Suppose that the amplitude A is an infinitely differentiable function vanishing outside the interval $[a,b]$ of integration.

When the phase function f does not have critical points on $[a,b]$, and hence is monotone, it can be taken for a new coordinate y on this interval. Then the integral can be rewritten as

$$I_\lambda=\int_\alpha^\beta B(y)\ e^{iy/\lambda}\ dy,$$

where $B(y)dy = A(x(y))dx(y)$ vanishes outside $[\alpha,\beta]$. Integrating by parts, and taking into account that $B, B', B'', \dots$ vanish at $y = \alpha, \beta$,we find

$$\int_\alpha^\beta B(y)e^{iy/\lambda}dy=i\lambda\int_\alpha^\beta B'(y)e^{iy/\lambda}dy=(i\lambda)^2\int_\alpha^\beta B''(y)e^{iy/\lambda}dy,$$

and so on. That is, $I_\lambda = o(\lambda^n)$ where positive integer n can be made as large as one wishes.

Suppose now that the phase function in the oscillating integral has a non-degenerate critical point at $x = 0$, and no other critical points on the interval $[-a, a]$. We expand f at the critical point as $\alpha + \beta x^2 + \gamma x^3 + \cdots$, where $\alpha = f(0)$, $\beta = f''(0)/2$, $\gamma = f'''(0)/6$, etc., and the amplitude also expands as $A + Bx + \cdots$. In the following computation, familiar to all physicists, we first make the change $x = \lambda^{1/2} y$. Then the range of integration expands from $[-a, a]$ for x into $[-a\lambda^{-1/2}, a\lambda^{-1/2}]$ for y, and tends to $(-\infty, \infty)$ as λ approaches 0. Next, we rearrange the integrand in such a way that the structure of the integral as a power series in λ becomes apparent (but for simplicity of notation we retain only the principal terms):

$$\begin{aligned}\int_{-a}^{a} A(x) e^{2\pi i f(x)/\lambda} dx = & \int_{-a}^{a} (A + Bx + \cdots)\ e^{i(\alpha+\beta x^2+\gamma x^3+\cdots)/\lambda} dx \\ =& \lambda^{1/2} e^{i\alpha/\lambda} \int_{-a/\lambda^{1/2}}^{a/\lambda^{1/2}} (A + B\lambda^{1/2} y + \cdots)\ e^{i(\beta y^2 + \gamma y^3 \lambda^{1/2} + \cdots)} dy \\ =& \lambda^{1/2} e^{i\alpha/\lambda} \int_{-\infty}^{\infty} e^{i\beta y^2} [A + O(\lambda^{1/2})]\ dy \\ =& (i\lambda)^{1/2} \frac{e^{2\pi i f(0)/\lambda}}{\sqrt{f''(0)}} [A + O(\lambda)]\,.\end{aligned}$$

In the $\lambda^{1/2}$-series expansion, the terms with the half-integer powers of λ integrate to 0 due to parity considerations. Note that the factor $e^{2\pi i f(0)/\lambda}$ has absolute value 1, so that the whole expression behaves as $O(\lambda^{1/2})$.

This computation is easily adapted to the case of integrals and non-degenerate critical points of phase functions in $d > 1$ variables. Near such a point the phase function can be transformed by a linear change of variables to $\alpha + \beta_1 y_1^2 + \cdots + \beta_d y_d^2 + o(|y|^2)$ so that the integral can be factored into d one-dimensional ones, which results in the asymptotics $I_\lambda = \lambda^{d/2}[A + O(\lambda)]$ as promised.

Note that near a simplest *degenerate* critical point, where $f(x) = f(0) + f'''(0)x^3/6 + \cdots$, a similar one-variable computation begins with the change $dx = dy\lambda^{1/3}$, and respectively ends with $I_\lambda = \lambda^{1/3}[A + O(\lambda)]$. This describes the asymptotical behavior of the wave field near a typical point of the caustic, and justifies its name. Indeed, $\lambda^{1/3}$ tends to 0 infinitely slower than $\lambda^{1/2}$.

The wave equation. It would be an enthralling enterprise to derive a combinatorial formula (the so-called *Wick's theorem*) for the coefficients c_k of the entire asymptotical expansion for the oscillating integral near a non-degenerate critical point of the phase function, which can be written in the form

$$I_\lambda(q) = (i\lambda)^{d/2} e^{2\pi i (f''(x_{cr}(q))/\lambda + c_0(q) + c_1(q)\lambda + c_2(q)\lambda^2 + \cdots)}.$$

We will do this in the last section, but our current goal of collecting the necessary clues before sinking into quantum mechanics is more modest. So, we would like to make our last enlightening observation.

It is reasonable to argue that as a function of time t and an observer's position q, the asymptotical expansion $u(t,q) = e^{i\omega t} I_\lambda(q)$ of the wave field should satisfy the same equation as the field itself, namely the *wave equation*:

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = \Delta u, \quad \text{where } \Delta := \frac{\partial^2}{\partial q_1^2} + \frac{\partial^2}{\partial q_2^2} + \frac{\partial^2}{\partial q_3^2}$$

is the *Laplace operator*.

The actual story behind this equation is quite interesting. When James Clerk Maxwell formulated his theory of electromagnetism, he derived the wave equation as its consequence, and found that it indeed has wave-like solutions, e.g. the harmonic running waves $u(t,q) = e^{i\omega t + ik\cdot q}$, where $|k|^2 = \omega^2/c^2 = (2\pi/\lambda)^2$. Thereby Maxwell discovered electromagnetic waves "at the tip of his pen", and as the coefficient c in his equation agreed with the experimental value of the speed of light, one could say that he therefore established that light consists of such waves. Apparently Maxwell himself downplayed the last connection, and dismissed electromagnetic waves as a purely mathematical artifact. However, among his students there was Heinrich Hertz, who later generated radio waves in a lab, thereby opening the door into our age of telecommunication.

Returning to the short-wave asymptotics, take a function of the form

$$u(t,q) = e^{i\omega t} e^{2\pi i (S(q) + O(\lambda))/\lambda},$$

and feed it to the wave equation. The time derivatives yield $-\frac{\omega^2}{c^2}u$, which is the same as $-\frac{4\pi^2}{\lambda^2}u$. The first spatial derivative yields

$$\lambda\frac{\partial}{\partial q_k}\ e^{2\pi i(S+O(\lambda))/\lambda} = \left(2\pi i \frac{\partial S}{\partial q_k} + O(\lambda)\right)\ e^{2\pi i (S+O(\lambda))/\lambda}.$$

Applying $\lambda \partial/\partial q_k$ again, we would obtain two terms according to the product rule. Note however — and this is our key observation! — that the derivative of the first factor contributes only to $O(\lambda)$, and the derivative of the second factor brings down $2\pi i \, \partial S/\partial q_k$ once again. Taking this into account, we obtain:

$$-4\pi^2 u = -4\pi^2 \left[\left(\frac{\partial S}{\partial q_1}\right)^2 + \left(\frac{\partial S}{\partial q_2}\right)^2 + \left(\frac{\partial S}{\partial q_3}\right)^2 \right] u + O(\lambda).$$

Thus, the short-wave asymptotical expansion $e^{i\omega t} e^{i(S(q)+O(\lambda))/\lambda}$ of the wave field, in order to satisfy the wave equation, must have the phase term S — which is in fact the optical distance $f(x_{cr}(q))$ along a critical ray arriving to the observation point q — satisfying the eikonal equation:

$$|\nabla S|^2 = 1.$$

This mechanism of how the linear differential equation, such as the wave equation, leads in the short-wave limit $\lambda \to 0$ to a non-linear 1-st order differential equation which through the method of characteristics is equivalent to a Hamilton equation (with the Hamilton function $H = p \cdot p$ in this case) will prove to be most illuminating in the forthcoming subject of quantization.

4 The wizardry of quantization

Classical and quantum observables. In his 1924 PhD thesis, Louis de Broglie suggested that matter, which was usually thought of as consisting of particles, should exhibit wave-like properties at the microscopic scale. In 1927, Clinton Davisson and Lester Germer confirmed this prediction in their *electron diffraction experiment* by observing an interference pattern resulted from scattering electrons on a crystal. This won de Broglie a Nobel Prize of 1929, and marked the pinnacle (rather than the beginning) of the heroic era in the history of quantum mechanics. We present now a physicist's new worldview which seems to be much more logical than that amazing succession of fantastic ideas and clever experiments that led to the discovery of it during the first quarter of the 20th century.

The key aspect of the discovery is that, pretty much the same way as electromagnetic waves in the short-wave limit $\lambda \approx 0$ are perceived as pencils of light rays, our macroscopic picture of matter as consisting of moving particles is merely the limit of a certain wave field, an illusion caused by the property of that field to have very short wavelength by our macroscopic standards. The differences from optics, however, are that the wave field propagates not through our physical 3-dimensional space, but lives on the configuration space (or even the phase space) of a Hamiltonian system, reduces to pencils of its classical trajectories when certain universal quantity $\hbar$ (instead of λ), called the *Planck constant*, tends to 0, and prior to the limit, obeys not the wave equation, but the so-called *Schrödinger equation.*

Before proceeding to the equation, let us briefly discuss the nature of the Planck constant. It was Albert Einstein, who in his 1905 paper on photoelectric effect suggested that the period $\tau = \lambda/c$ of oscillations of a light wave, and the energy E carried by this wave cannot assume arbitrary independent values, but are related by *quanization condition*

$$E\tau = 2\pi\hbar n, \quad n = 1, 2, 3, \ldots$$

The proposal was motivated by the success of the earlier work of Max Planck on the *black-body radiation*, but also by the ideas of relativity theory Einstein was developing that same year. The physical dimension of the product is known as *action*, and thereby Einstein proposed that light of a fixed wavelength or frequency comes in discrete chunks, with the minimal value of action equal to some universal constant

$$2\pi\hbar \approx 6.62607004 \times 10^{-34} kg\ m^2/s.$$

Another way to represent Einstein's relation is by saying that *photon's* energy, angular frequency ω, and wavelength λ satisfy

$$E = \hbar\omega \text{ or } E\lambda = 2\pi\hbar c, \quad c \approx 3 \times 10^8 m/s.$$

Since the photon moves with the lightspeed, c (which is another universal constant), it is a *relativistic particle*, and the reasons why the energy and the wavelength of a photon must be so related should be understood in the context of relativity theory. We postpone an introduction into this subject till a later section, as we will mostly deal with non-relativistic quantum mechanics. But the Planck constant (which is indeed very small on our macroscopic scale) will persist as the discrete unit of action.

In order to understand where the Schrödinger equation comes from, let us examine the nature of various physical quantities — the *observables* — using the notation of Hamiltonian mechanics. Among such observables, there are: the coordinates q_k on the configuration space (they determine the positions of the constituents of the mechanical system); the components p_k of the momenta (possibly generalized ones); the potential energy $V(q)$ of the mechanical system; the kinetic energy (often in the form $\sum_k p_k^2/m_k$); the components $q_k p_l - q_l p_k$ of the angular momentum vector $q \times p$ (in the case of a particle in the 3-space, when $k, l = 1, 2, 3$); the total energy $H(p, q)$. In other words, physical quantities are represented by functions on the phase space of the Hamiltonian mechanical system. Any such functions are *classical observables.*

Imagine now that we have a "wave field", defined on the configuration space (and traditionally denoted in quantum mechanics by the Greek Ψ, the notorious "*psi-function*"), which is presented in the form arising from the short-wave asymptotics:

$$\Psi(q) = e^{i(S(q)/\hbar + c_0(q) + c_1(q)\hbar + \cdots)}.$$

The problem we encounter here is that of "reverse engineering": What kind of equation should Ψ satisfy which in the limit $\hbar \to 0$ would assure that the phase function S obeys the Hamilton-Jacobi equation $H(\nabla S, q) = const$, i.e. $H(p(q), q) = const$ where $p_k(q) = \partial S/\partial q_k$?

From the example of the wave/eikonal equations, we know that modulo terms of order $O(\hbar)$, each differentiation $-i\hbar\partial/\partial q_k$ would bring down the factor $p_k(q) = \partial S/\partial q_k$ in front of Ψ. Inspired by this observation, we can continue and say that a component

$q_k p_l(q) - q_l p_k(q)$ of the angular momentum can be obtained by applying to Ψ the operator $-i\hbar q_k \partial/\partial q_l + i\hbar q_l \partial/\partial q_k$. Likewise, the equation $H(p(q), q) = 0$ will be satisfied whenever $\hat{H}\Psi = 0$ is a linear differential equation assembled from differentiations $-i\hbar\partial/\partial q_k$ and multiplications by q_l in such a way that it turns into the function $H(p, q)$ after replacing each differentiation with p_k (and multiplication with q_l). We can summarize these examples by saying that in the wave mechanics, *quantum observables* are represented by (linear) differential operators.

Time-dependent hamiltonians. In fundamental physics, we usually assume that laws of Nature don't change with time, and respectively deal with time-independent Hamilton functions. However it is possible to consider time-dependent systems, such as e.g. a pendulum suspended on a chain of periodically varying length (in fact on a swing, you yourself become such a pendulum by periodically raising and lowering your center of gravity). Such mechanical systems can be described by time-dependent Hamilton functions:

$$\dot{q}_k = \partial H(p, q, t)/\partial p_k, \quad \dot{p}_k = -\partial H(p, q, t)/\partial q_k, \quad k = 1, \dots, n.$$

The energy conservation law in such a system may be broken:

$$\frac{d}{dt} H(p(t), q(t), t) = \frac{\partial H}{\partial t} + \sum_k \frac{\partial H}{\partial p_k}\dot{p}_k + \sum_k \frac{\partial H}{\partial q_k}\dot{q}_k = \frac{\partial H}{\partial t}.$$

By the way this shows that the energy conservation law is due to the "homogeneity of time", i.e. time-independence ($\partial H/\partial t = 0$) of H.

There is a simple mathematical trick that reduces the study of time-dependent systems to time-independent ones. Namely, one can consider t as a new $n+1$-st configuration variable, τ, introduce the corresponding momentum variable $-E$ (the sign is traditional), and define a new Hamilton function $\mathcal{H}(p, q, \tau, E) := H(p, q, \tau) - E$ on the $2n+2$-dimensional phase space (called *extended*). The new Hamilton equations will have the form:

$$\dot{\tau} = -\frac{\partial \mathcal{H}}{\partial E} = 1, \quad \dot{E} = \frac{\partial \mathcal{H}}{\partial \tau} = \frac{\partial H}{\partial t}(p, q, \tau),$$

while the equations for p_k and q_k remain unchanged. The total energy conservation law is now restored in the form $d\mathcal{H}/dt = 0$. In particular the $2n+1$-dimensional surface in the extended phase space given by the equation $E = H(p, q, \tau)$ (i.e. $\mathcal{H} = 0$) consists of trajectories which when projected to the pq-space by forgetting τ become the trajectories of the original (time-dependent) Hamiltonian system. The Hamilton-Jacobi equation $\mathcal{H} = 0$ of the extended Hamiltonian system assumes the form

$$-\partial S/\partial t = H(\partial S/\partial q_1, \dots, \partial S/\partial q_n, q_1, \dots, q_n, t).$$

But even when the original hamiltonian is time-*in*dependent, this approach encompasses in one go the solutions to the Hamilton-Jacobi equation $H(\nabla S, q) = E$ for all values of the "energy level" E.

Schrödinger equations. We are ready now to formulate and appreciate the Hamiltonian counterpart of the wave equation:

$$i\hbar \frac{\partial \Psi(q,t)}{\partial t} = \hat{H}\left(\frac{\hbar}{i}\frac{\partial}{\partial q_1}, \cdots, \frac{\hbar}{i}\frac{\partial}{\partial q_n}, q_1, \dots, q_n\right) \Psi(q,t).$$

Note that the operator $i\hbar\partial/\partial t$ in our description of quantum observables represents the classical observable E. The *Hamilton operator* $\hat{H}$ on the right is built of the multiplications by q_k and differentiations $-i\hbar\partial/\partial q_k$ representing the classical observables q_k and p_k respectively. The hamiltonian here is assumed time-independent. Yet the "wave field" Ψ is a complex-valued function on the configuration space extended by the time variable t. Here is an example which for a long time will be sufficient for us: the Schrödinger equation corresponding to a particle of mass m moving on a line in the force field with the potential energy V:

$$i\hbar \frac{\partial \Psi(q,t)}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2 \Psi(q,t)}{\partial q^2} + V(q)\Psi(q,t).$$

This example, where the operator $\hat{H}$ reproduces the Hamilton function $p^2/2m + V(q)$ quite literally, conceals the following important feature of the quantum formalism: While p_k and q_k commute, the corresponding operators (with the same k) don't:

$$\frac{\hbar}{i}\frac{\partial}{\partial q_k} q_k \Psi(q) = \frac{\hbar}{i}\Psi(q) + q_k \frac{\hbar}{i}\frac{\partial}{\partial q_k}\Psi(q).$$

In other words, denoting by $\hat{p}_k$ and $\hat{q}_k$ the operators corresponding to the classical observables p_k and q_k, we arrive at *Heisenberg's commutation relations*

$$\hat{p}_k \hat{q}_k - \hat{q}_k \hat{p}_k = \frac{\hbar}{i}.$$

Note that the discrepancy on the right is $O(\hbar)$. This leads to the conclusion, that in the limit $\hbar \to 0$ (called the *quasi-classical limit*) a quantum Hamilton operator $\hat{H}$ turns into a well-defined classical Hamilton function of p and q. Yet the reverse procedure, called *quantization*, of reconstructing the operator $\hat{H}$ from the classical hamiltonian H is ambiguous. There are many operators $\hat{H}$ (e.g. those which differ by an ordering of the letters $\hat{p}_k$ and $\hat{q}_k$), which at $\hbar = 0$ become the same H.

This ambiguity explains why we call quantization wizardry. Strictly speaking, one cannot "derive" the Schrödinger equation of a quantum system from its classical description, but must apply some guess work, perhaps aided by some symmetry considerations, and eventually backed up by a real-world experiment.

Determinism vs. randomness. The last remark brings up the question: What does the Ψ-function actually represent?

Speaking abstractly, at a fixed t, $\Psi(\cdot, t)$ represents a *quantum state* of the system, and the Schrödinger equation controls the change of this state over time. The situation is fundamentally the same as in classical mechanics, where one can say that the Hamilton-Jacobi equation $-\partial S/\partial t = H(\nabla S, q)$ determines the evolution of the classical state $S(\cdot, t)$. Both equations are deterministic: In a *closed mechanical system*, classical or quantum, the future and the past are uniquely determined (by the Hamilton-Jacobi or Schrödinger equations respectively) *if* the current state of the system is known.

In fact this is a big "if". How can a current state be known (measured), when the system is closed, i.e. banned from interacting with the rest of the Universe?

In classical mechanics, this issue is addressed by making the measurement non-invasive. Plainly put, in order to locate the Moon in the skies, one needs to exchange some photons with it, but this interaction disturbs the Moon's trajectory so little, that one can safely assume that the observed system remained closed.

In quantum mechanics, intended to be applied to microscopic objects, the situation is different. As we will later see, it is impossible to measure the state of a quantum system without altering the state. Any measurement is invasive, disruptive of the deterministic evolution of the quantum state prescribed by the Schrödinger equation.

Yet, what is an act of measuring? Here we encounter the most paradoxical aspect of quantum theory. In fact the interpretation of the Ψ-function which makes it accessible experimentally is that the *square of the absolute value* $|\Psi(q)|^2$ of the complex-valued function describes the *probability density* of finding the system in the configuration q.

What is probability? Well, it is the *frequency* of certain events in the limit when the number of trials tends to infinity. What kind of events? Here is an example: Hitachi's *double-slit experiment*. There a "cannon" shoots electrons one-at-a-time toward a detector through a region with strong magnetic field forcing them to "choose" between two routes. After 30 minutes of shooting, the bright spots on a computer's screen showing the detected electrons accumulate into a series of bright and dim fringes. One could interpret the result by saying that each electron, interacting with the detector, produced a bright spot, whose location is distributed randomly according to the interference pattern described by the function $|\Psi(q)|^2$. What is probably most striking here, is that an experimental observation of fundamental quantum behavior of a microscopic conservative system involves a classical macroscopic detector, time-irreversible medium (such as *youtube*), and eventually us in the role of the observers.

We began our discussion of geometrical optics with the dilemma whether light consists of waves or particles, and based on Huygens' principle answered "both". Returning to this question in the context of quantum mechanics, we should answer "neither". The Universe was not designed to cater to the intuition of cavemen dropping pebbles into water and watching the circles spread. The reality is better captured by such a mathematical abstraction as complex numbers $\Psi(q)$. The relative probabilities of detecting a "pebble" are encoded by their magnitudes, but due to their phases, adding complex numbers may result in the magnitude's decrease, resembling the interference between superimposed waves.

14 Several experiments and one history

Black body radiation. Thanks to Thomas Kuhn's 1962 book *The structures of scientific revolutions*, science historians realized that the image of a steadily growing bank of scientific ideas is at odds with reality. The latter is better described by Hegel's dialectical double negation: the succession of crisis-caused revolutions (*paradigm shifts* in Kuhn's terminology), each fruitful, but progressing inevitably into a new crisis, and so on, and so on. The paradox with Zeno's arrow, resolved by Newton's introduction of instantaneous positions and velocities, yet reinstated by Heisenberg's uncertainly principle is a good illustration. Another is the transition from classical to quantum mechanics, and then to the "second quantization" and quantum field theory.

One implication is that the ways how the classical discoveries are usually fitted into the worldview after a paradigm shift may have little to do with the actual motivation of the discoverers, who operated within the outdated paradigm. The discovery of quantum mechanics is in no way an exception, and the stories about its emergence as they are represented in the textbooks reflect not so much the real history of the subject but rather mythology circulating in the physics community.

In this sense, the present brief outline of the key events and ideas which governed the emergence of quantum mechanics should also not be confused with the genuine history of it. Our main intention here is to make some sense, and hopefully in straightforward ways, of a few basic steps, such as Planck's theory of black body radiation, Einstein's work on the photoelectric effect, and a few others which led to the concept of "quanta". As it was discovered by no one else but Kuhn himself, the real trajectory of the ideas was not straightforward at all, and for the actual history we refer the reader to his 1978 book *Black-body theory and the quantum discontinuity, 1894-1912.*

The term *black body radiation* is perhaps confusing. The problem was posed around 1860 by Gustav Kirchhoff. He studied radiation emitted from a body, and argued that by subtracting the light reflected by the body's surface, one ends up with the purely thermal radiation whose properties should not depend on the material or the surface, but only on the wavelength and the temperature. Such radiation in the form of thermal oscillations of the electromagnetic field exists inside any closed space ("cavity"), and becomes observable when it is allowed to escape through a small hole in the cavity's wall.

At this point one is usually advised to take a day-time look at a neighbor's window — to find out that it looks dark. Daylight rays entering the room through the window, before finding their way out, make enough reflections inside the cavity to blend with the inner radiation to thermal equilibrium. At room temperatures, the thermal radiation has infrared spectrum, invisible to our eye. That's why the window looks dark, and the radiation called "black-body".

The classical approach to problem of understanding thermal oscillations, posed by Kirchhoff, leads to a contradiction known as *ultraviolet catastrophe.* It predicts that the thermal energy accumulated in a cavity is infinite due to infinitely many modes of oscillation of shorter and shorter wavelengths, each allocated by the thermostat the same average energy.

In more detail, oscillations of the electromagnetic waves inside a cavity (you may think of a rectangular box, but any other shape would do just as well) is the superposition of *standing waves* — the eigenfunctions of time-independent Maxwell's equations with appropriate boundary conditions at the walls of the cavity. There are infinitely many such eigenfunctions. For example, taking a cube of size L and imposing (for simplicity) periodic boundary conditions, we get a two-dimensional space of eigenfunctions

$$(E,B)=(E_0,B_0)e^{2\pi i(k\cdot q/L-\nu t)},\ \ \nu=\frac{c|k|}{L},\ E_0\cdot k=0,\ B_0=\frac{k}{|k|}\times E_0,$$

for every *integer* wave vector k. Each eigenfunction is a degree of freedom of the field, and can be considered as a classical harmonic oscillator. In thermal equilibrium with the thermostat, such an oscillator is allocated the same average energy, kT. For large ν, the number of integer vectors k inside the ball of radius $\nu L/c$ (as in the free-electron model) approaches the volume of the ball. Therefore the distribution of energy over frequencies should be given by the density (known as *Rayleigh-Jeans law*)

$$kT\times 2\times\frac{4\pi L^3}{3c^3}\,d\nu^3=kT\,\frac{8\pi V}{c^3}\,\nu^2\,d\nu.$$

Here $V=L^3$ is the volume of the cube, and the factor 2 is the dimension of the space of polarizations of the above harmonic wave.

Of course, this result, predicting the unlimited growth of energy, at high frequencies disagrees with experiments, which show that the radiated energy peaks at a finite frequency ν_{max}.

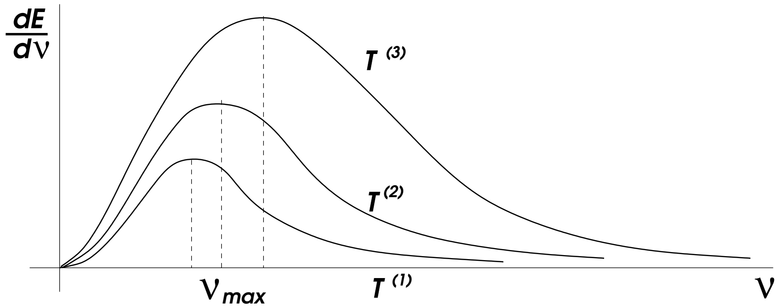


Figure 31: Wien's displacement law

Moreover, in 1893, Wilhelm Wien argued from very basic thermodynamic principles that the whole distribution must scale proportionately to the absolute temperature, and in particular, the peak frequencies ν_{max} are proportional to T.

In 1900, Max Planck conjectured that the energy of a harmonic oscillator ("*resonator*") takes values which are discrete multiples of some unit, ϵ_0. Then the average energy of the oscillator is given by

$$\bar{\epsilon} = -\frac{d}{d\beta}\log \mathcal{Z} = -\frac{d}{d\beta}\log\sum_{n=0}^{\infty} e^{-\beta n\epsilon_0} = \frac{\epsilon_0}{e^{\epsilon_0/kT}-1}.$$

To have the consequences compliant with Wien's displacement law, the energy unit had to be universally proportional to the frequency: $\epsilon_0 = h\nu$. This approach leads to the energy density

$$\frac{8\pi V\nu^2}{c^3}\frac{h\nu}{e^{h\nu/kT}-1}\,d\nu = (kT)^4\frac{8\pi V}{h^3c^3}\frac{X^3}{e^X-1}\,dX, \text{ where } X := \frac{h\nu}{kT}.$$

It agrees with experiments perfectly, as soon as the value of h compatible with the observed value of Wien's universal constant ν_{max}/T is taken.

By the way, the formula implies that the total thermal energy is proportional to T^4 (the so-called *Stephan-Boltzmann law*). To find the proportionality coefficient a, we compute

$$\int_0^\infty \frac{X^3dX}{e^X-1} = \sum_{n=1}^{\infty}\int_0^\infty X^3e^{-nX}dX = \sum_{n=1}^{\infty}\frac{1}{n^4}\int_0^\infty y^3e^{-y}dy = \zeta(4)3!$$

In fact one (non-trivial) job of Bernoulli numbers is to represent some values of the Riemann zeta-function:

$$\zeta(2n) = (-1)^{n+1} B_{2n} \frac{(2\pi)^{2n}}{2(2n)!}.$$

For $n = 2$, we find $\zeta(4) = \pi^4/15 \approx 1.0823\ldots$ Therefore, per unit of cavity's volume

$$\text{Total radiation energy} \;=\; \frac{8\pi^5 k^4}{15c^3h^3}\, T^4.$$

Photoelectric effect. Analyzing Planck's hypothesis about the discreteness of the energy levels of electromagnetic oscillations inside a cavity, Einstein and independently Ehrenfest concluded that it has far-reaching consequences as it cannot be easily explained within classical physics. Planck himself had hoped that the phenomenon can be a consequence of the way resonators (e.g. cavity's walls) generate electromagnetic waves. More specifically, the discreteness of electric charge, which as it had been recently discovered consisted of electrons, was a plausible cause. Einstein went further, and conjectured in his work on the photoelectric effect that light *per se* consists of discrete "bundles of energy" ΔE universally proportional to the frequency: $\Delta E = h\nu$ (or $\hbar\omega$ is our notation throughout the book, where $\omega = 2\pi\nu$ is angular frequency, and $h = 2\pi\hbar$ is the constant introduced by Planck).

The photoelectric phenomenon consists in ejecting electrons from a metal by a beam of light. Applying voltage, one engages the ejected electrons into a current that can be measured by a galvanometer. If the reverse voltage is applied, only the electrons with kinetic energy, sufficient to overcome the voltage barrier, will be able to participate in the current. Thus, the maximal kinetic energy of emitted electrons can be measured as the value of *stopping potential*: the size of the barrier which reduces the photocurrent to 0.

From the classical point of view, the stopping potential should depend on the intensity of light. From Einstein's point of view, an electron in the metal can absorb from the beam one "bundle of energy" $h\nu$. In the case $h\nu$ exceeds the energy W tying the electron to the metal, the electron is left with the kinetic energy $K = h\nu - W$. We have encountered W under the name *work function* in our discussion of conduction levels of metals (see Figure 25a). Thus, the stopping potential depends on the light frequency linearly with the slope h. The work function $W = h\nu_0$, where ν_0 is the frequency of incident light below which the photoelectric effect ceases entirely.

Note that "intensity of light" of a given frequency can be characterized as the state of energy $h\nu(n + \frac{1}{2})$ of an abstract harmonic oscillator. However, this energy is interpreted in quantum mechanics (or rather quantum theory of the electromagnetic field) as the energy $h\nu/2$ of the vacuum state plus the energy of n indistinguishable photons, each carrying energy $h\nu$ and absorbable by electrons only as one photon at a time. By the way, Bose-Einstein's statistics of these photons coincides (not coincidentally though) with the statistical sum $\mathscr{Z} = 1/(1 - e^{-h\nu/kT})$ describing Planck's resonator.

In experiments of 1914, Robert Millikan confirmed Einstein's predictions, and also managed to accurately determine the value of the Planck constant. In 1921, A. Einstein was awarded a Nobel Prize in Physics, yet not for relativity theory (which some still considered controversial), but "for his services to theoretical physics, and especially for his discovery of the law of the photoelectric effect."

Compton's scattering. Controversially or not, consider a resting electron of relativistic energy mc^2 scattering a photon with the momentum vector p and relativistic energy $c|p|$. The event results in a scattered photon with the momentum $\tilde{p}$ and energy $c|\tilde{p}|$, and a recoiled electron with the momentum vector $p_e = p - \tilde{p}$ (by the momentum conservation law) and relativistic energy E,

$$E^2 = m^2c^4 + c^2|p_e|^2, \text{ or } |p_e|^2 = \left(\frac{E}{c} - mc\right)\left(\frac{E}{c} + mc\right).$$

From the energy conservation law we have $E - mc^2 = c|p| - c|\tilde{p}|$. Together with the cosine theorem $|p_e|^2 = |p|^2 + |\tilde{p}|^2 - 2|p||\tilde{p}|\cos\theta$, this yields:

$$|p|^2 + |\tilde{p}|^2 - 2|p||\tilde{p}|\cos\theta = |p|^2 + |\tilde{p}|^2 - 2|p||\tilde{p}| + 2mc(|p| - |\tilde{p}|).$$

From here, using the quantum hypothesis of Planck and Einstein, one obtains a relation between the wavelengths $\tilde{\lambda}$ and λ of the scattered and incident photons, and the cosine of the scattering angle θ:

$$(1 - \cos\theta) = \frac{mc}{|\tilde{p}|} - \frac{mc}{|p|} = \frac{mc}{2\pi\hbar}(\tilde{\lambda} - \lambda).$$

The expression $\lambda_0 = 2\pi\hbar/mc$ is called the *Compton wavelength* corresponding to mass m. A photon of this wavelength carries energy $\hbar\omega = 2\pi\hbar c/\lambda_0 = mc^2$ equal to the relativistic energy of the particle of mass m at rest. For an electron, the Compton wavelength comes out as $\lambda_0 = 0.0024$ nm. The above computation predicts the shift

$\Delta\lambda = \lambda_0(1 - \cos\theta)$ between the wavelengths of the scattered and incident beams, depending on the angle of scattering.

In 1923, Arthur Compton observed this effect by scattering X-rays of wavelength 0.0709 nm ($\approx$ 71 pm) on a graphite target (Figure 32). The wavelengths were measured using *Bragg's diffraction* of the scattered X-rays on a crystal. The intensities of radiation accumulated at given wavelengths were determined using ionization chambers.

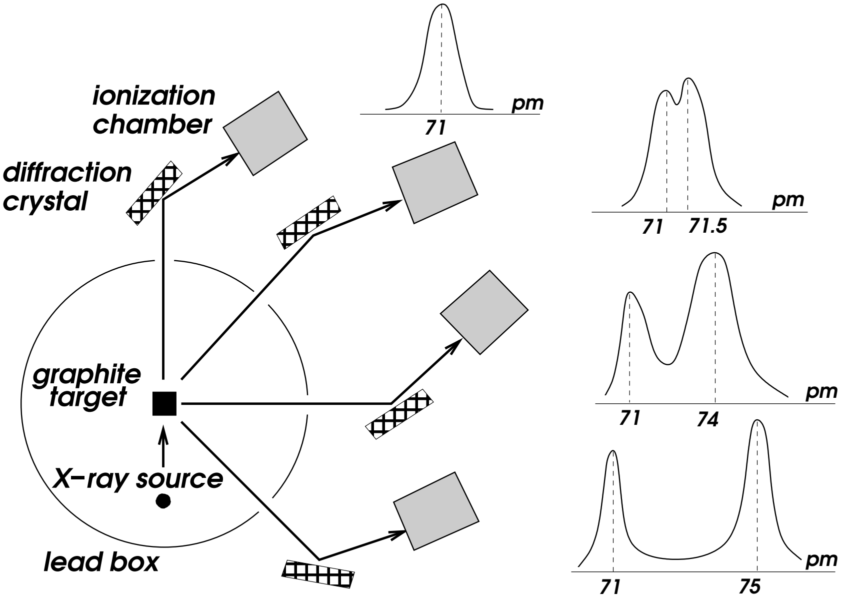


Figure 32: Compton's scattering

The results show two peaks: at the wavelength λ and $\tilde{\lambda}$. According to the classical Thomson theory of scattering of electromagnetic waves on electrons, the source waves of frequency ν would cause the electrons to oscillate with the same frequency (rather then recoil) and, working as vibrators, re-emit electromagnetic waves of the initial wavelength in all directions. This explains the left maxima on the graphs. But the presence of the right maxima at a greater wavelength cannot be explained without the quantum hypothesis stipulating the rigid relationship $\lambda|p| = 2\pi\hbar$ between a photon's wavelength λ and momentum $|p|$ via the Planck constant. The recoiled electrons were directly detected in subsequent experiments, making Compton's effect the turning point in convincing the skeptics (such as e.g. Millikan) that light indeed consists of quanta.

Electron diffraction. The method of measuring the wavelengths in Compton's experiment demonstrating particle-like properties of photons was based, rather ironically, on their wave-like properties. In 1913, following Max von Laue's 1912 discovery of diffraction patterns resulting from scattering X-rays on a crystal, W. Lawrence Bragg proposed a simple formula (the so-called *Bragg's condition*) predicting the directions of maximum intensity of the scattered beam. He then tested the prediction in the experimental setting devised by his father W. Henry Bragg.

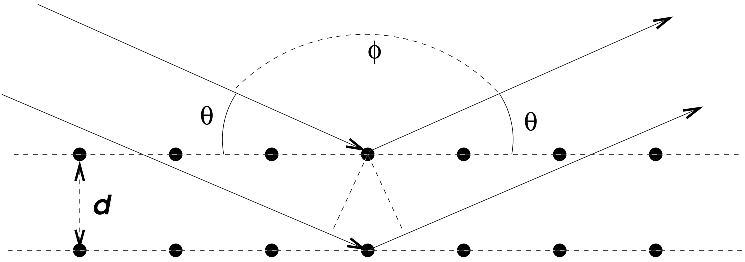


Figure 33: Bragg's condition

Namely, as it is clear from Figure 33, the difference in the optical distance between the rays reflected at an angle θ from two consecutive layers of atoms of a crystalline lattice with the distance d between the layers is equal to $2d\sin\theta$. The interference between the two reflected beams of wavelength λ is to produce maximum intensity — physicists say: is *additive* or *constructive* (as opposed to *subtractive* or *destructive*, occurring when the beams cancel each other) — when *Bragg's condition* is met:

$$2d\sin\theta = n\lambda, \quad \text{where } n \text{ is an integer.}$$

In particular, X-rays of different wavelengths will experience constructive diffraction at different angles. This phenomenon of *dispersion* was used by Compton to measure the intensity of the scattered beam as a function of wavelength (Figure 32). The diffraction crystal disperses the beam into the ionization chamber at an angle θ varying with the wavelength.

In 1924, Louis de Broglie conjectured that the relationship $p\lambda = 2\pi\hbar$, postulated by Einstein for photons of momentum p and wavelength λ in his theory of photoelectricity, should remain valid for particles of matter, which therefore should possess the properties of a wave of *de Broglie wavelength* $\lambda = 2\pi\hbar/p$.

In 1921-25, unaware of this conjecture, Clinton Davisson and Lester Germer were conducting experiments (at what later became Bell Labs) of shooting electrons at a nickel plate in vacuum. Inadvertently they created on the surface of the plate a mono-crystal of nickel sufficient in size to produce a diffraction pattern in the scattered electron beam.

In 1926, Davisson, attending a conference in Oxford, was surprised by a talk of Max Born who used their early results on electron diffraction as an evidence in favor of de Broglie's conjecture. More focused experiments ensued, resulting in good agreement between the conjectural values of de Broglie wavelengths and the observed values found on the basis of Bragg's formula.

Namely, the thermally excited electrons were accelerated by static voltage of 54 V and scattered at the nickel mono-crystal, producing the highest intensity of the reflected electron beam (see Figure 33) at the angle $\phi = 50^\circ$ with the incident beam. From Bragg's condition with $n = 1$, $\theta = 90^\circ - \phi/2$, and $d = 91$ pm for nickel (found from X-ray scattering), one computes $\lambda = 165$ pm. The theoretical value of de Broglie wavelength $2\pi\hbar/\sqrt{2m_e E}$ for electrons of kinetic energy $E = p^2/2m_e = 54$ eV is about 167 pm.

This success, together with the same experiment performed about the same time and independently by George Thomson, son of the electron's discoverer John Thomson (1897), are considered the first direct confirmations of the existence of "waves of matter", generally described by the Schrödinger and in the case of electron by the Dirac equation.

Zeeman's normal and anomalous effects. The conjecture of de Broglie about the wave nature of matter didn't come out of the blue. In 1909, Ernest Rutherford, based on the so-called *gold foil experiments*, demonstrating occasional strong deflections of alpha-particles typically penetrating a metal foil without scattering, concluded that matter consists of tiny positively charged *nuclei* surrounded by a cloud of orbiting electrons.

The problem of instability of atoms arising from Rutherford's model (his orbiting charges would radiate electromagnetic waves and, loosing energy, quickly fall to the nucleus) was addressed in 1913 by Niels Bohr. He conjectured that the electrons would *not* emit photons when they orbit the nucleus along a *circular* orbit characterized by a discrete value of sectorial velocity $|L| = n\hbar$, $n = 1, 2, \ldots$ In the classical Kepler problem, these assumptions lead to the formula for the electron's energy levels in the hydrogen atom

$$E_n = -\frac{me^4}{(4\pi\epsilon_0)^2 2\hbar^2 n^2}.$$

Bohr's model was improved by Arnold Sommerfeld who allowed elliptic Keplerian orbits possessing the same quantized values $n\hbar$ of their *action*. In the quantum Kepler problem, this corresponds to introducing the orbital angular momentum quantum number $l = 0, \dots, n-1$.

Bohr-Sommerfeld's model of the hydrogen atom agrees with Rydberg's empirical formula for the hydrogen spectral series: $\lambda^{-1} = (E_n - E_m)/2\pi\hbar c = R(m^{-2} - n^{-2})$. These are the positions, albeit approximate, on the scale of wavelengths λ, for the spectral lines of light re-emitted by hydrogen. Moreover, by extending the model, Sommerfeld managed to explain the *Zeeman effect* of splitting of each spectral line into several close lines in the presence of a weak external magnetic field.

To start with a bit of theory, denote by q the 3-dimensional radius-vector of a particle of mass m and charge Q moving in a potential $V(q)$ in the presence of a magnetic field B, which for simplicity we will assume constant. The Newton equation

$$m\ddot{q} = Q(\dot{q} \times B) - \nabla V(q)$$

can be interpreted (check this!) as the Euler-Lagrange equation of the following Lagrangian:

$$L(q, \dot{q}) = m\frac{\dot{q} \cdot \dot{q}}{2} + Q(A(q) \cdot \dot{q}) - V(q),$$

where $A = (B \times q)/2$ is the *vector-potential* of the magnetic field in our special case of constant B. In general, it is defined, somewhat ambiguously, as any vector field such that $\nabla \times A = B$. It exists at least locally due to the Maxwell equation $\nabla \cdot B = 0$.

Let us compute the corresponding Hamiltonian. The generalized momentum $p := \partial L/\partial \dot{q} = m\dot{q} + QA(q)$. From this, we express $\dot{q} = p - QA/m$, and substitute into $p \cdot \dot{q} - L$ to find

$$H(p, q) = \frac{1}{2m}(p - QA) \cdot (p - QA) + V(q).$$

Note that when the magnetic field is weak, the terms of order $|A|^2$ can be neglected. The linear terms (for constant B) have the form

$$-\frac{Q}{2m} p \cdot (B \times q) = -\frac{Q}{2m} B \cdot (q \times p).$$

Thus, in the limit of weak constant field B, the Hamiltonian can be expressed as

$$H(p,q) \approx \frac{p \cdot p}{2m} - \frac{Q}{2m} B \cdot L + V(q), \quad \text{where } L = q \times p$$

is the angular momentum. This tells us how to quantize the hydrogen atom problem in the presence of a weak magnetic field (say, directed along the z-axis). The hamiltonian operator $\hat{H}_K$ of the quantum Kepler problem should be changed into

$$\hat{H}_Z := \hat{H}_K + \frac{e|B|}{2m_e} \hat{L}_z.$$

Here we took $Q = -e$ to be the charge of an electron, and $m = m_e$ to be its mass.

Let us recall that the eigenspace of $\hat{H}_K$ with the eigenvalue E_n splits into the direct sum of irreducible SO_3-representations V_l according to the orbital angular momentum quantum number $l = 0, 1, \ldots, n-1$. In each V_l, the operator $\hbar^{-1}\hat{L}_z$ acts with integer eigenvalues ranging from $-l$ to $+l$. The eigenvectors of $\hat{L}_z$ form therefore the basis of eigenstates of the modified hamiltonian operator $\hat{H}_Z$. The eigenvalues are

$$E_n + \frac{e|B|}{2m_e} k\hbar = E_n + \mu_B |B| k, \quad \text{where } k = 0, \pm 1, \ldots, \pm l,$$

and μ_B is a universal constant known as the *Bohr magneton*:

$$\mu_B = \frac{e\hbar}{2m_e} = 5.79 \times 10^{-5} \text{ eV/T(esla)}.$$

To put the numbers in context: the strength of a typical refrigerator magnet is about 10^{-3} T, and for the hydrogen atom, $E_1 = -13.6\,\text{eV}$.

Within the semi-classical Bohr-Sommerfeld theory, the interpretation of the terms $\mu_B|B|k$ was as follows. An electron orbiting the nucleus effectively forms a tiny loop of electric current. It generates therefore a magnetic field identical to that of a tiny magnet — magnetic *dipole* (Figure 34). The dipole is characterized by the *magnetic moment* vector μ. In this case it is equal to $\mu_B L/\hbar$, where L is the angular momentum vector of the orbiting electron. The dot-product $B \cdot \mu$ represents the energy of interaction between the external magnetic field and the dipole.

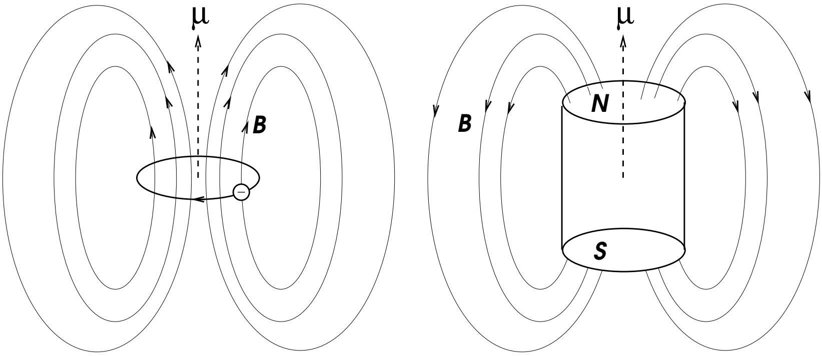


Figure 34: Magnetic dipole

The splitting of the energy levels in the presence of an external magnetic field explains the respective splitting of spectral lines, with spacings proportional to the intensity of the applied magnetic field. The phenomenon, somehow expected by Hendrik Lorentz, and first observed in 1896 by Pieter Zeeman, is known as the *normal Zeeman effect*.

The explanation suggests an odd number of lines arising from a single spectral line in the absence of the magnetic field. The *anomalous Zeeman effect* first reported in 1897 by Thomas Preston, shows, in particular, that the number of lines can be even.

The effect is due to the spin phenomenon. Namely, for the electron, which is a particle with spin 1/2, the eigenspaces of $\hat{H}_K$ have the form $V_l \otimes V_{1/2} = V_{l+1/2} \oplus V_{l-1/2}$. It turns out that the external magnetic field interacts with the electron through both its *orbital magnetic moment* $\hbar^{-1}\mu_B(\hat{L} \otimes 1)$ and *spin magnetic moment* $2\hbar^{-1}\mu_B(1 \otimes \hat{S})$. The factor 2 in the latter case (added to fit spectroscopic data) was at first the source of doubts and controversy, but turned out to be a relativistic effect. The accounting for the spin magnetic moment shifts the expected energy levels and splits them into “doublets”. This improves the agreement of the expected numbers and exact positions of the spectral lines with the experiment.

The “anomalous” splitting of some spectral frequencies into pairs, shifted by $\pm|B|\mu_B/\hbar$ comparing to the positions, expected on the basis of Bohr-Sommerfeld theory, played a role in the discovery of spin by Samuel Goudsmit and George Uhlenbeck in 1925. Namely, their bold idea was preceded by Goudsmit’s purely numerological observation of how one could account for the anomalous Zeeman effect by replacing some integers in Sommerfeld’s formulas with half-integer values.

Stern-Gerlach’s experiment. Though this experiment, conducted in 1922, is considered to be the most direct evidence of spin, it played no role in its discovery. Otto Stern and Walter Gerlach were trying to corroborate Bohr-Sommerfeld’s theory by sending a beam of silver atoms through a strongly inhomogeneous magnetic field (with the gradient of over $1\,\mathrm{T/cm}$) transversal to the beam.

Atoms of silver have 47 electrons, but 46 of them occupy states with pairwise opposite spins. Thus, only the valence electron contributes to the overall magnetic moment of the atom. It contributes so only through its spin magnetic moment, since the orbital angular momentum quantum number l of this electron turns out to be 0.

Of course, the whole atom is a boson, with the nucleus containing 47 protons each carrying the charge $+e$ and spin $1/2$. Note however, that the formula for the magneton contains the mass m of the particle in the denominator. Since protons are much heavier than electrons, their spin magnetic moment is much smaller than the magnetic moment of the valent electron. The latter is therefore largely responsible for the magnetic properties of the whole atom.

Passing through a uniform magnetic field, the atom, thought of as a classical magnetic dipole, would experience torque, but no overall magnetic force (since it is also eclectically neutral). However, if the field is inhomogeneous, the gradient of the magnetic force acting on the dipole’s North and South poles could deflect the atom from the straight trajectory.

From the point of view of classical physics, the deflected atoms would create a continuous distribution on the screen with the maximum at the center due to the random orientations of the atom’s magnetic moments. Stern and Gerlach incorrectly thought that the valence electron of silver has the orbital angular momentum quantum number $l = 1$. In this case the orbital magnetic moment would take on three discrete values corresponding to $k = 0, 1, -1$. Respectively, they expected the beam to split into three: one undeflected and two deflected in opposite directions.

Instead, they observed the beam split into two oppositely deflected ones (Figure 35).

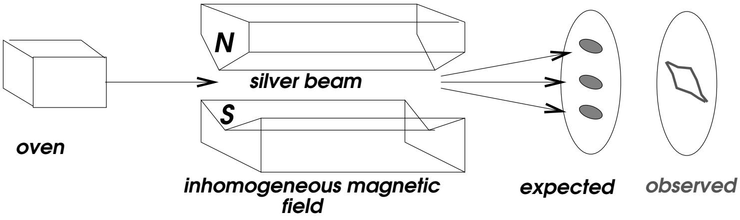


Figure 35: Stern-Gerlach experiment

This result was at odds with Bohr-Sommerfeld's theory, and was only explained with the idea of spin. The two deflected beams of atoms with zero orbital magnetic momentum ($l = 0$) differ by the spin magnetic momentum ($k = \pm 1/2$), that is, to be more precise, by the eigenvalue of one of the components of spin (say S_z).

The experiment can be iterated as follows. Separating the beam with the value of S_z equal to, say, $+\hbar/2$, one can send it through inhomogeneous magnetic field again. If the magnetic field is oriented the same way as before, then the beam does not split: all the atoms deflect the same way as they have the same value of S_z. However, when the magnetic field is oriented in, say, x-direction, the beam splits again into two: with the value of S_x equal to $+\hbar/2$ and $-\hbar/2$.

This illustrates the nature of quantum measurements. The operators S_z an S_x do not commute. So, they cannot simultaneously attain definite values. When the iterated experiment enforces the atoms to "choose" between the two values of S_x, the initially enforced value of S_z is "forgotten".

From the abstract point of view, what happened in the act of measuring is that the closed system, described by the deterministic evolution of the psi-function in accordance with the Schrödinger equation, fails to remain closed as a result of interaction with the measuring device. The consequence of measuring (often described as the "*collapse of the wave function*") is that the vector in the Hilbert space represented by the wave function is projected orthogonally to the subspace spanned by the states, consistent with the outcome of the measurement. In principle this is similar to how probabilities of possible events "collapse" into conditional probabilities as a result of the condition turning out certain. The difference is that the "collapsed" ψ is the complex-valued amplitude, while the probabilities are proportional to $|\psi|^2$.

Nobel laureates. Many of the aforementioned achievements were awarded Nobel Prizes in Physics. Otto Stern received one, but for the discovery of the magnetic moment of proton, and not for Stern-Gerlach's experiment. Ernest Rutherford made his contribution to our story only *after* receiving his Nobel Prize in Chemistry. Here is our list of the (mostly) familiar laureates, in the chronological order of their awards.

William Röntgen, 1901: for discovery of X-rays.

Hendrik Lorentz and Pieter Zeeman, 1902: for the Zeeman effect.

Wilhelm Wien, 1911: for Wien's law of black-body radiation.

Max von Laue, 1914: for diffraction of X-rays on crystals.

W. Henry Bragg and W. Lawrence Bragg, 1915: for Bragg's law.

Max Planck, 1918: for introducing quanta of energy.

Albert Einstein, 1921: for the theory of photoelectricity.

Niels Bohr, 1922: for his model of atom.

Robert Millikan, 1923: in part for the photoelectric effect.

Arthur Compton, 1927: for Compton's effect.

Louis de Broglie, 1929: for "the discovery of the wave nature of electrons."

Werner Heisenberg, 1932: for "the creation of quantum mechanics."

Erwin Schrödinger and Paul Dirac, 1933: for "the discovery of new productive forms of atomic theory."

Carl Anderson, 1936: for the discovery of positron.

Clinton Davisson and George Thomson, 1937: for the demonstration of electron diffraction.

Enrico Fermi, 1938: in fact for his discovery of nuclear reactions.

Otto Stern, 1943: for discovering the magnetic moment of proton.

Wolfgang Pauli, 1945: for the exclusion principle.

Max Born, 1954: for the statistical interpretation of psi-functions.

Richard Feynman, Julian Schwinger, Shinishiro Tomonaga, 1965: for Q.E.D.

Subrahmanyan Chandrasekhar, 1983: for the theory of white dwarfs.

Eric Cornell, Carl Wieman, Wolfgang Kettlerle, 2001: for demonstrating Bose-Einstein condensation.

Index