Introduction to Quantum Mechanics

Alexander Givental
Introduction
to
Quantum Mechanics
by
Alexander Givental

Sumizdat
University of California, Berkeley Cataloging-in-Publication Data
Givental, Alexander
Introduction to Quantum Mechanics / by Alexander Givental ;
xii, 165 p. : 35 ill. ; 23 cm.
Includes index
ISBN 978-0-9779852-2-7

©2019 by Alexander Givental
All rights reserved. Copies or derivative products of the whole work or any part of it may not be produced without the written permission from Alexander Givental (givental@math.berkeley.edu), except for brief excerpts in connection with reviews or scholarly analysis.

ISBN 978-0-9779852-2-7
Contents

Preface (page v)

1. Geometrical optics (1)
   Waves or particles? — The eikonal equation.

2. Hamiltonian mechanics (5)
   Hamilton equations. — Examples in one degree of freedom.
   — The Kepler problem. — Symplectic geometry. — Hamilton
   equations on $S^2$. — Poisson brackets. — Groups. — The
   Hamilton-Jacobi equation.

3. Short-wave optics (17)
   Newton’s rings. — Oscillating integrals. — Fresnel’s inte-
   grals. — Short-wave asymptotics. — The wave equation.

4. The wizardry of quantization (25)
   Classical and quantum observables. — Time-dependent hamil-
   tonians. — Schrödinger equations. — Determinism vs. random-
   ness.

5. One degree of freedom (31)
   A free particle, phase and group velocities, wave packets, and the
   Fourier transform. — Dirac’s delta-function, and Zeno-Heisenberg’s
   principle. — The infinite well potential. — Finite square well. —
   Delta-function shaped well. — Double-well and molecular bond-
   ing. — The step potential, scattering, and probability current. —
   The rectangular barrier and tunneling.

6. The operator formalism (45)
   Hermitian linear algebra. — The spectral theorem. — The
   abstract uncertainty principle. — Time evolution operators, and
   the energy-time uncertainty relation. — The harmonic oscillator.

7. Identical particles (57)
   Multiparticle systems. — Bosons and fermions. — Tensor
   algebra. — Symmetric and anti-symmetric tensors. — Two ideal
   gases: oscillators and qubits.

8. The hydrogen atom (65)
   Central fields and the orbital angular momentum. — Spherical
   harmonics. — The quantum Kepler problem. — Achievements
   and limitations of the quantum Kepler model.
9. Spin


10. A bit of relativity


11. Quantum statistics


12. Solid state physics


13. Summation over histories

- Fermat’s least time principle. — Calculus of variations. — Lagrangian mechanics. — The least action principle. — Feynman’s path integrals. — Wick’s theorem.

14. Several experiments and one history


Index
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 chapter, 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 chapters 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, Chapter 2 introduces Poisson brackets, and the idea of a group. Chapter 3 brings in asymptotics of oscillating integrals. Chapter 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. Chapter 6 teaches Hermitian linear algebra. In Chapter 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 Chapter 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 Chapter 9 in connection with spin. A minimalists introduction to special relativity in Chapter 10 reaches at least the formulation of the Dirac equation and links it to quaternions introduced earlier in Chapter 9. Chapter 11 on quantum statistics makes use of partition functions, as well as tensor algebra developed in Chapter 7. The band structure of solids is discussed in Chapter 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 Chapter 13, meant to be both the epilogue to this and prelude to a higher level course, 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
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).

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°.
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.

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](image)

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* $S(q) = \text{const}$. 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:

$$\begin{cases}
\dot{q} &= \frac{\partial H}{\partial p} \\
\dot{p} &= -\frac{\partial H}{\partial q}
\end{cases}$$

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 $m\ddot{q} = -\nabla_q V$. 
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 $\ddot{q} = p/m$, $\dot{p} = 0$, 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”.

Figure 5: One degree of freedom
The finite square well (Figure 5c):

\[ V(q) = \begin{cases} 
0 & |q| > a \\
-V_0 & |x| < a 
\end{cases} \]

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 \( \ddot{q} = \frac{p}{m}, \ \ddot{p} = -Kq \), i.e. \( \ddot{q} = -(K/m)q \). The trajectories on the \( pq \)-plane are ellipses \( H = \text{const} \) (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 \( \dot{q} = \frac{p}{m}, \ \dot{p} = -\frac{Gq}{|q|^3} \), and show that the angular momentum vector \( \mathbf{L} := q \times p \) is conserved:

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

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 \( |\mathbf{L}| \) 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 \( \theta \) 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
\[
m\ddot{q} = -\frac{Gq}{r^3},
\]
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 \( m\ddot{q} = -\frac{Gq}{|q|^3} \) 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 \textit{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 \textit{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 \textit{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 \textit{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 “\textit{energy conservation law}” easily follows from the chain rule and the form of the equations of motion:

\[
\frac{d}{dt} \frac{dH(p(t), q(t))}{dp} \dot{p} + \frac{dH(p(t), q(t))}{dq} \dot{q} = -\frac{\partial H}{\partial p} \frac{\partial H}{\partial q} \frac{\partial H}{\partial q} \frac{\partial H}{\partial p} = 0.
\]
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 \textit{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 anti-symmetric 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 non-trivial 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.
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 $z = \text{const}$ 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:

\[ L_1 = q_2 p_3 - q_3 p_2, \quad L_2 = q_3 p_1 - q_1 p_3, \quad L_3 = q_1 p_2 - q_2 p_1. \]

When a point \((p, q)\) of the phase space evolves in time according to the Hamilton equations \(\dot{q}_i = \partial H / \partial p_i, \quad \dot{p}_i = -\partial H / \partial q_i\), 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 \(\dot{q}_i = \{H, q_i\}, \quad \dot{p}_i = \{H, p_i\}\) 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 antisymmetry).
• **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

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

will lie on the spheres $x^2 + y^2 + z^2 = \text{const}$, 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

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

The first two equations describe on the $(x, y)$-plane clockwise rotations with constant angular velocity and period $2\pi$ (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 $r^2 = \text{const}$ 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 $(\alpha, \beta, \gamma)$. 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:

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

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 \( H = \text{const} \) 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 \( \nabla S \) 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 $\lambda$ is the wavelength. When the shift is close to 0, the beams reinforce each other’s intensity, and when it is close to $\pi$, 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 $k$th bright ring can be found from $d^2/R = 2\pi k$ as $d_k = \sqrt{2\pi kR\lambda}$, with the spacings $d_{k+1} - d_k$ between the rings getting smaller with $k$.

![Figure 8: Newton’s rings](image)

That’s the theory; now let’s get down to numbers. The wavelength $\lambda$ of visible light ranges between 390 and 700 nanometers (nm). For a lens of curvature radius $R = 10$ 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 $\lambda$ 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 $\lambda$.

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

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

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 $\lambda$ 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](image)

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 confirmation 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 $\lambda$ 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 $\lambda$, 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 $B, B', B'', \ldots$ 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 \( \lambda \) approaches 0. Next, we rearrange the integrand in such a way that the structure of the integral as a power series in \( \lambda \) becomes apparent (but for simplicity of notation we retain only the principal terms):

\[
\int_{-a}^{a} A(x)e^{2\pi i f(x)/\lambda} \, dx = \int_{-a}^{a} (A + Bx + \cdots) e^{i(\alpha+\beta x^2+\gamma x^3+\cdots)/\lambda} \, dx
\]

\[
= \lambda^{1/2} e^{i\alpha/\lambda} \int_{-a\lambda^{1/2}}^{a\lambda^{1/2}} (A + B\lambda^{1/2}y + \cdots) e^{i(\beta y^2+\gamma y^3\lambda^{1/2}+\cdots)} \, dy
\]

\[
= \lambda^{1/2} e^{i\alpha/\lambda} \int_{-\infty}^{\infty} e^{i\beta y^2} [A + O(\lambda^{1/2})] \, dy
\]

\[
= (i\lambda)^{1/2} e^{2\pi i f(0)/\lambda} \sqrt{f''(0)}^{-1} \, [A + O(\lambda)].
\]

In the \( \lambda^{1/2} \)-series expansion, the terms with the half-integer powers of \( \lambda \) 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

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

We will do this in the last chapter, 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:

$$\frac{1}{c^2} \frac{\partial^2 u}{\partial t^2} = \Delta u,$$

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 \( \lambda \)), 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 quantization 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

\[
2\pi \hbar \approx 6.62607004 \times 10^{-34} \text{kg m}^2/\text{s}.
\]
Another way to represent Einstein’s relation is by saying that photon’s energy, angular frequency $\omega$, and wavelength $\lambda$ satisfy

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

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 chapter, 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 $\Psi$, the notorious “psi-function”), which is presented in the form arising from the short-wave asymptotics:

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

The problem we encounter here is that of “reverse engineering”: What kind of equation should $\Psi$ satisfy which in the limit $\hbar \to 0$ would assure that the phase function $S$ obeys the Hamilton-Jacobi equation $H(\nabla S, q) = \text{const}$, i.e. $H(p(q), q) = \text{const}$ 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 $\Psi$. Inspired by this observation, we can continue and say that a component
\( q_k \dot{p}_k(q) - q_l \dot{p}_l(q) \) of the angular momentum can be obtained by applying to \( \Psi \) 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:

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

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, \( \tau \), 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 \( \tau \) 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

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

But even when the original hamiltonian is time-independent, 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:

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

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” $\Psi$ 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 $\Psi$-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 $\Psi$-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.
5 One degree of freedom

A free particle, phase and group velocities, wave packets, and the Fourier transform. We look for solutions $\Psi$ to the Schrödinger equation with a given initial condition $\psi$:

$$i\hbar \frac{\partial \Psi(q, t)}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \Psi(q, t)}{\partial q^2}, \quad \Psi(q, 0) = \psi(q).$$

Informed by our experience in classical mechanics, fix a level $E$ of total energy, and separate the variables in the Schrödinger equation:

$$i\hbar \frac{\partial \Psi}{\partial t} = E = -\frac{\hbar^2}{2m} \frac{\partial^2 \Psi}{\partial q^2}.$$

Solving the two ordinary differential equations, we find a solution in the form of a harmonic wave of arbitrary amplitude $A$:

$$\Psi(q, t) = Ae^{i(kq - \omega t)}, \text{ where } k^2 = 2mE/\hbar^2, \text{ and } \omega = E/\hbar = \hbar k^2/2m.$$

Rewriting $kx - \omega t = k(x - wt)$ we see that the wave is traveling with the so-called phase velocity $w = \omega/k$.

The probability density $|\Psi|^2 = |A|^2$ here is uniform. Note that if $\Psi$ is a superposition of such traveling waves with spatial frequencies $k$ concentrated near some $k_0$, then $|\Psi|^2$ would involve only the variations of the exponent:

$$\Delta(kq - \omega(k)t) = \left( q - \frac{d\omega}{dk}(k_0)t \right)(k - k_0) + o(k - k_0).$$

Consequently the probability density would travel (in this approximation) with the group velocity $v = d\omega/dk$ computed at $k = k_0$.

In general, to satisfy the initial condition, we look for $\Psi$ in the form of a wave packet, i.e. arbitrary superposition of harmonic waves:

$$\Psi(q, t) = \int_{-\infty}^{\infty} A(k) e^{i(kq - \omega(k)t)} dk.$$

Putting $t = 0$, we find that $A$ must be the Fourier transform of $\psi$:

$$\psi(q) = \int_{-\infty}^{\infty} A(k) e^{ikq} dq, \quad \text{i.e. } A(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi(q) e^{-ikq} dq.$$
Note that an application of the momentum operator \(-i\hbar \partial/\partial q\) to \(\psi\) results in the multiplication of the Fourier transform \(A\) by \(\hbar k\), while the multiplication by \(q\) results in the differentiation:

\[
q\psi(q) = -i \int_{-\infty}^{\infty} A(k) \frac{d}{dk} e^{ikq} dk = \int_{-\infty}^{\infty} i \frac{dA(k)}{dk} e^{ikq} dk.
\]

Here we’ve integrated by parts assuming that \(A(k) \to 0\) as \(k \to \pm \infty\).
(The latter is true provided that \(\psi\) is infinitely differentiable). We see that in the Fourier representation of quantum states, the phase coordinates \(q\) and \(p\) almost switch their roles: \(\hat{p}\) becomes the operator of multiplication by \(p := \hbar k\), while

\[
\hat{q} = i \frac{\partial}{\partial k} = i\hbar \frac{\partial}{\partial p}.
\]

In fact this is a consequence of the symmetry \((q, p) \mapsto (p, -q)\) of the symplectic structure on the phase plane. It contains an important (though perhaps cryptic) clue: A quantum state is represented by a function on the configuration space only because we chose to represent it this way. The actual quantum state lives somehow on the phase space. It can be represented by functions in many different ways, e.g. as a function of \(p\) via the Fourier transform (the so-called momentum representation).

**Dirac’s delta-function, and Zeno-Heisenberg’s principle.**
Here mathematics of the Fourier transform meets the fundamental aspect of quantum physics: the uncertainty principle. While \(|\Psi(q)|^2 dq\) is interpreted as the distribution density of the particle’s position, \(|A(p/\hbar)|^2 dp\) is interpreted as the distribution density of the particle’s momentum. The exponent \(ikq = ipq/\hbar\) present in the definition of both the direct and inverse Fourier transform indicates that rescaling \(q\) in \(\Psi\) is compensated by the reciprocal rescaling of \(p\) in \(A\): the more narrow is the distribution of the position \(q\) of the particle, the wider becomes the distribution of its momentum, and *vice versa*. Physicists often express this rule by

\[
\Delta p \Delta q \approx \hbar,
\]

meaning that it is impossible to determine simultaneously the exact position and exact momentum of a quantum particle, as the product of the errors must have the scale of \(\hbar\).

The uncertainty principle doesn’t imply that the precise determination of, say, position \(q\) alone is fundamentally impossible. When
it is known that $q = q_0$, the distribution density is given by Dirac’s delta-function. By definition it possesses the properties:

$$
\delta(q - q_0) = \begin{cases}
  +\infty & q = q_0 \\
  0 & \text{elsewhere}
\end{cases}, \quad \int_{-\infty}^{\infty} \delta(q - q_0) \phi(q) dq = \phi(q_0)
$$

for any continuous function $\phi$. Well, of course, there is no such a function, but the operation of evaluation of continuous functions $\phi$ at a point $q_0$ is well-defined. This operation is an element of the linear space dual to the linear space of continuous functions. In this capacity Dirac’s delta-function does exist.

As the Fourier transform acts on genuine functions, its transpose acts on the dual space. For example, if $\psi(q) = \delta(q - q_0)$, its Fourier transform

$$
A(k) := \frac{1}{2\pi} \int_{-\infty}^{\infty} \delta(q - q_0) e^{-ikq} dq = \frac{e^{-i\omega q_0 k}}{2\pi}
$$

is the exponential function. The corresponding distribution density $|A|^2 = 1/4\pi^2$ is constant. Thus, when the position $q = q_0$ is certain, the momentum is totally uncertain, and vice versa. In a way this is a modern formulation of the arrow paradox discovered by the ancient Greek philosopher Zeno of Elea.

What happens if nonetheless we try to narrow down the range of $p$ by measurement for a particle located at $q$? Well, this will require some interaction of our measuring device with the particle, which will disturb the position of it at least to the extent that the uncertainty $\Delta q \geq \hbar/\Delta p$, where $\Delta p$ is the error of our measurement: the better is the measurement, the worse must be the disturbance.

**The infinite well potential.**

$$
i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \Psi}{\partial q^2} + V(q) \Psi, \quad V(q) = \begin{cases}
  0 & 0 < q < L \\
  \infty & \text{elsewhere}
\end{cases}
$$

Separating the variables, i.e. looking for solution in the product form $\Psi(q, t) = f(t) \psi(q)$, we find $f(t) = e^{-i\omega t}$, we arrive at the stationary or time-independent version of the Schrödinger equation

$$
-\frac{\hbar^2}{2m} \frac{d^2 \Psi}{dq^2} + V(q) \Psi = E \Psi, \quad E = \hbar \omega.
$$

Due to the specific form of the potential, we are therefore looking for $\psi$ on the interval $[0, L]$ such that $-\hbar^2 \psi''(q)/2m = E\psi$ and $\psi(0) = \psi(L)$. The solution is

$$
\psi(q) = \sqrt{\frac{2}{L}} e^{-i\sqrt{E/\hbar} q}.
$$

For $E < m\omega$, the energy levels are discrete, and for $E > m\omega$, the energy levels are continuous. The corresponding wavefunctions are

$$
\psi(q) = \sqrt{\frac{2}{L}} \sin \frac{\sqrt{E/\hbar} q}{2}.
$$

In this case, the wavefunction is not normalizable, and the energy levels are not discrete.
\(\psi(L) = 0\). The solutions exist only for discrete values of energy \(E\) found from 
\[\hbar^2 n^2 \pi^2 / L^2 = 2mE\]
\[\psi_n(q) = A_n \sin \frac{n\pi q}{L}, \quad E_n = \frac{n^2 \hbar^2 \pi^2}{2mL^2}, \quad n = 1, 2, 3, \ldots\]

It is said that the energy is quantized, and this phenomenon, which reflects basic properties of solutions to differential equations, happened to give the whole theory its name.

The general solution has the form of superposition:
\[
\Psi(q,t) = \sum_{n=1}^{\infty} e^{-iE_n t / \hbar} A_n \sin \frac{n\pi q}{L},
\]
where the coefficients \(A_n\) are to be found from the initial condition \(\Psi(q,0) = \psi(q)\). Since \(\int_0^L \sin \frac{n\pi q}{L} \sin \frac{m\pi q}{L} dq = 0\) for \(n \neq m\), we have:
\[
\psi(q) = \sum_{n=1}^{\infty} A_n \sin \frac{n\pi q}{L}, \quad \text{where} \quad A_m = \frac{2}{L} \int_0^L \psi(q) \sin \frac{m\pi q}{L} dq.
\]

This expression is known as a Fourier series, and the whole process of solving the Schrödinger equation is an instance of the so-called Fourier method of solving partial differential equations.

Note that there is no quantum state here with energy \(E = 0\); the minimal energy \(E_1 = \hbar^2 \pi^2 / 2mL^2\). This can be explained by the uncertainty principle: The particle’s location is known with finite uncertainty \(\Delta q = L/2\) (the half-size of the “well”), and if the particle were at rest, its momentum would have value 0 with no uncertainty.

In the ground state, since \(\sin(\pi q / L) = (e^{i\pi q / L} - e^{-i\pi q / L}) / 2i\), the momentum \(p = k\hbar = \pm \pi \hbar / L\) is uncertain with \(\Delta p = \hbar \pi / L\), so that \(\Delta p \Delta q = \pi \hbar / 2 = O(\hbar)\) in agreement with the Heisenberg relation.

As an example of the particle’s dynamics, consider the following superposition of the ground state with the first excited state:
\[
\Psi(q,t) = e^{-i\frac{E_1}{\hbar} t} \psi_1 + e^{-i\frac{E_2}{\hbar} t} \psi_2 = e^{-i\frac{E_1}{\hbar} t} \left( \sin \frac{\pi q}{L} + e^{i\frac{E_1-E_2}{\hbar} t} \sin \frac{2\pi q}{L} \right).
\]

Let’s sketch it (ignoring the overall phase factor) together with \(|\Psi|^2\) at \(t = 0\) and the half-period \(t_0\) when \(t_0(E_1 - E_2) / \hbar = \pi\) (Figure 10). We see that the likelihood of finding the particle in the left half or the right half of the well changes periodically similarly to how the classical particle bounces between the walls \(q = 0\) and \(q = L\).
Finite square well. Following the Fourier method, we are looking for solutions $\psi$ to the stationary Schrödinger equation

$$\frac{-\hbar^2}{2m} \frac{d^2 \psi}{dq^2} + V(x)\psi = E, \quad V(q) = \begin{cases} 0 & |q| < a/2 \\ V_0 & |q| > a/2 \end{cases}$$

assuming that the energy level $E$ lies between 0 and $V_0$. Inside the well, we find $\psi(q)$ to be a linear combination of $\cos kq$ and $\sin kq$ with $k = \sqrt{2mE/\hbar}$. Outside, we have $d^2\psi/dq^2 = \kappa^2 \psi$, where $\kappa = \sqrt{2m(V_0 - E)/\hbar} > 0$.

The density distribution $\psi$ is required to decay at $q = \pm \infty$ (so that $\int |\psi(q)|^2 dq$ converges), or at least stay bounded (as in the case of harmonic waves). For this, $\psi$ must be proportional to $e^{\kappa q}$ for $q \leq -a/2$, and to $e^{-\kappa q}$ for $q \geq a/2$. We now need to sew these formulas at $q = \pm a/2$. For this, we represent $\psi$ as the sum of $\psi_+ + \psi_-$ of its even and odd parts. We have

$$\psi_+(q) = \begin{cases} A \cos kq & |q| \leq a/2 \\ Ce^{-\kappa|q|} & |q| \geq a/2 \end{cases} \quad \psi_-(q) = \begin{cases} B \sin kq & |q| \leq a/2 \\ \pm De^{-\kappa q} & |q| \geq a/2 \end{cases}$$

Here $\pm = \text{sign}(q)$. For the reasons of continuity of $\psi_\pm$ and their 1st
derivatives we also need

\[ A \cos ka/2 = C e^{-\kappa a/2}, \quad A k \sin ka/2 = C \kappa e^{-\kappa a/2} \text{ for } \psi_+ \]

\[ B \sin ka/2 = D e^{-\kappa a/2}, \quad B k \cos ka/2 = -D \kappa e^{-\kappa a/2} \text{ for } \psi_- \]

Taking the ratios eliminates the arbitrary coefficients:

for \( \psi_+ : \kappa = k \tan ka/2 \), and for \( \psi_- : \kappa = -k \cot ka/2 \).

On Figure 11, the blue and green lines show the branches of respectively \( \tan \mu \sqrt{E} \) and \( -\cot \mu \sqrt{E} \) with \( \mu = \sqrt{\frac{ma}{\hbar}} = \frac{ka}{2\sqrt{E}} \). The graph of \( \kappa/k = \sqrt{V_0/E - 1} \) is shown in red. Each intersection point of the red graph with blue or green corresponds to a solution \((k(E), \kappa(E))\) of the above equations, and respectively to a stationary quantum state \( \psi_+ \) (blue) or \( \psi_- \) (green). We see that the total number of such states with energies \( E < V_0 \) is finite, and equal to the number of integer values of \( n \) satisfying \( 0 \leq \pi(n - 1)/2 < \mu \sqrt{V_0} \).

![Figure 11: The square well spectrum](image)

When the depth \( V_0 \) of the well grows, the number of energy values increases, and the \( n \)th value \( E_n \) tends to \( \pi^2 n^2/4\mu^2 = \pi^2 n^2 \hbar^2 / 2ma^2 \) as in the case of the infinite well of width \( a \).

The shapes of the first three \( \psi \)-functions are shown on Figure 12. Within the range \(|q| < a/2\) they resemble the sine functions of the infinite well, but also have the exponentially decaying “tails” in the range \(|q| > a/2\). That is the region where the classical particle with energy \( E < V_0 \) cannot intrude, but there is a non-zero probability of finding the quantum particle with such small energy. This phenomenon is an instance of quantum tunneling.
One might be curious to know what happens when $E > V_0$. We’ll discuss this later in the context of scattering.

**Delta-function shaped well.** In the operator $-\frac{\hbar^2}{2m} \frac{d^2}{dq^2} - K\delta(q)$, the $\delta$-function has the dimension of inverse length (since $\int \delta(q)dq = 1$ is dimensionless). Therefore the arbitrary coefficient $K$ should be taken in the form $\frac{\hbar^2}{2md}$, where $d$ is an arbitrary length. In the resulting stationary Schrödinger equation

$$\frac{d^2\psi}{dq^2} + \frac{\delta(q)}{d}\psi = -\frac{2mE}{\hbar^2}\psi,$$

the middle term is zero for $q < 0$ and $q > 0$. So for the energy levels $E < 0$ (i.e. inside the infinitely deep — yet infinitely thin! — well), the respective solutions $\psi_-$ and $\psi_+$ have the form

$$\psi_\pm(q) = C_\pm e^{\mp\kappa q}, \quad \kappa = \frac{\sqrt{-2mE}}{\hbar}.$$

From the differential equation, we find however:

$$\frac{d\psi}{dq}(0^+) - \frac{d\psi}{dq}(0^-) = \int_{0^-}^{0^+} \frac{d^2\psi}{dq^2}dq = -\frac{1}{d} \int_{0^-}^{0^+} \delta(q)\psi(q)dq = -\psi(0).$$

In other words, the $\psi$ function is still continuous at $q = 0$, but must experience a jump between the left and right derivatives (that’s
the effect of the 2nd derivatives being infinite!) whose magnitude is controlled by \( \psi(0) \) through this relation. So, we have

\[
C_+ = C_- = \psi(0), \quad C_+(-\kappa) - C_- (+\kappa) = -\frac{\psi(0)}{d}, \quad \text{i.e.} \quad \kappa = -\frac{1}{2d}.
\]

Thus, \( E = -\hbar^2/8md^2 \) is the only available negative energy level. The corresponding state \( \psi \) is shown on Figure 13.

![Figure 13: The bound state in the delta-well](image)

Note that the \( \delta \)-well potential can be understood as a limit of finite well potentials, when depth \( V_0 \) tends to infinity, and width \( a \) to zero in such a way that \( 2mV_0a/\hbar^2 \) remains constant equal \( 1/d \). Then the quantity \( \mu \sqrt{V_0} = \sqrt{mV_0a^2/2\hbar^2} = \sqrt{a/4d} \) tends to 0. As Figure 11 shows, in this limit, indeed, only one energy level \( E_1 \) remains.

**Double-well and molecular bonding.** To get a glimpse of how a particle (such as an electron) behaves in the attracting field of two fixed particles (such as two protons in the hydrogen molecule \( H_2^+ \)), consider a double \( \delta \)-well potential:

\[
\frac{d^2\psi}{dq^2} + \frac{\alpha}{a} \left[ \delta \left( q - \frac{a}{2} \right) + \delta \left( q + \frac{a}{2} \right) \right] \psi = -\frac{2mE}{\hbar^2} \psi.
\]

Consider separately (Figure 14a) the case of even (\( \psi_+ \)) and odd (\( \psi_- \)) solution. Up to a scalar factor, they are:

\[
\psi_+(q) = \begin{cases} 
  e^{\mp \kappa (q \mp a/2)} \cosh \kappa a/2 & \text{if } \pm q \geq a/2 \\
  \cosh \kappa q & \text{if } |q| \leq a/2 
\end{cases}
\]

\[
\psi_-(q) = \begin{cases} 
  e^{\mp \kappa (q \mp a/2)} \sinh (\mp \kappa a/2) & \text{if } \pm q \geq a/2 \\
  \sinh \kappa q & \text{if } |q| \leq a/2 
\end{cases}
\]

where \( \kappa = \sqrt{-2mE/\hbar} \). The jump condition at \( q = a/2 \) reads
\[ \psi'(\frac{a-2}{2}) - \psi'(\frac{a+2}{2}) = \frac{\alpha}{a} \psi(\frac{a}{2}), \text{ i.e.} \]

\[ \psi_+ : \kappa \cosh \frac{\kappa a}{2} + \kappa \sinh \frac{\kappa a}{2} = \frac{\alpha}{a} \cosh \frac{\kappa a}{2}, \quad \text{or} \quad \frac{\alpha - \kappa a}{\kappa a} = \tanh \frac{\kappa a}{2} \]

\[ \psi_- : \kappa \sinh \frac{\kappa a}{2} + \kappa \cosh \frac{\kappa a}{2} = \frac{\alpha}{a} \sinh \frac{\kappa a}{2}, \quad \text{or} \quad \frac{\kappa a}{\alpha - \kappa a} = \tanh \frac{\kappa a}{2}. \]

The values of \( \kappa a \) satisfying these conditions are shown on Figure 14b in blue (\( \psi_+ \)) and green (\( \psi_- \)). We see that ground state \( \psi_+ \) has energy \( E = -\kappa^2 \hbar^2 / 2m \), where \( \kappa a > \alpha / 2 \). For \( \alpha/a = 1/d \) this implies that \( E \) is smaller than \( -\hbar^2 / 8md^2 \), the energy in a single \( \delta \)-shaped well. This explains the molecular bonding: the electron saves energy when it binds two hydrogen atoms comparing to when it sticks to one of them.

The excited state \( \psi_- \) exists only when \( \alpha \geq 2 \) and corresponds to higher energy (and smaller \( \kappa a < \alpha \)).

Let us compare \( \psi_\pm \) with the superposition of two copies of the states \( \psi \) in the single delta-well potential centered at \( q = \pm a/2 \). When the inter-atomic distance \( a \) is large (i.e. \( a \to \infty \), while \( \alpha/a = \))
1/d is kept constant), any linear combination of the functions $\psi(q - a/2)$ and $\psi(q + a/2)$ almost satisfies the Schrödinger equation for the double well. Indeed, since $\psi$ decays exponentially as $q \to \pm \infty$, the discrepancy is exponentially small in $a$. Thus, for all practical purposes, the situation is indistinguishable from having a two-dimensional space of states with the same energy $E \approx -\hbar^2/8md^2$. This can also be seen from Figure 14b: As $\alpha = a/d \to \infty$, the blue and green hyperbolas stretch to the right and intersect the black graph at almost the same point near the “tip”. The splitting between the energy levels of $\psi_+$ and $\psi_-$ becomes exponentially small in $a$ (because that’s how $\tanh \kappa a/2$ approaches its asymptotical value 1). However, as the distance $a$ between the ions decreases, the approximation to $\psi_\pm$ by the simple superposition $\psi(q - a/2) \pm \psi(q + a/2)$ gradually becomes worse, the (negative) energy level of the ground state $\psi_+$ becomes noticeably smaller (and that of the excited state noticeably greater) than $-\hbar^2/8md^2$, and the splitting between the energy levels of $\psi_+(q)$ and $\psi_-$ increases. When $a = \alpha d$ becomes smaller than $2d$, the excited state disappears.

This analysis will become helpful in the context of the band structure of solids.

The step potential, scattering, and probability current. The potential has the form $V(q) = 0$ for $q < 0$ and $V(q) = V_0 > 0$ for $q > 0$ (Figure 15). For $E > V_0$, the stationary Schrödinger equation has solutions

$$\psi(q) = A_+ e^{ikq} + A_- e^{-ikq} \ (q < 0), \ \psi(q) = B_+ e^{ik_0q} + B_- e^{-ik_0q} \ (q > 0).$$

Here $\hbar k = \sqrt{E/2m}$ and $\hbar k_0 = \sqrt{(E - V_0)/2m}$. The continuity conditions for $\psi$ and $d\psi/dq$ at $q = 0$ read:

$$A_+ + A_- = B_+ + B_- , \ ik(A_+ - A_-) = ik_0(B_+ - B_-).$$

Reminding ourselves about the time factor $e^{-itE/\hbar}$, we see that a general solution will have the form of two traveling waves, one in each direction, in each of the regions $q < 0$ and $q > 0$, with the amplitudes somehow redistributed between them at the barrier $q = 0$. For example, setting $B_- = 0$, we can interpret the situation as a wave $A_+ e^{ik(q-vt)}$ traveling from $q = -\infty$ with speed $v = Ek/\hbar$ toward the barrier, and then partly passing the barrier in the form of the wave $B_+ e^{ik_0(x-v_0t)}$ traveling toward $q = +\infty$ and partly reflected back as $A_- e^{ik(x+vt)}$. Taking also $A_+ = 1$ (for normalization’s sake), we find
the transmission and reflection coefficients:

\[ B_+ = \frac{2k}{k + k_0}, \quad A_- = \frac{k - k_0}{k + k_0}. \]

An attempt to give a probabilistic interpretation here stumbles at the following general issue: the probability density \( |\psi|^2 dq \) of a harmonic wave \( e^{ik_0q} \) equals \( dq \) and is not normalizable (since \( \int_0^\infty dq = \infty \)). In order to handle the problem, introduce the notion of probability current.

Namely, from the Schrödinger equations

\[
\begin{align*}
    i\hbar \frac{\partial \Psi}{\partial t} & = -\frac{\hbar^2}{2m} \frac{\partial^2 \Psi}{\partial q^2} + V(x) \Psi, \\
    -i\hbar \frac{\partial \Psi^*}{\partial t} & = -\frac{\hbar^2}{2m} \frac{\partial^2 \Psi^*}{\partial q^2} + V(x) \Psi^*,
\end{align*}
\]

for \( \Psi \) and its complex conjugate \( \Psi^* \), we compute

\[
\frac{\partial}{\partial t} |\Psi|^2 = \Psi^* \frac{\partial \Psi}{\partial t} + \Psi \frac{\partial \Psi^*}{\partial t} = \frac{i\hbar}{2m} \left( \Psi^* \frac{\partial^2 \Psi}{\partial q^2} - \Psi \frac{\partial^2 \Psi^*}{\partial q^2} \right)
\]

\[
= \frac{\partial}{\partial q} \left[ \frac{i\hbar}{2m} \left( \Psi^* \frac{\partial \Psi}{\partial q} - \Psi \frac{\partial \Psi^*}{\partial q} \right) \right] = -\frac{\partial j}{\partial q}.
\]

Here \( j = (\hbar/2mi)(\Psi^* \partial \Psi / \partial q - \Psi \partial \Psi^* / \partial q) \) is what’s called the probability current. In the integral form, we have

\[
\frac{d}{dt} \int_a^b |\Psi(q,t)|^2 dq = j(a,t) - j(b,t).
\]

This shows that the value \( j(q,t) \) equals the time rate with which probability density \( |\Psi|^2 dq \) is carried across the point \( q \).
One consequence of this computation is “conservation of probability”: assuming that $j(q,t) \to 0$ as $q \to \pm \infty$, we have

$$\frac{d}{dt} \int_{-\infty}^{\infty} |\Psi(q)|^2 dq = -j(q,t)|_{-\infty}^{\infty} = 0.$$ 

In the case of our traveling wave, this assumption is broken. Yet, we can compute the probability current:

$$\psi = \begin{cases} 
A e^{ikq} + A e^{-ikq} & q < 0 \\
B e^{ik_0q} + B e^{-ik_0q} & q > 0 
\end{cases}, \quad j = \begin{cases} 
\frac{\hbar k}{m} (|A_+|^2 - |A_-|^2) \\
\frac{\hbar \kappa}{m} (|B_+|^2 - |B_-|^2)
\end{cases}.$$ 

In our case $B_- = 0$, we derive the reflection ($R$) and transmission ($T$) probabilities (check that $R + T = 1$):

$$R = j_{\text{ref}} / j_{\text{inc}} = |A_-|^2 / |A_+|^2 = (k - k_0)^2 / (k + k_0)^2,$$

$$T = j_{\text{trans}} / j_{\text{inc}} = k_0 |B_+|^2 / k |A_+|^2 = 4kk_0 / (k + k_0)^2.$$ 

Note that in contrast with previous examples, where the allowed energy levels were discrete, here $E > V_0$ can be arbitrary. In such a situation, one speaks of the continuous spectrum of the Hamilton operator $\hat{H}$. In fact a linear differential operator can have both, the discrete part of the spectrum and the continuous one. As an example, in the case of the finite well potential, in addition to the discrete spectrum with $-V_0 < E_n < 0$, we would encounter continuous spectrum in the region $E > 0$.

**The rectangular barrier and tunneling.** The potential $V(q) = V_0 > 0$ for $0 < q < a$ and $V = 0$ elsewhere would present an impassable obstacle for a classical particle with energy $E < V_0$. As we shall see, for the quantum particle there is a positive probability of transmission through the barrier: the phenomenon known as tunneling.

We have:

$$\psi(q) = \begin{cases} 
A e^{ikq} + A e^{-ikq} & q < 0 \\
B e^{\kappa q} + B e^{-\kappa q} & 0 < q < a \\
C e^{ikq} & a < q
\end{cases}.$$ 

Here $\hbar k = \sqrt{2mE}$ and $\hbar \kappa = \sqrt{2m(V_0 - E)}$. The continuity conditions for $\psi$ at $q = 0$ and $q = a$ are:

$$A_+ + A_- = B_+ + B_-, \quad ik(A_+ - A_-) = \kappa(B_+ - B_-).$$
We use them to eliminate $B_\pm$ from the continuity conditions for $\psi'$:

$$B_+ e^{\kappa a} + B_- e^{-\kappa a} = C e^{ika}, \quad \kappa (B_+ e^{\kappa a} - B_- e^{-\kappa a}) = i k C e^{ika}.$$  

With notation $\zeta = \kappa + ik$, $\zeta^* = \kappa - ik$ and some algebra, we get:

$$A_+ \zeta + A_- \zeta^* = 2 C \zeta e^{-a \zeta^*}, \quad A_+ \zeta^* + A_- \zeta = 2 C \zeta^* e^{a \zeta}.$$  

Normalizing $C$ to be real, we conclude (since $\zeta$ is not) that $A_\pm$ are real too. Resorting to probability currents, we solve this as a linear system for $A_+$, and find:

$$\frac{A_+}{C} = e^{ika} \frac{\zeta^2 e^{-\kappa a} - \zeta^* 2 e^{\kappa a}}{\zeta^2 - \zeta^* 2}.$$  

After some more algebra, this yields the transmission coefficient:

$$T = \frac{|C|^2}{|A_+|^2} = \left[ \frac{4 k^2 \kappa^2}{4 k^2 \kappa^2 + (k^2 + \kappa^2)^2 \sinh^2 \kappa a} \right].$$  

When $\kappa a \gg 1$, the transmission coefficient

$$T \approx \left( \frac{4 k \kappa}{k^2 + \kappa^2} \right)^2 e^{-2\kappa a} = 16 \frac{E}{V_0} \left( 1 - \frac{E}{V_0} \right) e^{-2a \sqrt{2m(V_0-E)/\hbar}}$$

is very sensitive to the thickness $a$ of the barrier, and to the energy gap $V_0 - E = \hbar^2 \kappa^2 / 2m$. This has interesting applications e.g. in STM, the scanning tunneling microscopy. In a few words, the microscope's electrode retraces nanoscale details of a surface by keeping the electric current, formed by the electrons tunneling from the surface into the electrode through the vacuum gap, constant by means of a feedback loop.
6 The operator formalism

Hermitian linear algebra. After looking at a number of one-dimensional examples of quantum mechanical systems, it is time for us to outline the general picture. In classical mechanics, physical quantities (the observables) are real-valued functions on a (generally non-linear) symplectic phase space. In quantum mechanics, the phase space is a linear complex space, $\mathcal{H}$ (often called Hilbert space), equipped with an Hermitian inner product, $\langle \cdot | \cdot \rangle$, while the observables are Hermitian operators $A : \mathcal{H} \to \mathcal{H}$. These are complex analogues of the dot-product in real Euclidean geometry and symmetric operators respectively. More precisely, the inner product takes values in the field $\mathbb{C}$ of complex numbers, and must satisfy the following conditions. It must be Hermitian symmetric (here and everywhere the superscript “$\ast$” is the physicist’s notation for complex conjugation)

$$\langle \psi_2 | \psi_1 \rangle = \langle \psi_1 | \psi_2 \rangle^\ast$$

for all $\psi_1, \psi_2 \in \mathcal{H}$.

It must be $\mathbb{C}$-linear with respect to the ket argument $|\cdot\rangle$:

$$\langle \psi_3 | \lambda_1 \psi_1 + \lambda_2 \psi_2 \rangle = \lambda_1 \langle \psi_3 | \psi_1 \rangle + \lambda_2 \langle \psi_3 | \psi_2 \rangle$$

for all $\psi_i \in \mathcal{H}, \lambda_i \in \mathbb{C}$.

It must be anti-linear ($\frac{1}{2}$-linear) with respect to the bra argument $\langle \cdot | \cdot \rangle$:

$$\langle \lambda_1 \psi_1 + \lambda_2 \psi_2 | \psi_3 \rangle = \lambda_1^* \langle \psi_1 | \psi_3 \rangle + \lambda_2^* \langle \psi_2 | \psi_3 \rangle$$

for all $\psi_i \in \mathcal{H}, \lambda_i \in \mathbb{C}$.

And it must be positive in the sense that $||\psi||^2 := \langle \psi | \psi \rangle$ (which is real due to the Hermitian symmetry) is positive unless $\psi = 0$.

A complex linear operator $A : \mathcal{H} \to \mathcal{H}$ is a mapping satisfying

$$A(\lambda_1 \psi_1 + \lambda_2 \psi_2) = \lambda_1 A\psi_1 + \lambda_2 A\psi_2$$

for all $\psi_i \in \mathcal{H}, \lambda_i \in \mathbb{C}$.

It is called Hermitian if $A = A^\ast$, i.e.

$$\langle \psi_2 | A\psi_1 \rangle = \langle \psi_1 | A\psi_2 \rangle^\ast$$

for all $\psi_1, \psi_2 \in \mathcal{H}$.

Taking advantage of this Hermitian property, physicists often use the bra-ket notation: $|\psi\rangle$ for vectors in $\mathcal{H}$, $\langle \psi |$ for $\mathbb{C}$-linear functions $\mathcal{H} \to \mathbb{C}$, and $\langle \psi_1 | A | \psi_2 \rangle = \langle \psi_2 | A | \psi_1 \rangle^\ast$ for the Hermitian-symmetric inner product (not necessarily positive) associated this way with $A$.

For a finite-dimensional model of $\mathcal{H}$ we can take the space $\mathbb{C}^n$ of complex column $n$-vectors equipped with the inner product

$$\langle z | w \rangle = z_1^* w_1 + \cdots + z_n^* w_n.$$
An Hermitian operator in this space is given by an \( n \times n \)-matrix \( A = [a_{ij}] \) satisfying \( A = A^* \), where the superscript * stands for complex conjugation and transposition: \( a_{ji} = a^*_{ij} \) for all \( i, j = 1, \ldots, n \).

**The spectral theorem**, which is an Hermitian version of the orthogonal diagonalization theorem of linear algebra, says that any Hermitian operator \( A : \mathbb{C}^n \to \mathbb{C}^n \) has an orthonormal basis of eigenvectors:

\[
A|\psi_i\rangle = \lambda_i|\psi_i\rangle, \quad \langle \psi_i|\psi_j\rangle = 0 \quad \text{if} \quad i \neq j, \quad \langle \psi_i|\psi_i\rangle = 1, \quad i, j = 1, \ldots, n.
\]

Note that eigenvalues of an Hermitian operator are always real, and the eigenvectors corresponding to different eigenvalues are orthogonal automatically. Indeed,

\[
\lambda_i^* \langle \psi_j|\psi_i\rangle^* = \langle \psi_j|A|\psi_i\rangle^* = \langle \psi_i|A|\psi_j\rangle = \lambda_j \langle \psi_i|\psi_j\rangle,
\]

which for \( i = j \) implies \( \lambda_i^* = \lambda_i \), and for \( \lambda_i \neq \lambda_j \) implies \( \langle \psi_i|\psi_j\rangle = 0 \).

As we shall see, the Fourier method of solving Schrödinger equations is simply an application of the spectral theorem in infinite dimensions. In the examples we studied so far, the Hermitian inner product can be defined by the integral

\[
\langle \psi_1|\psi_2\rangle = \int_{-\infty}^{\infty} \psi_1^*(q)\psi_2(q) \, dq.
\]

The space \( \mathcal{H} \) then consists of complex-valued functions \( \psi \) on the line for which \( \int_{-\infty}^{\infty} |\psi(q)|^2 \, dq < \infty \). Note that the operator \( \hat{q} \) of multiplication by \( q \) (and actually by any real-valued function of \( q \)) is Hermitian. And so is the operator \( \hat{p} = -i\hbar d/dq \):

\[
\langle \psi_1|\hat{p}|\psi_2\rangle = \frac{\hbar}{i} \int_{-\infty}^{\infty} \psi_1^* \frac{d\psi_2}{dq} \, dq = -\frac{\hbar}{i} \int_{-\infty}^{\infty} \psi_2 \frac{d\psi_1^*}{dq} \, dq = \langle \psi_2|\hat{p}|\psi_1\rangle^*.
\]

Here we’ve integrated by parts assuming that the functions vanish at infinity (for otherwise \( \int |\psi_i|^2 dq \) would have no reason to converge).

Given two Hermitian operators, \( A^* = A \) and \( B^* = B \), their sum \( A + B \), as well as the expressions \( AB + BA \) and \( i(AB - BA) \) are Hermitian (check it!). The product is Hermitian only when the operators commute: \( AB \neq (AB)^* = B^*A^* = BA \) unless \( AB = BA \). But any operator commutes with itself, and so any real coefficient polynomial expression of one Hermitian operator is Hermitian.

In particular the hamiltonian

\[
\hat{H} = \frac{\hat{p}^2}{2m} + V(q) = -\frac{\hbar^2}{2m} \frac{d^2}{dq^2} + V(q)
\]
is Hermitian. The stationary Schrödinger equation

$$\hat{H}|\psi\rangle = E|\psi\rangle$$

has the abstract meaning of the eigenvalue problem: that of finding the eigenvalues $E$ for which a non-zero solution $|\psi\rangle$ exists. As we have seen in examples, the answer can look more complicated than the finite dimensional spectral theorem suggests: the spectrum of eigenvalues could have discrete and continuous ranges. We have no intention here to discuss the general spectral theory (studied in functional analysis), and will merely illustrate some aspects of it in two familiar examples: a free particle on a line and on a circle.

In the first case, $\psi_k(q) = e^{ikq}$ is an eigenfunction of the momentum operator $-i\hbar d/dq$ with the eigenvalue $\hbar k$, and therefore of the energy operator $\hat{H} = \hat{p}^2/2m$, with the eigenvalue $E_k = \hbar^2 k^2/2m$. Here the spectrum of $\hat{p}$ (resp. $\hat{H}$) consists of all real numbers $\hbar k$ (resp. all non-negative numbers). Any state $|\psi\rangle \in \mathcal{H}$ can be written as the superposition of the "orthonormal basis" $|\psi_k\rangle$ with coefficients $A(k)$, but the basis and the superposition are continual:

$$|\psi\rangle = \int_{-\infty}^{\infty} A(k)|\psi_k\rangle dk,$$

i.e.

$$\psi(q) = \int_{-\infty}^{\infty} A(k) e^{ikq} dk.$$ 

This is the familiar Fourier transform.

In the second example, consider a circle of length $L$, i.e. assume that functions $\psi$ are periodic with the period $L$. The above function $\psi_k = e^{ikq}$ is $L$-periodic only when $kL = 2\pi n$ for some integer $n$. We have therefore a discrete spectrum of the momentum operator

$$\psi_n = e^{2\pi i q n/L}, \quad \hat{p}|\psi_n\rangle = \hbar \frac{2\pi n}{L} |\psi_n\rangle, \quad n = 0, \pm 1, \pm 2, \ldots$$

The eigenfunctions $\psi_n$ form a basis in the space of complex-valued $L$-periodic functions in the sense that such a function $\psi$ can be written as their infinite linear combination:

$$|\psi\rangle = \sum_{n=-\infty}^{\infty} A_n |\psi_n\rangle.$$ 

Note that the basis $\{\psi_n\}$ is orthonormal with respect to the inner product

$$\langle \psi|\psi' \rangle = \frac{1}{L} \int \psi^*(q)\psi'(q) \, dq.$$ 

47
Therefore the coefficients $A_m$ can be found as the inner products

$$A_m = \langle \psi_m | \psi \rangle = \frac{1}{L} \oint \psi(q) e^{-2\pi mq/L} \, dq.$$  

In other words, periodic functions can be decomposed into Fourier series.

For general Hermitian operators, the spectral theorem guarantees a similar representation of elements of the Hilbert space by (possibly continual) superposition, of the (properly understood) eigenvectors. For simplicity, let us stick to the discrete spectrum case. Consider an observable, i.e. an Hermitian operator $A$. Its eigenvectors $|\psi_n\rangle$, $n = 1, 2, \ldots$, are interpreted as quantum states in which the observable assumes a definite value equal to the corresponding eigenvalue $a_n$. An arbitrary quantum state can be represented as a superposition

$$|\psi\rangle = \sum_{n=1}^{\infty} c_n |\psi_n\rangle, \quad c_n = \langle \psi_n | \psi \rangle,$$

the Fourier series in the abstract sense. Now Parseval’s identity

$$\|\psi\|^2 = \langle \psi | \psi \rangle = \sum_{n=1}^{\infty} |c_n|^2$$

allows one to interpret the normalized Fourier coefficient $|c_n|^2/\|\psi\|^2$ as the probability that in the quantum state $\psi$, the value of the observable $A$ equals $a_n$.

In fact this interpretation is consistent with our earlier interpretation of $|\psi(q)|^2 dq$ as the probability density of finding a particle in the state $\psi$ at a point $q$. Namely, the Hermitian operator $\hat{q}$ acts as multiplication $\psi(q) \mapsto q \psi(q)$. Its spectrum is continuous: the eigenfunctions are the Dirac delta-functions $\delta(q - q_0)$, with $q_0$ being the corresponding eigenvalues. Indeed, for any continuous $\phi$

$$\int \phi(q) \, q \delta(q - q_0) \, dq = q_0 \phi(q_0) = q_0 \int \phi(q) \delta(q - q_0) \, dq.$$  

The corresponding “Fourier representation” of $\psi$ is tautological:

$$\psi(q) = \int \psi(q_0) \delta(q - q_0) \, dq_0,$$

i.e. $\psi(q_0)$ is the “generalized Fourier coefficient” of $\psi$. Yet, the probability density of the event, that in the quantum state $\psi$ the value
of the observable \( \hat{q} \) equals \( q_0 \), is described by the squared normalized absolute value of this “coefficient”:

\[
\frac{|\psi(q_0)|^2}{\|\psi\|^2} dq_0, \quad \text{where} \quad \|\psi\|^2 = \int |\psi(q_0)|^2 dq_0.
\]

**The abstract uncertainty principle.** Consider two observables, \( A \) and \( B \). When the operators \( A \) and \( B \) have a common basis of eigenvectors \( \{ \psi_n \} \), they have diagonal matrices in this basis, and hence commute. The converse is also true. Namely, if two operators commute, then each eigenspace of \( A \) is \( B \)-invariant:

\[
A \psi = a \psi \quad \Rightarrow \quad A(B \psi) = B(A \psi) = B(a \psi) = a(B \psi).
\]

Taking in each eigenspace of \( A \) (for simplicity let us assume that \( A \) and \( B \) have discrete spectra) an orthonormal basis of eigenvectors of \( B \), we obtain a common for \( A \) and \( B \) orthonormal basis of eigenvectors. This fact is interpreted as the possibility of measuring commuting observables simultaneously: when \( A \psi = a \psi \) and \( B \psi = b \psi \), the values of \( A \) and \( B \) in the state \( \psi \) are equal to \( a \) and \( b \) respectively.

What if \( AB - BA \neq 0 \)? Let \( \psi \in \mathcal{H} \) be a state, which we for convenience assume normalized, i.e. \( \|\psi\| = 1 \). The expectation value \( \overline{A} \) of \( A \) at this state is given by

\[
\overline{A} = \sum_n a_n |c_n|^2 = \left\langle \sum_n c_n \psi_n \left| \sum_m a_m c_m \psi_m \right\rangle = \langle \psi | A | \psi \rangle.
\]

Subtracting the expectation value (i.e. a scalar operator, which commutes with anything), we may assume that \( \overline{A} = 0 \). Then the standard deviation \( \Delta A \) of \( A \) in the state \( \psi \) is determined by \( (\Delta A)^2 = \overline{A^2} = \langle \psi | A^2 | \psi \rangle = \| A \psi \|^2 \). We claim that

\[
\text{if} \quad C = i(AB - BA), \quad \text{then} \quad \Delta A \Delta B \geq \frac{|\langle \psi | C | \psi \rangle|}{2}.
\]

In fact this is a version of Schwartz’ inequality. For any real \( \lambda \)

\[
0 \leq \langle (A + i \lambda B) \psi | (A + i \lambda B) \psi \rangle = \langle A \psi | A \psi \rangle + \langle \psi | C | \psi \rangle \lambda + \langle \psi | B \psi \rangle \lambda^2.
\]

A non-negative quadratic polynomial in \( \lambda \) must have non-positive discriminant, i.e. \( \langle \psi | C | \psi \rangle^2 \leq 4 \| A \psi \|^2 \| B \psi \|^2 \) as promised.

For example, the Heisenberg commutation relation \( i(\hat{p} \hat{q} - \hat{q} \hat{p}) = \hbar \) shows that \( \Delta p \Delta q \geq \hbar/2 \).
Time evolution operators, and the energy-time uncertainty relation. Consider now the time evolution of quantum state according to the Schrödinger equation $i\hbar \frac{\partial}{\partial t} |\Psi\rangle = \hat{H} |\Psi\rangle$. Let $|\psi_n\rangle$, $n = 1, 2, \ldots$, form an eigenbasis: $\hat{H} |\psi_n\rangle = E_n |\psi_n\rangle$, and let $\Psi(t)$ satisfy the initial condition $|\Psi(0)\rangle = |\psi\rangle = \sum_n c_n |\psi_n\rangle$. Then

$$|\Psi(t)\rangle = \sum_n e^{iE_n \hbar t} c_n |\psi_n\rangle = e^{i\hat{H} \hbar t} |\psi\rangle.$$ 

The operators $U(t) = e^{i\hat{H} \hbar t}$ are unitary, i.e. preserve the Hermitian inner product:

$$\langle U\phi | U\psi \rangle = \langle \phi | \psi \rangle \text{ for all } \phi, \psi \in \mathcal{H}.$$ 

When $\hat{H}$ is time-independent (as it has been in all examples so far) they form a one-parametric group: $U(t_1 + t_2) = U(t_1)U(t_2)$.

Let us examine now how the expectation values of observables behave under time evolution. For a time-independent observable $A$,

$$-i\hbar \frac{d}{dt} \langle \Psi(t) | A | \Psi(t) \rangle = \langle i\hbar \dot{\Psi} | A | \Psi \rangle - \langle \Psi | A | i\hbar \dot{\Psi} \rangle = \langle \hat{H} \Psi | A | \Psi \rangle - \langle \Psi | A | \hat{H} \Psi \rangle = \langle \Psi | [\hat{H}, A] | \Psi \rangle,$$

where $[\hat{H}, A]$ denotes the commutator $\hat{H}A - A\hat{H}$. Thus,

$$\frac{d}{dt} A = \frac{i}{\hbar} [\hat{H}, A].$$

For example, taking $H = p^2/2m + V(q)$, we compute

$$[\hat{H}, \hat{p}] = i\hbar \frac{\partial V}{\partial q}, \quad [\hat{H}, \hat{q}] = -i\hbar \frac{\dot{p}}{m}.$$

These are (almost) Hamilton’s classical equations for the expectation values:

$$\frac{d}{dt} q = \frac{\overline{p}}{m}, \quad \frac{d}{dt} p = -\frac{\partial V}{\partial q}.$$

This result is known as Ehrenfest’s theorem.

In fact our computation with Hermitian forms $\langle \Psi | A | \Psi \rangle$ yields the abstract Schrödinger equation for the corresponding observables $A$:

$$\hbar \frac{d}{i dt} A = [\hat{H}, A].$$
In other words, instead of moving vectors in the Hilbert space by $|\psi\rangle \mapsto U(t)|\psi\rangle$, one can transform the observables:

$$\langle U(t)|A|U(t)\psi\rangle = \langle \psi|U(-t)AU(t)|\psi\rangle,$$

i.e. $t \mapsto U(-t)AU(t)$.

Differentiating in $t$, and using $i\hbar dU(t)/dt = \hat{H}U(t)$, one obtains $-i\hbar dA/dt = [\hat{H}, A]$.

Combining this with the abstract uncertainty relation, we find $\Delta E \Delta A \geq \hbar |dA/dt|/2$, where $\Delta E$ is the standard deviation of the energy operator $\hat{H}$ at a given state $\psi$. Let us use this relation to characterize the uncertainty $\Delta t$ of time, by defining $\Delta t := \Delta A/|dA/dt|$, and obtain the energy-time uncertainty relation:

$$\Delta E \Delta t \geq \hbar / 2.$$

The meaning of this uncertainty becomes more clear if we recall the example of time-evolving superposition of two states of a particle in the infinite well:

$$\Psi(q, t) = e^{-it\frac{E_1}{\hbar}}\psi_1 + e^{-it\frac{E_2}{\hbar}}\psi_2 = e^{-i\frac{E_1}{\hbar}t} \left( \sin \frac{\pi q}{L} + e^{it\frac{E_2 - E_1}{\hbar}} \sin \frac{2\pi q}{L} \right).$$

At $t = 0$, the two energy states $E_1$ and $E_2$ are mixed with 50% probability, so that the expected value is $(E_1 + E_2)/2$, and $\Delta E = (E_2 - E_1)/2$. The time interval $\Delta t$ needed to noticeably change the initial distribution is found by equating the exponent $2\Delta E \Delta t/\hbar$ to $\approx 1$. Equivalently $\Delta E \Delta t \approx \hbar / 2$.

Another illustration: In an eigenstate, $\hat{H}|\psi_0\rangle = E_0|\psi_0\rangle$, the energy value $E_0$ is certain ($\Delta E = 0$), but the system is bound to remain in this state forever: $U(t)|\psi_0\rangle = e^{tE_0/\hbar}|\psi_0\rangle$ (i.e. $\Delta t = \infty$).

Finally, let us recall that in classical Hamiltonian mechanics we introduced the extended phase space and the Hamilton function $H(p, q, \tau) = E$, where $-E$ is the momentum corresponding to the time variable $\tau$. The Schrödinger equation was obtained as the quantum analog of the Hamilton-Jacobi equation $E = H(p, q, \tau)$. From this point of view, $\hat{E} = i\hbar \partial / \partial t$ and $\hat{\tau} = t$ should satisfy the usual uncertainty relation $\Delta E \Delta t \geq \hbar / 2$. This is exactly the energy-time uncertainty relation we derived earlier in an ad hoc fashion.

Let us conclude this section with a fundamental example of a system with one degree of freedom where our understanding of the operator formalism will prove useful.
The harmonic oscillator. Consider a classical mechanical system, whose Hamilton function is the sum of kinetic and potential energy, near a stable equilibrium configuration. Such a configuration is inevitably a local minimum of the potential energy function. Retaining in the Taylor expansion of the potential energy at such an equilibrium the first nontrivial (i.e. quadratic) terms, we obtain an approximation describing well the small oscillations of the system near the equilibrium. This small oscillation model involves therefore two quadratic forms: the potential energy \( V(q) \) and kinetic energy \( T(\dot{q}) \). The first one can be assumed positive definite if the local minimum of the original potential energy is non-degenerate. The kinetic energy is always a positive definite quadratic function of velocities. In this situation, the orthogonal diagonalization theorem (from real linear algebra) says that by a suitable linear change of coordinates both quadratic forms can be simultaneously transformed to the following normal form:

\[
T = \frac{1}{2}(\dot{q}_1^2 + \cdots + \dot{q}_n^2), \quad V = \frac{1}{2}(\omega_1^2 q_1^2 + \cdots + \omega_n^2 q_n^2).
\]

Here \( n \) is the number of degrees of freedom, and \((\omega_1^2, \ldots, \omega_n^2)\) are positive coefficients (the spectrum of eigenvalues) unambiguously determined up to reordering by the pair of quadratic forms. Consequently the mechanical system falls apart into \( n \) independent 1-dimensional harmonic oscillators: \( \ddot{q}_i = -\omega_i^2 q_i \). Because of this, physicists say that near an equilibrium, any conservative system can be described as an ideal gas of harmonic oscillators. This explains the universal importance of the example we are about to study: the eigenvalue problem

\[
-\frac{\hbar^2}{2m} \frac{d^2}{dq^2} + \frac{\omega^2 q^2}{2} \psi = E \psi.
\]

Choose the dimensionless \( x = \sqrt{\frac{m \omega}{\hbar}} q \), and introduce operators

\[
a_\pm = \pm \frac{d}{dx} - x.
\]

They are not Hermitian, as they are adjoint to each other, \( a_+^* = a_- \):

\[
\langle \psi_1 | a_+ \psi_2 \rangle = \langle \psi_2 | a_- \psi_1 \rangle^* \quad \text{where} \quad \langle \psi_1 | \psi_2 \rangle := \int_{-\infty}^{\infty} \psi_1^*(x) \psi_2(x) \, dx.
\]

The property of \( a_\pm \) important to us is:

\[
(\pm \frac{d}{dx} - x)(\mp \frac{d}{dx} - x) = -\frac{d^2}{dx^2} + x^2 \mp 1.
\]
Consequently, denoting \(-\frac{d^2}{dx^2} + x^2\) by \(D\), we have:

\[
D = \frac{a_-a_+ + a_+a_-}{2}, \quad \frac{a_-a_+ - a_+a_-}{2} = 1.
\]

These commutation relations have the following implications.

Suppose that \(\psi_0\) is a state annihilated by \(a_-\), i.e. \(a_-\psi_0 = 0\). Then

\[
D\psi_0 = \frac{1}{2}a_-a_+\psi_0 = \frac{1}{2}(a_-a_+ - a_+a_-)\psi_0 = \psi_0.
\]

Next, \(Da_+ = a_+(D + 2)\). Indeed, applying \(a_-a_+ = a_+a_- + 2\) twice, we get:

\[
2Da_+ = (a_-a_+ + a_+a_-)a_+ = (a_+a_- + 2)a_+ + a_+(a_+a_- + 2) = a_+(a_-a_+ + a_+a_- + 4) = 2a_+(D + 2).
\]

Therefore if \(\psi\) is an eigenvector of \(D\) with an eigenvalue \(\lambda\), then \(a_+\psi\) is also an eigenvector, with the eigenvalue \(\lambda + 2\).

This yields the sequence \(\psi_n := a_+^n\psi_0\) of eigenvectors of \(D\) with the eigenvalues \(2n + 1\).

In fact \(\psi_0 = A_0e^{-x^2/2}\), as can be found by solving the 1-st order differential equation \(a_-\psi_0 = 0\), i.e. \(d\psi_0/dx = -x\psi_0\). Consequently

\[
\psi_n = A_n \left(-\frac{d}{dx} + x\right)^n e^{-x^2/2} = A_nH_n(x)e^{-x^2/2}.
\]

Here \(H_n\) are Hermite polynomials. For example,

\[
\begin{align*}
H_0 &= 1, \\
H_1 &= 2x, \\
H_2 &= 4x^2 - 2, \\
H_3 &= 8x^3 - 12x, \\
H_4 &= 16x^4 - 48x^2 + 12.
\end{align*}
\]

The operator relation \(\frac{d}{dx} - x = e^{x^2/2}\frac{d}{dx}e^{-x^2/2}\) shows that

\[
H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2}, \quad n = 0, 1, 2, \ldots
\]

For \(m \leq n\), integrating by parts \(n\) times, we find

\[
\int_{-\infty}^{\infty} H_m(x)H_n(x)e^{-x^2}dx = \int_{-\infty}^{\infty} H_m(x) \left(-\frac{d}{dx}\right)^n e^{-x^2}dx = \cdots
\]

\[
\cdots = \int_{-\infty}^{\infty} \left(\frac{d^n}{dx^n}H_m(x)\right)e^{-x^2}dx = \begin{cases} 0 & m < n \\ 2^n n! \sqrt{\pi} & m = n \end{cases}.
\]
In other words, $\langle \psi_m | \psi_n \rangle = 0$ for $m \neq n$ (in agreement with the spectral theorem), and $\| \psi_n \|^2 = 2^n n! \sqrt{\pi} |A_n|^2$. Thus, we obtain the orthonormal system of eigenvectors of $D$

$$\psi_n = \frac{H_n(x)e^{-x^2/2}}{(2^n n! \sqrt{\pi})^{1/2}}, \quad n = 0, 1, 2, \ldots,$$

$$\left(-\frac{d^2}{dx^2} + x^2 \right) \psi_n = (2n+1)\psi_n.$$

Returning from dimensionless to physicists’ notation, we have

$$\psi_n(q) = \frac{1}{\sqrt{2^n n!}} \left( \frac{m\omega}{\pi \hbar} \right)^{1/4} H_n \left( \sqrt{\frac{m\omega}{\hbar}} q \right) e^{-m\omega q^2/2\hbar}, \quad E_n = (n+\frac{1}{2}) \hbar \omega.$$

On Figure 16a, the first four eigenfunctions are sketched, and 16b

![Figure 16](image)

**Figure 16:** The harmonic oscillator

shows the corresponding level curves of the Hamilton function on the classical phase plane:

$$\frac{p^2}{2m} + \frac{m \omega^2 q^2}{2} = \frac{2n+1}{2} \hbar \omega.$$

In this example, several general phenomena can be observed. First, for any potential growing at infinity (quadratically in this case), the spectrum of the Hamilton operator is going to form an infinite discrete sequence $E_n$ increasing indefinitely with $n = 0, 1, 2, \ldots$
In one degree of freedom the $n$th eigenfunction, according to the so-called Sturm’s theorem, is going to have $n$ zeroes (in the case at hands they are the roots of Hermite polynomial $H_n$, which happens to have degree $n$ and only real roots). The ground energy level $E_0 = \hbar \omega / 2$ is non-zero (i.e. above the equilibrium level of the classical system) — in fact in agreement with the uncertainty principle, for at the equilibrium, both values: of the position ($q = 0$) and of the momentum ($p = 0$) are certain. In fact the product $\Delta p \Delta q$ of the principal sizes of the innermost ellipse on Figure 16b is exactly $\hbar / 2$ — as if quantum properties of the oscillator could be measured by examining the classical trajectory of the phase point over many revolutions!

It is probably more interesting to examine the phase areas $A_n$ enclosed by the ellipses. The reason is that $A(E) := \iint_{H(p,q) \leq E} dp \wedge dq$ is intrinsically defined in terms of the symplectic structure of the phase space. Moreover, the so-called $\textit{WKB-theory}$, similar to short-wave analysis in optics, provides approximate formulas for the eigenfunctions $\psi_n$ in terms of the classical trajectories. The trajectories are selected by certain “quantization rules” formulated in terms of such areas. For the oscillator, the WKB-theory prescribes to use the energy levels $E_n$ which enclose the areas $A_n = \pi \hbar (n + 1/2)$. (By the way, both the Planck constant and the symplectic area have physical dimension of action, mass $\times$ length$^2$/time, so it is not an accident that $A_n$ are expressed in the units of $\hbar$.) In our example, the quantization rule happens to yield not an approximate (for large $n$) but an exact (for all $n$) formula $E_n = (n + 1/2) \hbar \omega$.

One more remark in connection with the harmonic oscillator is in order: about its role in the quantum theory of fields. Fields have infinitely many degrees of freedom, but near an equilibrium (i.e. in “vacuum”) they still can be described by an “ideal gas of harmonic oscillators”. In this context, the lowest state $\psi_0$ of the harmonic oscillator is considered as the vacuum state. The operators $a_+$ and $a_-$, which transform $\psi_n$ to $\psi_{n+1}$ and $\psi_{n+1}$ to $\psi_n$ respectively, are interpreted as the operators of creation and annihilation of particles associated with the field. What makes this interpretation meaningful is the constant spacing $E_{n+1} - E_n = \hbar \omega$ between the nearby energy levels. Indeed, the state $\psi_n$ of the field is then understood as the presence of $n$ identical particles carrying the energy bit $\hbar \omega$ each. Thus, the example we have just explored serves as the staring point of quantum field theory.
7 Identical particles

Multiparticle systems. Consider, for example, a system of 3 non-interacting particles of masses \(m_1, m_2, m_3\), each moving on its own line with coordinate \(q_i, i = 1, 2, 3\), in its own field specified by the potential energy function \(V_i\). Then the Hamiltonian of the whole system

\[
H = \left(\frac{p_1^2}{2m_1} + V_1(q_1)\right) + \left(\frac{p_2^2}{2m_2} + V_2(q_2)\right) + \left(\frac{p_3^2}{2m_3} + V_3(q_3)\right).
\]

This illustrates the general principle: the total phase space of a system consisting of several subsystems is the Cartesian product of their individual phase spaces, while the total Hamiltonian is the sum of the Hamiltonians of the subsystems, provided that they don’t interact with each other.

The corresponding Schrödinger equation \(i\hbar \partial \Psi / \partial t = \hat{H} \Psi\) in our example has the form

\[
i\hbar \frac{\partial}{\partial t} \Psi(q_1, q_2, q_3, t) = \sum_{i=1}^{3} \left(-\frac{\hbar^2}{2m_i} \Delta q_i + V_i(q_i)\right) \Psi(q_1, q_2, q_3, t).
\]

It can be solved by the method of separation of variables: if \(\psi^{(i)}_n\), \(n = 1, 2, \ldots\), are eigenfunctions of \(\hat{H}_i\) with eigenvalues \(E^{(i)}_n\), then the products

\[
\psi_n := \psi^{(1)}_{n_1}(q_1)\psi^{(2)}_{n_2}(q_2)\psi^{(3)}_{n_3}(q_3)
\]

run the entire collection of eigenfunctions of \(\hat{H}\) as \(n = (n_1, n_2, n_3)\) runs all possible triples of positive integers. The corresponding eigenvalues are the sums:

\[
E_n = E^{(1)}_{n_1} + E^{(2)}_{n_2} + E^{(3)}_{n_3}.
\]

For instance, when the potential \(V_i\) is an infinite well of length \(L_i\), then the energy levels of the system are

\[
E_{n_1, n_2, n_3} = \frac{\pi^2 \hbar^2 n_1^2}{2m_1 L_1^2} + \frac{\pi^2 \hbar^2 n_2^2}{2m_2 L_2^2} + \frac{\pi^2 \hbar^2 n_3^2}{2m_3 L_3^2}.
\]

The corresponding eigenfunctions are

\[
\psi_{n_1, n_2, n_3} = \sqrt{\frac{8}{L_1 L_2 L_3}} \sin \frac{\pi n_1 q_1}{L_1} \sin \frac{\pi n_2 q_2}{L_2} \sin \frac{\pi n_3 q_3}{L_3}
\]
inside the rectangular box $0 \leq q_i \leq L_i$, and 0 outside of it.

By the way, by taking all the three $m_i = m$, we obtain the eigen-functions for a single particle of mass $m$ moving in 3-space inside the infinite well of the shape of a rectangular box of sizes $L_1 \times L_2 \times L_3$.

**Bosons and fermions.** Our goal, however, is to consider even more narrow situation, when not only the masses $m_i$, but also the potentials $V_i$ are the same, because we are interested in systems formed by several *identical* subsystems (“particles” in our example). From this point of view, the previous formulas (with $m_i = m$ and $L_i = L$) still describe correctly the system of 3 identical particles in the infinite potential well of length $L$ assuming that the particles are *distinguishable*.

We can imagine a situation when the 3 entities behave the same way in our potential, but are distinguished by some other property (say, color) irrelevant for the current problem. Still the state $\psi = \psi_1(q_1)\psi_2(q_2)\psi_2(q_3)$ with the energy $E_1 + 2E_2$ is different from the state $\psi' = \psi_2(q_1)\psi_1(q_2)\psi_2(q_3)$ with the same energy. Indeed, although probability density $|\psi(q_1, q_2, q_3)|^2$ of finding the particles 1, 2, 3 in the positions $(q_1, q_2, q_3)$ coincides with probability density $|\psi'(q_2, q_1, q_3)|^2$ of finding the particles 2, 1, 3 in the same positions, the events described by these probabilities are not the same, as long as the particles 1 and 2 differ in color.

However — and this is one of the main features of multiparticle quantum systems — quantum entities of genuinely the same nature, such as all photons, or all electrons, are *truly indistinguishable*, i.e. there is no way to detect a difference between the list $(1, 2, 3)$ and $(2, 1, 3)$ of such entities. Any two electrons are identical twins, and even themselves do not know which one carries which of the names, 1 or 2, we attempted to assign to them for the sake of our discussion.

More formally: A system of $N$ identical particles is *symmetric* with respect to permutations of the particles. Thus, if $\psi(q_1, \ldots, q_N)$ is the psi-function representing some quantum state of the system, then

$$\sigma \psi := \psi(q_{\sigma(1)}, \ldots, q_{\sigma(N)}), \quad \text{where} \quad \sigma : \{1, 2, \ldots, N\} \to \{1, 2, \ldots, N\}$$

is any invertible function, must represent the same quantum state. Interestingly, this does not mean that $\sigma \psi = \psi$. Experimentally measurable properties of the state are expressed by the probability density $|\psi|^2$, and so $\psi$ and $\sigma \psi$ still can differ by a complex factor $\epsilon$ of unit absolute value. Thus, interchanging two of the identical particles, say 1 and 2, we could get $\psi(q_2, q_1, \ldots) = \epsilon \psi(q_1, q_2, \ldots)$. But
then interchanging them again, we get $\epsilon^2 = 1$, i.e. inevitably either $\epsilon = 1$ or $\epsilon = -1$.

Thus, one can imagine two sorts of identical particles: *fermions*, interchanging any two of which causes the sign change in the psi-function of the system, and *bosons*, for which the interchange leaves the psi-function unchanged. That is,

$$\sigma \psi = \psi \text{ for bosons, and } \sigma \psi = \epsilon(\sigma)\psi \text{ for fermions,}$$

where $\sigma$ is any permutation, and $\epsilon(\sigma) = \pm 1$ is the *sign* of the permutation.

We will see that the Nature exploits both opportunities, bosons and fermions, and while the photons, for example, are bosons, the electrons are fermions. This becomes apparent as one considers the behavior of such multi-electron systems as atoms.

Returning to our example of 3 identical particles on the line, we should say that if the particles are indistinguishable, then the Hilbert spaces of all possible quantum states of the system is not the whole space of complex-valued functions $\psi(q_1, q_2, q_3)$ but a subspace of it, consisting of either *totally symmetric* functions (when the particles are bosons), or *totally anti-symmetric* functions (when they are fermions).

Let us examine the consequences for the eigenstates of our system. Assuming that $\psi_n$ are one-particle eigenfunctions with the eigenvalues $E_1 < E_2 < E_3 < \ldots$, we find that in the bosonic case, the ground state (the one with the minimal energy $3E_1$) is represented (the same as for distinguishable particles) by

$$\psi_{1,1,1} = \psi_1(q_1)\psi_1(q_2)\psi_1(q_3).$$

It is symmetric, as well as any eigenstate $\psi_{n,n,n}$, with energy $3E_n$. The next available energy level, $E_2 + 2E_1$, is represented by

$$\psi_2(q_1)\psi_1(q_2)\psi_1(q_2) + \psi_1(q_1)\psi_2(q_2)\psi_1(q_3) + \psi_1(q_1)\psi_1(q_2)\psi_2(q_3).$$

Similarly, from any $\psi_{n,n,m}$ with $m \neq n$, a single symmetric eigenstate $\psi_{m,n,n} + \psi_{n,m,n} + \psi_{n,n,m}$ with energy $E_m + 2E_n$ can be formed. Likewise, from $\psi_{l,m,n}$ with pairwise unequal $l, m, n$, a single symmetric eigenstate with $E = E_l + E_m + E_n$ is formed by taking the sum over all 6 permutations:

$$\psi_{l,m,n} + \psi_{m,l,n} + \psi_{m,n,l} + \psi_{n,m,l} + \psi_{n,l,m} + \psi_{l,n,m}.$$
In the fermionic case, such a sum taking in account the signs of permutations yields the determinant expression:

\[ \psi_{l,m,n} + \psi_{m,n,l} + \psi_{n,l,m} - \psi_{m,l,n} - \psi_{n,m,l} - \psi_{l,n,m} = \begin{vmatrix} \psi_l(q_1) & \psi_l(q_2) & \psi_l(q_3) \\ \psi_m(q_1) & \psi_m(q_2) & \psi_m(q_3) \\ \psi_n(q_1) & \psi_n(q_2) & \psi_n(q_3) \end{vmatrix} . \]

Note, however, that if \( l, m, n \) are not distinct, the determinant vanishes. This is a manifestation of the general rule known as Pauli’s exclusion principle:

No two identical fermions can occupy the same quantum state.

In our case, the lowest energy level available for the 3 identical fermions is \( E_1 + E_2 + E_3 \), corresponding to the ground state of the system when one of the fermions (but we can’t say which one!) occupies the level \( E_1 \), one \( E_2 \), and one \( E_3 \). The first excited state will have energy \( E_1 + E_2 + E_4 \), the transition into which from the ground state can be naively described by saying that the highest energy particle (which one is it?) moves from \( E_3 \) to the next available level \( E_4 \).

**Tensor algebra.** Let us now move our discussion of bosons and fermions to the next level, namely into the abstract realm of linear algebra and Hilbert spaces. When we say that quantum states of a system of 3 particles are represented by functions \( \psi(q_1, q_2, q_3) \) of 3 variables, we build a new Hilbert space, \( \mathcal{H} \), out of the three Hilbert spaces \( \mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3 \), each consisting of complex-valued functions \( \psi(q) \) in one variable. This new space has a notation in linear algebra, \( \mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3 \), and is called the tensor product of the three spaces. Generally, when we have configuration spaces \( C_i \) of classical mechanical systems, and form their common configuration space, we take their Cartesian product (say, \( C = C_1 \times C_2 \times C_3 \)). All functions on the Cartesian product of spaces form the tensor product of the respective spaces of functions. Of course, speaking of spaces of functions, one should be more specific and say what kind of functions (polynomial, continuous, differentiable, etc.) one is talking about. We are not going to concern ourselves here with analytical details, and will manipulate with Hilbert spaces as if they were finite dimensional. All we want to achieve is to get used to the tensor product notation by looking at a few examples.

To begin with, the space \( \mathbb{C}^n \) of coordinate vectors \( (x_1, \ldots, x_n) \) can be identified, using a slight change in syntax: \( (x(1), \ldots, x(n)) \),
with the space of complex-valued functions $x : \{1, \ldots, n\} \to \mathbb{C}$ on the finite set of $n$ objects. Thus, $\mathbb{C}^m \otimes \mathbb{C}^n$ is the space of functions on the Cartesian product of those finite sets. We usually call such functions \textit{matrices}, and write $A_{ij}$ instead of $A(i, j)$. Thus, $m \times n$-matrices form the tensor product space $\mathbb{C}^m \otimes \mathbb{C}^n$.

The dimension of the tensor product space is the product $mn$ of the dimensions of the factors. Given some bases $\{e_i\}$ and $\{f_j\}$ in spaces $\mathcal{V}$ and $\mathcal{W}$, the basis in $\mathcal{V} \otimes \mathcal{W}$ whose elements can be denoted by $e_i \otimes f_j$ is defined. In our example with matrices, this is the basis of delta-functions of points in the Cartesian product, or, equivalently the elementary matrices: $A(i, j) = 1$ for one pair $(i, j)$ and 0 for all others. Another example: when we build eigenfunctions $\psi_{n_1, n_2, n_3} = \psi_{n_1}(q_1)\psi_{n_2}(q_2)\psi_{n_3}(q_3)$ for the system of 3 identical but distinguishable particles, we actually take the basis $\{\psi_n\}$ of eigenfunctions in the Hilbert space $\mathcal{H}$ for one particle, and use

$$\psi_{n_1, n_2, n_3} := \psi_{n_1} \otimes \psi_{n_2} \otimes \psi_{n_3}$$

for the basis in the space $\mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H}$. This is another way to say that the three functions are multiplied not in the usual, pointwise way as in $\psi_{n_1}(q)\psi_{n_2}(q)\psi_{n_3}(q)$, but by using in each factor its own variable.

To summarize: the space of states of the system formed by $N$ identical but distinguishable copies of the initial system (“particles”) with the state space $\mathcal{H}$ each is the $N$th tensor power

$$\mathcal{H} \otimes N := \mathcal{H} \otimes \cdots \otimes \mathcal{H} \quad (N \text{ times}).$$

If $\hat{H} : \mathcal{H} \to \mathcal{H}$ is the hamiltonian of the one-particle system, the hamiltonian of the $N$-particle system is the sum of operators acting in each factor separately:

$$\sum_{i=1}^{N} I \otimes \cdots \otimes I \otimes \hat{H} \otimes I \otimes \cdots \otimes I \quad (\hat{H} \text{ on the } i\text{th place}).$$

When $\psi_n$ form a basis of eigenstates of $\hat{H}$ of the one-particle system with the energies $E_n$, the eigenstates of the $N$-particle system, in our new notation, are

$$\psi_{n_1, \ldots, n_N} := \psi_{n_1} \otimes \cdots \otimes \psi_{n_N}, \quad \text{and the eigenvalue is } E_{n_1} + \cdots + E_{n_N}.$$
Symmetric and anti-symmetric tensors. Let us now describe the machinery that accounts for the states of \( N \) bosonic particles. An element \( \psi \in \mathcal{H} \) can be written as a linear combination of the basis: \( \psi(x) = \sum_n x_n \psi_n \). We are going to treat \( x_n \) not as numbers but as variables, and write

\[
\psi(x)^\otimes N = \left( \sum x_n \psi_n \right) \otimes \left( \sum x_n \psi_n \right) \otimes \cdots \otimes \left( \sum x_n \psi_n \right) \quad (N \text{ times}).
\]

Now do the multiplication, and write the result as a polynomial in \((x_1, x_2, \ldots, x_n, \ldots)\) with coefficients in \( \mathcal{H}^\otimes N \), for instance (\( N = 3 \)):

\[
\begin{align*}
(x_1 \psi_1 + x_2 \psi_2 + \cdots) \otimes (x_1 \psi_1 + x_2 \psi_2 + \cdots) \otimes (x_1 \psi_1 + x_2 \psi_2 + \cdots) &= x_1^3 (\psi_1 \otimes \psi_1 \otimes \psi_1) + x_2^3 (\psi_2 \otimes \psi_2 \otimes \psi_2) + \cdots \\
&+ x_1^2 x_2 (\psi_2 \otimes \psi_1 \otimes \psi_1 + \psi_1 \otimes \psi_2 \otimes \psi_1 + \psi_1 \otimes \psi_1 \otimes \psi_2) \\
&+ x_1 x_2^2 (\psi_1 \otimes \psi_2 \otimes \psi_2 + \psi_2 \otimes \psi_1 \otimes \psi_2 + \psi_2 \otimes \psi_2 \otimes \psi_1) + \cdots
\end{align*}
\]

As we see, due to the commutativity of the variables \( x_n \), only the symmetric combinations of the tensor products \( \psi_{n_1, \ldots, n_N} \) occur. They form a basis in the part of \( \mathcal{H}^\otimes N \) invariant under permutations of all the \( N \) factors. This part is denoted \( S^N(\mathcal{H}) \), and called the \( N \)th symmetric tensor power of \( \mathcal{H} \). If \( \mathcal{H} \) is \( m \)-dimensional, the dimension of \( \mathcal{H}^\otimes N \) equals \( m^N \). The dimension of \( S^N(\mathcal{H}) \) is equal to the number of degree \( N \) monomials in \( m \) variables: \( x_1^{n_1} x_2^{n_2} \cdots x_m^{n_m} \) such that \( n_1 + n_2 + \cdots n_m = N \). It is given by the binomial coefficient \( \binom{N+m-1}{N} \). Furthermore, one can say that the space of degree \( N \) polynomial functions on \( \mathcal{H} \) is identified by this construction with the dual, \( S^N(\mathcal{H})^* \), i.e. the space of linear functions on the space \( S^N(\mathcal{H}) \) of states for \( N \) bosonic particles.

A similar approach to fermions is also possible, and it goes back to 19th century mathematician Hermann Grassmann who invented what is called now Grassmann or exterior algebra. At the hands of the 20th century mathematician Felix Berezin it got transformed into an alternative version of geometry: super-geometry. While in ordinary geometry, variables commute: \( x_i x_j = x_j x_i \), in super-geometry, there are two sorts of variables, including the new sort, odd variables \( \xi_i \). They commute with the usual, even ones: \( \xi_i x_j = x_j \xi_i \), but anti-commute with each other: \( \xi_i \xi_j = -\xi_j \xi_i \). In particular, \( \xi_i^2 = 0 \). These are the variables we currently need.

Thus, we write a generic element of \( \mathcal{H} \) as the linear combination \( \psi(\xi) = \sum_n \xi_n \psi_n \), and form the tensor product:

\[
\left( \sum \xi_n \psi_n \right) \otimes \left( \sum \xi_n \psi_n \right) \otimes \cdots \otimes \left( \sum \xi_n \psi_n \right) \quad (N \text{ times}).
\]
In multiplying out and reducing terms similar in $\xi$s. we should re-
member that turning $\xi_n\xi_m$ into $\xi_m\xi_n$ is accompanied with the change
of sign, and terms containing $\xi^2_n$ disappear. For example ($N = 2$)

$$(\xi_1\psi_1 + \xi_2\psi_2 + \xi_3\psi_3 + \cdots) \otimes (\xi_1\psi_1 + \xi_2\psi_2 + \xi_3\psi_3 + \cdots) =$$

$$\xi_1\xi_2 (\psi_1 \otimes \psi_2 - \psi_2 \otimes \psi_1) + \xi_1\xi_3 (\psi_1 \otimes \psi_3 - \psi_3 \otimes \psi_1) +$$

$$\xi_2\xi_3 (\psi_2 \otimes \psi_3 - \psi_3 \otimes \psi_2) + \cdots$$

Likewise, for $N$ fermions, each monomial in $\xi$s can be uniquely writ-
ten in the form $\xi_{n_1}\xi_{n_2}\cdots\xi_{n_N}$ with monotonically increasing sub-
scripts $n_1 < n_2 < \cdots < n_N$. The “coefficient” will be the totally
anti-symmetric eigenstate with the energy $E_{n_1} + \cdots + E_{n_N}$, repre-
sented by the $N \times N$-determinant expression

$$\sum_{\text{permutations } \sigma} \epsilon(\sigma) \psi_{n_{\sigma(1)}} \otimes \cdots \otimes \psi_{n_{\sigma(N)}}.$$ 

These eigenstates form a basis in the space, denoted $\Lambda^N(\mathcal{H})$ and
called the $N$th exterior power of $\mathcal{H}$. When $\mathcal{H}$ has finite dimension
$m$ (so that only $m$ letters $\xi$ are available), the total number of the
monomials is $\binom{m}{N}$. In particular, no states with more than $m$ fermions
exist. Similar to the bosonic case, the space of states of the system of
$N$ fermions is identified with the dual, $\Lambda(\mathcal{H})^*$, of the space of degree
$N$ polynomials, but in the anti-commuting, Grassmann coordin-
ates $\xi_n$ on the space $\mathcal{H}$. Thus, the space $\mathcal{H}$ of states of a fermionic particle
is, in terms of super-geometry, an odd super-space.

Two ideal gases: oscillators and qubits. Intruding once
again into the domain of quantum field theory, let us derive now a
conclusion from our description of the phase spaces of multi-particle
systems of bosons and fermions, which will prove useful in the context
of quantum statistical mechanics.

Consider the system containing an indeterminate number $N = 0, 1, 2, \ldots$ of identical bosons. Since the events of having different
numbers of particles are mutually disjoint, the total quantum phase
space of such a system can be described as the direct sum:

$$S(\mathcal{H}) = \bigoplus_{N=0}^{\infty} S^N(\mathcal{H}) = \mathbb{C} \oplus \mathcal{H} \oplus S^2(\mathcal{H}) \oplus \cdots.$$ 

The dual of it can be therefore identified with the total space

$$\mathbb{C}[x_1, x_2, \ldots, x_n, \ldots]$$

of all polynomials in the coordinates of $\psi = \sum_n x_n \psi_n$. But $\mathcal{H}$ is the
direct sum (= Cartesian product) of 1-dimensional spaces (the lines
spanned by \( \psi_n \)). Functions on a Cartesian product of spaces form the tensor product:

\[
\mathbb{C}[x_1, x_2, \ldots, x_n \ldots] = \mathbb{C}[x_1] \otimes \mathbb{C}[x_2] \otimes \cdots \otimes \mathbb{C}[x_n] \otimes \cdots.
\]

In simple terms, the monomials \( x_k^n, k = 0, 1, 2, \ldots \) form a basis of one-variable polynomials \( \mathbb{C}[x_n] \), and their products \( x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n} \cdots \) (where however all but finitely many \( k_n = 0 \)) form a basis of polynomials in all the variables. This means that the space of states of the system with an indeterminate number of identical bosonic particles can be rearranged into the phase space of the system of \textit{distinguishable} fictitious “particles”. Each fictitious particle corresponds to one state, \( \psi_n \), of a single boson, and has the sequence of states \( \psi_n^\otimes k \) (of energies \( kE_n \)), \( k = 0, 1, 2, \ldots \). In a way, one can think of the whole system as an ideal gas of (distinguishable) harmonic oscillators of frequencies \( \omega_n = E_n/\hbar \).

In the fermionic case, the story is similar. The sum

\[
\Lambda(\mathcal{H}) = \bigoplus_{N=0}^{\infty} \Lambda^N(\mathcal{H}) = \mathbb{C} \oplus \mathcal{H} \oplus \Lambda^2(\mathcal{H}) \oplus \cdots
\]

describes the quantum phase space of the system of an indeterminate number \( N = 0, 1, 2, \ldots \) of identical fermions. The dual space can be identified with the entire \textit{Grassmann algebra} \( \mathbb{C}[\xi_1, \xi_2, \ldots, \xi_n, \ldots] \) of “polynomial” functions on \( \mathcal{H} \) considered as an odd super-space. The Grassmann algebra on the direct sum of lines is the tensor product:

\[
\mathbb{C}[\xi_1, \xi_2, \ldots, \xi_n, \ldots] = \mathbb{C}[\xi_1] \otimes \mathbb{C}[x_2] \otimes \cdots \otimes \mathbb{C}[\xi_n] \otimes \cdots.
\]

The difference with the bosonic case is that the “polynomial” algebra in one Grassmann variable is 2-dimensional (since \( \xi^2 = 0 \)):

\[
\mathbb{C}[\xi] = \{ a + b\xi \mid a, b \in \mathbb{C} \}.
\]

Thus, one can say that the phase space of a system consisting of an indeterminate number of identical fermions can be rearranged into the phase space of distinguishable fictitious entities, which would be appropriate to call \textit{qubits}. Each qubit corresponds to one state, \( \psi_n \), of the original fermion, and has only two states: \( \psi_n^\otimes k \) (of energy \( kE_n \)), where \( k = 0 \) or \( 1 \). The value 0 signifies that the energy level \( E_n \) is vacant, and \( k = 1 \) that it is occupied by a fermion. For example, when the total number \( m \) of the fermion’s states is finite, the number \( N \) of fermions in the system can vary only from 0 to \( m \). The total dimension of the phase space \( \Lambda(\mathcal{H}) \) is then \( \sum_{N=0}^{m} \binom{m}{N} = 2^m \). This space is rearranged as the tensor product of \( m \) planes \( \mathbb{C}^2 \). One can interpret this system as an “ideal gas” of \( m \) qubits.
8 The hydrogen atom

Central fields and the orbital angular momentum. The time-independent Schrödinger equation for a particle in 3-space moving in a central potential field has the form

\[-\frac{\hbar^2}{2m} \Delta \psi + V(r) \psi = E \psi, \quad \Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.\]

Here \( V \) is the potential energy depending only on the distance \( r = \sqrt{x^2 + y^2 + z^2} \) to the origin.

When \( V \) is proportional to \(-1/r\), this is the quantum analog of the Kepler problem — except that instead of describing planetary motions in the gravitational field of the Sun, it can be used to model an electron in the electrostatic field of proton in the hydrogen atom, or as a matter of fact in any atom. Although the atoms are multi-body objects, the same way as solving the classical Kepler problem unlocks the door into the mechanics of the Solar system (which is also multi-body), the theory of one electron in the electrostatic field of the atomic nucleus is the key to understanding the atom. Having this application in mind, we will take \( V = -Ze^2/4\pi\varepsilon_0 r \), where \( -e \) is the charge of the electron, and \( Ze \) is the charge of the nucleus (e.g. \( Z = 1 \) for hydrogen, \( Z = 2 \) for helium, etc.)

The same as in the case of the classical Kepler problem, it is the rotational symmetry, i.e. the invariance of the equation under the group \( SO_3 \) of rotations about the origin, that makes this problem approachable. Namely, for a classical particle moving in a central force field, the angular momentum vector \( L = q \times p \) is conserved. Let \( L_z = xp_y - yp_x \) be the \( z \)-component of \( L \). Its quantization has the form \( \hat{L}_z = i\hbar(y\partial/\partial x - x\partial/\partial y) \). The corresponding evolution equation \( i\hbar\partial\Psi/\partial t = \hat{L}_z \Psi \) can be solved by the method of characteristics. The latter are solutions to \( \dot{x} = y, \dot{y} = -x, \dot{z} = 0 \) which simply describe rotations of \( q = (x, y, z) \) about the \( z \)-axis. Consequently

\[ \Psi(t, x, y, z) = \Psi(0, x \cos t + y \sin t, x \sin t - y \cos t, z). \]

In particular, any function \( f(r) \) depending only on the radius \( r = \sqrt{x^2 + y^2 + z^2} \) is annihilated by \( \hat{L}_z \). Therefore the operator of multiplication by \( f \) commutes with \( \hat{L}_z \) (and likewise with \( \hat{L}_y \) and \( \hat{L}_x \)).

For similar reasons, the Laplacian \( \Delta \) (which is rotation-invariant) also commutes with \( \hat{L}_x, \hat{L}_y, \hat{L}_z \), and therefore the central field hamiltonian \( \hat{H} = -\hbar^2 \Delta/2m + V(r) \) does too. This is convenient, because
commuting Hermitian operators have a common basis of eigenfunctions. So, we could begin with looking for solutions to $\hat{L}_z \Psi = \lambda_z \Psi$, the eigenfunctions of $\hat{L}_z$. This equation involves only the dependence of $\Psi$ on the polar (cylindrical) angle $\theta$ about the $z$-axis, and can be written as $-i\hbar d\Psi / d\theta = \lambda_z \Psi$. Thus, $\Psi(\theta) = \Psi_0 e^{i\lambda_z \theta / \hbar}$. But $\Psi$ must be $2\pi$-periodic in $\theta$. This shows that

$$\lambda_z = k \hbar, \ k = 0, \pm 1, \pm 2, \ldots,$$

i.e. in quantum mechanics, the values of the components of the orbital angular momentum are always integer multiples of $\hbar$.

For example, examine $\Psi$ polynomial in $x, y, z$, possibly with complex coefficients. Let us use complex coordinates $w = x + iy$, $\bar{w} = x - iy$ instead of $x = (w + \bar{w})/2$ and $y = (w - \bar{w})/2i$. Since in polar notation, $w^a \bar{w}^b = |w|^{a+b} e^{i(a-b)\theta}$, we conclude that $(x + iy)^a (x - iy)^b z^c$, $a, b, c \geq 0$, are the eigenfunctions with the eigenvalues $\lambda_z = (a - b) \hbar$, and form the eigenbasis of $\hat{L}_z$ in the space $\mathcal{P} := \mathbb{C}[x, y, z]$ of all polynomial functions.

This success comes with complications: by exploiting it we would forfeit the $SO_3$-symmetry. Indeed, the angular momentum operators don’t commute, e.g. $[\hat{L}_z, \hat{L}_x] = i\hbar \hat{L}_y$. So the eigenfunctions of $\hat{L}_z$ won’t be also eigenfunctions of $\hat{L}_x$ (unless they are eigenfunctions of $\hat{L}_y$ too, i.e. have the form $f(r)$).

One way out is to use instead the operator of the total orbital angular momentum $\hat{L}^2 = \hat{L}_x^2 + \hat{L}_y^2 + \hat{L}_z^2$. It is rotation-symmetric, and commutes with all the three components, and with $\hat{H}$ as well. So, let’s compute it. We have

$$\left( y \frac{\partial}{\partial x} - x \frac{\partial}{\partial y} \right)^2 = y^2 \frac{\partial^2}{\partial x^2} + x^2 \frac{\partial^2}{\partial y^2} - 2xy \frac{\partial^2}{\partial x \partial y} - x \frac{\partial}{\partial x} - y \frac{\partial}{\partial y}.$$

Adding the other two squares, and using some algebra, we get

$$-\frac{1}{\hbar^2} \hat{L}^2 = r^2 \Delta - (\vec{r} \cdot \nabla)^2 - (\vec{r} \cdot \nabla).$$

By the way, this is the quantum version of the geometric identity

$$|L|^2 := |q \times p|^2 = (q \cdot q)(p \cdot p) - (q \cdot p)^2.$$

Here

$$\vec{r} \cdot \nabla = x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z}.$$

66
is the Euler operator, which multiplies each homogeneous functions by its homogeneity degree. This suggest the way of looking for eigenfunctions of the hamiltonian $\hat{H}$ by separating the radial and the spherical directions: Look for solutions to $-\hbar^2 \Delta \psi / 2m + V(r)\psi = E\psi$ in the form $\psi(\vec{r}) = f(r)\phi$, where $\phi$ is a function of homogeneity degree 0, and hence determined by its values on the unit sphere $|\vec{r}| = 1$. Moreover, we may assume that $\phi$ is an eigenfunction of $\hat{L}^2$ with the eigenvalue $-\hbar^2 \mu$. Now substitute for $\Delta$ its expression via $\vec{r} \cdot \nabla$ (which acts trivially on $\phi$) and $\hat{L}^2$ (which acts trivially on $f(r)$). Writing $f$ for further convenience in the form $f(r) = \frac{u(r)}{r}$, we obtain:

$$-\frac{\hbar^2}{2m} \frac{d^2 u}{dr^2} + \left[ \frac{\mu \hbar^2}{2mr^2} + V(r) \right] u = Eu,$$

and $\hat{L}^2 \phi = \mu \hbar^2 \phi$.

Namely, the “Eulerian” terms $rf' + rf'$ have simplified into $ru''$. Indeed, $rf' = r(u/r)' = u' - u/r$ and $rf'' = r(u' - u/r) = ru'' - u' + u/r$. So the sum is $ru''$.

The left equation has the form of a time-independent Schrödinger equation on the line with the effective potential $\frac{\mu \hbar^2}{2mr^2} + V(r)$. The term depending on the eigenvalue of the angular momentum operator is called centrifugal, as it creates a “barrier” keeping the particle away from the origin $r = 0$. Before attempting to solve the equation, we need to study the eigenvalues of the angular momentum operator in the space of functions on the unit sphere.

Spherical harmonics. We will solve here a more general problem, namely decompose the space $\mathcal{P} = \mathbb{C}[x, y, z]$ of complex coefficient polynomials into eigenspaces of $\hat{L}^2$, and find the corresponding eigenvalues. The restrictions of such polynomials to the unit sphere will then give us the solution of the previous problem of finding $\phi$.

First, the space $\mathcal{P}$ can be represented as the direct sum

$$\mathcal{P} = \bigoplus_{l \geq 0} \mathcal{P}_l,$$

where $\mathcal{P}_l$ is the space of homogeneous polynomials of degree $l$. For example, $\mathcal{P}_0 = \mathbb{R}$ consists of constant functions, $\mathcal{P}_1$ is spanned by $x, y, z$, $\mathcal{P}_2$ by $x^2, y^2, z^2, xy, yz, zx$, etc. In general the dimension $\dim \mathcal{P}_l$ equals the number $\binom{l+2}{2} = (l + 1)(l + 2)/2$ of monomials $x^ay^bz^c$ of total degree $a + b + c = l$. Clearly, the components of $\hat{L}$ and respectively $\hat{L}^2$ preserve the degrees: $\hat{L}^2 : \mathcal{P}_l \to \mathcal{P}_l$. So, we want to decompose $\mathcal{P}_l$ into the eigenspaces of $\hat{L}^2$. The Laplacian $\Delta$ and
the operator of multiplication by \( r^2 = x^2 + y^2 + z^2 \), which commute with \( \hat{L}^2 \), will help.

Namely, for \( l > 1 \), the operator of multiplication \( r^2 : \mathcal{P}_{l-2} \to \mathcal{P}_l : p \mapsto r^2p \) embeds \( \mathcal{P}_{l-2} \) as the subspace \( r^2\mathcal{P}_{l-2} \subset \mathcal{P}_l \). This subspace is invariant under \( \hat{L}^2 \) (for \( \hat{L}^2 r^2p = r^2\hat{L}^2 p \)). Hence it decomposes into eigenspaces of \( \hat{L}^2 \) the same way as \( \mathcal{P}_{l-2} \) does. The image \( r^2\mathcal{P}_{l-2} \) has codimension in \( \mathcal{P}_l \) equal \( (l+2) - (1) = 2l + 1 \).

We claim that in \( \mathcal{P}_l \), the operator \( \hat{L}^2 \) has exactly one eigenspace, \( \mathcal{H}_l \), complementary to \( r^2\mathcal{P}_{l-2} \), and that the eigenvalue of \( \hat{L}^2 \) on \( \mathcal{H}_l \) equals \( l(l+1)\hbar^2 \). Moreover, we claim that \( \mathcal{H}_l \) is the null-space of the Laplacian \( \Delta : \mathcal{P}_l \to \mathcal{P}_{l-2} : H_l = \{ p \in \mathcal{P}_l \mid \Delta p = 0 \} \).

Traditionally (as any functions annihilated by the Laplacian) elements from \( \mathcal{H}_l \) are called harmonic polynomials, or spherical harmonics (not to be confused with harmonic waves), and \( \mathcal{H}_l \) is called the space of spherical harmonics of degree \( l \).

Of course, if \( \Delta p = 0 \), then \( \Delta \hat{L}^2 p = \hat{L}^2 \Delta p = 0 \), i.e. the space \( \mathcal{H}_l \) is invariant under \( \hat{L}^2 \). Moreover, recalling from the previous subsection the identity

\[
\hat{L}^2 = \hbar^2 (\vec{r} \cdot \nabla)^2 + \hbar^2 (\vec{r} \cdot \nabla) - \hbar^2 r^2 \Delta,
\]

and applying the right hand side to a harmonic polynomial \( p \) of degree \( l \), we see that \( \hat{L}^2 p = \hbar^2 (l^2 + l) \) as promised. It remains to show that \( \mathcal{P}_l = r^2\mathcal{P}_{l-2} \oplus \mathcal{H}_l \), i.e. that \( \mathcal{H}_l \) is complementary to \( r^2\mathcal{P}_{l-2} \).

One way to do this would be, by using a bit of complex analysis, to show that the operators \( r^2 \) and \( \Delta \) are adjoint with respect to some inner product on the space \( \mathcal{P} \) of all polynomials, and conclude from this that the null-space of \( \Delta \) is the orthogonal complement to the range of \( r^2 \).

To keep the matters more elementary though, let us resort to pure algebra, and find out how \( r^2 \) commutes with \( \Delta = \nabla \cdot \nabla \). Here \( \nabla \cdot \) denotes the divergence operator, and the right \( \nabla \) the gradient. We have:

\[
\Delta r^2 p = \nabla \cdot (2p\vec{r} + r^2 \nabla p) = 6p + 4\vec{r} \cdot \nabla p + r^2 \Delta p.
\]

Applying this to a homogeneous \( p \), we find:

\[
p \in \mathcal{P}_{l-2} \implies \Delta (r^2 p) = r^2 \Delta p + (4l - 2)p.
\]
One consequence of this is that a harmonic polynomial divisible by $r^2$ is zero. Indeed, assuming that $r^{2k}p$ is harmonic, of degree $l$, and applying the commutation relation inductively, we compute

$$0 = \Delta(r^{2k}p) = r^2\Delta r^{2k-2}p + (4l-2)r^{2k-2}p = \cdots = r^{2k}\Delta p + Nr^{2k-2}p,$$

where $N$ is some positive integer. This shows that $p$ is divisible by $r^2$, i.e. $r^{2k}p$ is divisible by a higher power of $r^2$ than was initially assumed. This leads to a contradiction unless $p = 0$.

Thus, $\mathcal{H}_l \cap r^2\mathcal{P}_{l-2} = \{0\}$, so dim $\mathcal{H}_l$ cannot exceed codim $r^2\mathcal{P}_{l-2} = \dim \mathcal{P}_l - \dim \mathcal{P}_{l-2}$. But by the rank theorem, the null-space of $\Delta$ has dimension at least $\dim \mathcal{P}_l - \dim \mathcal{P}_{l-2}$. Consequently, dim $\mathcal{H} = \text{codim } r^2\mathcal{P}_{l-2} = 2l+1$, $\Delta \mathcal{P}_l = \mathcal{P}_{l-2}$, and $\mathcal{P}_l = r^2\mathcal{P}_{l-2} \oplus \mathcal{H}_l$ as claimed.

Applying this fact inductively, we conclude that $\mathcal{P}_l$ decomposes into the direct sum of the eigenspaces $r^{2k}\mathcal{H}_{l-2k}$ of $\hat{L}^2$ with the eigenvalues $\hbar^2(l - 2k)(l+1 - 2k)$:

$$\mathcal{P}_l = \mathcal{H}_l \oplus r^2\mathcal{H}_{l-2} \oplus r^4\mathcal{H}_{l-4} \oplus \cdots, \text{ dim } \mathcal{H}_{l-2k} = 2l + 1 - 4k.$$ 

In other words, a degree $l$ polynomial $p$ can be uniquely written as $p = p_l + r^2p_{l-2} + r^4p_{l-4} + \cdots$, where all $p_{l-2k} \in \mathcal{H}_{l-2k}$ are harmonic.

For example ($l = 2$) subtracting from $ax^2 + by^2 + cz^2 + dxy + eyz + fzx$ the multiple $\text{const}(x^2 + y^2 + z^2)$ of $r^2$ with $\text{const} = (a + b + c)/3$, we obtain a harmonic polynomial of degree 2.

Returning to our problem about functions on the unit sphere, let us call the restriction $p|_S$ of a polynomial $p(x,y,z)$ to $S = \{(x,y,z) \mid x^2 + y^2 + z^2 = 1\}$ a spherical polynomial. Since $p|_S = r^2p|_S$, we find that a spherical polynomial can be uniquely written as the sum of spherical harmonics:

$$p|_S = (p_0 + p_1 + p_2 + \cdots + p_l + \cdots)|_S, \quad \hat{L}^2 p|_S = \hbar^2 \sum_{l \geq 0} l(l+1)p_l|_S.$$

For each spherical polynomial, the sum is finite. For instance, the total dimension of the eigenspaces of $\hat{L}^2$ with the eigenvalues $< \hbar^2 n^2$ (i.e. with $l < n$) equals $1 + 3 + \cdots + 2n - 1 = n^2$. However, convergent infinite sums of spherical harmonics can be used to represent more general functions on the sphere in a manner similar to representing periodic functions by Fourier series.
In fact the problem we have solve belongs to the branch of mathematics known as harmonic analysis, or group representation theory. Namely, the group $SO_3$ of rotations acts on the sphere $S$ and therefore on the space $\mathcal{H}$ of (say, polynomial) functions on the sphere. The problem of harmonic analysis is to find linear subspaces invariant under this action, and moreover, decompose the space $\mathcal{H}$ into the direct sum of such invariant subspaces which cannot be split further into smaller invariant subspaces (they are called irreducible representations of the group). We will address these notions more thoroughly in the next section, where all irreducible representations of $SO_3$ will be classified up to equivalence.

Here is the connection with the eigenvalue problem for $\hat{L}^2$. It is built from the operators $\hat{L}_x, \hat{L}_y, \hat{L}_z$ which actually are infinitesimal rotations and generate the action of $SO_3$. So $\hat{L}^2$ acts in any representation space of $SO_3$. But $\hat{L}^2$ is rotation-invariant, i.e. commutes with the action of $SO_3$. This implies that in each irreducible space, it must act by a scalar. This is the consequence of the so-called Schur’s lemma: Any non-trivial eigenspace of $\hat{L}^2$ would be $SO_3$-invariant in conflict with assumed irreducibility. Thus, finding the eigenspaces of $\hat{L}^2$ (which are $SO_3$ invariant) helps decomposing $\mathcal{H}$ into irreducible representations, and knowing the eigenvalues helps to identify their equivalence types. In the next chapter, we will show that the spaces $\mathcal{H}_l$ are irreducible and realize all possible irreducible representations of $SO_3$. Thus, our results about spherical harmonics actually mean that $\mathcal{H} = \bigoplus_{l\geq 0} \mathcal{H}_l$ contains each type of irreducible representations of $SO_3$ exactly once.

**The quantum Kepler problem.** Having found the eigenvalues $\mu = l(l+1)$ of the total orbital angular momentum $\hbar^2 \hat{L}^2$, and their multiplicities $\text{dim } \mathcal{H}_l = 2l + 1$, we return now to studying the eigenfunctions $\psi$ of the Schrödinger equation with the Kepler potential $V(r) = -Ze^2/4\pi\epsilon_0 r$ in the form $\psi = \phi u(r)/r$ where $\phi$ is any degree $l$ spherical harmonic (constant in the radial direction), and $u$ satisfies the equation

$$-\frac{\hbar^2}{2m} \frac{d^2 u}{dr^2} + \left[ \frac{l(l+1)\hbar^2}{2mr^2} - \frac{Ze^2}{4\pi\epsilon_0 r} \right] u = Eu.$$
there are infinitely many for each $l$, but surprisingly, the energy levels do not depend on $l$, and are numbered on the rightmost graph uniformly by $E_1, E_2, \ldots$. When several independent eigenfunctions have the same eigenvalue, physicists call it {degenerate}. This degeneracy actually agrees with the property of classical Keplerian orbits to have periods of revolution and the major axes of the ellipses depend on the energy only. The value of the angular momentum of an orbit affects only the orbit’s shape — the {eccentricity}. In fact one could even “predict” the exact values of $E_n$ from the following quantization rule: Take $l = 0$, and find such $E_n$ that the {phase area} enclosed by the trajectory on the phase plane equals $2\pi \hbar n$. Let’s try this: From $p^2/2m - Q/r = E$, where $Q = Ze^2/4\pi\epsilon_0$, we find $r = Q/(-E + p^2/2m)$. So, for $E < 0$ the phase area on the $(p, r)$-plane is given by

$$\int_{-\infty}^{\infty} \frac{Qdp}{-E + p^2/2m} = \sqrt{-2mE} \int_{-\infty}^{\infty} \frac{dx}{1 + x^2} = \pi \sqrt{\frac{2m}{-E}} Q.$$ 

Equating this to $2\pi\hbar n$, we find

$$E_n = -\frac{mQ^2}{2n^2} = -\frac{mZ^2e^4}{(4\pi\epsilon_0)^22\hbar^2n^2}.$$ 

We will now make this prediction, known as {Bohr–Sommerfeld theory} come true with Schrödinger’s approach.

For this, we substitute into the radial Schrödinger equation functions $u = r^\alpha e^{-\lambda r/2}$, and drop the common exponential factor $e^{-\lambda r/2}$:

$$-\alpha(\alpha - 1)r^{\alpha - 2} + \lambda \alpha r^{\alpha - 1} + \frac{\lambda^2}{4} r^\alpha + l(l + 1)r^{\alpha - 2} - \frac{2mQ}{\hbar^2} r^{\alpha - 1} - \frac{2mE}{\hbar^2} r^\alpha.$$
We see that it is plausible that the equation has solutions \( u \) of the form \( p(r)e^{-\lambda r/2} \) with \( p \) growing at \( r \to \infty \) not faster than some power \( r^\alpha \). But for this, to have the highest order terms canceled, we must have (regardless of the value of \( l \)):

\[
E = -\frac{\lambda^2 \hbar^2}{8m} \quad \text{and then} \quad \lambda \alpha = \frac{2mQ}{\hbar^2}.
\]

Now recall that we are looking for eigenfunctions \( \psi = \phi u(r)/r \), and in order to guarantee at least convergence of \( \int \int \int |\psi|^2 \, dx \, dy \, dz \) near the origin, we must have \( \int \frac{|u|^2}{r} r^2 \, dr \) converge near \( r = 0 \). For this it is necessary that the terms \( r^\alpha \) present in \( u \) have \( \alpha > \frac{-1}{2} \).

Now let \( \alpha \) be the lowest exponent present in \( p \). Then, to have cancellations in the Schrödinger equation, regardless of the value of \( E \), we must have \( \alpha (\alpha - 1) = l(l + 1) \). This implies that \( \alpha = l + 1 \), since the other solution, \( \alpha = -l \) conflicts \( \alpha > \frac{-1}{2} \) when \( l > 0 \). In fact for \( l = 0 \) too, since the value \( \alpha = 0 \) leaves the other term of the same order, \(-2mQ/\hbar^2 r\), uncanceled.

Granted that the lowest order term in \( p \) is \( r^{l+1} \), the Schrödinger equation serves as a recursion for finding all coefficients of the polynomial

\[
p = \sum_{k=l+1}^{n} p_k r^k.
\]

Here \( n > l \) is the greatest exponent \( \alpha \) present in \( p \), which by the previous satisfies \( \lambda n = 2mQ/\hbar^2 \). The corresponding energy level \( E_n \) is thus found to be

\[
E_n = -\frac{4m^2 Q^2}{\hbar^4 n^2} \frac{\hbar^2}{8m} = -\frac{mQ^2}{2\hbar^2 n^2} = -\frac{mZ^2 e^4}{(4\pi \epsilon_0)^2 2\hbar^2 n^2}
\]

as predicted.

The polynomial \( p = p_{n,l}(r) \) is uniquely determined up to a scalar multiple from the recursion which is not hard to write down explicitly. The explicit answer is also known and can be expressed in terms of the so-called Laguerre’s polynomials. Note by the way, that \( \psi = \phi e^{-\lambda r/2} p(r)/r \) comes out differentiable at the origin, since \( \phi = h/r^l \), where \( h \) is a degree-\( l \) harmonic polynomial, while the lowest term in \( p \) is \( r^{l+1} \).

We admit that our analysis was not water-tight rigorous (e.g. one can imagine bounded growth functions \( p(r) \) more general than sums of power function \( r^\alpha \)), but the result we’ve obtained turns out to withstand higher scrutiny.
To summarize, according to this model, bound states of an electron in the field of the nucleus have energy levels $E_n$ proportional to $-1/n^2$. For each $n = 1, 2, 3, \ldots$ (known as the principal quantum number), and for each value of $l = 0, 1, \ldots, n-1$ (known as the orbital angular momentum quantum number), there is an $2l + 1$-dimensional space of eigenfunctions of the form

$$\psi = r^{-1}p_{n,l}(r)e^{-\lambda_n r/2}\phi,$$

where $\phi \in \mathcal{H}_l$ is any spherical harmonic of degree $l$. The total dimension of the eigenspace at the energy level $E_n$ is $1 + 3 + \cdots + (2n-1) = n^2$.

**Achievements and limitations of the quantum Kepler model.** In fact the accomplishment is nothing less than spectacular. Just think of it: In early 17th century, Johannes Kepler formulated three empirical laws of planetary motion. In 1687, Isaac Newton published his *Principia* where he proved that the inverse-square law of gravitation is the one responsible for Kepler’s elliptic orbits, and also laid down the foundations of mechanics. This led to great advances in industry and science, including the theory of electricity and magnetism, and the empirical classification of chemical elements. Some 250 years after Newton, in the early 20th century, that same Kepler’s problem, but solved at the quantum level, and applied to the electric interactions within the atom, was about to explain the entire periodic table and all other fundamental properties of matter.

One of them is the discreteness of the wavelengths $\lambda$ of light emitted by hydrogen (and other chemical elements), and described by the **Rydberg’s empirical formula** (1888):

$$\frac{1}{\lambda} = RZ^2 \left( \frac{1}{n_1^2} - \frac{1}{n_2^2} \right).$$

Here $Z$ is the atomic number specifying the position of the element in the periodic table, $R$ is the Rydberg constant, $1/R \approx 91.13\text{nm}$, and $n_1 < n_2$ are arbitrary integers. The explanation is that the transition of an electron from a higher energy level $E_{n_2}$ to a lower level $E_{n_1}$ is accompanied by the emission of a photon carrying the released energy

$$2\pi\hbar c/\lambda = E_{n_2} - E_{n_1} = \frac{me^4}{(4\pi\epsilon_0)^22\hbar^2Z^2} \left( \frac{1}{n_1^2} - \frac{1}{n_2^2} \right).$$

This yields the theoretical expression for Rydberg’s constant:

$$R = \frac{me^4}{8\epsilon_0^2(2\pi\hbar)^3c}.$$
This success should be taken, however, with a grain of skepticism, because single spectral lines of hydrogen under better resolution exhibit what is called the fine structure: close but multiple lines. This suggests that the eigenspaces $\mathcal{H}_l$ of the operator $\hat{L}^2$, populating, in the quantum Kepler model, the same energy level $E_n$, in reality carry slightly different energy values. As it turned out, this phenomenon is fully captured by the relativistic quantum model of the hydrogen atom, invented by Paul Dirac.

Speaking of the periodic table, one stark feature of it is that the rows (the periods) have lengths $2, 8, 8, 18, 18, 32, 32, ??$, a sequence in which one can recognize a pattern of squares $1, 4, 9, 16, 25(?)$, albeit doubled, and duplicated.

A well-known modern explanation of it starts with assuming that the elements are positioned in the table in the increasing order of the number of electrons (and hence protons) in the atom. The electrons occupy energy levels $E_1, E_2, E_3, \ldots$ (shells), labeled by $1, 2, 3, \ldots$, each consisting of subshells labeled by the letters $s, p, d, f$. In the Kepler model, they correspond to the spaces $\mathcal{H}_0, \mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3$ of spherical harmonics of which the eigenspaces at each energy level are composed:

<table>
<thead>
<tr>
<th>$n$</th>
<th>$l$</th>
<th>$\mathcal{H}_0$</th>
<th>$\mathcal{H}_1$</th>
<th>$\mathcal{H}_2$</th>
<th>$\mathcal{H}_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>$\mathcal{H}_0$</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>$\mathcal{H}_0$</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>$\mathcal{H}_0$</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$\mathcal{H}_0$</td>
<td>4</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$\mathcal{H}_0$</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$\mathcal{H}_0$</td>
<td>6</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

On the right, the table of dimensions $\dim \mathcal{H}_l$ is shown. As the atomic number increases, the subshells are filled up by electrons (apparently, up to two electrons per each unit of dimension), and not by rows but by diagonals, in the order $1s \rightarrow 2s \rightarrow 2p \rightarrow 3s \rightarrow 3p \rightarrow 4s \rightarrow 3d \rightarrow 4p \rightarrow 5s \rightarrow 4d \rightarrow 5p \rightarrow 6s \rightarrow 4f \rightarrow 5d \rightarrow 6p \rightarrow 7s \rightarrow 5f \rightarrow 6d \rightarrow 7p$.

Namely (Figure 18):

1. $1s$ in the 1st period, so we have $2\dim \mathcal{H}_0 = 2$ elements here, then
2. $2s \rightarrow 2p$ in the 2nd period, $2(\dim \mathcal{H}_0 + \dim \mathcal{H}_1) = 2(1 + 3) = 8$,
3. $3s \rightarrow 3p$ in the 3rd period, $2(1 + 3) = 8$,
4. $4s \rightarrow 3d \rightarrow 4p$ in the 4th period, $2(1 + 5 + 3) = 18$,
5. $5s \rightarrow 4d \rightarrow 5p$ in the 5th period, $2(1 + 5 + 3) = 18$,
6. $6s \rightarrow 4f \rightarrow 5d \rightarrow 6p$ in the 6th period, $2(1 + 7 + 5 + 3) = 32$,
7. $7s \rightarrow 5f \rightarrow 6d \rightarrow 7p$ in the 7th period, $2(1 + 7 + 5 + 3) = 32$. 

74
This picture demonstrates the significance of the quantum Kepler problem for the law of periodicity, but it also leaves many questions unanswered. By the way, some of them might be parallel to the issues with the classical Kepler problem.

First, in the gravitational two-body problem, the planet revolves not about the Sun, but about their common center of gravity, with the total momentum of the two bodies equal zero. Thus, in the quantum problem too, the kinetic energy of the system should be taken in the form $p^2/2m_e + p^2/2m_n$. This can be accounted for by setting $m$ to be the reduced mass $m = \frac{m_em_n}{m_e + m_n} = \frac{m_e}{1 + m_e/m_n}$,

where $m_e$ is the mass of the electron, and $m_n$ of the nucleus. In the hydrogen atom, $m_e$ is $1/1837$th of the mass $m_n$ of proton, so the change is rather small.

Next, it is one thing to solve the two-body problem, and another to apply the result to many-planet system. Interactions between the planets, however small, should be taken into account. With
many electron system, the effect will be even more dramatic because of “screening”: in naive terms, the inner electrons would shield the outer ones from the electric charge of the nucleus. This might explain why filling up the shells in a strange order could provide energy savings.

Yet, it seems inconceivable why the number of the electron states is doubled compared to our model, nor why wouldn’t all the electrons sink into the same state of minimal energy $E_1$ — given that more than one electron can be found in the same state.

The resolution of both problems came from the discovery of spin, a new degree of freedom of quantum systems not visible from the classical point of view. For an electron, its spin states form a 2-dimensional space. This changes the dimension of the eigenspace corresponding to the energy $E_n$ in the hydrogen model from $n^2$ to $2n^2$ — in agreement with the periodic table. Then it becomes clear that the electrons are fermions, and hence obey Pauli’s exclusion principle. According to this principle, the number of electrons in an atom with the same values of both: the principal quantum number $n$ and the orbital quantum number $l$, cannot exceed $2 \dim \mathcal{H}_l = 2(2l + 1)$. This explains why the electron subshells $s, p, d, f$ have respectively 2, 6, 10, 14 “seats”, and why in their minimal energy configurations, the electrons still have to populate several energy levels.

However, the theory of spin brings up another mystery: It turns out that the property of particles to be bosons or fermions is exactly determined by the parity of the number of their spin states: odd for bosons, and even for fermions.

Our next step is an introduction into the brilliant mathematics the Nature chose to employ in the phenomenon of spin.
9 Spin

Quaternions. Perhaps the best accomplishment of Sir William Hamilton was not his invention of the formalism of classical mechanics named after him, but an act of vandalism he committed on October 16, 1843 when, passing across Broome Bridge in Dublin, Ireland, he carved on its stone the formulas which crossed his mind that very moment:

\[ i^2 = j^2 = k^2 = ijk = -1. \]

Here 1, i, j, k form a basis in \( \mathbb{R}^4 \), and the formulas provide a multiplication table making the space a division algebra, \( \mathbb{H} \). It is associative but not commutative, as the formulas imply

\[ ij = k = -ji, \quad jk = i = -kj, \quad ki = j = -ik. \]

Yet, the product of a quaternion \( q = a + bi + cj + dk \) with the conjugate quaternion \( q^* = a - bi - cj - dk \) is the same real:

\[ qq^* = a^2 + b^2 + c^2 + d^2 = q^*q. \]

This shows that \( q^{-1} = q^*/\sqrt{qq^*} \) is well-defined whenever \( q \neq 0 \).

Another point of view on \( \mathbb{H} \) comes from identifying it with \( \mathbb{C}^2 \):

\[ q = a + bi + cj + dk = (a + bi) + (c + di)j = z + wj, \text{ where } z, w \in \mathbb{C}. \]

Then \( q'q = (z' + w'j)(z + wj) = (z'z - w'w^*) + (z'w + w'z^*)j, \) where \( z^*, w^* \) are complex conjugates of \( z \) and \( w \). Taking \( q' = q^* = z^* -wj \), we find \( q^*q = |z|^2 + |w|^2 = a^2 + b^2 + c^2 + d^2 \) once again.

One more, “object-oriented” point of view, associates to a quaternion \( q = z + wj \) the operator of multiplication \( \mathbb{C}^2 \to \mathbb{C}^2 : x + yj \mapsto (x + yj)q \). This leads to the matrix description:

\[
\begin{bmatrix}
x \\
y
\end{bmatrix} \mapsto \begin{bmatrix} z & -w^* \\
w & z^*
\end{bmatrix}
\begin{bmatrix}
x \\
y
\end{bmatrix}.
\]

Thus \( \mathbb{H} \) consists of \( 2 \times 2 \)-complex matrices of the above form, and the product \( q_1q_2 \) of two quaternions is expressed as the matrix product (though in the opposite order, according to our notation). The basis \( 1, i, j, k \) is represented by \( I, i\sigma_z, i\sigma_y, i\sigma_x \), where

\[
\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}
\]

are the Pauli matrices.
Groups $Sp_1$, $SU_2$, and $SO_3$. Defining the norm of a quaternion $q$ by $\|q\| = \sqrt{qq^*}$, we find from $\|q_1q_2\|^2 = q_1q_2q_2^*q_1^* = q_1\|q_2\|^2q_1^* = \|q_1\|^2\|q_2\|^2$ that the norm is multiplicative. In particular, the product of quaternions of unit norm has unit norm again. This shows that the unit 3-sphere

$$S^3 = \{a + bi + cj + dk \mid a^2 + b^2 + c^2 + d^2 = 1\} \subset \mathbb{H}$$

is a group with respect to the product of quaternions. This group is denoted $Sp_1$ (the name is the compact symplectic group of the 1-dimensional quaternionic space, but names don’t matter). It is the quaternionic analog of the unitary group $U_1 = \{e^{i\theta}\}$ of unit complex numbers, and of the orthogonal group $O_1 = \{\pm 1\}$ of unit real numbers. Using our “object-oriented” notation, we can identify $Sp_1$ with the group of matrices

$$\begin{bmatrix} z & -w^* \\ w & z^* \end{bmatrix}$$

such that $|z|^2 + |w|^2 = 1$.

Note that the expression $|z|^2 + |w|^2$ here has triple meaning: it is the Hermitian inner square of the 1st column, of the 2nd column, and also the determinant of the matrix. Besides, the Hermitian inner product of the 1st and 2nd column vanishes, i.e. the matrix is unitary, with determinant 1. By definition such matrices form the special unitary group $SU_2$ (where “special” stands for the det = 1 property, and the subscript for the matrix’ size).

In fact Hamilton introduced quaternions in order to better describe the group of special orthogonal transformations of the Euclidean 3-space, i.e. the group $SO_3$ of rotations of $\mathbb{R}^3$. Here is the connection.

To a (unit) quaternion $q$, associate the transformation $x \mapsto qxq^{-1}$ of conjugation by it on the space of all quaternions $x \in \mathbb{H}$. Since $1 \in \mathbb{H}$ commutes with $q$, the real axis remains invariant under such conjugation. Therefore its orthogonal complement $\mathbb{R}^3 = \{bi + cj + dk\}$ of purely imaginary quaternions stays also invariant under this transformation. Indeed, the conjugation preserves the sphere $S^3$, and hence Euclidean lengths of all 4-vectors — and hence all angles between the vectors (by the SSS-criterion of congruence of triangles, if you wish) — and hence the property of two vectors to be orthogonal to each other. Therefore conjugation by $q$ acts on this $\mathbb{R}^3$, and still preserves distances and hence angles there. Thus, we obtain a group homomorphism from $SU_2$ to $O_3$, the group of orthogonal transformations of the Euclidean 3-space. In fact $O_3$ consists of 2 connected
components: $SO_3$ formed by transformations with det (real $3 \times 3$-determinant) = $+1$, and the rest with det = $-1$. Since $SU_2$ is geometrically the sphere, it is connected, and so the homomorphism lands where the unit element lands: in $SO_3$. That is, the conjugation of purely imaginary quaternions by a unit quaternion $q$ defines a rotation $R_q$. Both groups are 3-parametric, and with a bit of multivariable calculus one could conclude that the map is onto. However, it is not a bijection. Namely, the conjugation by $\{\pm I\}$ (they form the center of $SU_2$, which by definition consists of the elements commuting with all elements of the group) acts trivially. Thus, we have a 2-to-1 homomorphism $q \mapsto R_q$:

$$(bi + cj + dk) \mapsto q(bi + cj + dk)q^{-1} : SU_2 \rightarrow SO_3 = SU_2/(\pm I).$$

The last equality means that $SO_3$ is obtained from $SU_2$ by identifying each $q$ with $-q$, the diametrically opposite point on $S^3$.

**The spin-statistics theorem.** The idea that “electrons have spin” emerged in the middle of 1925 in a discussion between two graduate students, Samuel Goudsmit and George Uhlenbeck, and was published on the insistence and authority of their adviser Paul Ehrenfest. The last step seemed as bold as the idea itself: as it soon transpired, a year earlier Wolfgang Pauli discouraged Ralph Kronig from pursuing a similar proposal. Indeed, physicists thought of an electron as a point-like object, but the authors of the spin concept envisioned a small electrically charged ball spinning in space, and creating thereby a tiny magnet. The magnet was to have a 2-dimensional space of quantum states, responsible for the duplication of quantum states in the hydrogen atom model.

While the picture of the spinning electric ball looks too naive, it turns out that the phenomenon does exist, and moreover, uncovers a fundamental property of the quantum mechanical Universe.

The “fabric of space” is isotropic, i.e. invariant under rotation of space about any point. Of course, the $SO_3$-symmetry of an atom can be broken by exposing it to, say, a magnetic field directed along the $z$-axis. But the behavior of the system won’t change if both the atom and the magnetic field are rotated simultaneously.

It turns out that due to this universal isotropy property, the space of quantum states of any physical system carries the action of the rotation group $SO_3$. In other words, the Hilbert space $\mathcal{H}$ of any system is not simply an Hermitian vector space, but carries additional $SO_3$-symmetry.
However, this does not necessarily mean the $\mathcal{H}$ carries an action of $SO_3$ by unitary transformations. We should not forget that complex-proportional vectors $\psi \in \mathcal{H}$ represent the same state of the quantum system. Consequently, symmetries of the system must be expressed by unitary transformations of $\mathcal{H}$, but two transformations, which differ only by a scalar factor $e^{i\phi}$ define the same symmetry of the system. Thus, the space $\mathcal{H}$ could carry a unitary action of a bigger group $G$, containing the group $U_1 = \{e^{i\phi}\}$ of complex scalars, and such that $SO_3 \cong G/U_1$ is obtained as the quotient group, i.e. by identifying those elements of $G$ which differ only by factors from $U_1$. 

In fact there are two candidates for the group $G$. One is the product $U_1 \times SO_3$, in which case $\mathcal{H}$ carries the action of $SO_3$ by unitary transformations. The other is $U_2$, the group of unitary $2 \times 2$-matrices. It contains $U_1$ as the subgroup of scalar $2 \times 2$-matrices. The latter intersects the subgroup $SU_2$ at two points, $\pm I$. So, the quotient $U_2/U_1 \cong SU_2/(\pm I) \cong SO_3$. This 2nd case results in $SU_2$ acting on $\mathcal{H}$ by unitary transformations. It actually contains the 1st case, when $-I$ happens to act as multiplication by $+1$. The verdict is: the isotropy of space implies that the Hilbert space $\mathcal{H}$ of any physical quantum system must carry a unitary action of $SU_2$.

This does not suggest that we should redo all our previous work, because one of the possibilities is that the action is trivial: all elements of $SU_2$ act as multiplications by 1. But it could be non-trivial. One immediate consequence of this is that generally speaking the space $\mathcal{H}$ splits into the direct sum $\mathcal{H} = \mathcal{H}_+ \oplus \mathcal{H}_-$ according to the eigenvalues $\pm 1$ of the central element $-I \in SU_2$: it acts by $+1$ on $\mathcal{H}_+$ and by $-1$ on $\mathcal{H}_-$. All meaningful quantum observables, i.e. Hermitian operators $A: \mathcal{H} \to \mathcal{H}$, should respect this symmetry, i.e. transform $\mathcal{H}_\pm$ separately. Physically this means that there are two unrelated quantum systems, and so in real systems either $\mathcal{H} = \mathcal{H}_+$ or $\mathcal{H} = \mathcal{H}_-$.

We have already encountered a similar situation when we discussed bosons and fermions, and mentioned super-geometry. A general vector super-space consists of two parts: $\mathcal{V} = \mathcal{V}_0 \oplus \mathcal{V}_1$, called even and odd. The coordinates, if multiplied, behave differently, namely

$$ab = (-1)^{\bar{a}\bar{b}}ba,$$

where $\bar{a}, \bar{b}$ are parities of $a$ and $b$, equal to $0$ mod 2 for coordinates on $\mathcal{V}_0$ and $1$ mod 2 for coordinates on $\mathcal{V}_1$. On a super-space, the parity operator $P$ is defined by $P(\mathcal{V}) = (-1)^{\bar{a}}$. That is, $\mathcal{V}_d, d = 0, 1,$ are the eigenspaces of $P$ with the eigenvalues $(-1)^d$. The difference between
bosons and fermions can be explained by saying that the Hilbert space for a boson is purely even ($V_1 = 0$), and for a fermion purely odd ($V_0 = 0$). This distinction manifests in describing the states of indistinguishable particles. What the symmetric tensor power $S^N(H)$ for the even space $H$ is, turns out to be anti-symmetric, i.e. $\Lambda^N(H)$, when $H$ is odd. Due to the properties of permutation groups, these were the only two logical possibilities for multi-particle systems, but it leaves out the question whether both possibilities occur in the nature.

It turns out that the answer is yes, and the action of the intrinsic symmetry $-I \in SU_2$ on the Hilbert space $H$ of the system is the parity operator $P$. That is, every quantum system is either bosonic or fermionic, and it is bosonic when $H$ carries the action of $SO_3$, and fermionic when the action of $SU_2$ descends to that of $SO_3$ only when one stops distinguishing each $\psi$ from $-\psi$ as quantum states. This principle is the content of the so-called “spin-statistics theorem”.

The word “theorem” comes from theoretical attempts to explain the connection between the action of $SU_2$ on $H$ ("spin") and permutations on $H \otimes N$ ("statistics"). An elementary explanation seems to be lacking. It is logically possible to consider fermionic systems where $-I$ acts by $+1$ (as we did by considering no $SO_3$ action), or the other way around. A theoretical explanation would have to show that accepting such a possibility would eventually lead to absurd physical conclusions. For instance, in quantum field theory, one encounters undue negative energies if particles with $-I$ acting as $-1$ are treated as bosons. Or, under some hypotheses of relativistic quantum mechanics, an expected symmetry is broken unless the spin-statistics connection holds true. But in any case, the empirical fact is that $P = -I \in SU_2$ is the universal parity operator of the quantum world.

Thus, the Hilbert space $H$ of any real quantum system carries an action of $SU_2$, and is either bosonic or fermionic. Joining two fermionic systems, we obtain a bosonic one, since the action of $-I$ in $H' \otimes H''$ is given by $(-1) \times (-1) = 1$. For instance, protons and electrons are fermions, and so the hydrogen atom as a whole is a boson.

Returning to an electron, we should realize that the Hilbert space $H$ we studied in the hydrogen atom problem does carry the action of $SO_3$ due to the rotational symmetry of the potential energy. But this action has nothing to do with the action of $SU_2$ which should reflect the isotropy of the intrinsic world of the electron, from its own point of view, so to speak, which does not depend on whether
the external force field is applied. Perhaps the simplest action of $SU_2$ with $-I$ acting by $-1$ is the natural action in $\mathbb{C}^2$ by $2 \times 2$-matrices. Therefore the tensor product $\mathcal{H} \otimes \mathbb{C}^2$ is the candidate for the phase space in the hydrogen atom model taking into account the fermionic nature of the electron. Elements of $\mathcal{H} \otimes \mathbb{C}^2$ can be thought of as vectors from $\mathcal{H}$ “with components in $\mathbb{C}^2$”, i.e. in our case as psi-functions $\psi(q_1, q_2, q_3)$ with values not in $\mathbb{C}$ but in $\mathbb{C}^2$ (quaternion-valued functions, if you wish). The hamiltonian $\hat{H}$ doesn’t know anything about the 2-dimensional degree of freedom (i.e. commutes with the $SU_2$-action, and acts the same way on both components of the vector-function $\psi$). Thus, the eigenspaces remain the same, only doubled in dimension. The same applies to the orbital angular momentum operator $\hat{L}$, since the rotations of space acting in $\mathcal{H} \otimes \mathbb{C}^2$ via $\mathcal{H}$ commute with the action of $SU_2$ via $\mathbb{C}^2$. This resolves both difficulties in the explanation of the periodic table: being fermions, the electrons cannot populate the same state twice, and we now have two such states per each eigenfunction of the quantum Kepler model.

In this explanation, the action of $SU_2$ on $\mathbb{C}^2$ is used. This is the simplest non-trivial irreducible representation of $SU_2$, but not the only one, and the full-fledged theory of spin depends on their classification provided by the theory of representations.

**Representation theory of $SU_2$.** By a (complex) representation of a given abstract group $G$ one means an action of (elements of) $G$ by linear transformations on a complex vector space $V$. Two such representations, say $V$ and $W$, of the same group $G$ are called equivalent if one can identify the spaces by a linear isomorphism, $V \cong W$, so that the actions become the same. Our current goal is to describe all possible representations of $SU_2$ up to equivalence.

It is a general fact of the theory of representations of compact groups (and $SU_2$, being a closed bounded subset $S^3$ in $\mathbb{R}^4$, is compact) that if a representation space $V$ has a non-trivial (i.e. $\neq 0$ and $\neq V$) $G$-invariant subspace $W \subset V$, then it has a $G$-invariant complementary subspace, $W'$, so that $V$ is equivalent to the direct sum $W \oplus W'$ of smaller representation spaces. Yet another general fact about compact groups is that any irreducible representation of $G$, i.e. representation which has no non-trivial invariant subspaces, is finite dimensional (a property the reader probably assumed anyway). Thus, our problem consists in describing all irreducible representations of $SU_2$ up to equivalence, understanding that any other finite dimensional representation is obtained by taking direct sums of the irreducible representations as building blocks.
The answer to our problem is this. For every \( d = 0, 1, 2 \ldots \) there exists exactly one, up to equivalence, irreducible representation of \( SU_2 \) of complex dimension \( d + 1 \). Physicists’ notations for these representations are \( V_0, V_{1/2}, V_1, V_{3/2}, \ldots \), where \( V_l \) has dimension \( 2l + 1 = 1, 2, 3, \ldots \).

It is said that a quantum particle has spin \( l \), which can assume half-integers values \( l = 0, 1/2, 1, 3/2, 2, \ldots \), if the Hilbert space of its quantum states has, as a representation of \( SU_2 \), the form \( \mathcal{H} \otimes V_l \) (where \( \mathcal{H} \) is some Hilbert space). In the sense in which \( \mathcal{H} \) can be represented as the direct sum of 1-dimensional subspaces, \( L_\alpha \), the representation \( \mathcal{H} \otimes V_l \) can be represented as the direct sum of copies of \( V_l \cong L_\alpha \otimes V_l \). In any case, any irreducible subspace of \( \mathcal{H} \otimes V_l \) carries a representation equivalent to \( V_l \) — no other types of representations of \( SU_2 \) can occur there.

In fact \( -I \) acts in \( V_l \) as \( e^{2\pi il} \), i.e. \(-1\) for half-integer values of \( l \) and \(+1\) for integer ones. Therefore, by the spin-statistics theorem, particles of integer spins are bosons, and of half-integer spins are fermions. (E.g. proton and electron have spin 1/2, and photon has spin 1).

For a general quantum system, its phase space \( \mathcal{H} \) carrying a representation of \( SU_2 \) can be decomposed into irreducible representation of different types, possibly each occurring multiple times. This can be expressed by writing

\[
\mathcal{H} = \bigoplus_l \mathcal{H}_l \otimes V_l,
\]

where \( \mathcal{H}_l \) are some Hilbert spaces (“multiplicity spaces”). The sum can be infinite, with the summands known as isotypical components, meaning that each contains only one kind of representations, \( V_l \). This is an expression of the general statement, quoted earlier, about representation of compact groups. However, in quantum physics, the decomposition of \( \mathcal{H} \) involves either only half-integer \( l = 1/2, 3/2, \ldots \) (for fermions) or only integer \( l = 0, 1, 2 \ldots \) (for bosons). That is, a general quantum system does not have to have a definite value of spin, but it has parity.

**Irreducible representations.** The model representations \( V_l \) are easy to describe. Namely, \( SU_2 \) comes with the natural action on the complex plane \( \mathbb{C}^2 \) with coordinates \((x, y)\) by the above \( 2 \times 2 \) matrices. For each \( d \), denote by \( V_{d/2} \) the space of homogeneous degree-\( d \) polynomials in \((x, y)\):

\[
a_0 x^d y^0 + a_1 x^{d-1} y^1 + \cdots + a_{d-1} x^1 y^{d-1} + a_d x^0 y^d, \quad a_0, \ldots, a_d \in \mathbb{C}.
\]
Clearly $SU_2$ acts on $V_{d/2}$ by the linear changes of the variables $(x,y)$. So, the claim is that: (a) each $V_{d/2}$ is irreducible, and (b) every irreducible representation of $SU_2$ is equivalent to one of these. In what follows we outline the proof in physicists’ notation.

Let us think of $\mathbb{C}^2$ as (albeit, 2-dimensional) Hermitian space of quantum states $\psi = (x,y)$, with $\langle \psi | \psi' \rangle := x^*x' + y*y'$ as the Hermitian inner product. Note that the Pauli matrices are Hermitian. So, we introduce quantum observables, the components of the spin (akin to quantum angular momentum) vector $\hat{S}$:

$$S_x = \frac{\hbar}{2} \sigma_x, \quad S_y = \frac{\hbar}{2} \sigma_y, \quad S_z = \frac{\hbar}{2} \sigma_z.$$  

Their commutation relations $[S_x, S_y] = i\hbar S_z$ etc. come, of course from those for the quaternions $i,j,k$ (e.g. $ij - ji = 2k$). Let us consider the respective “Schrödinger equations”, starting with $S_z$: $i\hbar \frac{d\psi}{dt} = S_z \psi$. The evolution is described by the one parametric group of unitary transformations

$$U_z(t) \psi = e^{-it\sigma_z/2} \psi = \left[ \begin{array}{cc} e^{-it/2} & 0 \\ 0 & e^{it/2} \end{array} \right] \left[ \begin{array}{c} x \\ y \end{array} \right],$$

which actually lie in $SU_2$. We could say that $S_z$ is an infinitesimal $SU_2$-transformation, as it can be recovered from $U_z(t)$ by differentiation:

$$i\hbar \left. \frac{dU_z(t)}{dt} \right|_{t=0} = S_z.$$

Similarly, one can define and compute the 1-parametric subgroups $U_x, U_y$ in $SU_2$ generated by $S_x$ and $S_y$. It is clear that the three generate the whole action of $SU_2$ (in terms of $SO_3$ they correspond to the rotations of the 3-space about the $x,y,z$-axes - which happen to currently have names $d,c,b$). But it seems easier to deal with $S_z$ since in our coordinates the matrix is diagonal.

Now we compute the infinitesimal action of $S_z$ on $p \in V_{d/2}$:

$$S_z p = \left. i\hbar \frac{d}{dt} \right|_{t=0} p(e^{-it/2}x, e^{it/2}y) = \frac{\hbar}{2} \left( y \frac{\partial}{\partial y} - x \frac{\partial}{\partial x} \right) p(x,y).$$

Clearly, the monomials $x^k y^{d-k}$ are the eigenvectors of $S_z$ with the eigenvalues $\hbar(d/2 - k)$. That is, the spectrum of $S_z$ is an arithmetic sequence of half-integer multiples of $\hbar$ ranging between $-\hbar d/2$ and $+\hbar d/2$ with step $\hbar$. We could similarly describe the action of $S_x$ and

84
which would have the same spectrum but in different bases of eigenvectors. Yet, it is more useful to introduce the ladder operators similar to the creation/annihilation operators in the harmonic oscillator problem:

\[
J_+ := S_x + iS_y = \hbar \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad J_- := S_x - iS_y = \hbar \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.
\]

They are not Hermitian, but still generate 1-parametric subgroups \(G_\pm(t)\) of linear transformations which solve the differential equations

\[
i\hbar \frac{d\psi}{dt} = J_\pm \psi.
\]

The solutions are very simple:

\[
G_+(t) = e^{-iJ_+ t} : (x, y) \mapsto (x - ity, y),
\]

\[
G_-(t) = e^{-iJ_- t} : (x, y) \mapsto (x, y - itx).
\]

Respectively, on the degree \(d\) polynomials \(p\), \(J_\pm\) act this way:

\[
i\hbar \frac{d}{dt} \bigg|_{t=0} p(x - ity, y) = \hbar y \frac{\partial}{\partial x} p, \quad i\hbar \frac{d}{dt} \bigg|_{t=0} p(x, y - itx) = \hbar x \frac{\partial}{\partial y} p.
\]

By the way, it is easy to find from this the action of \(S_x = (J_+ + J_-)/2\) and \(S_y = (J_+ - J_-)/2i\), but we won’t need it. We see that \(J_+\) raises, and \(J_-\) lowers the degree of \(y\) in the monomials \(x^ky^{d-k}\). Thus, the action of \(J_\pm\) and \(S_z\) on \(V_{d/2}\) can be described by the ladder diagram (Figure 19), where the integers over the vertices show the eigenvalues of \(S_z\) in units of \(\hbar/2\). It is no accident that the eigenvalues’ increment

\[
\begin{align*}
S_x &= (J_+ + J_-)/2 \\
S_y &= (J_+ - J_-)/2i
\end{align*}
\]

is \(2(\hbar/2)\). Indeed, the following commutation relation holds:

\[
[S_z, J_\pm] = [S_z, S_x \pm iS_y] = i\hbar(S_y \mp iS_x) = \pm\hbar(S_x \pm iS_y) = \pm\hbar J_\pm.
\]

For an eigenvector \(v\) of \(S_z\) with the eigenvalue \(\lambda\hbar/2\), they imply that \(J_\pm v\) (unless \(= 0\)), is an eigenvector with the eigenvalue \((\lambda \pm 2)\hbar/2\):

\[
S_zJ_\pm v = J_\pm S_z v \pm \hbar J_\pm v = \frac{\hbar}{2}(\lambda \pm 2)J_\pm v.
\]
Now we apply this information to an arbitrary finite dimensional irreducible representation \( V \) of \( SU_2 \). Let \( v_0 \) be the eigenvector of \( S_z \) with the greatest eigenvalue \( \lambda_0 \hbar/2 \). Then \( J^+ v_0 \) must be 0, and \( v_k := J^- v_0 \) must be eigenvectors of \( S_z \) with the eigenvalues \((\lambda_0 - 2k)\hbar/2\). Besides, from the commutation relation

\[
[J^+, J^-] = [S_x + iS_y, S_x - iS_y] = 2\hbar S_z,
\]

one can deduce by induction on \( k \) that

\[
J^+ v_{k+1} = \hbar^2 (k + 1)(\lambda_0 - k)v_k.
\]

But \( V \) won’t be finite dimensional unless some \( v_{d+1} = 0 \). Thus, it is necessary that \( \lambda_0 = d \) for some integer \( d \geq 0 \). This possibility leads to the basis \( \{v_k, k = 0, 1, \ldots, d\} \) spanning an invariant subspace \( W \) with the action of \( S_z, J^\pm \) exactly as on the ladder diagram of Figure 19. Thus, \( W \) is equivalent to \( V_{d/2} \) as a representation of \( SU_2 \). Since we assumed \( V \) irreducible, \( W \) must be the whole of \( V \), as claimed.

This result also provides a complete description of irreducible representations of \( SO_3 \). Namely, every such representation space \( V \) is also a representation of \( SU_2 \) where, however, \( -I \) acts trivially. In the spaces of polynomials of degree \( d \) this happens only when \( d \) is even. Thus, \( SO_3 \) has exactly one, up to equivalence, irreducible complex representation in each \( \text{odd dimension } 2l+1 \), namely \( V_l \) for \( l = 1, 2, 3, \ldots \). Such representations are sometimes referred to as \textit{vector representations} of \( SU_2 \), while the remaining \( V_l \), with the half-integer values of \( l \), as \textit{spinor representations}.

\textbf{Some examples.} \( l = 0 \): The space \( V_0 \) is 1-dimensional and carries the \textit{trivial representation} of \( SU_2 \): all elements of the group act as the multiplication by 1.

\( l = 1/2 \): By our model, \( V_{1/2} \) is the space of degree-1 polynomials on the \((x,y)\)-plane \( \mathbb{C}^2 \), i.e. the \textit{dual space} \((\mathbb{C}^2)^*\). One may ask how \( \mathbb{C}^2 \) itself fits the classification. Well, since it is irreducible, it must be equivalent to \( V_{1/2} \). Actually defining an isomorphism \( A : V \to V^* \) between a vector space and its dual is the same as defining a non-degenerate bilinear form \( B \) on \( V \): \( B(u, v) := A(u)(v) \) (= the value of the linear form \( A(u) \in V^* \) on the vector \( v \in V \)). Thus, \( \mathbb{C}^2 \) must carry such a form, and in order to provide an equivalence of representations, it must be \( SU_2 \)-invariant. And indeed, the letter “S” in the notation of \( SU_2 \) stands for \( \det = 1 \), meaning that the operators from the group preserve “complex area”. It happens to be an \textit{anti-symmetric} bilinear form on \( \mathbb{C}^2 \): \( B((x,y),(x',y')) = xy' - yx' \).
$l = 1$: It must be an irreducible representation of $SO_3$ of dimension 3. We know one, namely the natural action of $SO_3$ by rotations in the Euclidean 3-space. It is $\mathbb{R}^3$, but no one can prohibit us to apply real rotation matrices to complex 3-vectors (officially this trick is called complexification), and thus obtain an $SO_3$-representation in $\mathbb{C}^3$. It must be equivalent to $V_1$, realized as the space of quadratic polynomials $ax^2 + bxy + cy^2$ on $\mathbb{C}^2$. By the above argument, $V_1^*$ must also be equivalent to $V_1$, so there has to exist a non-degenerate $SO_3$-invariant bilinear form on $V_1$. And indeed, in our Euclidean model, it is obvious that the ordinary dot-product $B(u_1, u_2, u_3), (v_1, v_2, v_3)) = u_1 v_1 + u_2 v_2 + u_3 v_3$ fits the description.

By the way, this bilinear form is symmetric $B(u, v) = B(v, u)$. This is not an accident, and in fact each $V_l$ is equivalent to its dual, $V_l^*$, and for integer $l$ carries a non-degenerate $SO_3$-invariant bilinear form, which is symmetric, while for half-integer $l$, the requisite $SU_2$-invariant form is anti-symmetric: $B(u, v) = -B(v, u)$.

Actually bilinear forms themselves form a vector space. For example, all bilinear forms on $\mathbb{C}^2$ have the form: $B((x, y), (x', y')) = \alpha xx' + \beta xy' + \gamma yx' + \delta yy'$, and form a 4-dimensional space. Matrices from $SU_2$ act on it by simultaneously changing the coordinates $(x, y)$ and $(x', y')$ of both arguments. This is an instance of the tensor product of two representations, $V_{1/2} \otimes V_{1/2}$ in this case. So, how does it fit in our classification? It cannot be $V_{3/2}$ which also has dimension 4 but it is spinor, while $V_{1/2} \otimes V_{1/2}$ is vector (since $(-1) \otimes (-1)$ is still +1). The answer comes easily once we examine the spectrum of $S_z$. The eigenvectors are the monomial bilinear forms $xx', xy', yx', yy'$, and if $x$ and $y$ have eigenvalues $-\hbar/2$ and $\hbar/2$ in $V_{1/2}$, then spectrum of $S_z$ in $V_{1/2} \otimes V_{1/2}$ consists of the pairwise sums:

$$(-\hbar/2 - \hbar/2, -\hbar/2 + \hbar/2, \hbar/2 - \hbar/2, \hbar/2 + \hbar/2) = (-\hbar, 0, 0, \hbar).$$

Indeed, remember that $S_z$ is obtained by differentiation $-i\hbar dU/dt$ from the one-parametric family of unitary transformations $U(t) = e^{itS_z/\hbar}$. Therefore, by Leibniz’ rule

$$(S_z B)(u,v) = -i\hbar \frac{d}{dt} \bigg|_{t=0} B(U(t)u, U(t)v) = B(S_z u, v) + B(u, S_z v).$$

Recall now that in each $V_{d/2}$, the spectrum of $S_z$ is the arithmetic sequence $(-\hbar d/2, -\hbar d/2 + \hbar, \ldots, \hbar d/2 - \h\hbar d/2)$. There is only one way to split $(-\hbar, 0, 0, \hbar)$ into a union of such sequences: $(0)$ and $(-\hbar, 0, \hbar)$. This means that $V_{1/2} \otimes V_{1/2} \cong V_0 \oplus V_1$. 87
The Clebsch–Gordan formula. Here is an afterthought. A bilinear form can be uniquely written as the sum of symmetric and anti-symmetric ones:

$$B(u,v) = \frac{1}{2}[(B(u,v) + B(v,u)] + \frac{1}{2}[B(u,v) - B(v,u)].$$

We know one $SU_2$-invariant anti-symmetric form on $\mathbb{C}^2$, the determinant. And we know that symmetric bilinear forms (which can be matched with quadratic polynomials $B(u,u)$) form $V_1$. Thus we should’ve known beforehand that $V_{1/2} \otimes V_{1/2} \cong V_0 \oplus V_1$.

However, the method we applied first is more general. Consider the tensor product of two irreducible representations, and decompose it into the sum of irreducible ones. To avoid abstract definitions, let us think of the space of all bilinear forms $B(v,w)$ where $v \in V$ and $w \in W$. This is a vector space of dimension equal to the product of $\dim V$ and $\dim W$. The products $v_iw_j$ of coordinates in $V$ and $W$ form a basis in it. The official notation for this space $V^* \otimes W^*$, the tensor product of the dual spaces. But in our situation, when each $V_l \cong V_l^*$, we don’t need to make a distinction between $V_l \otimes V_{l'}$ and the space of bilinear functions $B : V_l \times V_{l'} \rightarrow \mathbb{C}$. Generalizing our previous argument, we see that the spectrum of $S_z$ on $V_l \otimes V_{l'}$ consists of the sums $\hbar(\lambda + \mu)$, where the half-integers $\lambda$ come from the ladder diagram for $V_l$ and $\mu$ for $V_{l'}$. The sums $\lambda + \mu$ form a rectangular table (shown for $l \geq l'$):

| $-l + l'$ | $-l + l' + 1$ | $\ldots$ | $l + l' - 1$ | $l + l'$ |
| $-l + l' - 1$ | $-l + l'$ | $\ldots$ | $\ldots$ | $l + l' - 2$ | $l + l' - 1$ |
| $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ |
| $-l - l'$ | $-l - l' + 1$ | $-l + l'$ | $\ldots$ | $l - l'$ |

The table of the eigenvalues can be partitioned in an essentially unique way into the union of arithmetic sequences shown in the table as “hooks” in black, blue, $\ldots$, red colors. They represent the spectra of $S_z$ in $V_{l+l'}$ (black), $V_{l+l'-1}$ (blue), $\ldots$, $V_{l-l'}$ (red). Thus

$$V_l \otimes V_{l'} \cong V_{l-l'} \oplus V_{l-l'+1} \oplus \cdots \oplus V_{l+l'-1} \oplus V_{l+l'}.$$

This result is known as the Clebsch-Gordan formula. It completely describes the tensor algebra of $SU_2$-representations. For instance,

$$V_l \otimes V_l = V_0 \oplus V_1 \oplus V_2 \oplus \cdots \oplus V_{2l-1} \oplus V_{2l}.$$
In a sense, Clebsch-Gordan formulas describe the rule of “addition of spins”, but shows that a system of two particles of spins \( l \geq l' \) does not have a particular value of spin, since its Hilbert space decomposes into isotypical components corresponding to spin values \( l - l', l - l' + 1, \ldots, l + l' - 1, l + l' \).

**The total spin operator.** The last example should remind us of our promise to identify the spaces \( \mathcal{H}_l \) of degree \( l \) spherical harmonics with irreducible representations of \( SO_3 \) currently denoted \( V_l \).

To this end, let us compute the eigenvalues in \( V_{d/2} \) of the total spin operator \( \hat{S}^2 := S_x^2 + S_y^2 + S_z^2 \). It is rotationally-symmetric, i.e. commutes with \( S_x, S_y, S_z \). Then, by Schur’s lemma, it must act in each irreducible representation \( V_{d/2} \) by a scalar. So, it suffices to compute the scalar on the eigenvector \( v_0 = y^d \) of \( S_z \) annihilated by \( J_+ \). We have

\[
J_- J_+ = (S_x - iS_y)(S_x + iS_y) = S_x^2 + S_y^2 + i[S_x, S_y] = S_x^2 + S_y^2 - \hbar S_z.
\]

Therefore \( \hat{S}^2 = J_- J_+ + S_z^2 + \hbar S_z \). Applying this to \( v_0 \), we find that

\[
\hat{S}^2 \text{ acts on } V_{d/2} \text{ as multiplication by } \hbar^2 \frac{d}{2} \left( \frac{d}{2} + 1 \right).
\]

We see that: (i) the values of \( \hat{S}^2 \) are different in different irreducible representations, (ii) for even \( d = 2l \), this value \( \hbar^2 l(l + 1) \) coincides with the value of \( \hat{L}^2 \) on \( \mathcal{H}_l \).

In particular, for \( V_1 \), which is the ordinary 3-dimensional representation of \( SO_3 = SU_2/(\pm I) \) by rotations, this coincidence verifies that the normalization of the operators \( \hat{L}^2 = \hat{L}_x^2 + \hat{L}_y^2 + \hat{L}_z^2 \) and \( \hat{S}^2 = S_x^2 + S_y^2 + S_z^2 \) (which \textit{a priori} could differ by a scalar factor) is actually the same. Therefore \( \hbar^2 l(l + 1) \) is the value of \( \hat{S}^2 \) on \( \mathcal{H}_l \), which (together with \( \dim \mathcal{H}_l = \dim V_l \)) identifies it as a representation of \( S\hat{O}_3 \) equivalent to \( V_l \). In particular \( \mathcal{H}_l \) are irreducible and as \( l = 0, 1, 2, \ldots \) varies, represent all equivalence classes of irreducible representations of \( SO_3 \).

**The classical limit.** Since we are talking about \( V_l \) as spaces of quantum states moved around by elements of \( SU_2 \), it is natural to ask to what classical mechanical systems they could correspond. This should also be a good exercise in undoing the wizardry of quantization and exploring the routine of passing to the classical limit. So, let’s try.
The Hermitian operators $S_x, S_y, S_z$ defining infinitesimally the action of $SU_2$ represent some quantum observables. Typically the product of two Hermitian operators, say $A$ and $B$, are not Hermitian, but the anti-commutator $(A, B) := (AB + BA)$ and $i[A, B]$ are. In the classical limit $\hbar \to 0$, the operators should turn into some classical observables, i.e. functions on some classical phase space (what could it be?) while $(\cdot, \cdot)$ and $i[\cdot, \cdot]$ should turn into some meaningful operations with the functions. It is not hard to guess that the usual product of functions should come from $(AB + BA)/2$. This is how we treated the products in the process of quantization: to the product of functions, e.g. $pq$, we tried to associate the products of the corresponding operators, i.e. $(-i\hbar \partial/\partial q)q$. Except that $pq$ and $qp$ coincide as functions, but don’t quite coincide as operators: $\hat{p}\hat{q} = \hat{q}\hat{p} - i\hbar$. However, in the limit $\hbar \to 0$, i.e. if we neglect with the terms containing $\hbar$, the difference disappears. In particular, the operation $i[\hat{p}, \hat{q}] = \hbar$ yields 0 in this limit. The important clue though is that, since it is always 0, we can divide the result by $\hbar$ and only then pass to the limit. What we will obtain this way is the Poisson bracket! Let’s try it with $p^2$ and $q^2$:

$$i[\hat{p}^2, \hat{q}^2] = -i\hbar^2 \frac{\partial^2}{\partial q^2} q^2 + i\hbar^2 q^2 \frac{\partial^2}{\partial q^2} = -2i\hbar^2 \frac{\partial}{\partial q} - 2i\hbar^2 = 2\hbar \hat{q} \hat{p} - 2i\hbar^2.$$

Thus, dividing by $\hbar$ and passing to the limit $\hbar = 0$, we obtain $2qp$ which is the Poisson bracket $\{p^2, q^2\}$. This example is not hard to upgrade into a general statement about differential operators, but our goal now is to apply it to $SU_2$.

So, introduce functions $x, y, z$ to represent classical observables corresponding to $S_z, S_y, S_z$ respectively. The latter are understood not as $2 \times 2$-matrices by which they act in $V_{1/2}$, but as quantum observables acting by Hermitian operators in all representations $V_1$. For instance, $(S_x, S_y)$ is a new quantum observable, which in the classical limit turns into $2xy$. As an operator in $V_{1/2}$, it is zero (because $ij + ji = 0$ in $\mathbb{H}$), but it is non-zero in $V_1$ (as one can check). Consequently the quantum observables represented by anti-commutators of $S_x, S_y, S_z$, by anti-commutators of their anti-commutators, etc., in the limit $\hbar = 0$, would simply give us arbitrary polynomials $F(x, y, z)$. Thus, our phase space seems to be $\mathbb{R}^3$, with all polynomials on it in the role of classical observables.

Now, the commutation relations

$$i[S_x, S_y] = -\hbar S_z, \quad i[S_y, S_z] = -\hbar S_x, \quad i[S_z, S_x] = -\hbar S_y,$$

90
are the same in all $V_l$ as they reflect the properties of the Lie group $SU_2$. More precisely, $iS_x, iS_y, iS_z$ span its Lie algebra, i.e. represent infinitesimal $SU_2$-transformations. Various products of their exponentials comprise the whole group. So, these commutation relations tell us what the Poisson brackets should be:

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

The total spin operator $\hat{S}^2$ in the limit $\hbar = 0$ becomes the Casimir function $x^2 + y^2 + z^2$, which has zero Poisson brackets with $x, y, z$. Thus, the classical limit is the familiar rotational dynamics on the spheres $x^2 + y^2 + z^2 = const$. 

10 A bit of relativity

A relativistic Hamilton equation. As the common wisdom has it, relativistic effects take place only when objects move at speeds comparable to the speed of light. There is, however, a relativistic phenomenon familiar to every child: the magnets. Indeed, according to the Galilean principle, laws of Nature would look alike in two reference systems moving with a constant velocity with respect to one another. That is, the fundamental forces are proportional to accelerations (Newton’s Second Law), but cannot depend on velocities. This principle is manifestly broken by a magnetic field, \( B \), which acts on an electric charge \( Q \) moving with the velocity \( v \) with the force proportional to \( Q(v \times B) \).

To assess the Galilean principle in Hamiltonian terms, consider a free particle of mass \( m \) in 3-space. In the extended phase space (it has dimension \( 2(3 + 1) = 8 \)) the particle is described by the hypersurface \( E = (p \cdot p)/2m \). This simple equation has many symmetries: it is translation-invariant with respect to time variable \( \tau \); it is translation-invariant with respect to space variables \( q = (q_1, q_2, q_3) \); and it is rotation-invariant with respect to rotations of space (which also rotate momentum vectors \( p = (p_1, p_2, p_3) \)), as the dot-product \( p \cdot p \) has this property by definition. In accordance with Noether’s theorem, these properties of homogeneity of time and space, and the isotropy of space are responsible for the energy, momentum, and angular momentum conservation laws.

Consider now the Galilean transformation to a reference system moving with the velocity \( v_0 \). It results in the momentum shift: \( p \mapsto p + mv_0 \), causing the above hypersurface to transform into \( E = (p \cdot p)/2m + p \cdot v_0 + m(v_0 \cdot v_0)/2 \). Therefore, if the change \( E \mapsto E + p \cdot v_0 + m(v_0 \cdot v_0)/2 \) is also included into the Galilean transformation of the extended phase space, the hypersurface is preserved. Such extended Galilean transformations together with the symmetries mentioned earlier generate the Galilean group of symmetries of space-time in Newtonian mechanics.

A relativistic free particle of rest mass \( m \) is described in the extended phase space by the hypersurface

\[
E^2 - c^2(p \cdot p) = m^2c^4.
\]

On Figure 20a, the hypersurfaces corresponding to \( m = 0 \) (the light cone) and \( m > 0 \) (the 2-sheeted hyperboloid) are shown, though in dimension \( 1 + 2 \) rather than \( 1 + 3 \), and not in space-time, but rather
in the space with the dual, energy-momenta coordinates. Note that when \(|p| \ll mc\), we get
\[
E = c\sqrt{m^2c^2 + (p \cdot p)} = mc^2 \sqrt{1 + \frac{p \cdot p}{m^2c^2}} \approx mc^2 + \frac{p \cdot p}{2m},
\]
i.e. the sum of Einstein’s famous rest energy expression and the non-relativistic kinetic energy. Similarly to the non-relativistic case, the

![Figure 20: Lorentz’ boosts in Minkowski’s space](image)

hypersurface is invariant with respect to the translations in space-time, and rotations in space (visualizable in terms of Figure 20a as rotations about the \(E\)-axis). However, the place of Galilean transformations is taken by Lorentz’ boosts, or hyperbolic rotations, given in dimension \(1 + 1\) (Figure 20b) by the formulas:

\[
\begin{bmatrix}
E \\
p
\end{bmatrix} \mapsto \frac{1}{\sqrt{1 - \frac{v_0^2}{c^2}}} \begin{bmatrix}
\frac{1}{c^2} & v_0 \\
1 & 1
\end{bmatrix} \begin{bmatrix}
E \\
p
\end{bmatrix}.
\]

Together with the other transformations, they generate the Lorentz group, the symmetry group of space-time in relativity theory.

**Klein–Gordon.** We have enough experience to immediately quantize the Hamilton equation of the free relativistic particle of the rest mass \(m\). We arrive at the Klein-Gordon equation:

\[
\frac{\partial^2}{\partial t^2} \Psi(t, q) - c^2 \Delta \Psi(t, q) = -\frac{m^2c^4}{\hbar^2} \Psi(t, q),
\]

where \(\Delta\) is the Laplace operator in \(q = (q_1, q_2, q_3)\). A solution of it is a superposition (i.e. a Fourier integral) of traveling harmonic waves:

\[
A(k)e^{i(k \cdot q - \omega t)}, \quad \text{where } \hbar^2 \omega^2 = c^2 \hbar^2(k \cdot k) + m^2c^4.
\]
That is, $\hbar \omega$ equals the relativistic energy $E = c\sqrt{m^2c^2 + (p \cdot p)}$ of the particle with the momentum $p = \hbar k$. As a parenthetical remark (or an exercise) we note that the phase velocity $\omega/|k|$ of such harmonic waves exceeds $c$, but their group velocity $d\omega/d|k|$ doesn’t.

The special case $m = 0$ of the Klein-Gordon equation describes a relativistic particle of zero rest mass. In fact this is simply the wave equation
\[
\frac{\partial^2 \Psi}{\partial t^2} = c^2 \Delta \Psi.
\]
It describes superpositions of harmonic waves $e^{i(k \cdot q - \omega t)}$ traveling with the speed $\omega/|k| = c$. Ironically, the equation, known to describe the propagation of waves long before relativity theory and quantum mechanics entered the scene, is in fact both relativistic and quantum-mechanical. Note that the Planck constant falls out of the equation completely. Nevertheless the harmonic wave represents a massless spin-0 particle carrying energy $E = \hbar \omega$ and momentum $p = \hbar k$.

**Maxwell.** While the Klein-Gordon is to describe a free massive particle of spin 0 (such as the “God particle”, the *Higgs boson*), no massless spin-0 particles are known. The truly massless particle, the photon, has spin 1, and is better described as a perturbation of the electromagnetic field. The latter consists of 6 components which can be organized into two vector fields, $E$ (electric) and $B$ (magnetic), varying in space and time:

\[
E(x, y, z, t) = E_x i + E_y j + E_z k, \quad B(x, y, z, t) = B_x i + B_y j + B_z k.
\]
They satisfy the *Maxwell equations*, written below in the “empty” space-time, i.e. in the absence of electric currents and/or charges:

\[
\nabla \cdot E = 0, \quad \nabla \cdot B = 0, \quad \frac{1}{c} \frac{\partial E}{\partial t} = \nabla \times B, \quad \frac{1}{c} \frac{\partial B}{\partial t} = -\nabla \times E.
\]

Here $c$ is the speed of light, $(i, j, k)$ an orthonormal basis in $\mathbb{R}^3$, and

\[
\nabla := i \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} + k \frac{\partial}{\partial z},
\]

i.e. $\nabla$, $\nabla \cdot$ and $\nabla \times$ are respectively the gradient, divergence and curl:

\[
\nabla \psi = \frac{\partial \psi}{\partial x} i + \frac{\partial \psi}{\partial y} j + \frac{\partial \psi}{\partial z} k, \quad \nabla \cdot V = \frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y} + \frac{\partial V_z}{\partial z},
\]

\[
\nabla \times V = \left( \frac{\partial V_z}{\partial y} - \frac{\partial V_y}{\partial z} \right) i + \left( \frac{\partial V_x}{\partial z} - \frac{\partial V_z}{\partial x} \right) j + \left( \frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y} \right) k.
\]
These explicit formulas are to help the reader to verify the identity
\[ \nabla (\nabla \cdot V) - \nabla \times (\nabla \times V) = \Delta V. \]

It implies that the components of \( E \) and \( B \) satisfy the wave equation. For instance, differentiating \( \partial E/\partial t = c \nabla \times B \) in \( t \) and substituting for \( \partial B/\partial t \) its value \(-c \nabla \times E\) from Maxwell’s other equation, we get \( \partial^2 E/\partial t^2 = c^2 \Delta E \) due to \( \nabla \cdot E = 0 \).

It is not true, however, that any 6 functions satisfying the wave equation form a solution to the Maxwell system; the latter is more stringent. For example, a harmonic traveling wave
\[ E = E_0 e^{\sqrt{-1}(k \cdot q - \omega t)}, \quad B = B_0 e^{\sqrt{-1}(k \cdot q - \omega t)}, \]
to represent a photon, i.e. to satisfy the Maxwell equations, must not merely have \( \omega^2 = c^2 |k|^2 \) (as the wave equation requires), but have the constant 3-vectors \( B_0 \) and \( E_0 \) satisfy \( \omega B_0 = c(k \times E_0) \) and \( k \cdot E_0 = 0 \) (or equivalently \( \omega E_0 = c(B_0 \times k) \) and \( k \cdot B_0 = 0 \)).

Thus, for each energy \( \hbar \omega \) and each direction of propagation \( k/|k| \), there is a two-dimensional space of photon states (determined by vectors \( E_0 \) perpendicular to \( k \)). The direction of \( E_0 \) determines the photon’s polarization. The associated electric and magnetic fields oscillate in perpendicular planes, one spanned by \( E_0 \) and \( k \), the other by \( B_0 \) and \( k \).

**Dirac.** The Dirac equation was introduced in 1928 as a quantum relativistic model of a massive spin 1/2 particle, such as an electron. Apparently, Paul Dirac was inspired by the relationship between the Maxwell system of first order equations and the wave equation as its 2nd order consequence, and was looking for a similar first order system that would imply the Klein-Gordon equation. For this, he needed a “square root” of the Laplacian, and he invented one by borrowing Hamilton’s recipe:
\[
\left( i \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} + k \frac{\partial}{\partial z} \right)^2 = i^2 \frac{\partial^2}{\partial x^2} + j^2 \frac{\partial^2}{\partial y^2} + k^2 \frac{\partial^2}{\partial z^2} + \\
(ij + ji) \frac{\partial^2}{\partial x \partial y} + (jk + kj) \frac{\partial^2}{\partial y \partial z} + (ki + ik) \frac{\partial^2}{\partial z \partial x} = \\
- \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \quad \text{provided that } i^2 = j^2 = k^2 = ijk = -1.
\]

Here \( i, j, k \) are not basis elements of \( \mathbb{R}^3 \), but operators of multiplication by the quaternions. They differ by the factor \( \sqrt{-1} \) from Pauli.
matrices. Therefore the operator $\nabla$ here is to be applied not to 3-vectors, as in the Maxwell equations, but to complex 2-vectors. Consequently, the Dirac equations take the form

$$\frac{i\hbar}{c} \frac{\partial \Psi_\pm}{\partial t} = mc\Psi_\pm - i\hbar \nabla\Psi_\mp, \quad \frac{i\hbar}{c} \frac{\partial \Psi_\mp}{\partial t} = -mc\Psi_\mp - i\hbar \nabla\Psi_\pm,$$

where $\nabla := \sigma_x \frac{\partial}{\partial x} + \sigma_y \frac{\partial}{\partial y} + \sigma_z \frac{\partial}{\partial z}$, $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} = \Delta$.

Here $i = \sqrt{-1}$ again, and $\Psi_\pm(x, y, z, t)$ are a pair of spinors (i.e. functions each taking values in $\mathbb{C}^2$, the natural representation of $SU_2$), and thus totally four complex-valued scalar functions. They all satisfy $(mc^2 - i\hbar \partial/\partial t)(mc^2 + i\hbar \partial/\partial t)\Psi_\pm = c^2 \Delta \Psi_\pm$, which is the componentwise Klein-Gordon equation. It follows from the Dirac equation pretty much the same way as the wave equation follows from the Maxwell system. However, as solutions of Klein-Gordon equation, $\Psi_+$ and $\Psi_-$ would be independent, while in the Dirac system they essentially determine each other.

There are more economical ways of writing down the Maxwell and Dirac equations so that their Lorentz invariance (which we are not going to verify here) would be manifest. We note though that the doubling of the number of components (6 instead of 3 for spin 1, and 4 instead of 2 for spin 1/2) is a by-product of the Lorentz (rather than $SO_3$) symmetry in the concept of spin. By the way, Dirac was particularly content that spinors enter his theory of electron not artificially but out of necessity.

**Fine structure.** One aim of the theory was to revisit the hydrogen atom model at the new, relativistic level. The eigenvalue problem for the Dirac electron in the electrostatic field of the nucleus has the form

$$mc^2 \psi_+ - i\hbar c \nabla \psi_- - \frac{Z \alpha c}{r} \psi_+ = E \psi_+,$$

$$-mc^2 \psi_- - i\hbar c \nabla \psi_+ - \frac{Z \alpha c}{r} \psi_- = E \psi_-,$$

where $m = m_em_n/(m_e + m_n)$ is the reduced mass, and

$$\alpha = \frac{e^2}{4\pi \varepsilon_0 \hbar c} \approx \frac{1}{137}$$

is the dimensionless (and famously mysterious) fine structure constant.
To examine the non-relativistic limit of the equation, we assume that $E = mc^2 + E'$, where $|E'| \ll mc^2$. Then we express $\psi_-$ through $\psi_+$ from one of the equations:

$$\left(2mc^2 + E' + \frac{Z\alpha c}{r}\right)\psi_- = -ihc\nabla\psi_+ \implies \psi_- = \frac{-ihc\nabla\psi_+}{2mc^2 + E' + Z\alpha c/r}.$$ 

Then we substitute the result into the other equation:

$$\left(E' + \frac{Z\alpha c}{r}\right)\psi_+ = -\frac{\hbar^2 c^2}{2mc^2}\nabla \left(\frac{\nabla\psi_+}{1 + \frac{E'}{2mc^2} + \frac{Z\alpha c}{2mc^2 r}}\right) \approx -\frac{\hbar^2}{2m}\Delta\psi_+.$$ 

The approximation (which yields the Kepler quantum equation for the spinor-valued function $\psi_+$) is obtained by discarding the terms in the denominator under the assumption that the scale of both $E'$ and $Z\alpha c/r$ is small compared to $mc^2$.

Note that in the non-relativistic limit, the equation is invariant to both rotations in space, and $SU_2$-transformations of spinors. This is not the case for the relativistic equation. When written explicitly, it would contain the term $r(\nabla r)\nabla\psi_+ = (x\sigma_x + y\sigma_y + z\sigma_z)\nabla\psi_+ \neq \vec{r} \cdot \nabla\psi_+$. This operator does not act on the spinor $\psi_+$ componentwise, and is $SU_2$-invariant only if the natural action of $SU_2$ on $\mathbb{C}^2$ is accompanied by its simultaneous action on $\vec{r} \in \mathbb{R}^3$ through the 2-to-1 map $SU_2 \rightarrow SO_3$. Thinking ahead about exploiting the rotational symmetry as in the Kepler quantum problem, and therefore about decomposing the Hilbert space of spinors $\psi_+$ into the eigenspaces of the angular momentum operator, we can foresee the following crucial change.

The space of degree $n$ harmonic polynomials, when tensored with the representation $\mathbb{C}^2 = V_{1/2}$ of $SU_2$, will decompose into irreducible representations of $SU_2$ according to the Clebsch-Gordan formulas. Namely, each eigenspace $V_l$, $l = 0, 1, \ldots, n-1$, of the orbital angular momentum operator in the non-relativistic theory will generate two eigenspaces: $V_l \otimes V_{1/2} = V_{l-1/2} \oplus V_{l+1/2}$. Consequently, the eigenvalue $\hbar^2 l(l+1)$ of the orbital angular momentum operator will be replaced with the eigenvalues $\hbar^2 j(j+1)$ where $j = l \pm 1/2$ is the total angular momentum quantum number. The name reflects the fact that the total (vector) angular momentum operator now is the sum $\hat{L} \otimes I + I \otimes \hat{S}$ of the orbital one, $\hat{L}$, with the spin operator $\hat{S}$. 

98
Next, for \( j = 1/2, 3/2, \ldots, n - 3/2 \), the same \( j \) comes from \( l = j + 1/2 \) and \( j - 1/2 \), and hence the dimension of the eigenspace with the quantum number \( j \) is \( 2 \dim V_j = 2(2j + 1) \). For \( j = n - 1/2 \) (which comes only from \( l = n - 1 \), the dimension is \( 2j + 1 = 2n \)).

Finally, it turns out (though we are not going to enter here into any details of solving the eigenvalue problem, and merely quote some reputable sources), that the energy level \( E_{n,j} \) of the states with the principal quantum number \( n \) and total angular momentum quantum number \( j \) is no longer independent of \( j \), and is given by the formula

\[
E_{n,j} = mc^2 \left[ 1 + \left( \frac{Z\alpha}{n - j - 1/2 + \sqrt{(j + 1/2)^2 - (Z\alpha)^2}} \right)^2 \right]^{-1/2}
= mc^2 \left[ 1 - \frac{(Z\alpha)^2}{2n^2} - \frac{(Z\alpha)^4}{n^4} \left( \frac{n}{2j + 1} - \frac{3}{8} \right) + O((Z\alpha)^6) \right].
\]

In the expansion, the term following Einstein’s \( mc^2 \) has order \((Z\alpha)^2\) and coincides with the non-relativistic value \( E_n \) of energy independent of \( j \). The next term shows the fine splitting of order \((Z\alpha)^4\) between the energy levels of the states with different values of the total angular momentum. This result agrees with the observable splittings in spectral lines, and signifies a major success of Dirac’s theory.

**Positrons, field theory, and second quantization.** In our approach to the relativistic hydrogen model, we assumed that \( E \approx mc^2 \). Instead, we could assume that \( E \approx -mc^2 \), express \( \psi_+ \) via \( \psi_- \), and proceed to describing eigenstates with the roles of \( \psi_+ \) and \( \psi_- \) interchanged, and the sign of the energy spectrum reversed. The presence of such parasite solutions, which theoretically allows the electrons to sink into the lower and lower energy levels, was initially viewed as a fundamental flaw of Dirac’s theory. In his 1931 truly imaginative response to the critique, Dirac conjectured that almost all the unwanted energy states are in fact occupied by infinitely many electrons, and are therefore mostly inaccessible due to Pauli’s exclusion principle. The only accessible ones would be those few from which the electron has been accidentally excited into a higher energy level, leaving beneath a **hole**. The holes would then behave like **anti-electrons**, fictional particles of exactly the same mass and spin as those of the electrons, but with the exactly opposite electric charge. The act of a normal electron sinking into the hole would look like the annihilation of the two mirror particles, and accompanied by the emission of photons carrying away the released energy.
In 1932, [Carl David Anderson](http://example.com) discovered *positrons* by identifying in cosmic rays some particles that behaved in the magnetic field exactly as electrons except for bending their trajectories in the opposite direction. This discovery, awarded a Nobel prize in 1936, as it turned out, was preceded by several earlier observations of positrons, including the one in cosmic rays by [Dmitry Skobeltsyn](http://example.com) in 1929, and as an electron-positron annihilation, by Anderson’s classmate at Caltech [Chung-Yao Chao](http://example.com) in 1930.

While the story with positron marked another great success of Dirac’s theory, it also showed that the former view of matter as consisting of indestructible particles is inadequate. The particles can *annihilate*, like electron-positron pairs turning into photons, and hence can also be created from photons in the reverse process. Eventually this led to a new paradigm, according to which not only photons are viewed as mere excitations of an electromagnetic field, but also electrons, protons, etc. are to be considered as excitations of the appropriate fields. Respectively, the Klein-Gordon, or Dirac equations, which were introduced as quantum models of classical particles, are to be rather reinterpreted, similarly to the Maxwell equations, as describing *classical fields*. Such a field can be considered as a classical mechanical system, relativistic or not, albeit with infinitely many degrees of freedom. For example, what evolves in time according to the Dirac equation, is not a point \((p, q)\) in a finite dimensional phase space, but the pair \((\psi_+, \psi_-)\) of spinor-valued functions of \((x, y, z)\). The genuine theory needs to be obtained by *quantizing* such classical mechanical systems, i.e. by constructing the quantum theory of the corresponding field. This procedure is often referred to as “second quantization”.

The starting point would be to consider the classical field near the *vacuum state*, i.e. the zero equilibrium solution to the equations. The same as any mechanical system near an equilibrium, the field can be viewed as an ideal gas of harmonic oscillators, although with an infinite number of the oscillators. For instance, representing solutions to the Klein-Gordon equation by Fourier integrals, we thereby decompose the trajectories of the mechanical system into the superposition of elementary harmonic traveling waves:

\[
A(k)e^{i(kq-\omega(k)t)} + A^*(k)e^{i(kq+\omega(k)t)}.
\]

Quantizing each of the harmonic oscillators, we end up with the creation and annihilation operators \(\hat{A}(k), \hat{A}^*(k)\), one pair for each value of \(k\). All these operators together generate an algebra of quantum observables of the field.
This approach is merely a beginning. The actual development of the theory of quantum fields, that would not only consider a “free field” but include interactions and make verifiable predictions, is another story, and a quite heroic one. Among the outcomes is \(\text{QED}\), the quantum electrodynamics, whose predictions of several different phenomena are so accurate, that the values of the inverse fine structure constant \(1/\alpha\) compatible with both the experimental data and the QED theory of these phenomena agree to \(8\) decimal places.

**The quantum hypothesis.** The above abstract discussion has, however, one concrete by-product: we can finally understand the nature of Einstein’s quantum hypothesis.

At our first mention of the Planck constant, we indicated that the hypothesis that a photon’s energy is rigidly related to the angular frequency of light by \(E = \hbar\omega\) can be elucidated in the context of relativity theory.

By now we understand that photons should be considered as excitations of the electromagnetic field, i.e. traveling wave solutions to the Maxwell equations, which are Lorentz-invariant indeed. However, the Planck constant plays no role there until the second quantization is invoked.

Namely, from the point of view of field theory, the Maxwell equations describe the classical electromagnetic field. The latter can be interpreted as an ideal gas of harmonic oscillators: two such oscillators per each wave vector \(k\). The frequency \(\omega\) is determined by the wave vector, but the amplitude of oscillation, and hence the energy can still be arbitrary.

However, the passage from classical to quantum harmonic oscillators results in the allowed energy levels to become discrete, with the spacings \(\Delta E\) equal to \(\hbar\omega\), and the states with energies \(E_n = \hbar\omega(n + 1/2)\) interpreted as the presence of \(n\) indistinguishable photons of the given frequency.

Yet, why the proportionality coefficient between \(\Delta E\) and \(\omega\) has the particular value \(\hbar \approx 6.626 \times 10^{-34} \text{m}^2\text{kg/s}\) remains a mystery of all time and space.
11 Quantum statistics

Statistical mechanics. When a system is not simply multi-particle, but many-particle (and in practice “many” could be of the order of $10^{23}$), it becomes hopeless to track the detailed behavior of the system. So, one resorts to probabilistic methods.

The assumptions one usually makes in doing so are contradictory. One considers a system which can be found in a number of states, $\psi_n$, each characterized by a particular energy level, $E_n$, possibly the same or close for many different states. While an isolated system would stay forever at the same energy level, one assumes instead that our system can exchange energy (or even particles) with a large reservoir called thermostat. For example, think of a quantum harmonic oscillator, which typically stays at the lowest energy level, but is occasionally excited by photons arriving from the environment. Strictly speaking, one should then consider the model consisting of the oscillator and the environment, and taking into account their interaction. Instead, one assumes that the states and energies of the oscillator do not change, but it spends part time in every state, and the question is to find the portion of time spent in each. Alternatively, but somewhat equivalently, one can think of $10^{10}$ identical (but distinguishable) oscillators (or atoms, or better degrees of freedom), which can somehow randomly exchange energy with the remaining $10^{23} - 10^{10}$ (which serve as the thermostat), and ask about the portion of them found at a fixed moment at each particular state.

The problem of putting statistical mechanics on more rigorous foundations and deriving its conclusions from the “first principles” of classical or quantum mechanics seems hard and largely open. However the theory works well when the systems are in thermal equilibrium with the thermostat. Here is how Richard Feynman describes this notion at the beginning of his Statistical Mechanics:

If a system is very weakly coupled to a heat bath at a given “temperature,” if the coupling is indefinite or not known precisely, if the coupling has been on for a long time, and if all the “fast” things have happened, and all the “slow” things not, the system is said to be in thermal equilibrium.

Perhaps, the conclusions of such a vaguely defined theory work well, not because the reasoning that leads to them is flawless, but because it is based on very simple, and hence robust assumptions.
The main principle of statistical mechanics stipulates that the thermostat does not discriminate between states with the same energy, i.e. the probability of finding a system in a particular state depends only on the energy of the state.

To be more accurate, since the energy is defined only up to an additive constant, let $\Delta E = E - E_0$ be the energy by which a state $\psi$ of the system differs from the energy of some reference state $\psi_0$. Then the non-discrimination principle says that the probability ratio $P(\psi)/P(\psi_0) = \rho(\Delta E)$, i.e. depends not on the states, but only on the energy $\Delta E$, and is given by some function $\rho$ which depends only on the state of the thermostat.

Such non-discrimination principle actually implies that $\rho$ must be exponential. Indeed, since the thermostat doesn’t know which state we chose as the reference one, the principle means that for any three states $\psi_0, \psi_1, \psi_2$ with energy levels $E_0, E_1, E_2$ we would have

$$\rho(E_2 - E_0) = \frac{P(\psi_2)}{P(\psi_0)} = \frac{P(\psi_2) P(\psi_1)}{P(\psi_1) P(\psi_0)} = \rho(E_2 - E_1) \rho(E_1 - E_0).$$

That is, $\rho$ is multiplicative: $\rho(A + B) = \rho(A)\rho(B)$. The only continuous functions possessing this property are

$$\rho(E) = e^{-\beta E}.$$

Thus, if a system has states $\psi_n$ with energies $E_n$, then in the thermal equilibrium with the thermostat the probability of finding the system in state $\psi_n$ is

$$P(\psi_n) = \frac{e^{-\beta E_n}}{Z}, \quad Z = \sum_n e^{-\beta E_n}.$$

Here $Z$ is the normalizing constant, and $\beta$ characterizes the state of the thermostat. This distribution is known as the Gibbs distribution, or the canonical ensemble.

The thermodynamical meaning of the parameter $\beta$, which is better be positive since the energies are assumed to be unbounded above, can be inferred from the following example.

As we saw, the energy levels of a one-dimensional harmonic oscillator are equally spaced with the increment $\hbar \omega$. In the quasi-classical limit, when the average energy is high compared to the $\hbar \omega$ scale, it can be considered continuous, with the Gibbs probability distribution $\beta e^{-\beta E}$. Here the normalizing factor $Z = \int_0^\infty e^{-\beta E} dE = \beta^{-1}$. 

104
The expectation value

$$\overline{E} = \beta \int_{0}^{\infty} E e^{-\beta E} \, dE = -E e^{-\beta E} \bigg|_{0}^{\infty} + \int_{0}^{\infty} e^{-\beta E} \, dE = \beta^{-1}.$$ 

Thus, $\beta^{-1}$ has the meaning of the average energy per degree of freedom of any system of (distinguishable) harmonic oscillators in equilibrium with the thermostat.

The latter can be characterized by its absolute temperature $T$ measured in kelvins $K$. The energy passed to the system is proportional to the temperature:

$$\beta^{-1} = kT, \quad \text{where } k \approx 1.38064852 \times 10^{-23} \text{ J/K}$$

is the Boltzmann constant, also denoted $k_B$. It was introduced by Planck and named after Ludwig Boltzmann. Here $J = m^2 \text{kg}/s^2$ stands for the unit of energy joule.

Heat is commonly associated with kinetic energy, and the energy of a harmonic oscillator is evenly divided between kinetic and potential ones. So, it is usually said that $kT/2$ is the measure of kinetic energy per degree of freedom in a system at temperature $T$.

The normalizing factor

$$\mathcal{Z}(\beta) = \sum_{n} e^{-\beta E_n} = \sum_{n} e^{-E_n/kT}$$

of the canonical ensemble is known as the statistical sum or the partition function of the system. As a function of $\beta$ (or $T$), of $E_n$, and of other parameters of the system possibly hidden in the values of $E_n$, it encodes all important thermodynamical properties of the system (the same way as generating functions in combinatorics do). For example the expectation value

$$\overline{E} = \mathcal{Z}^{-1} \sum_{n} E_n e^{-\beta E_n} = -\frac{d}{d\beta} \log \mathcal{Z}.$$ 

By the way, the statistical sum of the quantum harmonic oscillator computed “honestly”, without the continual approximation, is

$$\mathcal{Z} = e^{-\beta \hbar \omega / 2} \sum_{n=0}^{\infty} e^{-n\beta \hbar \omega} = \frac{e^{-\beta \hbar \omega / 2}}{1 - e^{-\beta \hbar \omega}}.$$
This leads to the average energy

$$\overline{E} = -\frac{d}{d\beta} \log Z = \frac{\hbar \omega}{2} + \frac{\hbar \omega}{e^{\hbar \omega/kT} - 1} = kT \sum_{k=0}^{\infty} \frac{B_{2k}}{(2k)!} \left( \frac{\hbar \omega}{kT} \right)^{2k}.$$  

Here $B_{2k}$ are Bernoulli numbers, $B_0 = 1$, $B_2 = 1/6$, $B_4 = -1/30$, ... The expression $\hbar \omega / (e^{\hbar \omega/kT} - 1)$ for the average energy in the system with discrete energy levels $E_n = n \hbar \omega$ plays the crucial role in Planck’s theory of black-body radiation.

The following general (and intriguing) observation is in order. If $E_n$ are the eigenvalues of a Hamilton operator $\hat{H}$ on the Hilbert space $\mathcal{H}$ of quantum states, then

$$Z(\beta) = \text{trace} \left( e^{-\beta \hat{H}} \right).$$

Note that the abstract Schrödinger equation $i\hbar \partial \Psi / \partial t = \hat{H} \Psi$ has the abstract solution $\Psi(t) = U(t) \Psi(0)$. Here $U(t) = e^{t\hat{H}/i\hbar}$ is the one-parametric group of unitary operators describing the evolution of quantum states. Thus,

$$Z \left( \frac{it}{\hbar} \right) = \text{trace} U(t), \text{ or } Z(\beta) = \text{trace} U \left( -\frac{i\hbar}{kT} \right).$$

**Chemical potentials.** The same logic that led us to the Gibbs distribution applies not only to energy, but to any additive characteristic of system’s states, such as e.g. volume of the system, if it is variable. For us, the important additive characteristic will be the number of particles in the system.

Consider a system of distinguishable identical entities (“particles”) whose number $N$ can vary through exchange with the reservoir. Let $\rho(N, E)$ denote the probability ratio of finding the system in a state $\psi$ with $N$ particles and with energy $E$. (As before, the system can have many states with the same values of $N$ and $E$, and $\psi$ is just one of them.) By the previous logic, the function $\rho$ is multiplicative:

$$\rho(N_A + N_B, E_A + E_B) = \rho(N_A, E_A) \rho(N_B, E_B),$$

implying that $\rho = e^{(\mu N - E)/kT}$. Thus, if the system can be in states $\psi_\alpha$ with $N_\alpha$ particles and energy $E_\alpha$, then

$$P(\psi_\alpha) = \frac{e^{(\mu N_\alpha - E_\alpha)/kT}}{Z_{\text{grand}}}, \quad Z_{\text{grand}}(\mu, \beta) = \sum_{\alpha} e^{(\mu N_\alpha - E_\alpha)/kT}. $$

106
Here $\mu$ is the thermodynamical parameter of the thermostat known as chemical potential. This distribution is often called the grand canonical ensemble. In terms of its partition function $Z_{\text{grand}}$, the expected number of particles

$$\overline{N} = kT \frac{\partial}{\partial \mu} \log Z_{\text{grand}}.$$  

The meaning of chemical potential, which seems less intuitive than the meaning of temperature, can be informally described as the “willingness” to supply particles of the given species. (In the more general case when there are several species of particles, each one will have its own chemical potential.) In any case, to reach the state of thermal equilibrium with the thermostat, the system will exchange heat and particles with it until the values of $T$ and $\mu$ of the system become the same as the ones set by the thermostat. This exchange won’t alter the values of $T$ and $\mu$ of the thermostat itself, because it is considered infinite comparing to our system. However, if two systems of comparable size are put in contact, they will exchange heat and particles until the values of $T$ and $\mu$ of both systems equalize.

**The Maxwell–Boltzmann statistics.** Consider the system of identical but distinguishable particles which can occupy states $\psi_n$ with energy levels $E_n$. In a canonical ensemble of such particles, the probability of a particle to be in the state $\psi_n$ is given by the Gibbs formula $e^{-E_n/kT}/Z$. Since different particles occupy this state (or not) independently, it is clear that in a system of $N$ such particles, the share $N(\psi_n)/N$ of them in the state $\psi_n$ coincides with this probability:

$$\frac{N(\psi_n)}{N} = \frac{e^{-E_n/kT}}{Z}, \quad Z = \sum_n e^{-E_n/kT}.$$  

This result is called Maxwell-Boltzmann statistics. This was easy, but let us re-derive it in a hard way, using more machinery.

As $\psi_n$ form a basis in the space $\mathcal{H}$ of states of one particle, the multi-states $\psi_m = \psi_{m_1} \otimes \cdots \otimes \psi_{m_N}$ of the $N$-particle system have energies $E_m = E_{m_1} + \cdots + E_{m_N}$ and form a basis in $\mathcal{H}^\otimes N$. The corresponding partition function, which is actually equal to $Z^N$, can be rewritten by contributions of the multi-states:

$$Z(\beta; E_1, \ldots, E_n, \ldots)^N = \sum_{m_1, \ldots, m_N} e^{-(N_1(m)E_1 + \cdots + N_n(m)E_n + \cdots)/kT}.$$
Here $N_n(m)$ is number of $ns$ among $m_1, \ldots, m_N$, i.e. the number of particles in the multi-state state $\psi_m$ which occupy the state $\psi_n$. It is clear now that the expectation value $\overline{N(\psi_n)} := \sum_m N_n(m) P(\psi_m)$ can be found by differentiating log $Z^N$ in the variable $E_n$. Thus,

$$\overline{N(\psi_n)} = - kT \frac{\partial}{\partial E_n} \log Z^N = - kTN \frac{\partial}{\partial E_n} \log Z = N \frac{e^{-E_n/kT}}{Z}.$$ 

Yet, let’s see what we get by using the grand canonical ensemble, with the phase space $T(\mathcal{H}) = \bigoplus_{N=0}^{\infty} \mathcal{H}^\otimes N$, i.e. by allowing the number of particles vary. The “grand” partition function is computed as the geometric sum:

$$Z_{\text{grand}} = \sum_{N=0}^{\infty} e^{\beta\mu N} Z^N = \frac{1}{1 - e^{\beta\mu}} Z.$$ 

The expected total number of particles and those in the state $\psi_n$ are respectively

$$\overline{N} = \beta^{-1} \frac{\partial}{\partial \mu} \log Z_{\text{grand}} = \frac{e^{\beta\mu} Z}{1 - e^{\beta\mu} Z} = Z_{\text{grand}} - 1$$

$$\overline{N(\psi_n)} = - \beta^{-1} \frac{\partial}{\partial E_n} \log Z_{\text{grand}} = e^{\beta(\mu - E_n)} Z_{\text{grand}}$$

We conclude that

$$\frac{\overline{N(\psi)}}{\overline{N} + 1} = \frac{1}{e^{(E(\psi) - \mu)/kT}},$$

where $\mu$ is the chemical potential. This is the form in which Maxwell-Boltzmann statistics is usually presented — except for the +1 in the denominator, which makes little difference when the typical count of particles is in powers of 10.

**The statistics of Bose-Einstein and Fermi-Dirac.** Consider the situation when our identical particles with states $\psi_n$ and energies $E_n$ are indistinguishable bosons. The phase space now is the symmetric algebra

$$S(\mathcal{H}) = \bigoplus_{N=0}^{\infty} S^N(\mathcal{H}).$$

Fortunately, we have managed to reinterpret the system as an “ideal gas” of distinguishable fictitious “particles”, one per state $\psi_n$, each possessing a sequence of states $\psi_n^\otimes L$ with the energy levels $LE_n$, 
where \( L = 0, 1, 2, \ldots \). In simple terms, the multi-state of the entire system is determined now not by which state is occupied by each boson (they are indistinguishable!) but by the numbers \( L_n \) of bosons in each state \( \psi_n \). The partition function for the \( n \)th fictitious “particle” is

\[
Z_n = \sum_{L=0}^{\infty} e^{L(\mu - E_n)/kT} = \frac{1}{1 - e^{(\mu - E_n)/kT}}.
\]

Not surprisingly, the result is the same as \( Z_{\text{grand}} \) in the Maxwell-Boltzmann case with only one state \( \psi_n \) (i.e. one-dimensional \( \mathcal{H} \)). Note that as a function of energy \( E_n \) it is the same for all \( n \). Consequently the expected number \( \overline{N(\psi)} \) of bosons in a state \( \psi \) of energy \( E \) will be given by the universal formula, known as Bose-Einstein’s statistics:

\[
\overline{N(\psi)} = kT \frac{\partial}{\partial \mu} \log \frac{1}{1 - e^{(\mu - E)/kT}} = \frac{1}{e^{(E - \mu)/kT} - 1}.
\]

The same approach works for the system of identical indistinguishable fermions. In this case, the phase space is the exterior algebra, which we’ve earlier rearranged as a tensor product to represent an ideal gas of qubits:

\[
\Lambda(\mathcal{H}) = \bigoplus_{N=0}^{\infty} \Lambda^N \mathcal{H} = \bigotimes_n \Lambda(\mathbb{C}\psi_n).
\]

Each fictitious qubit here has only two states each: \( \psi_n^{\otimes 0} \) with energy 0, and \( \psi_n^{\otimes 1} \) with energy \( E_n \). For the \( n \)th qubit, the partition function

\[
Z_n = \left(1 + e^{(\mu - E)/kT}\right)_{E=E_n}.
\]

Consequently the expected number \( \overline{N(\psi)} \) of fermions in each state \( \psi \) with energy \( E \) is given by what is known as Fermi-Dirac’s statistics:

\[
\overline{N(\psi)} = kT \frac{\partial}{\partial \mu} \log \left(1 + e^{(\mu - E)/kT}\right) = \frac{1}{e^{(E - \mu)/kT} + 1}.
\]

The chemical potential \( \mu \) is usually denoted in this context by \( E_F \) and is called the Fermi level. A state at the energy level \( E = E_F \) has equal probability to be occupied or vacant: \( \overline{N(\psi)} = 1/2 \). At \( T = 0 \), all states with \( E < E_F \) are occupied, and with \( E > E_F \) are vacant.
Example: the free-electron model. In this model, one considers an electron in a crystal as if it were a free particle of mass $m$ confined in the volume of the solid, which we will assume to have the shape of a cube of size $L$. As a by-product of our discussion of 3-particle systems, we have found the eigenfunctions and eigenvalues of this system:

$$E_{n_1,n_2,n_3} = \frac{(n_1^2 + n_2^2 + n_3^2) \hbar^2 \pi^2}{2mL^2}, \quad n_1, n_2, n_3 = 1, 2, 3, \ldots$$

The number of integer points $(n_1, n_2, n_3)$ in the 1st octant of the integer lattice, found within the (dimensionless) radius $r$, is approximately equal (and the more accurately, the greater is the radius) to the volume of the region $n_1^2 + n_2^2 + n_3^2 \leq r^2$, $n_1, n_2, n_3 > 0$, i.e. $\pi r^3/6$. The differential $\pi dr^3/6 = \pi r^2 dr/2$ of this function provides an approximate number of lattice points within the spherical shell between $r$ and $r + dr$. Therefore the density $D(E) \, dE$ of the distribution of the electron’s stated by energies can be computed in this model as

$$D(E) \, dE = 2 \times \frac{\pi}{6} dr^3 = \frac{\pi}{3} d \left( \frac{2mL^2 E}{\hbar^2 \pi^2} \right)^{3/2} = \frac{(2m)^{3/2} V}{2\hbar^3 \pi^2} \sqrt{E} dE.$$

Here the factor 2 accounts for the number of spin states per eigenfunction, and $V$ stands for the volume $L^3$ of the crystal.

At temperatures $T \approx 0$, the Fermi level $E_F$ of the crystal can be determined by equating the total number of states available at $E \leq E_F$ to the total number $N$ of electrons in the system:

$$N = \int_0^{E_F} D(E) \, dE = \frac{(2m)^{3/2} V}{3\hbar^3 \pi^2} E_F^{3/2}, \quad \text{i.e. } E_F = \frac{\hbar^2}{2m} \left( \frac{3\pi^2 N}{V} \right)^{2/3}.$$

The total energy of electrons in this model is

$$E_{\text{total}} = \int_0^{E_F} E \, D(E) \, dE = \frac{(2m)^{3/2} V}{5\hbar^3 \pi^2} E_F^{5/2} = \frac{3}{5} NE_F.$$

A decrease of $V$ increases $E_{\text{total}}$ and hence creates electron degeneracy pressure, $P$. Namely, $dE_{\text{total}} = -P \, dV$ (the minus is here because the work of pressure forces increases the energy). Thus,

$$P = -\frac{dE_{\text{total}}}{dV} = \frac{2}{5} (3\pi^2)^{2/3} \left( \frac{\hbar^2}{2m} \left( \frac{N}{V} \right) \right)^{5/3}.$$
Such pressure, caused by the inability of electrons to occupy lower energy levels due to Pauli’s exclusion principle, can account for substantial portion of metal’s resistance to compression.

In the case $T > 0$, the computation relating $N$ and $E_F$ needs to be modified to take into account Fermi-Dirac statistics:

$$N = \int_0^{E_F} \frac{1}{e^{(E-E_F)/kT} + 1} \frac{(2m)^{3/2}V}{2\hbar^3\pi^2} \sqrt{E} \, dE.$$ 

Assuming $N$ constant results in a decrease of the value of $E_F$ with temperature. However, as estimates show, under normal conditions the approximation $T \approx 0$ is justified. For instance, the Fermi temperature $T_F = E_F/k$ for metallic sodium is about $36 \times 10^3$ K, i.e. about 120 times the room temperature ($\approx 300$ K). This means that at $T \approx 300$ K only a small portion of electrons will be excited.

The relativistic theory, due to Subrahmanyan Chandrasekhar, of the electron degeneracy pressure predicts the eventual fate of the Sun. Large stars, when exhaust their fuel, cannot resist gravitational collapse, and either turn into black holes, or explode as supernovae. But according to Chandrasekhar’s theory, stars of up to approximately 8 solar masses, due to the electron degeneracy pressure, withstand the collapse, and turn into white dwarfs.

**Bose-Einstein condensation.** This is a phenomenon predicted by Einstein in 1925, but confirmed experimentally only in 1995. It is a phase transition of a dilute gas at temperatures close to 0 K into a state where almost all atoms are trapped at the lowest energy level.

Let this level be $E_0$, the next energy level of atoms be $E_1 > E_0$, and let us first disregard all higher states (which makes sense as $T \to 0$). By Bose-Einstein’s statistics, in the ground state, the number of atoms

$$N_0 = \frac{1}{e^{(E_0-\mu)/kT} - 1}.$$ 

At low temperatures, and as the total number of particles grow, this should tend to $\infty$, meaning that $\mu \to E_0$. (By the way, in Bose-Einstein’s statistics, $\mu < E_0$, since the denominator must remain positive.) In this limit, the number of atoms in the next state

$$N_1 = \frac{1}{e^{(E_1-\mu)/kT} - 1} \to \frac{1}{e^{\Delta E/kT} - 1}$$ 

i.e. does not depend on the total number of particles. This means that any extra particle added to the system will have to condense
to the ground state. Note that in the Maxwell-Boltzmann case, the fraction \( N_1/N_0 = e^{-\Delta E/kT} \) would remain fixed, i.e. the number \( N_1 \) would grow proportionately to the total number of atoms.

In order to find the temperature at which the transition to this phase occurs, we will estimate the total number \( N_+ \) of atoms found in all non-ground states, i.e. all states with energies \( E > E_0 \), using the previous continual (and more realistic) model of a gas in the box \( L \times L \times L \).

Assuming for simplicity that the atoms have spin 0, we find

\[
N_+ = \int_{E=E_1 \approx E_0}^{\infty} \frac{1}{e^{(E-E_0)/kT} - 1} \frac{(2m)^{3/2}V}{4\hbar^3\pi^2} \sqrt{E - E_0} \, dE \approx \frac{(2mkT)^{3/2}V}{4\hbar^2\pi^2} \int_0^\infty \frac{\sqrt{x} \, dx}{e^x - 1} = \frac{(2mkT)^{3/2}V}{4\hbar^2\pi^2} \times \text{Const.}
\]

The critical temperature \( T_C \) of condensation is then defined by equating \( N_+ \) to the total number of particles \( N \):

\[
\frac{N}{V} = \text{Const} \times \frac{(2mkT_C)^{3/2}}{4\hbar^2\pi^2}.
\]

By the way, \( \text{Const}/\sqrt{\pi} = \zeta(3/2) \approx 2.6 \). Indeed,

\[
\int_0^\infty \frac{x^{1/2} \, dx}{e^x - 1} = \sum_{n=1}^{\infty} \int_0^\infty e^{-nx} x^{1/2} \, dx = \left( \sum_{n=1}^{\infty} \frac{1}{n^{3/2}} \right) \times \int_0^\infty e^{-y} y^{1/2} \, dy.
\]

The remaining integral is Gaussian:

\[
\int_0^\infty e^{-y} y^{1/2} \, dy = \frac{1}{2} \int_0^\infty e^{-y} y^{-1/2} \, dy = \int_{-\infty}^\infty e^{-z^2/2} \, dz = \sqrt{\pi}.
\]

Thus,

\[
\frac{N_+}{N} = \left( \frac{T}{T_C} \right)^{3/2} \quad \text{for} \quad T < T_C = \frac{2\pi\hbar^2}{mk} \left( \frac{N/V}{\zeta(3/2)} \right)^{2/3}.
\]

This means that when the temperature is below the critical one, the fraction \( 1 - (T/T_C)^{3/2} \) of all atoms condenses to the ground state, while above the critical temperature the condensate “evaporates” from the ground state completely.
12 Solid state physics

An engineer approaches a famous physicist with a puzzling experimental result. The physicists immediately gives a convincing explanation. ‘But your reasoning suggests a positive effect,’ objects the engineer, ‘and in my experiment it was negative!’ — ‘Ha!’ replies the physicist. ‘This is even easier to explain...’

One-dimensional periodic potentials. We are about to embark on a thorough study of the behavior of electrons in crystals. As a model of the situation, we may assume that atoms’ nuclei form an ideal lattice, and the electrons move in a perfectly periodic electrostatic potential created by the nuclei. Moreover, in the simplest model, we assume that the crystal is 1-dimensional rather than 3-dimensional. This way we arrive at the familiar situation: the stationary Schrödinger equation on the line, where however the potential energy function \( V \) is periodic with the period \( a \) equal to the distance between the nuclei in the crystalline lattice:

\[
-\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2} + V(x)\psi = E\psi, \quad V(x + a) = V(x).
\]

Taking \( V = 0 \) identically, we find ourselves in the situation of a free particle on a line, where we know the spectrum is continuous, with the eigenfunctions \( \psi = e^{ikx}, \ k = \pm \sqrt{2mE}/\hbar \). Our current goal is to understand, for arbitrary periodic \( V \), the structure of the spectrum of this Schrödinger operator, i.e. of the set of those values of \( E \) for which the equation has solutions \( \psi \) remaining bounded as \( x \to \pm \infty \).

It might be beneficial at this point to realize that mathematical equations don’t know how we interpret them, and to take a look at the equation as dynamical, with the space variable \( x \) playing the role of time. When \( V = 0 \), and \( E > 0 \), we have the equation of a harmonic oscillator \( \psi'' = -\frac{2mE}{\hbar^2} \psi \). After turning \( V \) on, we obtain an oscillator affected by an additional “time”-periodic linear force field proportional to \( V(x)\psi \). It is Hamiltonian, with the “time”-periodic Hamilton function (where \( q \) stands for \( \psi \) and \( p \) for \( \psi' \)):

\[
H(q, p, x) = \frac{p^2}{2} + \frac{2m}{\hbar^2} (E - V(x)) \frac{q^2}{2}.
\]

In any case, for a fixed value of \( E \), a solution \( \psi \) is uniquely determined by the initial conditions \( (\psi(0), \psi'(0)) \), and all solutions form a real 2-dimensional vector space \( \mathbb{R}^2 \). Namely, since the equation...
is linear in $\psi$, if $\psi_1, \psi_2$ are two solutions, and $\lambda_1, \lambda_2$ are arbitrary scalars, then $\psi = \lambda_1 \psi_1 + \lambda_2 \psi_2$ is a solution too.

The fact that the coefficients of the equation are $a$-periodic does not mean that each solution is $a$-periodic. It means instead that, given a solution $\psi(x)$, its shift $\psi(x + a)$ is a solution again. Thus, the shift $x \mapsto x + a$ defines a linear transformation in the plane $\mathbb{R}^2$ of solutions. It is called the monodromy transformation, or monodromy matrix, and it will help us to efficiently analyze the behavior of solutions at $x \to \pm \infty$.

Explicitly, let $\psi_1, \psi_2$ be the basis of solutions with the initial conditions $(\psi_1(0), \psi'_1(0)) = (1, 0)$, and $(\psi_2(0), \psi'_2(0)) = (0, 1)$. Then the monodromy matrix

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} : = \begin{bmatrix} \psi_1(a) & \psi_2(a) \\ \psi'_1(a) & \psi'_2(a) \end{bmatrix}.$$ 

For any solution $\psi$, we have

$$\begin{bmatrix} \psi(na) \\ \psi'(na) \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^n \begin{bmatrix} \psi(0) \\ \psi'(0) \end{bmatrix}, \quad n = 0, \pm 1, \pm 2, \ldots.$$

The fact that the equation is Hamiltonian has its consequences for $M$. Namely, the monodromy transformation preserves the symplectic structure, i.e. the signed area on the phase plane $\mathbb{R}^2$: $\det M = 1$. Without fancy words (and in classical terms) the Wronskian of two solutions $\psi_1, \psi_2$ is constant:

$$\frac{d}{dx} \begin{vmatrix} \psi_1(x) & \psi_2(x) \\ \psi'_1(x) & \psi'_2(x) \end{vmatrix} = \begin{vmatrix} \psi_1'(x) & \psi_2'(x) \\ \psi'_1(x) & \psi'_2(x) \end{vmatrix} + \begin{vmatrix} \psi_1''(x) & \psi_2''(x) \\ \psi_1'(x) & \psi_2'(x) \end{vmatrix} = 0 + 0,$$ 

where the rightmost determinant is 0 because, due to the equation $\psi''_i = 2m\hbar^{-2}(V(x) - E)\psi_i$, the row of 2nd derivatives is proportional to the top row.

Thus, monodromy matrices belong to the special linear group $SL_2(\mathbb{R})$ of real invertible $2 \times 2$-matrices with determinant $ad - bc = 1$.

To understand the behavior of the iterations $M^n$, we examine the characteristic equation of $\lambda^2 - (a + d)\lambda + 1 = 0$ of the matrix $M$. It has two roots $\lambda_\pm$ satisfying $\lambda_+ \lambda_- = \det M = 1$, which can be:

(i) distinct complex conjugate: $\lambda_\pm = e^{\pm i \theta}$;
(ii) distinct real positive $\lambda_+ > 1 > \lambda_- = 1/\lambda_+$ or distinct real negative $0 > \lambda_- > -1 > \lambda_+ = 1/\lambda_+$;
(iii) double roots: $\lambda_\pm = 1$ or $\lambda_\pm = -1$. 

114
In case (i), taking the complex conjugate eigenvectors of $M$ for initial conditions at $x = 0$, we obtain a basis of complex solutions $\psi_\pm$ satisfying $\psi_\pm(x + na) = e^{\pm in\theta} \psi_\pm(x)$. Such solutions are bounded, i.e. represent eigenfunctions of the Schrödinger operator. The eigenspace is 2-dimensional, and one can take a real basis in the form of the real and imaginary parts of $\psi_+$ (and assume that $\psi_- = \psi_+^*$). If we change slightly the value of energy $E$, the equation, the monodromy matrix $M$, the coefficients of its characteristic polynomial, and its roots $e^{\pm i\theta}$ will change a little (i.e. $\theta = \theta(E)$), but we will still remain in case (i). Thus, the value of $E$ in this case belongs to the continuous spectrum of the Schrödinger operator.

In case (ii), taking real eigenvectors for initial conditions, we obtain a real basis of solutions $\psi_\pm$ satisfying $\psi_\pm(x + na) = \lambda_n^{\pm} \psi_\pm(x)$. Since $|\lambda_\pm| \neq 1$, the powers of either eigenvalue will grow indefinitely in the absolute value either as $n \to +\infty$, or as $n \to -\infty$, and tend to 0 in the opposite direction. As a result, each of the functions $|\psi_\pm|$ will tend to $\infty$ in one direction, and to 0 in the other. Their linear combinations $\psi = A\psi_+ + B\psi_-$ will grow unbounded (typically in both directions). Thus, in case (ii), the energy value $E$ does not belong to the spectrum of the Schrödinger operator.

In case (iii), which is borderline between (i) and (ii), it is possible that the eigenspace corresponding to the single eigenvalue $\lambda(= \pm 1)$ is 2-dimensional. This results in the monodromy matrix $M = \pm I$, and the 2-dimensional space of eigenfunctions $\psi$, satisfying $\psi(x + a) = \psi(x)$ (i.e. $a$-periodic) when $\lambda = 1$, and $\psi(x + a) = -\psi(x)$ (anti-periodic) when $\lambda = -1$. However, as we can see later, this is a degenerate possibility, and the typical one is when the eigenspace of $M$ is 1-dimensional. When $\lambda = +1$, if we draw a rectangle in $\mathbb{R}^2$ with the base on the eigenline, then its image under $M$ will have the same base (since points on the eigenline don’t move) and the same altitude (since $M$ preserves areas), but will become a parallelogram. Thus, in some coordinate system, $M$ will look like the shear, and when $\lambda = -1$ this shear will be accompanied with the central symmetry:

$$M = \pm \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad M^n = (-1)^n \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix}, \quad n = 0, \pm 1, \pm 2, \ldots$$

We conclude that in this case, there is one periodic ($\lambda = +1$) or one anti-periodic ($\lambda = -1$) eigenfunction $\psi$ of the Schrödinger operator, defined by the choice of the initial condition $(\psi(0), \psi'(0))$ on the eigenline of $M$. All other initial conditions give rise to solutions which grow indefinitely in the absolute value as $x = na \to \pm \infty$ (though not exponentially but only linearly in $n$).
The band structure of the spectrum. We have understood so far how the place of an energy level \( E \) in the spectrum of the Schrödinger operator depends on the eigenvalues of the monodromy matrix \( M = M(E) \). The next question is what happens with these eigenvalues as \( E \) varies ranging from \(-\infty\) to \(+\infty\).

Our previous discussion suggests that intervals (bands) of continuous spectrum (case (i)) should alternate with intervals (gaps) with no eigenstates (case (ii)). The energy levels with \( a \)-periodic and \( a \)-anti-periodic eigenfunctions (case (iii)) should serve as the boundaries between the bands and gaps.

In fact one can say much more about the spectrum. Namely, there exists an infinite sequence of energy levels

\[
E_0 < E_1 \leq E_2 < E_3 \leq E_4 < \cdots
\]

\[
\cdots \leq E_{4k} < E_{4k+1} \leq E_{4k+2} < E_{4k+3} \leq E_{4k+4} < \cdots
\]

where \( E_0, E_3, E_4, E_7, \ldots, E_{4k}, E_{4k+3}, \ldots \) form the discrete spectrum of the Schrödinger operator in the class of \( a \)-periodic functions, while \( E_1, E_2, \ldots, E_{4k+1}, E_{4k+2}, \ldots \) form the discrete spectrum in the class of \( a \)-anti-periodic functions. The bands of continuous spectrum (in the class of arbitrary bounded functions on the line) are

\[
[E_0, E_1], [E_2, E_3], \ldots, [E_{4k}, E_{4k+1}], [E_{4k+2}, E_{4k+3}], \ldots
\]

Typically they are separated by the gaps

\[
(E_1, E_2), (E_3, E_4), \ldots, (E_{4k+1}, E_{4k+2}), (E_{4k+3}, E_{4k+4}), \ldots
\]

It is possible that some of these gaps shrink to a point (i.e. \( E_{2l-1} = E_{2l} \)). In this case the corresponding eigenspace of \( a \)-periodic or \( a \)-anti-periodic states becomes 2-dimensional (which is the degenerate possibility of case (iii)), so that the adjacent bands join into a single one (such as \( [E_{2l-2}, E_{2l+1}] \)).

Furthermore, when \( E < E_0 \), the trace \( \text{tr } M(E) = a + d = \lambda_+ + \lambda_- \) remains \( > 2 \), so the roots \( \lambda^2 \pm \sqrt{(\text{tr } M)\lambda_+ + 1} = 0 \) remain real positive. When \( E \) increases from \( E_0 \) to \( E_1 \), the trace decreases monotonically from \( 2 \) to \(-2\), and the roots \( \lambda_\pm = e^{\pm i\theta} \) move along the unit circle to “reconcile” again at \(-1\). While \( E \) ranges in the gap between \( E_1 \) and \( E_2 \), the roots remain real and negative (\( \text{tr } M < -2 \)), and return to \(-1\) at \( E = E_2 \). When \( E \) increases from \( E_2 \) to \( E_3 \), the pair of complex conjugate roots \( \lambda_\pm \) move from \(-1\) to \(+1\) each along its half of the unit circle. In the gap between \( E_3 \) and \( E_4 \), the roots stay real positive, and return to the value \(+1\) at \( E = E_4 \). From this point the whole process repeats on and on.
In the degenerate situation when a gap shrinks to a point, the two complex conjugate roots $\lambda_{\pm}$ "meet" at $+1$ or $-1$, and then path through each other and continue moving along the circle.

Moreover, the eigenfunctions $\psi_m$ corresponding to the eigenvalues $E_m$ can be characterized in terms of their zeroes.

Note that $\psi(x+a) = \pm \psi(x)$ implies that (in either case $\pm$) zeroes of $\psi$ form arithmetic sequences with the step $a$, each represented once on the interval $[0, a)$. It turns out that the $a$-periodic eigenfunctions $\psi_0, \psi_3, \psi_4, \psi_7$ etc. have on each period $[0, a)$ respectively no zeroes, 2 zeroes, 4 zeroes, 6 zeroes etc. The anti-periodic eigenfunctions $\psi_1, \psi_2$ have 1 zero, $\psi_5, \psi_6$ have 3 zeroes etc. on the interval $[0, a)$ (which for these functions is their half-period).

A simple illustration to this behavior can be obtained in the most degenerate case when $V(x) = 0$. Then all the gaps disappear. Namely, $E_0 = 0$, (with the eigenfunction $\psi_0 = 1$), while $E_{2l-1} = \frac{2m\pi^2a^2}{\hbar^2} = E_{2l}$ for all $l = 1, 2, \ldots$, and $\sin(\pi lx/a), \cos(\pi lx/a)$ can be taken for the eigenfunctions $\psi_{2l-1}, \psi_{2l}$ ($a$-anti-periodic for odd $l$ and $a$-periodic for even).

One way to obtain all this information would be to look at the phase $pq$-plane of the time-dependent Hamiltonian system with the hamiltonian $H(p, q, x) = p^2/2 + 2m\hbar^{-2}(E - V(x))q^2/2$ and with $x$ in the role of the time variable. The equations have the form

\[ q' = p, \quad p' = 2m\hbar^{-2}(V(x) - E)q, \]

where $q$ stands for $\psi$, and $p$ for $\psi'$. The system describes the dynamics of points on the phase plane, but since it is linear, it also describes the dynamics of 1-dimensional subspaces, i.e. lines passing through the origin.

Note that regardless of $V$ or $E$, the vertical line always moves clockwise (Figure 21a). At the "moment" $x$ when a phase point crosses the vertical line, the corresponding function $\psi$ has a zero. Since in the dynamical process distinct lines remain distinct, they cannot by-pass each other. This proves the so-called Sturm’s theorem: Between any two zeroes of one solution to the stationary Schrödinger equation on the line, any other solution also has a zero.

That is, on the $x$-interval when a line moves from one vertical position to the next one, any other line must also assume a vertical position — exactly once!
Furthermore, as the energy level changes from $E$ to $E + \Delta E$, the Hamilton function $H$ is corrected by $2m\hbar^{-2}(\Delta E)q^2/2$. The vector field on the $pq$-plane representing this correction is shown on Figure 21b. The picture indicates that all lines (except the vertical one) are rotated by this field clockwise. Assume now that the monodromy operator $M(E)$ has non-real eigenvalues $\lambda_\pm = e^{\pm i\theta}$. This means that in some (possibly non-Cartesian) coordinate system on the $pq$-plane, $M(E)$ looks like a rotation through the angle $\pm \theta \mod 2\pi$. That is, the line vertical at the moment $x = 0$ would make by the moment $x = a$ a certain number of full turns clockwise plus/minus $\theta$ radians (in this coordinate system). Therefore, at the energy level $E + \Delta E$ the rotation will happen even faster — due to the correction term. This explains the monotonic behavior of the eigenvalues on the unit circle, but also shows that as $E$ increases, the distances between consecutive zeroes of the solutions shrink. The precise count of the numbers of zeroes of the eigenfunctions $\psi_m$ can be derived from this observation.

In case the behavior of $\text{tr} \ M$ inside the gaps looks suspicious (for it first increases in the absolute value above 2 but then returns back to 2), we include Figure 22 showing what happens with the matrix $M(E)$ as a point in $SL_2(\mathbb{R})$ as $E$ grows monotonically.

Geometrically the group $SL_2(\mathbb{R})$ of real $2 \times 2$-matrices with determinant 1 can be visualized as the interior of the solid torus shown on the picture. The two black dots represent matrices $\pm I$. The regions enclosed by the four “horns” (there are two such regions inside the solid torus) represent the matrices with complex conjugate $\lambda_\pm$, and the space outside the horns with real $\lambda_\pm$. The surface of the horns corresponds to matrices with double-root $\lambda = +1$ (the right horns) and $\lambda = -1$ (the left horns), but with only 1-dimensional eigenspace. The equations of the rights and left horns are $\text{tr} = 2$ and $\text{tr} = -2$ respectively. There is no discontinuity of trace on the two dashed
vertical ellipses where the surfaces meet, because these ellipses are not in $SL_2(\mathbb{R})$ (which does not include the boundary points of the solid torus on our picture). The surfaces (which meet near $\pm I$ akin to two branches of a quadratic cone) also separate the space outside the horns into two components: right and left. They correspond to $\text{tr} > 2$ and $\text{tr} < -2$ respectively. The green ellipse on the picture consists of the matrices or rotation (in some coordinate system), and the arrows point in the direction of increasing rotation angle. The blue line represents the curve $E \leftrightarrow M(E)$ for a typical periodic potential $V$. It starts outside the horns, then enters (at $E = E_0$) the interior of the top right horn, and then proceeds as the picture shows, alternating the dashed intervals (bands) and solid intervals (gaps). Note that as $E$ increases, the blue curves behaves “logically”, spiraling (in some sense) along the axis of the solid torus in the same direction.

As a remark, let us mention that the gaps, which are the forbidden energy intervals in the quantum mechanical interpretation of the differential equation, in the dynamical interpretation of the same equation play a “constructive” role, and are known as zones of parametric resonance. The phenomenon should be familiar from the following experience. On a swing, trying to increase the amplitude of swinging, you apply periodic force, in such a way that the monodromy operator has an eigenvalue $\lambda$ with $|\lambda| > 1$. This makes the amplitude increase by a factor of $|\lambda|$ on each period.
The Kronig-Penney model. In this model example, the periodic potential is taken to be formed by delta-functions:

$$V(x) = \frac{\hbar^2 \alpha}{2ma} \left( \sum_{n=-\infty}^{\infty} \delta(x - na) \right).$$

The stationary Schrödinger equation has the usual form

$$-\frac{\hbar^2}{2m} \psi'' + V(x)\psi = E\psi,$$

where however $\psi$ could have discontinuous derivative at $x = na$ with the break controlled by the relation

$$\psi'(na^+) - \psi'(na^-) = \frac{\alpha}{a} \psi(na).$$

Recall that it is obtained from the differential equation by integration:

$$0 = \int_{na^-}^{na^+} \left[ \psi''(x) - \frac{2m}{\hbar^2} V(x)\psi(x) \right] dx = \psi'(x) \bigg|_{na^-}^{na^+} - \frac{\alpha}{a} \psi(na).$$

The monodromy matrix can be composed from two transformations defined by the solutions on the lattice’s period $[0^-, a^-]$: the transition $M_0$ from $x = 0^-$ to $x = 0^+$ followed by the evolution on the interval $(0, a)$, where $V = 0$. The latter is given by the rotation-like matrix

$$\begin{bmatrix}
\cos ka & \frac{1}{k} \sin ka \\
-k \sin ka & \cos ka
\end{bmatrix}, \text{ where } k^2 = \frac{2mE}{\hbar^2}.$$  

Namely, the top row is obtained by setting $x = a$ in the solutions $\psi_1 = \cos kx$ and $\psi_2 = \frac{1}{k} \sin kx$ satisfying the initial conditions $(\psi(0), \psi'(0)) = (1, 0)$ and $(0, 1)$ respectively. The bottom row is similarly obtained from the derivatives $\psi'_1$ and $\psi'_2$.

To find the matrix $M_0$, note that according to the breaking condition for solutions $\psi$, when $(\psi(x), \psi'(x)) \to (1, 0)$ as $x \to 0^-$, it must tend to $(1, \frac{\alpha}{a})$ as $x \to 0^+$. At the same time, the solution with $\psi(0) = 0$ will have continuous derivative.

Thus, the entire monodromy matrix is the product

$$M = \begin{bmatrix}
\cos ka & \frac{1}{k} \sin ka \\
-k \sin ka & \cos ka
\end{bmatrix} \begin{bmatrix}
\frac{1}{a} & 0 \\
\frac{\alpha}{a} & 1
\end{bmatrix}.$$
What we need is the trace of it:

\[
\text{tr } M = 2 \left( \cos ka + \frac{\alpha \sin ka}{2} \right).
\]

The graph of \( \text{tr } M/2 \) as a function of \( ka = \sqrt{2mE}/\hbar \) is sketched on Figure 23 for a positive value of \( \alpha \) (i.e. assuming that the intervals of \( V = 0 \) are separated by infinitely tall potential barriers at \( x = na \)).

The segments \([k_0a, k_1a]\), \([k_2a, k_3a]\), etc. are the bands, separated by the gaps \((k_1a, k_2a)\) etc. The solid dots on the graph correspond to the values \( ka = \pi l, l = 1, 2, \ldots \). The respective eigenfunctions in this model are very simple: \( \psi = \sin lx \). They happen to have continuous derivatives, and are \( a\)-anti-periodic for odd \( l \) and \( a\)-periodic for even.

![Figure 23: Graph of \( \frac{1}{2} \text{tr } M = \cos ka + \frac{\alpha \sin ka}{2} \)](image)

The above Kronig-Penney model of a lattice of delta-barriers is a good illustration to the general theory of energy bands for periodic potentials. However, another version of the Kronig-Penney model seems more adequate for representing properties of real crystals. It is a lattice of delta-shaped wells:

\[
V(x) = -\frac{\hbar^2 \alpha}{2ma} \left( \sum_{n=\text{-}\infty}^{\infty} \delta(q-na) \right),
\]

where \( \alpha \) is assumed positive.
In fact our previous computations cover this case: we only need to replace $\alpha$ with $-\alpha$. Thus,

$$\text{tr } M = 2 \left( \cos ka - \frac{\alpha \sin ka}{2} \right),$$

where $k^2 = \frac{2mE}{\hbar^2}$.

These formulas describe the bands of energy with $E > 0$, i.e. above the potential wells, but it turns out that the ground state of the system (as in the cases of a single or double delta-well potential we studied earlier) has negative energy. For $E < 0$, the “evolution” of $\psi$ between the wells is described by hyperbolic trigonometric functions. The monodromy matrix has the form

$$M = \begin{pmatrix} \cosh \kappa a & \frac{1}{\kappa} \sinh \kappa a \\ \kappa \sinh \kappa a & \cosh \kappa a \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\frac{\alpha}{a} & 1 \end{pmatrix},$$

and its trace is given by

$$\text{tr } M = 2 \left( \cosh \kappa a - \frac{\alpha \sinh \kappa a}{2 \kappa a} \right),$$

where $\kappa^2 = -\frac{2mE}{\hbar^2}$.

On Figure 24, the graph of $\text{tr } M/2$ is shown for the value $\alpha = 4$. This is a borderline case: when $\alpha < 4$, the lowest band $[E_0, E_1]$ lies between $E_0 < 0$ and $E_1 > 0$, while for $\alpha > 4$ both $E_0$ and $E_1$ are negative. For $\alpha = 4$, $E_1 = 0$, and the corresponding anti-periodic eigenfunction is described on its period $[-a, a]$ by the formula $\psi(q) = |q| - a/2$.
We have assumed so far that the crystal consists of infinitely many equally spaced atoms. What happens if the total number of atoms is finite? While some details of the answer depend on what kind of conditions are imposed on the boundary, the essence of the situation is captured well by assuming that the finite crystal of $N$ atoms is circular, i.e. the atom $N + 1$ coincides with the atom 1. In this case, the $\psi$-function should obey the periodicity rule: $\psi(x + Na) = \psi(x)$. Since the eigenfunctions in our problem also satisfy $\psi(x + a) = e^{i\theta} \psi(x)$, this implies that the complex eigenvalues $\lambda_{\pm} = e^{\pm i\theta}$ must be $N$th roots of unity: $N\theta \equiv 0 \mod 2\pi$. Thus, instead of (and inside of) each band of the continuous spectrum, the system will have $2N$ discrete (yet narrowly spaced) energy levels. In reality the value of $N \approx 10^7$ (per centimeter of crystal’s linear size). Thus, the energy levels fill the spectral bands so densely, that for many practical purposes they can be deemed continuous.

**Conductors and insulators.** The analogy between the nice mathematics of one-dimensional Schrödinger equations with periodic potentials on the one hand, and the properties of solid crystals on the other looks rather far-fetched. An accurate model should be based on the superposition of Coulomb-like potentials of the nuclei placed at the vertices of a 3-dimensional crystalline lattice. Yet, we encounter here the art of doing physics at its best (or worst, depending on the attitude), when the realistic description of a complicated system is obtained by cleverly combining the insights gained from several oversimplified (and often logically incompatible) theoretical models.

First, in a 3D periodic potential, bounded eigenfunctions of the stationary Schrödinger equation still have the quasi-periodic form. That is, $\psi(q) = e^{ik \cdot q} u(q)$, where $u$ is periodic with respect to the 3D-lattice, and $\theta = (\theta_1, \theta_2, \theta_3)$ is some lattice wave vector. Thus, if vector $a = (a_1, a_2, a_3)$ belongs to the lattice of periods, $\psi(q + a) = e^{i\theta \cdot a} \psi(q)$. This is the content of the so-called Bloch theorem (named after Felix Bloch).

Accepting the idea that other elements of the one-dimensional theory, including the band structure of the energy spectrum, also generalize to 3D, we obtain a simple explanation of why some crystals are electric conductors and some other insulators. Namely, if an energy band is only partially filled with electrons (it is then called conduction band), the crystal will be a good conductor. Indeed, since the band is densely filled with vacant energy levels, electrons can be easily excited into these levels, and that is what happens when voltage is applied to the crystal causing electric current. However,
if the band is filled to capacity (allowed by Pauli’s principle), the minimal energy needed to excite the electrons equals the energy gap between this and the higher band. When voltage is applied, this gap becomes the energy barrier for the current, making the crystal a good insulator.

“Wait a minute!” you should say. “And how the excitement of electrons is related to current?” Okay, a physicist explains, remember how the superposition of the ground and excited states of a particle in a well resulted in the time-dependent probability density $|\psi|^2$ behaving like a ball bouncing off the walls of the well? When we say that some electrons are excited, it means that the psi-function of the entire system of electrons in the crystal is similarly found in the superposition of two or more states, resulting, as it turns out, in a probability current in the direction of the applied voltage.

Furthermore, the probability of finding a state with the energy $E$ populated by an electron is described by the Fermi-Dirac statistics:

$$
\frac{1}{e^{(E-E_F)/kT} + 1}.
$$

Respectively, at temperatures $T \approx 0$ K, all energy levels below the Fermi level $E_F$ are filled with electrons, and all energy levels above it are vacant. Therefore, conductors are those crystals, where the Fermi level $E_F$ falls within the (conduction) energy band (as it happens for metals). Insulators (e.g. diamond) are those crystals where the Fermi level $E_F$ lies in the gap between two bands, called in this case conduction band (upper) and valence band (lower). In reality, there is an intermediate class of crystals (e.g. silicon), called semiconductors, for which the gap between the conduction and valence band is not too wide, and so they behave as insulators at low temperatures, but at higher temperature, when some electrons (according to the Fermi-Dirac statistics) are induced into the conduction band, they conduct electricity (though not as well as metals do).

“Wait a minute!” you should say again. “And what does the band right below the Fermi level have to do with chemical valence?” This time a physicist’s answer may take longer than a minute.

Let us recall the model of molecular bonding (the double delta-well potential). For a single delta-well potential, there is one bound state $\psi$ of the electron, which decays exponentially with the distance. Consequently, when two atoms are far from each other, an arbitrary linear combination of those bound states serves as a good approximation to the states of the double-well system. As the atoms become
closer, the energy of the ground state becomes lower. This mecha-
nism of saving energy by sharing an electron between the two atoms
is responsible for molecular bonding, called *covalent*, which is typical
for the crystals we are discussing.

Of course, the spectrum of an electron in a real atom is more
complicated than in the delta-well model. However, the eigenfunc-
tions in the quantum Kepler problem also decay exponentially with
the distance to the nucleus (due to the factor $e^{-\lambda n^2 r/2}$, remember?)
Therefore, when the crystal is still in the state of vapor (i.e. all
$\approx 10^{23}$ atoms of it are far from each other), the spectrum of the
entire system of electrons is the same as for one atom, but each is
$10^{23}$-fold degenerate. Arbitrary linear combinations of the copies
of the same eigenfunction centered at different nuclei will do well.
When the atoms move closer to form a crystal, the sharing of the va-
 lent electrons (i.e. the electrons of the outer subshell) by the entire
crystal creates the economy of energy that bind the crystal together.

Yet, the inter-atomic distance $a$ cannot become arbitrarily small
— in part because of the electron degeneracy pressure! As a re-
sult, at the optimal distance between the atoms, the inner subshells
still remain degenerate, and their energy levels are affected by the
crystallization rather little.

Thus, our theory of energy bands applies first and foremost to the
energy range of valent electrons in the crystal. The above Kronig-
Penney model with periodic delta-well (rather than delta-wall) po-
tential provides a good reference point for discussing metals and semi-
conductors.

![Figure 25: Conduction and valence bands](image)

On Figure 25a, the conduction bands of two metals are schemat-
ically shown as $[E_0, E_1]$ and $[E'_0, E'_1]$. They are filled with electrons
up to their Fermi levels $E_F$ and $E'_F$. Denoted by $W$ and $W'$ are each metal’s “work function”. It measures the energy (in electron-volts eV) needed to expel an electron from the crystal, i.e. the difference between the Fermi level and the brim of the potential well in this model. When the two metals are put in contact, the electrons will tunnel from one to the other until the Fermi levels equalize as shown on Figure 25b. This results in the contact potential $(W' - W)/e$ between the two metals. This simple arrangement known as [thermocouple] can be used for measuring temperatures by exploiting the property of Fermi levels, and hence of the contact potential, to vary with temperature.

On Figure 25c, the valence and conduction bands of insulators or semiconductors are schematically illustrated. In reference to the Kronig-Penney model, the energy band $[E_0, E_1]$ should be interpreted as the range populated by valency electrons, and is shown to be entirely filled. The conduction energy band $[E_2, E_3]$ at 0 K temperature is totally vacant, but it is separated from the valence band by the energy gap $[E_1, E_2]$. Assuming that the densities of states at the top of the valence band and bottom of the conduction band are the same, one can argue that the Fermi level $E_F$ is located exactly as it is shown: at the midpoint between $E_1$ and $E_2$. Indeed, at temperatures $T$ slightly above 0K, the electrons populating the bottom levels of the conduction band are exactly the ones which are missing from the top levels of the valence band. In terms of Fermi-Dirac statistics this equality between the probabilities: of finding an electron at a state with energy $E_2$, and not finding an electron (i.e. finding a hole) at a state with energy $E_1$, is expressed by

$$
\frac{1}{e^{(E_2 - E_F)/kT} + 1} = 1 - \frac{1}{e^{(E_1 - E_F)/kT} + 1} = \frac{1}{1 + e^{(E_F - E_1)/kT}}.
$$

This implies $E_2 - E_F = E_F - E_1$. In the next section we will discuss how the shifting of the Fermi level closer to the valence or to the conduction band by introducing impurities into semiconductors has revolutionized electronics industry.

**The silicon revolution.** Silicon $Si$ is the 14th element of the periodic table. It is located in the column $IV$, and has electronic structure $1s^22s^22p^63s^23p^2$, i.e. the 1st and 2nd shells complete, and the 3rd half-full, featuring 4 valent electrons. In a crystal, each atom of silicon bonds with 4 neighbors, forming a structure referred to as tetrahedral. In fact the spatial configuration is not so easy to imagine, but we shall try.
On Figure 26a, a fragment of the 3D chess board is shown. Each black dot in the lattice is a vertex of 3 black cubes and 3 white cubes. Each black cube has 4 black vertices positioned in the same way (and 4 blank vertices). Each white cube also has 4 black and 4 blank vertices, but the blank vertices are positioned there the way black vertices are positioned in black cubes. Now, at the center of each black cube, put one more black dot, and connect it with the black vertices by the 4 bold bonds as Figure 26b shows. Altogether, the black dots are the atoms of the lattice. Each black dot on Figure 26a is also connected by 4 bonds: with the centers of those 4 black cubes it is a vertex of.

By sharing its four valent electrons with four neighbors, each silicon atom thereby completes its valence shell to capacity. Well, as we understand by now, this wording about completing valence shells is somewhat naive, because in fact the atoms of the entire crystal share all the valence electrons. But anyway, all the valent electrons fill up the entire valence band of the crystal, with the gap to the conduction band (as it turns out) of about 1.1 eV, making silicon a semiconductor.

Figures 27ab show what happens with the electron bands of silicon when it is doped with impurities of one of two types. When a small fraction (on the scale of from $10^{-10}$ to $10^{-5}$) of silicon atoms is replaced with boron $B$, an element from column III of the periodic table, having respectively 3 valent electrons instead of silicon’s 4, this shortage of valence electrons creates new energy levels about 0.05 eV above the conduction band. These levels, which are ready to borrow electrons from the conduction band, are called acceptor levels. In a symmetric situation when silicon is doped with an element from column V of the periodic table, such as arsenic $As$, the excess of valent electrons results in the new band of donor levels arising
about 0.05 eV below the conduction band of silicon. Since the gap between the donor and conduction levels is so narrow, the electrons can be easily excited (by thermal energy or applied voltage) into the conduction band. This is the \textit{n-type semiconductor}, where “n” comes from “negative”, the sign of the charge available to conduct electricity. In the former case, electrons from the valence band can be easily excited into the acceptor band. While these electrons can be effectively tied to the rare boron atoms, and not capable of conducting electricity, the \textit{holes} left behind (or rather beneath) in the valence band work effectively as positive charges capable of moving in the crystal. Such doping with the acceptor type impurity results in the \textit{p-type semiconductor}, where “p” comes from “positive”, the sign of the conducting charge.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig27}
\caption{p-n-junction}
\end{figure}

The presence of vacant acceptor levels or populated donor levels shifts the Fermi level of the crystal accordingly, as marked on the pictures. Figure 27c shows what happens when a plate of p-type semiconductor is put into contact with a plate of n-type semiconductor. Although initially each plate is electrically neutral, the electrons from the p-type region, which has a lower Fermi level, begin to diffuse into the n-type region, which has a higher Fermi level, until the Fermi levels of both regions equalize. In the thin (nanoscale) layer between the two regions, this separation of charges creates a strong electric field (on the scale of $10^8 V/m$) in the direction \textit{toward} the p-region (now negatively charged with the arrived electrons) and \textit{away} from the p-region (positively charged with the holes left behind) in a manner of a parallel capacitor. This nanoscale layer, freed of mobile electric charges, is called the \textit{depletion region}. 

128
Now comes the best part, though somewhat resembling the epi-
graph to this chapter. Two oppositely directed processes counter-
balance each other. Due to the heat energy, the electrons on the 
p-side can be excited into the conduction band and then flow to 
the positively charged n-side. This is the \textit{thermal current}, which 
can be denoted by $I_0$. It measures in micro-amperes (mA) at room 
temperatures. In the opposite direction, the unfortunate (valence?) 
electrons of the n-side can nevertheless get thermally excited high 
足够的 to scale the potential barrier of the depletion zone, and end 
up on the p-side — only to get recombined with the holes there 
(which apparently were left behind by the electrons migrated to the 
n-side?) Anyway, this is called the \textit{recombination current}. The ther-
mal current does not depend on the exterior voltage $\phi$ applied to the 
p-n-junction, but only on the size of the gap between the conduction 
and valence bands. On the contrary, the recombination current de-

dpends on $\phi$ which heightens the potential barrier across the depletion 
region when $\phi < 0$ (\textit{reverse biasing}) and lowers it when $\phi > 0$ (\textit{for-
ward biasing}). As a result, the net current $I$ from the p-side toward 
the n-side across the junction is given by the formula

$$I = I_0(e^{q\phi/kT} - 1),$$

where $q$ is an electron’s charge if $\phi$ is measured in electron-volts, eV. This shows that the current, small in no-bias or reverse-bias volt-
age, grows exponentially when the forward-biased voltage is applied. Thus the p-n-junction functions as a \textit{diode}.

The p-n-junctions are key components of \textcolor{blue}{transistors}, \textcolor{blue}{solar cells} and [light-emitting diodes] (LED).

**Korteweg – de Vries.** This is a nonlinear partial differential 
equation,

$$\frac{\partial u}{\partial t} + 6u \frac{\partial u}{\partial x} + \frac{\partial^3 u}{\partial x^3} = 0,$$

named after Diederik Korteweg and Gustav de Vries who studied it 
in a paper of 1895. The equation has a solution in the form of a 
 solitary traveling wave (\textit{soliton}) finding which is a useful exercise. So, we look for a solution in the form $u(x,t) = \phi(x - vt)$, and obtain

$$-v\phi' + 6\phi\phi' + \phi''' = 0,$$

or

$$\phi'' = v\phi - 3\phi^2 + \alpha,$$

where $\alpha$ is a constant. The last equation can be interpreted as the 
Newton equation for a particle on a line $\phi$ moving with “time” $x$ in a 
conservative force field with potential energy $U(\phi) = \phi^3 - v\phi^2/2 - \alpha\phi$. 

129
On Figure 28a, the potential energy with $\alpha = 0$ is shown, while Figure 28b shows the corresponding phase portrait. The highlighted phase curve, called *separatrix*, represents a solution $\phi(x)$ which tends to the unstable equilibrium as “time” $x$ tends to $\pm \infty$. Since the equilibrium is at $\phi = 0$, it tends to 0 as shown on Figure 28c. In fact this is the solution we need. It exists only for $\alpha = 0$, since for other values of $\alpha$ the unstable equilibrium shifts to $\phi \neq 0$.

The solution corresponds to the zero energy level, and can be explicitly found from the ODE $(\psi')^2 + 2\phi^3 - v\phi^2 = 0$. Namely, by separating the variables, we have

$$\int dx = \int \frac{d\phi}{y(\phi)},$$

where $y^2 = v\phi^2 - 2\phi^3$.

A useful trick (known in algebraic geometry as *blow-up*) for computing the integral on the left consists in parameterizing the separatrix by the slope $k = y/\phi$ of the radius-vector on the phase plane. From the equation of the separatrix, we express both $y$ and $\phi$ via $k$:

$$k^2 \phi^2 = v\phi^2 - 2\phi^3$$

leads to $\phi = \frac{v - k^2}{2},$ $y = k\frac{v - k^2}{2}$.

Next, we integrate (taking in account that $k^2 < v$):

$$\int dx = \int \frac{d\phi}{y} = -\int \frac{2dk}{v - k^2} = \frac{1}{\sqrt{v}} \log \frac{\sqrt{v} - k}{\sqrt{v} + k}.$$ 

After some elementary algebra we finally find

$$\phi(x) = \frac{v/2}{\cosh^2 \sqrt{vx}},$$

and hence $u(x, t) = \frac{v/2}{\cosh^2 \sqrt{v(x - vt)}}$.

The last formula represents a wave traveling to the “right” with the speed $v$ depending on the magnitude $v/2$ of the wave.

The actual story behind the soliton goes back to August 1834, when a bright young mathematician and engineer [John Scott Russell](https://en.wikipedia.org/wiki/John_Scott_Russell) observed such a wave, formed inside a shallow and narrow canal of Edinburgh, following it on a horse for two miles.

In 1965, Norman Zabusky and Martin Kruskal, conducting a numerical experiment motivated by problems of plasma physics, discovered that contrary to their expectation, two colliding KdV solitons pass through each other largely unaffected by their interaction. This behavior suggested that the KdV equation, considered as a dynamical system in a space of functions, has hidden conservation laws.
number of theoretical insights due to many authors followed, resulting in a remarkably deep theory, which illuminates many properties of Schrödinger operators in one degree of freedom.

According to the invention by Peter Lax, the KdV equation for $u(x, t)$ is equivalent to the Lax equation $\dot{L} = [L, A]$ for the following Lax pair of differential operators:

$$ L := -\frac{d^2}{dx^2} - u, \quad A := 4 \frac{d^3}{dx^3} + 3 \left( u \frac{d}{dx} + \frac{d}{dx} u \right). $$

Indeed: for brevity, let us denote $d/dx$ by $\partial$, and use the commutation relation $\partial u = u' + u \partial$ (where $u'$ means the multiplication operator by the $x$-derivative of $u$) in the following computation:

$$ [L, A] = -(\partial^2 + u)(4\partial^3 + 3u\partial + 3\partial u) + (4\partial^3 + 3u\partial + 3\partial u)(\partial^2 + u) $$
$$ = -4\partial^5 - 4u\partial^4 - 3\partial^2 u\partial - 3\partial^3 u - 3u\partial u - 3u^2 \partial $$
$$ + 4\partial^5 + 3u\partial^3 + 3\partial u\partial^2 + 4\partial^3 u + 3u\partial u + 3\partial u^2 $$
$$ = -u\partial^3 - 3\partial u' \partial + \partial^3 u + 6uu'. $$

Since $\partial^3 u = u\partial^3 + 3(\partial u' \partial) + u''$, and since $\dot{L} := dL/dt = -\dot{u}$, we find that $\dot{L} = [L, A]$ is equivalent to $-\dot{u} = u''' + 6uu'$ as promised.
Note that $L$ can be considered as the one-dimensional Schrödinger operator with the potential energy function $V_t(x) := -u(x,t)$, depending on the parameter $t$, and written in such units that the Planck constant $\hbar = 1$. On the other hand, the operator $-i\partial$ of momentum (in these units) is self-adjoint. So is the operator $iA = 4(i\partial)^3 + 3(i\partial u + u\partial i)$, i.e. $A^* = -A$. Therefore, the family $U(t)$ of operators satisfying $dU/dt = AU$ and $U(0) = I$ is unitary. Indeed,

$$d(U^*U)/dt = (U^*A^*)U + U^*(AU) = U^*(A^* + A)U = 0,$$

i.e. $U^*(t)U(t) = U^*(0)U(0) = I$.

By the way, $U(t) \neq e^{tA}$ because $A$ itself depends on $t$, and the operators $A(t)$ with different $t$ have no reason to commute. Nevertheless, $L(t) = U^*(t)L(0)U(t)$ satisfies the Lax equation $\dot{L} = [L,A]$. This shows that all the Schrödinger operators $L(t) := -d^2/dx^2 - u(x,t)$ are obtained from each other by unitary transformations in the Hilbert space of psi-functions. This can be summarized as follows:

**When the Schrödinger potential $V_t(x)$ evolves with $t$ in such a way that $u(x,t) := -V_t(x)$ satisfies the KdV equation $\dot{u} + 6uu' + u''' = 0$, the spectrum of the Schrödinger operator $L(t) := -d^2/dx^2 + V_t(x)$ remains unchanged. In particular, the eigenvalues of the discrete spectrum of $L$ are conservation laws of the KdV equation.**

Furthermore, consider the KdV equation on a circle, i.e. assume $u$ periodic: $u(x + a,t) = u(x,t)$. In the class of periodic potentials, $V(x + a) = V(x)$, the spectrum $E_0 < E_1 \leq E_2 < E_3 \leq \cdots$ of the boundaries of the energy bands, as it turns out, forms a complete set of conservation laws of the KdV equation. That is, when an $a$-periodic function $u(x) = -V(x)$ evolves in time according to the KdV equation $\dot{u} = -uu' - u'''$, the energy bands remain unchanged, and moreover (according to Sergei Novikov and his collaborators) the entire evolution can be described by simple formulas in terms of further spectral data associated with the potentials.

In classical mechanics, the phenomenon of having a complete set of conservation laws sufficient in order to entirely describe the dynamics, is known as **complete integrability**. Thus, KdV turns out to be a completely integrable system with infinitely many degrees of freedom. This discovery triggered an avalanche of subsequent developments in the theory of completely integrable systems, and at the turn of the millennium, was voted to be one of five most influential mathematical discoveries of the 20th century.
13 Summation over histories

Fermat’s least time principle. We return now to the starting point of our exposition: the properties of light waves and ray systems, in order to address from our current, hopefully more advanced point of view some aspects of classical mechanics and short-wave optics that we tactfully overlooked at the first try, and in order to extract from them yet another fundamental insight about the nature of the quantum world.

In a classical geometry problem, James Bond, on his way from point A to point B (Figure 29a) has to meet an assistant driving on a straight road l. The problem asks to find a point C on the road where they should meet so that the walking distance |AC| + |CB| would be shortest. The solution, known to ancient Greeks, is shown on the figure: The path |AC| + |CB'| to the mirror-reflected point B' has the same length as |AC| + |CB|, and is shortest when ACB' is straight.

![Diagram](image)

Figure 29: Ferma’s least time principle

It was Pierre Fermat who saw in this example a general principle governing the behavior of light rays. Just like Mr. Bond in the problem, light particles minimize their travel time. This rule justifies the law of reflection, as well as the law of refraction of light at the border between two media where the speed of light changes.

The problem of how light particles learn beforehand in which direction the travel time would become the least (one suggestion was that they send forward some scout particles to explore) was resolved by Huygen’s principle. Instead of the scouts, each point of a wave front emits a spherical wavelet, with the ultimate effect that the rays perpendicular to the wave fronts minimize their travel time. The way the least time principle works in deriving the refraction law is illustrated on Figure 29b, where three consecutive wave fronts are shown.
Two lessons are to be learned from this discussion. In the chapter on short-wave optics, we described the wave field at an observation point (such as $B$ on Figure 29a) as the superposition of the wavelets arriving from each point of the initial wave front. But this was a convenient simplification, wasn’t it? Any of the intermediate wave fronts can be considered as the initial one. The wave field at a point of an intermediate wave front (say, at $t = 1$ on Figure 29b) is obtained as the superposition of wavelets arriving from every point of the front at $t = 0$. In its turn, this point emits a wavelet (of the amplitude and phase resulted from that superposition), and the superposition of these wavelets determines the wave field on the front at $t = 2$. Well, this two-step description is a simplification too. We could proceed by chopping the time interval into more segments of shorter duration, and describe the wave field at a given observation point as the superposition of the wavelets delivered to this point from the source in a many-step emission process. In fact — and that’s the first lesson — all conceivable trajectories of the light particles should contribute to the superposition, each bringing a wave of some phase and amplitude, depending on the optical distance along the trajectory.

Now, Fermat’s principle stipulates that in geometrical optics, i.e. in the limit when the wave length tends to 0, the contributions of all the conceivable trajectories cancel out except those which are critical points of the optical distance function. That is, they are not just shortest among all straight rays connecting the observation point to the initial front, but are time-minimizers among all conceivable trajectories from the source. So, the second lesson is that not only geometrical optics, but conservative classical mechanics in general should admit such a formulation, where trajectories are critical points of a certain function (the traditional term is functional) on the space of all conceivable trajectories. This also poses a mathematical problem we consider next: the problem of finding critical points of such functionals.

**Calculus of variations.** To set up a variational problem, one introduces a Lagrangian, i.e. a function $L(q, \dot{q}, t)$ of $2n + 1$ independent variables. Here $q$ denotes a point in the configuration space of a classical mechanical system with $n$ degrees of freedom, and $\dot{q}$ stands for another set of $n$ variables, which, geometrically speaking, are to specify a tangent vector to the configuration space at the point $q$. Given a differentiable trajectory $\gamma : t \mapsto q(t)$ in the configuration space, one can compute its velocity $dq(t)/dt$, substitute both into
the Lagrangian and integrate over a time interval $[a, b]$:

$$
\mathcal{F}(\gamma) := \int_{a}^{b} L \left( q(t), \frac{dq}{dt}(t), t \right) dt.
$$

The result is a number $\mathcal{F}(\gamma)$ which depends on the choice of the trajectory $\gamma$. Thus, we obtain a functional on the space of all differentiable trajectories in the configuration space defined on the time interval $[a, b]$.

The problem of Calculus of Variations is to characterize the trajectories $\gamma_0$, which minimize the functional, or more generally, are critical points of $\mathcal{F}$. The last condition means that all directional derivatives of $\mathcal{F}$ at the point $\gamma_0$ vanish, i.e.

$$
\frac{d}{d\epsilon} \bigg|_{\epsilon=0} \mathcal{F}(\gamma_\epsilon) = 0
$$

for every variation $\gamma_\epsilon$ of $\gamma_0$. As we will see soon, the problem is rather local with respect to the configuration space, which therefore can be considered to be a domain in $\mathbb{R}^n$. Thus, we take $\gamma_\epsilon : t \mapsto q(t) + \epsilon \delta(t)$, where $\delta(t)$ is an arbitrary parametric curve in $\mathbb{R}^n$, and compute the directional derivative (suppressing the possible explicit dependence of the Lagrangian on $t$, as it plays no role in the computation):

$$
\frac{d}{d\epsilon} \bigg|_{\epsilon=0} \mathcal{F}(q + \epsilon \delta) = \frac{d}{d\epsilon} \bigg|_{\epsilon=0} \int_{a}^{b} L \left( q + \epsilon \delta, \frac{dq}{dt} + \epsilon \frac{d\delta}{dt} \right) dt = \\
\int_{a}^{b} \left[ \frac{\partial L}{\partial q} \left( q, \frac{dq}{dt} \right) \cdot \delta(t) + \frac{\partial L}{\partial \dot{q}} \left( q, \frac{dq}{dt} \right) \cdot \frac{d\delta}{dt}(t) \right] dt = \\
\frac{\partial L}{\partial \dot{q}} \left( q, \frac{dq}{dt} \right) \cdot \delta(t) \bigg|_{a}^{b} - \int_{a}^{b} \left[ \frac{d}{dt} \frac{\partial L}{\partial \dot{q}} \left( q, \frac{dq}{dt} \right) - \frac{\partial L}{\partial q} \left( q, \frac{dq}{dt} \right) \right] \cdot \delta(t) dt.
$$

Here we first use the chain rule to differentiate the integrand with respect to $\epsilon$, and then apply integration by parts.

The conclusion to be derived from this result is that, even if we restrict ourselves to the class of trajectories with fixed endpoints, and hence to variations $\delta$ with $\delta(a) = \delta(b) = 0$ (so that the finite term vanishes), the only way how the directional derivatives of $\mathcal{F}$ in the directions of all such $\delta$ can vanish, is that the following Euler-Lagrange equation holds true:

$$
\frac{d}{dt} \frac{\partial L}{\partial \dot{q}} \left( q, \frac{dq}{dt}, t \right) = \frac{\partial L}{\partial q} \left( q, \frac{dq}{dt}, t \right).
$$
To appreciate what this equation entails, consider the example

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

where the Lagrangian is the difference of the kinetic and potential energies. We have: $\partial L/\partial q = -\partial V/\partial q$, $\partial L/\partial \dot{q} = m \dot{q}$, and the Euler-Lagrange equations assumes the form $m d^2 q/dt^2 = -\partial V(q)/\partial q$. Thus, the critical points of the functional are the trajectories of a Newtonian particle of mass $m$ in the force field with the potential energy $V$.

**Lagrangian mechanics.** In Lagrange’s formulation, classical mechanics is identical to calculus of variations. Generalizing the last example, one defines a *Lagrangian mechanical system* by specifying a Lagrange function $L$, and postulating that trajectories are solutions to the Euler-Lagrange equation, i.e. are critical points of the corresponding functional $F$.

The Lagrangian $L$ needs to satisfy some non-degeneracy condition to guarantee that the Euler-Lagrange equation is a well-posed system of 2nd order ordinary differential equations. Namely, when the differentiation $d/dt$ on the left of the equation is performed explicitly, the system takes on the form

$$\sum_j \frac{\partial^2 L}{\partial \dot{q}_i \dot{q}_j} \ddot{q}_j + \sum_j \frac{\partial^2 L}{\partial \dot{q}_i \dot{q}_j} \dot{q}_j + \frac{\partial^2 L}{\partial q_i \partial t} = \frac{\partial L}{\partial \dot{q}_i}, \quad i = 1, \ldots, n.$$

It needs to be resolved with respect to $\ddot{q}_j$. For this, the $n \times n$-matrix $[\partial^2 L/\partial q_i \partial q_j]$ needs to be invertible.

Under this condition, the Euler-Lagrange equation can be rewritten, at least locally, in the Hamiltonian form. Namely, one introduces the Hamilton function $H(p, q, t)$ as the Legendre transform of the Lagrange function $L(q, \dot{q}, t)$ with respect to velocity $\dot{q}$:

$$H(p, q, t) := \text{critical value in } \dot{q} \text{ of } p \cdot \dot{q} - L(q, \dot{q}, t).$$

To find the critical point (where the critical value is to be taken), we differentiate in $\dot{q}_i$ and obtain

$$p_i = \frac{\partial L}{\partial \dot{q}_i}, \quad i = 1, \ldots, n.$$

These are the *generalized momenta* (which we’ve mentioned twice but never defined). In geometric terms, if $\dot{q}$ is a tangent vector to the
configuration space at a point $q$, the momentum $p$ is the differential of $L$ with respect to $\dot{q}$, i.e. a tangent covector, a linear function $\sum p_i \dot{q}_i$ of the tangent vector $\dot{q}$.

In any case, by virtue of the Implicit Function Theorem, the non-degeneracy condition $\det[\partial^2 L/\partial \dot{q}_i \partial \dot{q}_j] \neq 0$ guarantees that the critical point can be expressed at least locally via $p$ as $\dot{q} = \dot{Q}(p, q, t)$:

$$\frac{\partial L}{\partial \dot{q}}(q, \dot{Q}(p, q, t), t) \equiv p, \quad H(p, q, t) = p \cdot \dot{Q}(p, q, t) - L(q, \dot{Q}(p, q, t), t).$$

Now, from the construction of $H$ as $(p \cdot \dot{q} - L)|_{\dot{q} = \dot{Q}}$, the Euler-Lagrange equation $\partial L/\partial q = d/dt (\partial L/\partial \dot{q})$ (where $\dot{q}$ is eventually taken to be $dq/dt$), the chain rule, and the critical point equation $p - \partial L/\partial \dot{q} = 0$, we find:

$$\frac{\partial H}{\partial p} = \frac{dq}{dt} + \left(p - \frac{\partial L}{\partial \dot{q}}\right) \cdot \frac{\partial \dot{Q}}{\partial p} = \frac{dq}{dt},$$  

$$\frac{\partial H}{\partial q} = \left(p - \frac{\partial L}{\partial \dot{q}}\right) \cdot \frac{\partial \dot{Q}}{\partial q} - \frac{\partial L}{\partial q} = -\frac{d}{dt} \frac{\partial L}{\partial \dot{q}} = -\frac{dp}{dt}.$$

This is the system of Hamilton equations with the hamiltonian $H$.

The Lagrangian approach to mechanics, where the role of $q$ and $p$ is asymmetric, seems less flexible than the Hamiltonian one. Nevertheless one can go back from the Hamiltonian to Lagrangian formulation, at least locally in the phase space. Namely, given $H(p, q, t)$ define $L(q, \dot{q}, t)$ by the Legendre transform of $H$ with respect to $p$:

$$L(q, \dot{q}, t) = \text{critical value in } p \text{ of } p \cdot \dot{q} - H(p, q, t).$$

To find the critical value, one needs to resolve the equation $\dot{q} = \partial H/\partial p$ with respect to $p$, i.e. express $p$ as a function $P(q, \dot{q}, t)$, and then compute the Lagrangian: $L(q, \dot{q}, t) = P \cdot \dot{q} - H(P, q, t)$.

**The least action principle.** Figuring out why (and under what non-degeneracy conditions) such a repetition of the Legendre transform yields the original Lagrange function is a useful exercise, which we leave to the reader. For us, the important consequence of this construction is that the Hamilton equations can be interpreted as the equations of critical points of a certain functional. Namely

$$\mathcal{F}(\gamma) = \int_a^b \left( p(t) \cdot \frac{dq(t)}{dt} - H(q(t), p(t), t) \right) dt = \int_a^b (pdq - Hdt).$$
In fact this functional, called *action*, has two interpretations. In one of them, equivalent to the Lagrangian approach, it is assumed that the functional is defined on the space of paths \( t \mapsto q(t) \) in the *configuration space* of the mechanical system, which are lifted to the phase space by taking

\[
p(t) := P\left(q(t), \frac{dq}{dt}(t), t\right), \quad \text{such that} \quad \frac{\partial H}{\partial p}(P(q, \dot{q}, t), q, t) \equiv \dot{q}.
\]

In the other, one considers parametric curves \( t \mapsto (p(t), q(t)) \) in the *phase space*, i.e. assumes that \( p(t) \) and \( q(t) \) are independent functions. The recipe of the calculus of variations for finding the critical trajectories leads us to considering the degenerate “Lagrangian” \( \mathcal{L} = p \cdot \dot{q} - H(p, q, t) \) (it does not involve \( \dot{p} \)). We arrive at the “Euler-Lagrange equation” \( 0 = \frac{\partial \mathcal{L}}{\partial p}, \quad \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}} = \frac{\partial \mathcal{L}}{\partial q} \). It coincides, perhaps unexpectedly, with the Hamilton system:

\[
0 = \frac{dq}{dt} - \frac{\partial H}{\partial p}, \quad \frac{dp}{dt} = -\frac{\partial H}{\partial q}.
\]

Thus, *Hamilton equations describe critical points of the action functional defined on the space of differentiable paths in the phase space. This is the Hamiltonian form of the least action principle.*

**Feynman’s path integrals.** One of the creators of quantum electrodynamics \[Freeman Dyson\] is famously quoted to say in 1979 about another one, \[Richard Feynman\]:

Thirty-one years ago, Dick Feynman told me about his ‘sum over histories’ version of quantum mechanics. “The electron does anything it likes,” he said. “It just goes in any direction, at any speed, forward and backward in time, however it likes, and then you add up the amplitudes and it gives you the wave function.” I said to him, “You’re crazy.” But he wasn’t.

The idea of Feynman, that Dyson found crazy at first, consists in extending to general Hamiltonian systems the observation we’ve made in connection with Fermat’s least time principle of light propagation. The same way as the wave field at a given point is obtained as the superposition of the wavelets delivered by all conceivable trajectories of the light particle, the complex amplitude \( \Psi \) of a quantum event (e.g. the amplitude \( \Psi(q) \) of finding the quantum system in a
particular configuration \( q \), according to Feynman, is the superposition of such amplitudes contributed by all conceivable paths \( \gamma \) in the phase space.

The amplitude contributed by the given path is equal to \( e^{i\mathcal{F}(\gamma)/\hbar} \), where the \( \mathcal{F}(\gamma) \) is the value of the action functional on this path, and \( \hbar \) is the Planck constant. Thus, the superposition assumes the form of the oscillating integral

\[
\Psi = \int e^{i\mathcal{F}(\gamma)/\hbar} d\gamma
\]

over the space of all contributing paths \( \gamma \) — Feynman’s path integral.

The idea of representing quantum amplitudes by path integrals has three remarkable features.

Firstly, it provides a universal quantization procedure of all classical Hamiltonian systems.

Secondly, it explains how our macroscopic worldview emerges in the classical limit \( \hbar \to 0 \). Namely, as we have seen in our study of oscillating integrals, in the limit \( \hbar = 0 \), contributions of the most trajectories cancel out. The only remaining contributions come from the critical points of the action functional (playing the role of the phase function), which are the classical trajectories of the Hamiltonian system.

Thirdly, path integrals don’t exist. It should be clear from dimensional considerations that infinite dimensional integrals cannot be defined; at least not in any intuitively familiar way. While a \( D \)-dimensional volume is measured in “\( D \)-dimensional feet” (square feet, cubic feet, etc.), when \( D = \infty \), the volume of the unit “cube” re-expressed in the units based on inches will be infinite, and in the units based on yards zero.

Consequently, the “summation over histories” approach to quantum mechanics plays an important role as a universal heuristic idea, but in every particular application it needs to be made sense of by some \textit{ad hoc} methods.

One of such methods, the \textit{theory of perturbations}, consists in replacing the integral by its asymptotical expansion in the powers of \( \hbar \) near a critical point of the phase function, and learning how to compute the coefficients of the series. A remarkable combinatorial answer to the last question is our next (and last) subject.
Wick’s theorem. The mathematical problem we are going to consider here is essentially the same as the one we studied in the section on short-wave asymptotics. The differences are that we have $\hbar$ in place of the wave length $\lambda$, want to compute the entire asymptotical expansion of the oscillating integral (rather than its principal term), and no longer assume the integral finite dimensional. Yet, the last distinction makes no difference in the combinatorics of the asymptotical expansion other than just the fact that the names of the variables in the phase function $F$ come from an infinite list.

Thus, we assume in our expressions that the variables are $x_k$, $k = 1, \ldots, D$, where $D$ could be $\infty$. So, the integral we are talking about has the form

$$\Psi = \int e^{iF(x_1, x_2, \ldots)/\hbar} dx_1 dx_2 \ldots$$

We assume that the phase function has the general form

$$F(x) = F(0) + \frac{1}{2} \langle x | A | x \rangle + \sum_a t_a x^a / a!,$$

where in $\langle x | A | x \rangle$ we use physicists’ notation to describe a quadratic form, $\sum_{kl} a_{kl} x_k x_l$, which we consider non-degenerate. Also, we use the shorthand $x^a / a!$ for arbitrary (finite degree) monomials:

$$x^a / a! := x_1^{a_1} x_2^{a_2} \ldots,$$

and denote by $t_a$ arbitrary independent formal parameters. That is, the function $F$, in addition to the integration variables $x$, also depends on infinitely many parameters $t_a$, and at $t = 0$ is a non-degenerate quadratic form with the critical point $x = 0$.

It should be clear from the analysis we’ve done in the section on short-wave asymptotics, that as $\hbar \to 0$ the integral behaves as

$$\Psi \sim \hbar^{D/2} e^{F(x_{\text{crit}})/\hbar} (A + B \hbar + C \hbar^2 + \ldots), \quad \text{where} \ A \neq 0.$$ 

This makes little sense when $D = \infty$. So, one takes the logarithm of $\Psi$, omits the ill-defined constant $D \log \sqrt{\hbar}$, and studies the $\hbar$-series expansion

$$\log \Psi \sim \hbar^{-1} F(x_{\text{crit}}) + \alpha + \beta \hbar + \gamma \hbar^2 + \ldots$$
This maybe the right moment to mention the analogy between path integrals and statistical sums. Per our experience with $Z$, all meaningful quantities of statistical mechanics were expressible as logarithmic derivatives of $Z$ with respect to various parameters. Likewise, a path integral $\Psi$, which can be viewed as “statistical sum”, albeit continual, and with $-i\hbar$ in place of $kT$, computes the complex amplitude as a function of observable parameters: energies of states, coordinates of configurations, etc. The probabilities predicted by the quantum theory are similarly expressible through the derivatives $\partial(\log \Psi)/\partial t_a = \Psi^{-1}\partial\Psi/\partial t_a$. So, the omitted ill-defined constant hardly matters.

Our first goal is to recast the multi-variable integration as power series differentiation. Namely, we claim that

$$\int e^{\frac{i\hbar}{2} \left( \langle x|A|x \rangle \right)} dx = \text{Const} \times e^{\frac{i\hbar}{2} \left( \sum_a \frac{\partial}{\partial x} \left( B_{\partial x} f \right) \right)} \bigg|_{x=0}.$$ 

To justify the claim, we need to exploit some properties of Fourier transforms of multi-variable functions $f : \mathbb{R}^D \to \mathbb{C}$. In presently convenient notation, the transform and its inverse are given by

$$\hat{f}(p) = \int_{\mathbb{R}^D} e^{-i(p \cdot x)/\hbar} f(x) dx, \quad f(x) = \frac{1}{(2\pi\hbar)^D} \int_{\mathbb{R}^D} e^{i(p \cdot x)/\hbar} \hat{f}(p) dp.$$ 

Differentiation of $f$ in $x_k$ translates into multiplication of $\hat{f}$ by $p_k$. More precisely, integrating by parts, we obtain

$$\int e^{-i(p \cdot x)/\hbar} \frac{\hbar}{i} \frac{\partial}{\partial x_k} f(x) dx = p_k \int e^{-i(p \cdot x)/\hbar} f(x) dx = p_k \hat{f}(p).$$

The finite terms were omitted not so much because we assume that $f$ tends to 0 at infinity, but because we intend to apply our computation to the asymptotics of the integrals as $\hbar \to 0$, where only the neighborhood of the critical point matters.

Let $\langle p|B|p \rangle$ denote the quadratic form $\sum_{kl} b_{kl} p_k p_l$ defined by a non-degenerate symmetric matrix $B$. By the previous rule, we have

$$e^{-i\langle p|B|p \rangle/2\hbar} \hat{f}(p) = \int e^{-i(p \cdot x)/\hbar} e^{-i\langle \frac{\hbar}{i} \frac{\partial}{\partial x} | B \frac{\hbar}{i} \frac{\partial}{\partial x} \rangle} f(x) dx.$$ 

From the formula for the inverse Fourier transform, it is clear that $f(0) = (2\pi\hbar)^{-D} \int \hat{f}(p) dp$. Therefore

$$\left[ e^{\frac{i\hbar}{2} \left( \frac{\partial}{\partial x} | B \frac{\partial}{\partial x} \right)} \right]_{x=0} = \frac{1}{(2\pi\hbar)^D} \int e^{-i\langle p|B|p \rangle/2\hbar} dp \int e^{-i(p \cdot x)/\hbar} f(x) dx.$$ 

141
Changing the order of integration on the right, we want to perform first the Gaussian (or rather Fresnelian) integral
\[
\frac{1}{(2\pi \hbar)^D} \int e^{-i (p \cdot x + \frac{1}{2} \langle p | B | p \rangle / \hbar)} dp.
\]

The function in the exponent becomes a (homogeneous) quadratic form in \( p \) after shifting the origin to the critical point. Note that in computing the critical value in the exponent, we are effectively performing the Legendre transform of the quadratic form \(-\langle p | B | p \rangle / 2\).

The critical point is found from \( x + Bp = 0 \), i.e. \( p = -B^{-1}x \). The critical value equals
\[
-(B^{-1} x) \cdot x + \frac{1}{2} \langle B^{-1} x | B | B^{-1} x \rangle = -\frac{1}{2} \langle x | B^{-1} | x \rangle.
\]

The Gaussian integral \textit{per se} yields a constant, \((-2\pi i \hbar)^{D/2} / \sqrt{\det B}\) (not relevant inasmuch as derivatives of \( \log \Psi \) are concerned). Thus
\[
\left[ e^{\frac{i \hbar}{2} \left\langle \frac{\partial}{\partial x} | B | \frac{\partial}{\partial x} \right\rangle f(x) \right]_{x=0} = \frac{1}{\sqrt{\det(2\pi i \hbar B)}} \int e^{i \langle x | B^{-1} | x \rangle / 2\hbar} f(x) dx.
\]

From here, we obtain the promised formula for \( \Psi \) by taking \( B^{-1} = A \) and \( f(x) = e^{\hbar^{-1} \sum_a t_a x^a / a!} \).

Multi-variable integration is reduced thereby to series differentiation at the origin, a task that can be treated as purely combinatorial. The outcome is a formula, known as \textit{Wick’s theorem}, which, loosely speaking, expresses the asymptotics of an oscillating integral via \textit{summation over all graphs}. A \textit{graph} is an entirely combinatorial gadget, consisting of finitely many \textit{vertices} connected by finitely many \textit{edges}. It is not hard to see why graphs come out in our situation. We intend to apply the infinite order differential operator to a function. The operator has the form
\[
\exp \left( \frac{i \hbar}{2} \left\langle \frac{\partial}{\partial x} | B | \frac{\partial}{\partial x} \right\rangle \right) = \sum_{N=0}^{\infty} \frac{1}{N!} \left( \frac{i \hbar}{2} \sum_{kl} b_{kl} \frac{\partial^2}{\partial x_k \partial x_l} \right)^N.
\]

The function it is to be applied to has the form
\[
\exp \left( \frac{i}{\hbar} \sum_a t_a x^a / a! \right), \quad \text{where} \quad \frac{x^a}{a!} = \prod_k \frac{x_k^{a_k}}{a_k!}.
\]
A first order differentiation $\partial/\partial x_k$ of this function brings down in front of the exponential expression a factor from every monomial with $a_k > 0$. Each subsequent differentiation can either bring down a factors from the exponent, or be applied to the mess that has already accumulated in front of the exponential expression.

One good news is that we intend to put all $x_k = 0$ at the end. So all the factors which were not entirely reduced to constants by subsequent differentiations will disappear from the answer (as well as the very exponential expression, which will turn into $\exp\left(\frac{it\alpha}{\hbar}\right)$).

Another good news is that the same way as an edge in a graph has two ends attached to two vertices (which may happen to be the same vertex though), each of the bi-differentiations $\partial^2/\partial x_k \partial x_l$ is to be applied to two monomials from the exponent (or perhaps to the same one, twice). Consequently, the entire mess of the terms still surviving after plugging $x = 0$ can be described as the sum of contributions expressible graphically as follows.

Take a graph, and decorate each edge of it with a bi-differentiation $i\hbar b_{kl} \frac{\partial^2}{\partial x_k \partial x_l}$ (deciding also which end of the edge differentiates in $x_k$ and which in $x_l$). Decorate each vertex with a monomial $t_a x^a/a!$ making sure that the list $a = (\ldots, a_k, \ldots)$ matches exactly the ends of the edges adjacent to that vertex. This is to mean that the number of derivations $\partial/\partial x_k$, brought to this vertex by the adjacent edges, equals $a_k$ for each $k$, thereby reducing the monomial $t_a x^a/a!$ to $t_a$, i.e. constant but not identically zero value.

After this informal outline, our aim is to accurately describe the graph summation formula. Let $\Gamma$ be a connected graph, $V(\Gamma)$ the set of its vertices, $E(\Gamma)$ the set of its edges. Denote by $|\text{Sym}(\Gamma)|$ the number of symmetries of the graph, i.e. the number of permutations of edges and/or vertices which preserve the graph. (Such symmetries form a group denoted here $\text{Sym}(\Gamma)$.) For example, the graph on Figure 30a with 5 vertices and 20 edges is very symmetric. It has 10 symmetries of the pentagon (5 reflections and 5 rotations, including the identity rotation), $3!$ permutations of each triple of the edges, 2 flippings (including the “non-flipping”) of each of the 5 loops attached to the vertices, and any compositions thereof. So, totally $|\text{Sym}(\Gamma)| = 10 \times (3!)^5 \times 2^5$.

Decorate a given graph $\Gamma$ by assigning to each edge $e$ the indices $k(e), l(e)$ of the derivations $\partial/\partial x_k$ and $\partial/\partial x_l$ at both ends of the edge, as shown in the example of Figure 30b. For each vertex $v$ of the graph, a decoration defines the string $a(v) = (\ldots, a_k, \ldots)$ indicating the number $a_k$ of times the index $k$ decorates an edge adjacent to $v$. 

143
To a decoration $\Delta$ of a graph $\Gamma$, assign the *Feynman amplitude*

$$W(\Gamma, \Delta) = \prod_{e \in E(\Gamma)} b_{k(e)}l(e) \prod_{v \in V(\Gamma)} t_{a(v)}.$$ 

For instance, for the graph on Figure 30b (where $D = 4$),

$$W(\Gamma, \Delta) = b_{12}^2 b_{23}^2 b_{13} b_{44} t_{2,1,1,0} t_{1,0,1,1} t_{0,3,0,1} t_{0,0,1,0}.$$

In terms of Feynman amplitudes, *Wick’s theorem* says:

$$\log \left( \frac{1}{\sqrt{\det(2\pi i\hbar A^{-1})}} \right) \int e^{\frac{i}{\hbar} \left( \frac{\langle x|A|x \rangle}{2} + \sum_a t_a x_a^a \right) d.x} \sim \sum_{\text{connected graphs } \Gamma} \frac{(i\hbar)^{|E(\Gamma)|}(i/\hbar)^{|V(\Gamma)|}}{|\text{Sym}(\Gamma)|} \sum_{\text{decorations } \Delta \text{ of } \Gamma} W(\Gamma, \Delta).$$

Here, as before, $B = (b_{kl})$ is the symmetric matrix inverse to $A$.

Note that the exponent in the $\hbar$-factor $h^{|E(\Gamma)|-|V(\Gamma)|}$ equals the negative *Euler characteristic* of the graph. For connected graphs, it takes values $\geq -1$.

Since $\hbar$ is small, the principal term is $\hbar^{-1}$. It consists of the amplitudes of *trees*. Truncating the rest of the $\hbar$-series is often referred to as the *tree-level approximation*. Comparing with our study of short-wave asymptotics, we conclude that this approximation yields the *critical value of the phase function*. The critical point here is perturbed away from the origin by the linear terms $t_{(1,0,...)}x_1$, $t_{(0,1,0,...)}x_2$, ... of the phase function.
The next approximation includes all terms of order $\hbar^0$, represented by graphs with one cycle, and is often nicknamed the one-loop approximation.

Imagine now that we exponentiate the right side of Wick’s formula to obtain the expansion of the integral on the left (rather than its logarithm). In combinatorics, if a sum $X + Y + Z + \cdots$ counts the numbers (or contributions) of some “connected objects”, then $e^{X+Y+Z+\cdots}$ counts all contributions of all (disconnected) collections of these objects, divided by the numbers of symmetries of such collections. For instance, $X^3Y^2Z$ occurs in the series $e^{X+Y+Z+\cdots}$ with the weight $1/3!2!1!$ indicating that the collection $X, X, X, Y, Y, Z$ is symmetric to $3!2!1!$ permutations of the identical items. From this point of view, it should be clear that the exponentiating the right hand side of Wick’s formula, we obtain the sum of Feynman amplitudes over all decorated graphs, not necessarily connected, with the weights $1/|\text{Sym}(\Gamma)|$ taking in account both the internal symmetries of the graph’s components and permutations of the identical components as well.

After this remark, the proof of Wick’s formula should become rather obvious. Indeed, the summands $t_ax^a$ in the exponent of the phase function represent all available contributions of decorated vertices, with the factor $1/a!$ accounting for the symmetries of each decoration. The summands $b_{kl}\partial^2/\partial x_k \partial x_l$ in the exponent of the differential operator represent all available contributions of decorated edges (while the weight $1/2$ accounts for the symmetry $k \leftrightarrow l$ of the edge). Exponentiating each sum one obtains all possible, correctly weighted, collections of decorated vertices on the one side, and of decorated edges on the other. By applying the operator to the function and setting $x = 0$, one glues the collections of vertices and edges into graphs according to the decorations. The ultimate formula is obtained by re-summing the result by undecorated graphs $\Gamma$. The factors $1/|\text{Sym}(\Gamma)|$ are to make sure that in the process of re-summing each gluing is counted only once.

The above graph summation formula is a combinatorial foundation of perturbation theory, one method of making sense of Feynman’s path integrals in quantum mechanics and quantum field theory. Note that the number $D$ of the variables of integration does not occur explicitly on the right hand side of the formula for $\log \Psi$, and is only lurking there by limiting the supply of available decorations. Should the number be infinite, the contribution $W(\Gamma, \Delta)$ of each decorated graph would still make sense.
Moreover, referring to our brief glimpse into quantum field theory, recall that it starts with a “free field” model. It is a linear mechanical system with infinitely many degrees of freedom. Namely it is an ideal gas of infinitely many harmonic oscillators — the Fourier modes of a classical field near the zero (“vacuum”) state. The Lagrangian action functional of such a system is a quadratic form, $\langle x|A|x \rangle$, though in infinitely many variables. Bringing into the model some interactions results in adding to the action functional new terms like $t_a x^a$. Interpreting the path integral via Wick’s formula, one can express various expectation values and correlation coefficients as some particular derivatives of $\log \Psi$ with respect to the parameters $t_a$, the Taylor coefficient of the action functional at the vacuum state. The derivatives take the form of contributions of particular graphs, thereby endowing the graphs with specific physical interpretations. This approach leads to the graphical language known in quantum electrodynamics as [Feynman diagrams](https://en.wikipedia.org/wiki/Feynman_diagram).

One should not think that the combinatorial formulation of quantum electrodynamics resolves all its difficulties. In the contrary, the difficulties of both conceptual and technical nature only begin here. What are these difficulties, and how they are resolved is a remarkable story. But it is already a different story, requiring a different book, and by a different author.
**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 uncertainty 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 *ultra-violet 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

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

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 $\nu$, 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 $\nu_{max}$. 148
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 $\nu_{\text{max}}$ 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, $\epsilon_0$. Then the average energy of the oscillator is given by

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

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

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

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 $\nu_{\text{max}}/T$ 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

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

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

**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” \( \Delta E \) universally proportional to the frequency: \( \Delta E = h\nu \) (or \( h\omega \) 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 \( \nu_0 \) 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 $Z = 1/(1 - e^{-h\nu/kT})$ 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 $E$,

$$E^2 = m^2 c^4 + c^2 |p_e|^2,$$

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 $\lambda$ of the scattered and incident photons, and the cosine of the scattering angle $\theta$:

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

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.

\[ \lambda \left| p \right| = 2\pi \hbar \]

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.

Figure 32: Compton’s scattering
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](image)

Namely, as it is clear from Figure 33, the difference in the optical distance between the rays reflected at an angle $\theta$ 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 $\lambda$ 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:

$$2d \sin \theta = n \lambda,$$

where $n$ 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 $\theta$ 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 $\lambda$ 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 $2\pi \hbar / \sqrt{2m_eE}$ 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 $|L| = nh$, $n = 1, 2, \ldots$. In the quantum Kepler problem, these assumptions correspond to setting the orbital angular momentum quantum number $l = 0$, and lead to
the formula for the electron’s energy levels in the hydrogen atom

\[ E_n = -\frac{me^4}{(4\pi \varepsilon_0)^22\hbar^2n^2}. \]

Bohr’s model was improved by [Arnold Sommerfeld] who allowed elliptic Keplerian orbits possessing the same quantized values \( nh \) of their action. In the quantum Kepler problem, this corresponds to including non-zero values of the orbital angular momentum quantum number \( l \).

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 \( \lambda \), 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

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

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

$$E_n + \frac{e|B|}{2m_e} k\hbar = E_n + \mu_B |B| k,$$

where $k = 0, \pm 1, \ldots, \pm l$,

and $\mu_B$ 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, $E_1 = -13.6$ 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 $\mu$. 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.
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 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 electrically 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](image)

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 \(S_z\) and \(S_x\) 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” \(\psi\) 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 the list of the mostly familiar laureates, in chronological order of the 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.
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.
## Index

<table>
<thead>
<tr>
<th>Term</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O_1$, $O_3$</td>
<td>78, 78</td>
</tr>
<tr>
<td>$SL_2(\mathbb{R})$, $SO_3$, $SU_2$, $S_n$, $Sp_1$, $U_1$</td>
<td>114, 15, 65, 78, 78, 15, 78, 178</td>
</tr>
<tr>
<td>absolute temperature</td>
<td>105</td>
</tr>
<tr>
<td>abstract Schrödinger equation</td>
<td>50</td>
</tr>
<tr>
<td>acceptor levels</td>
<td>127</td>
</tr>
<tr>
<td>action</td>
<td>138</td>
</tr>
<tr>
<td>angular momentum</td>
<td>7, 65</td>
</tr>
<tr>
<td>angular momentum operator</td>
<td>98</td>
</tr>
<tr>
<td>annihilate</td>
<td>100</td>
</tr>
<tr>
<td>annihilation operator</td>
<td>55, 100</td>
</tr>
<tr>
<td>anomalous Zeeman effect</td>
<td>157</td>
</tr>
<tr>
<td>anti-commutator</td>
<td>90</td>
</tr>
<tr>
<td>anti-electrons</td>
<td>99</td>
</tr>
<tr>
<td>arrow paradox</td>
<td>33</td>
</tr>
<tr>
<td>atomic number</td>
<td>74</td>
</tr>
<tr>
<td>bands</td>
<td>116</td>
</tr>
<tr>
<td>Bernoulli numbers</td>
<td>106, 150</td>
</tr>
<tr>
<td>bi-differentiation</td>
<td>12</td>
</tr>
<tr>
<td>black body differentiation</td>
<td>147</td>
</tr>
<tr>
<td>black-body radiation</td>
<td>25, 106</td>
</tr>
<tr>
<td>Bloch theorem</td>
<td>123</td>
</tr>
<tr>
<td>blow-up</td>
<td>130</td>
</tr>
<tr>
<td>Bohr magneton</td>
<td>156</td>
</tr>
<tr>
<td>Bohr–Sommerfeld theory</td>
<td>71</td>
</tr>
<tr>
<td>Bohr–Sommerfeld’s model</td>
<td>155</td>
</tr>
<tr>
<td>Boltzmann constant</td>
<td>105</td>
</tr>
<tr>
<td>Bose-Einstein condensation</td>
<td>111</td>
</tr>
<tr>
<td>Bose-Einstein’s statistics</td>
<td>109</td>
</tr>
<tr>
<td>boson</td>
<td>59, 83</td>
</tr>
<tr>
<td>Bragg’s condition</td>
<td>152</td>
</tr>
<tr>
<td>Bragg’s diffraction</td>
<td>152</td>
</tr>
<tr>
<td>calculus of variations</td>
<td>135</td>
</tr>
<tr>
<td>canonical ensemble</td>
<td>104</td>
</tr>
<tr>
<td>Cartesian product</td>
<td>60</td>
</tr>
<tr>
<td>Casimir function</td>
<td>13, 91</td>
</tr>
<tr>
<td>caustic</td>
<td>3</td>
</tr>
<tr>
<td>center</td>
<td>79</td>
</tr>
<tr>
<td>central field</td>
<td>65</td>
</tr>
<tr>
<td>centrifugal term</td>
<td>67</td>
</tr>
<tr>
<td>chemical potential</td>
<td>107</td>
</tr>
<tr>
<td>classical field</td>
<td>100</td>
</tr>
<tr>
<td>classical observable</td>
<td>26</td>
</tr>
<tr>
<td>classical observables</td>
<td>11</td>
</tr>
<tr>
<td>Clebsch-Gordan formula</td>
<td>88</td>
</tr>
<tr>
<td>closed mechanical system</td>
<td>29</td>
</tr>
<tr>
<td>collapse of the wave function</td>
<td>159</td>
</tr>
<tr>
<td>commutation relations</td>
<td>13</td>
</tr>
<tr>
<td>commutator</td>
<td>50</td>
</tr>
<tr>
<td>compact symplectic group</td>
<td>78</td>
</tr>
<tr>
<td>complete integrability</td>
<td>132</td>
</tr>
<tr>
<td>completely integrable system</td>
<td>132</td>
</tr>
<tr>
<td>complexification</td>
<td>87</td>
</tr>
<tr>
<td>Compton wavelength</td>
<td>151</td>
</tr>
<tr>
<td>conduction band</td>
<td>123</td>
</tr>
<tr>
<td>configuration space</td>
<td>9</td>
</tr>
<tr>
<td>conjugate quaternion</td>
<td>77</td>
</tr>
<tr>
<td>conjugation</td>
<td>78</td>
</tr>
<tr>
<td>constructive diffraction</td>
<td>153</td>
</tr>
<tr>
<td>contact potential</td>
<td>126</td>
</tr>
<tr>
<td>continuous spectrum</td>
<td>42, 115</td>
</tr>
<tr>
<td>covalent bonding</td>
<td>125</td>
</tr>
<tr>
<td>covector</td>
<td>137</td>
</tr>
<tr>
<td>creation operator</td>
<td>55, 100</td>
</tr>
<tr>
<td>critical temperature</td>
<td>112</td>
</tr>
<tr>
<td>crystalline lattice</td>
<td>113</td>
</tr>
<tr>
<td>curl</td>
<td>95</td>
</tr>
<tr>
<td>de Broglie wavelength</td>
<td>153</td>
</tr>
<tr>
<td>degeneracy</td>
<td>74</td>
</tr>
<tr>
<td>degrees of freedom</td>
<td>9</td>
</tr>
<tr>
<td>delta-well potential</td>
<td>37</td>
</tr>
<tr>
<td>depletion region</td>
<td>128</td>
</tr>
<tr>
<td>destructive diffraction</td>
<td>153</td>
</tr>
<tr>
<td>differential</td>
<td>137</td>
</tr>
<tr>
<td>diffraction</td>
<td>18</td>
</tr>
<tr>
<td>diode</td>
<td>129</td>
</tr>
<tr>
<td>Dirac equation</td>
<td>96</td>
</tr>
<tr>
<td>Dirac’s delta-function</td>
<td>33</td>
</tr>
<tr>
<td>direct sum</td>
<td>63</td>
</tr>
<tr>
<td>discrete spectrum</td>
<td>42</td>
</tr>
</tbody>
</table>

161
initial condition, 34
invertible mappings, 14
irreducible representations, 70
isotropy, 93
isotypical components, 83
Jacobi identity, 12
joule, 105
KdV equation, 129
kelvins, 105
Kronig-Penney model, 121
ladder operators, 85
Lagrangian, 134
Lagrangian mechanical system, 136
Lagrangian submanifolds, 16
Laguerre’s polynomials, 72
Laplace, 65
lattice wave vector, 123
Lax equation, 131
Lax pair, 131
least action principle, 138
least time principle, 133
Legendre transform, 136
Lie algebra, 12, 91
light cone, 93
linear operator, 45
Lorentz group, 94
Lorentz’ boost, 94
magnetic moment, 156
Maxwell equations, 95
Maxwell-Boltzmann statistics, 107
method of characteristics, 4
momentum, 5
momentum operator, 32
momentum representation, 32
monodromy matrix, 114
monodromy transformation, 114
n-type semiconductor, 128
Newton’s rings, 17
Noether’s theorem, 13
normal form, 52
normal Zeeman effect, 157
observables, 26, 45
odd super-space, 63
odd variables, 62
one-loop approximation, 145
one-parametric group, 15
optical distance function, 8
orbital angular momentum
quantum number, 73
orbital magnetic moment, 137
orientation, 15
orthogonal diagonalization
theorem, 46, 52
orthogonal group, 78
orthogonal transformations, 78
oscillating integral, 139
p-type semiconductor, 128
paradigm shift, 147
parity operator, 80
Parseval’s identity, 48
partition function, 105
path integral, 139
Pauli matrices, 77
Pauli’s exclusion principle, 60, 76
pencil of rays, 2
pencil of trajectories, 16
periodic table, 73
permutation group, 14, 81
phase function, 139, 140
phase space, 9
phase transition, 111
phase velocity, 34
phase volume, 11
photoelectric effect, 150
photon, 96
Planck constant, 25
Poisson bracket, 12, 90
Poisson’s theorem, 13
polarization, 96
positrons, 100
potential energy, 5
principal quantum number, 73
probability current, 41
probability density, 30
probability ratio, 104
psi-function, 26
purely imaginary quaternions, 78
quadratic form, 52
trivial representation, 86

tunneling, 36 42

ultraviolet catastrophe, 148
uncertainty principle, 32 49
unitary group, 78
unitary operator, 50 106

vacuum state, 55 100
valence band, 124
variation, 135
variational problem, 134
vector representation, 86
vector-potential, 155

vertices, 142

wave front, 1
wave packet, 31
wavelet, 1
white dwarf, 111
Wick’s theorem, 142
WKB-theory, 55
work function, 126 150
Wronskian, 114

Zeeman effect, 155
zones of parametric resonance, 119