

This Quantum World

Wikibooks.org

March 13, 2013

On the 28th of April 2012 the contents of the English as well as German Wikibooks and Wikipedia projects were licensed under Creative Commons Attribution-ShareAlike 3.0 Unported license. An URI to this license is given in the list of figures on page 137. If this document is a derived work from the contents of one of these projects and the content was still licensed by the project under this license at the time of derivation this document has to be licensed under the same, a similar or a compatible license, as stated in section 4b of the license. The list of contributors is included in chapter Contributors on page 135. The licenses GPL, LGPL and GFDL are included in chapter Licenses on page 143, since this book and/or parts of it may or may not be licensed under one or more of these licenses, and thus require inclusion of these licenses. The licenses of the figures are given in the list of figures on page 137. This PDF was generated by the L^AT_EX typesetting software. The L^AT_EX source code is included as an attachment (`source.7z.txt`) in this PDF file. To extract the source from the PDF file, we recommend the use of <http://www.pdflabs.com/tools/pdfkit-the-pdf-toolkit/> utility or clicking the paper clip attachment symbol on the lower left of your PDF Viewer, selecting **Save Attachment**. After extracting it from the PDF file you have to rename it to `source.7z`. To uncompress the resulting archive we recommend the use of <http://www.7-zip.org/>. The L^AT_EX source itself was generated by a program written by Dirk Hünniger, which is freely available under an open source license from http://de.wikibooks.org/wiki/Benutzer:Dirk_Huenniger/wb2pdf. This distribution also contains a configured version of the `pdflatex` compiler with all necessary packages and fonts needed to compile the L^AT_EX source included in this PDF file.

Contents

1 Atoms	3
1.1 What does an atom look like?	3
1.2 Quantum states	9
1.3 Fuzzy observables	14
2 Serious illnesses require drastic remedies	17
2.1 Planck	17
2.2 Rutherford	17
2.3 Bohr	18
2.4 de Broglie	20
2.5 Schrödinger	23
2.6 Born	26
2.7 Heisenberg	27
3 The Feynman route to Schrödinger	29
3.1 Two rules	29
3.2 An experiment with two slits	31
3.3 Bohm's story	35
3.4 Propagator for a free and stable particle	39
3.5 From quantum to classical	41
3.6 Schrödinger at last	51
4 The Schrödinger equation: implications and applications	55
4.1 How fuzzy positions get fuzzier	55
4.2 Time-independent Schrödinger equation	60
4.3 Why energy is quantized	61
4.4 A quantum bouncing ball	63
4.5 Atomic hydrogen	65
4.6 Observables and operators	78
4.7 Beyond hydrogen: the Periodic Table	82
4.8 Probability flux	84
5 Entanglement (a preview)	87
5.1 Bell's theorem: the simplest version	87
5.2 A quantum game	89
5.3 The experiment of Greenberger, Horne, and Zeilinger	90
6 Authors	93
7 Appendix	95
7.1 Probability	95

7.2 Mathematical tools	100
7.3 The ABCs of relativity	121
8 Contributors	135
List of Figures	137
9 Licenses	143
9.1 GNU GENERAL PUBLIC LICENSE	143
9.2 GNU Free Documentation License	144
9.3 GNU Lesser General Public License	145

1 Atoms

1.1 What does an atom look like?

1.1.1 Like this?

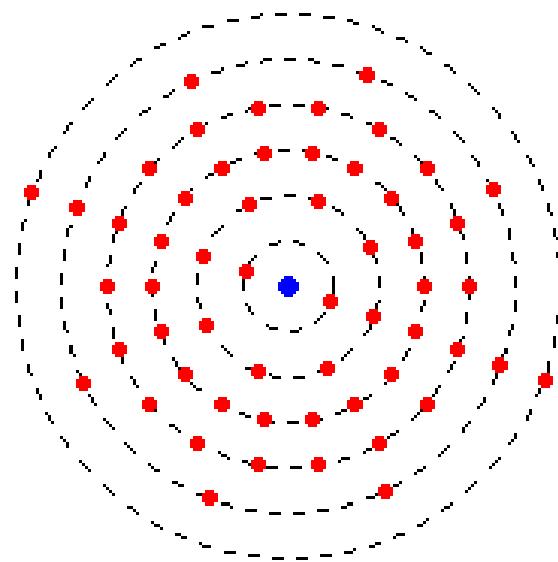


Figure 1

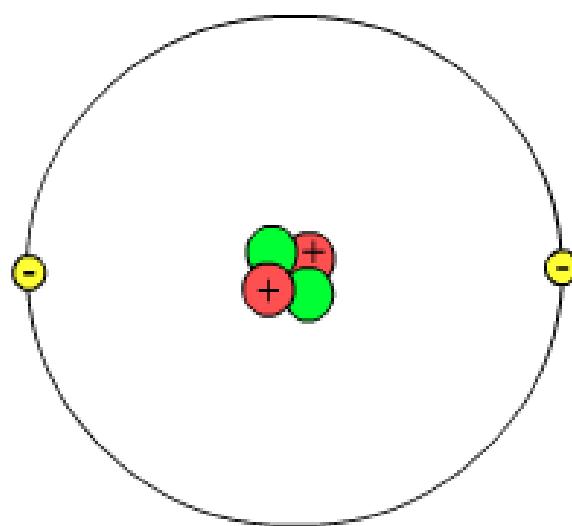


Figure 2

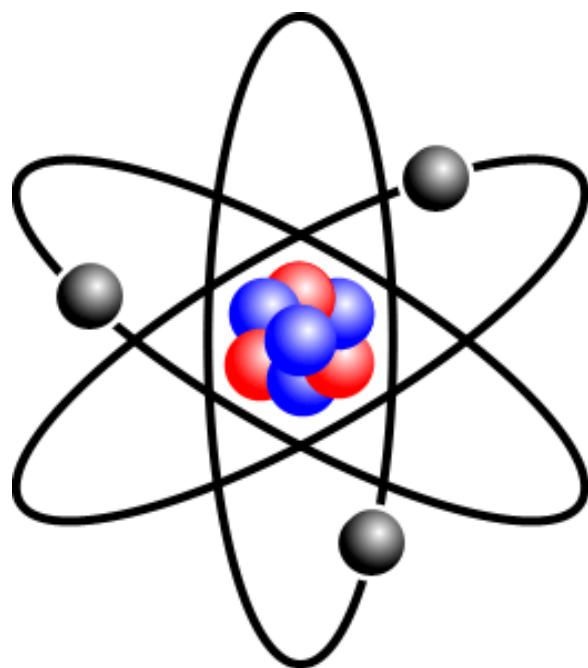


Figure 3

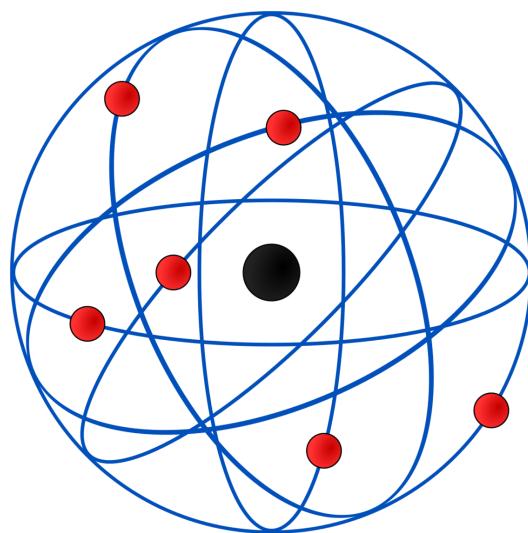


Figure 4

1.1.2 Or like this?

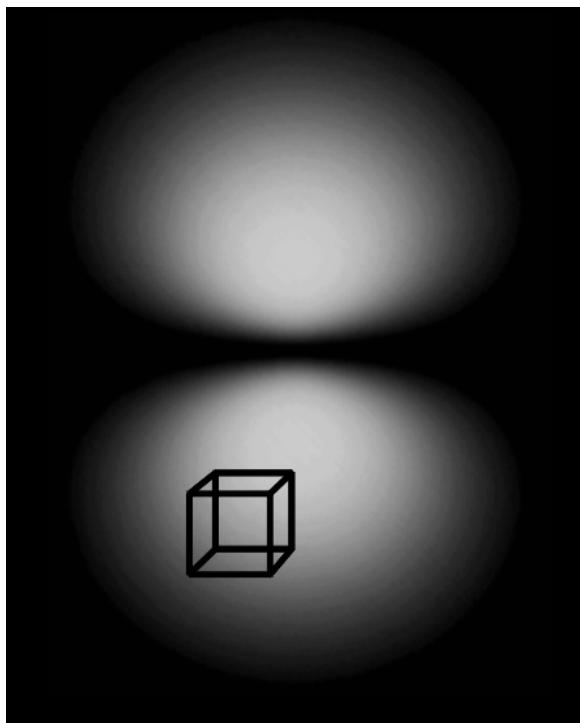


Figure 5 ρ_{2p0}

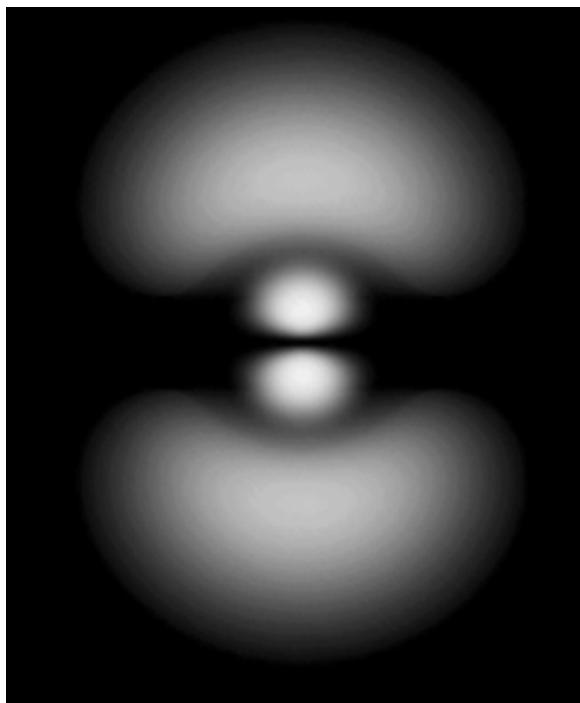


Figure 6 ρ_{3p0}

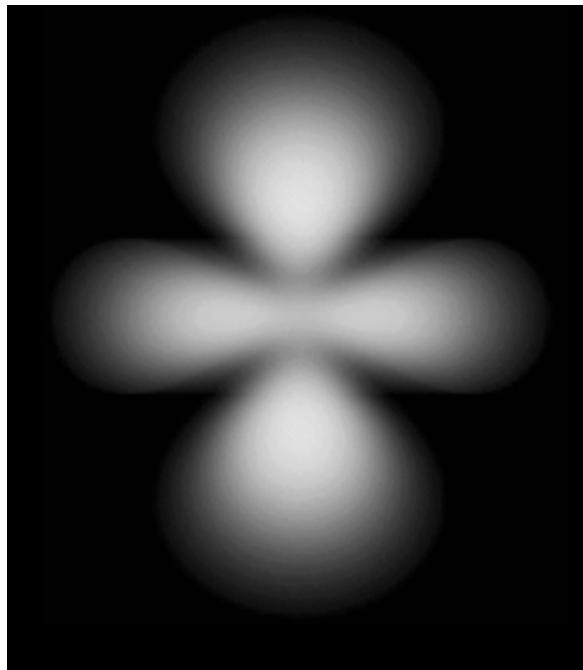


Figure 7 ρ_{3d0}

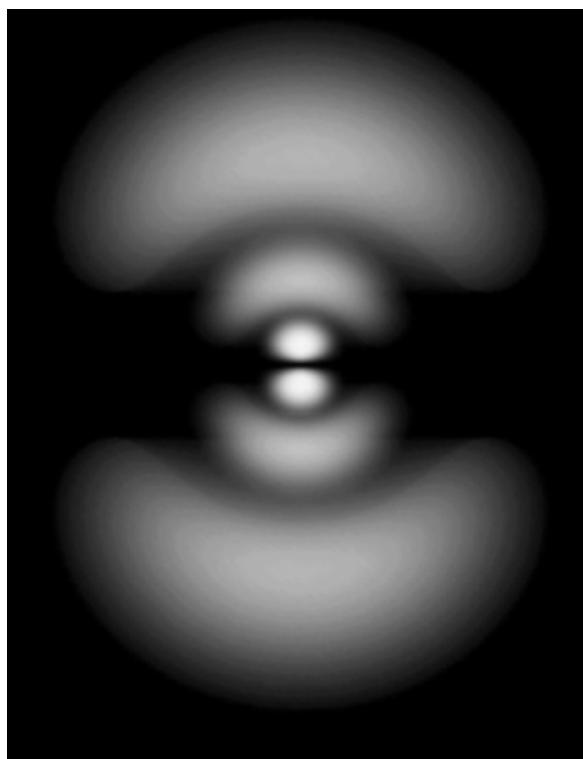


Figure 8 ρ_{4p0}

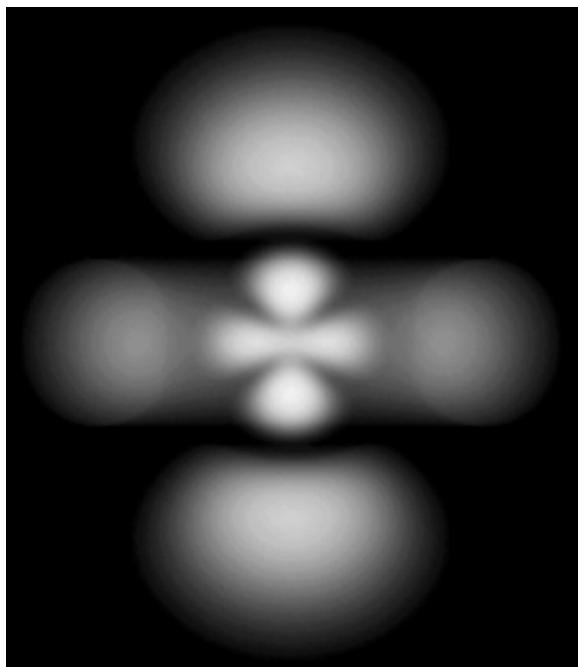


Figure 9 ρ_{4d0}

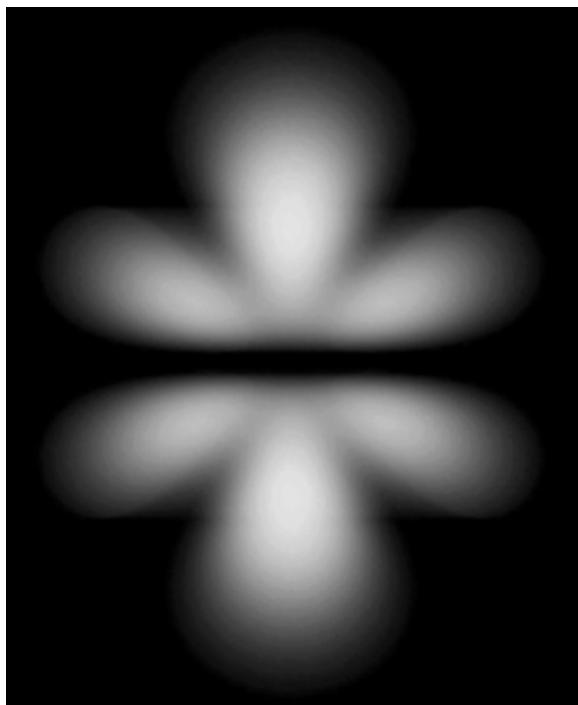


Figure 10 ρ_{4f0}

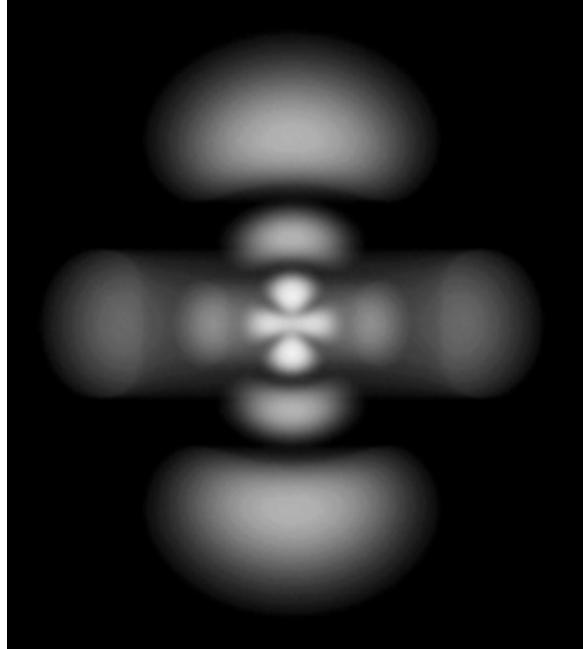


Figure 11 ρ_{5d0}

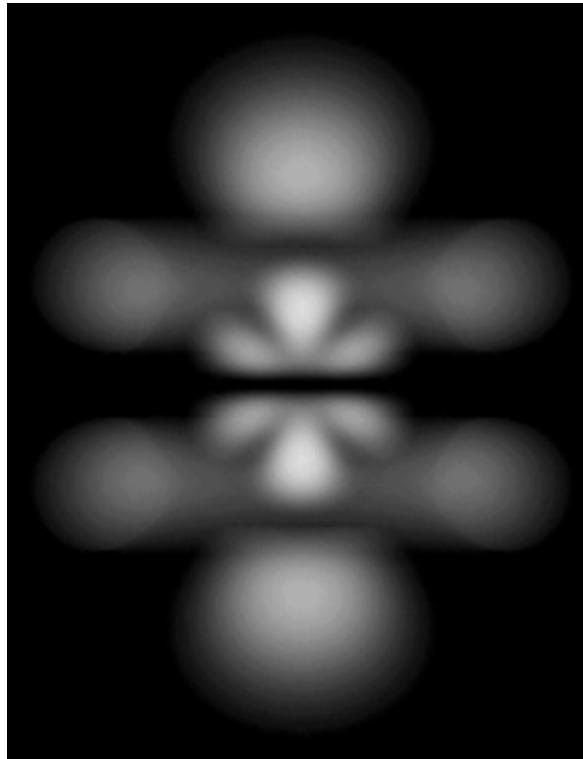


Figure 12 ρ_{5f0}

None of these images depicts an atom *as it is*. This is because it is impossible to even visualize an atom *as it is*. Whereas the best you can do with the images in the first row is to erase them from your memory, they represent a way of viewing the atom that is too simplified for the way we want to start thinking about it, the eight fuzzy images in the next

row deserve scrutiny. Each represents an aspect of a stationary state of atomic hydrogen. You see neither the nucleus (a proton) nor the electron. What you see is a fuzzy position. To be precise, what you see is a cloud-like blur, which is symmetrical about the vertical axis, and which represents the atom's internal relative position — the position of the electron relative to the proton *or* the position of the proton relative to the electron.

- What is the *state* of an atom?
- What is a *stationary* state?
- What exactly is a *fuzzy* position?
- How does such a blur represent the atom's internal relative position?
- Why can we not describe the atom's internal relative position *as it is*?

1.2 Quantum states

In quantum mechanics, **states**¹ are probability algorithms. We use them to calculate the probabilities of the possible outcomes of measurements² on the basis of actual measurement outcomes. A quantum state takes as its input

- one or several measurement outcomes,
- a measurement M,
- the time of M,

and it yields as its output the probabilities of the possible outcomes of M.

A quantum state is called **stationary** if the probabilities it assigns are independent of the time of the measurement to the possible outcomes of which they are assigned.

From the mathematical point of view, each blur represents a density function³ $\rho(\mathbf{r})$. Imagine a small region R like the little box inside the first blur. And suppose that this is a region of the (mathematical) space of positions relative to the proton. If you integrate $\rho(\mathbf{r})$ over R , you obtain the probability $p(R)$ of finding the electron in R , *provided* that the appropriate measurement is made:

$$p(R) = \int_R \rho(\mathbf{r}) d^3\mathbf{r}.$$

"Appropriate" here means capable of ascertaining the truth value of the proposition "the electron is in R ", the possible truth values being "true" or "false". What we see in each of the following images is a surface of constant probability density.

1 <http://en.wikipedia.org/wiki/Quantum%20state>

2 <http://en.wikipedia.org/wiki/Measurement%20in%20quantum%20mechanics>

3 <http://en.wikipedia.org/wiki/Probability%20density%20function>

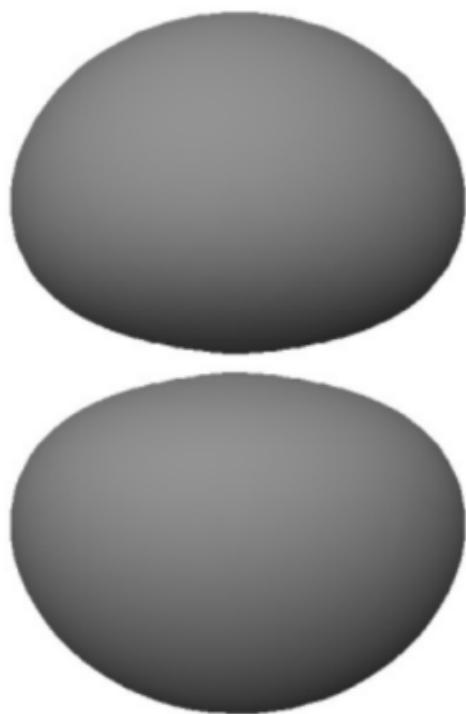


Figure 13 ρ_{2p0}

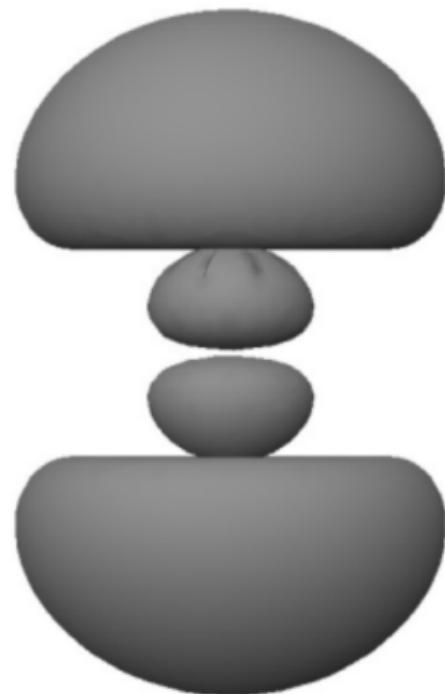


Figure 14 ρ_{3p0}

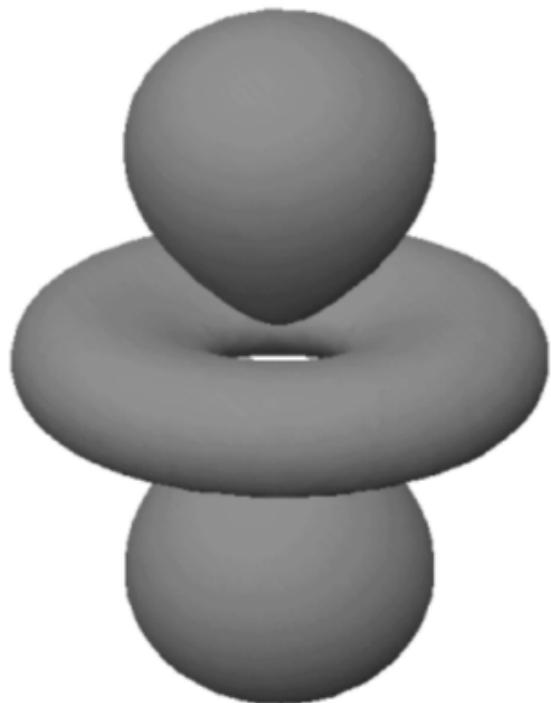


Figure 15 ρ_{3d0}

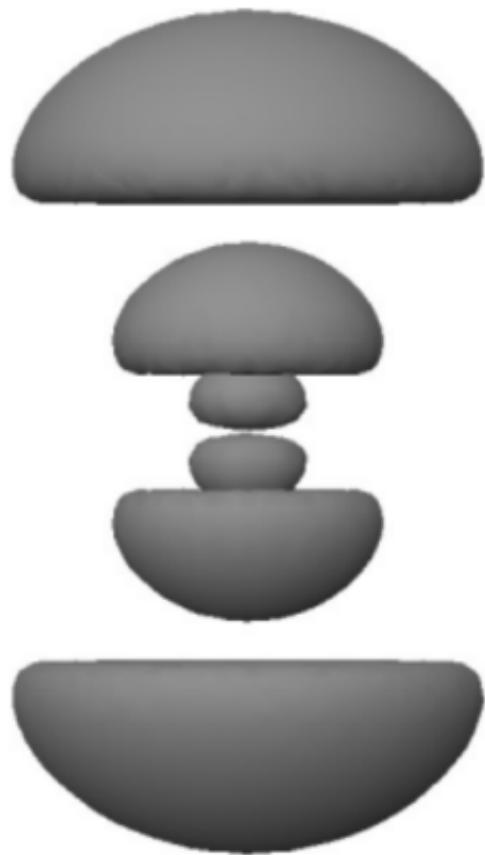


Figure 16 ρ_{4p0}

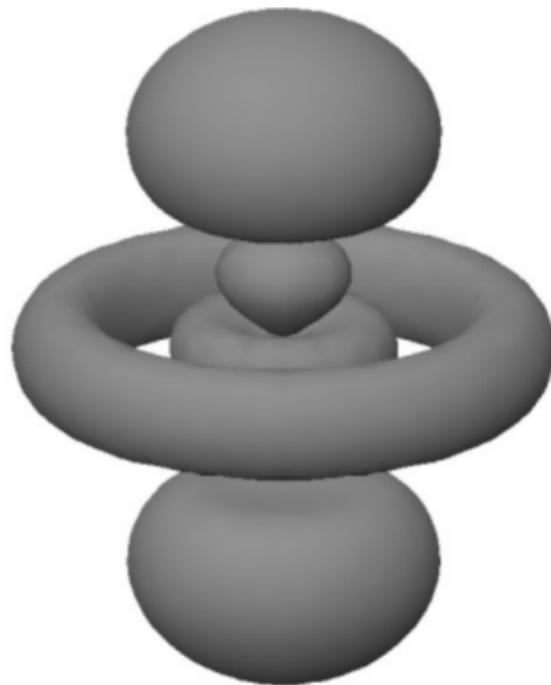


Figure 17 ρ_{4d0}

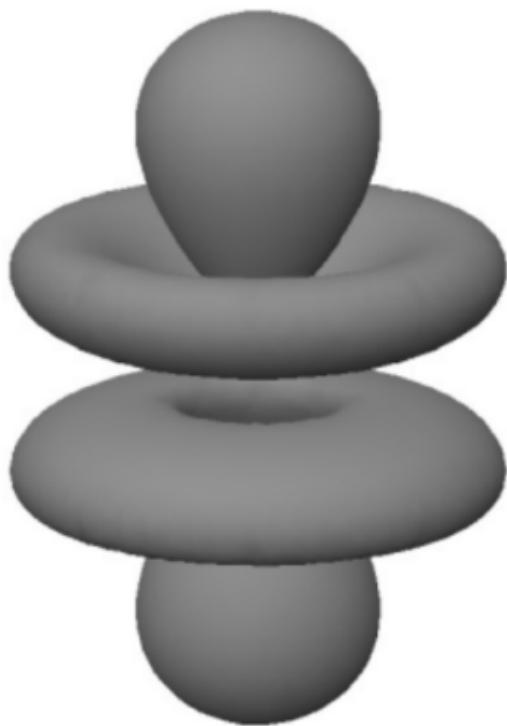


Figure 18 ρ_{4f0}

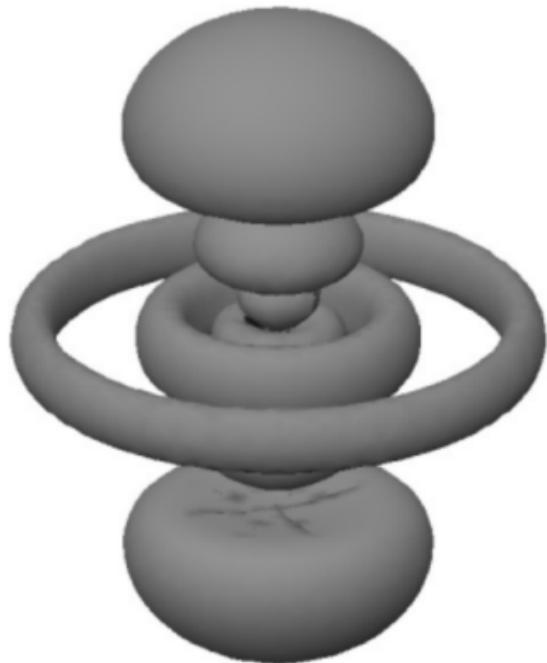


Figure 19 ρ_{5d0}

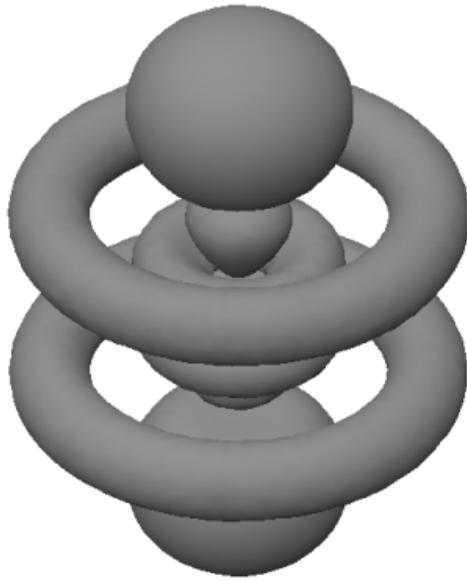


Figure 20 ρ_{5f0}

Now imagine that the appropriate measurement is made. *Before* the measurement, the electron is neither inside R nor outside R . If it were inside, the probability of finding it outside would be zero, and if it were outside, the probability of finding it inside would be zero. *After* the measurement, on the other hand, the electron is either inside or outside R .

Conclusions:

- Before the measurement, the proposition "the electron is in R " is neither true nor false; it lacks a (definite) truth value⁴.
- A measurement generally changes the state of the system on which it is performed.

As mentioned before, probabilities are assigned not only *to* measurement outcomes but also *on the basis of* measurement outcomes. Each density function ρ_{nlm} serves to assign probabilities to the possible outcomes of a measurement of the position of the electron relative to the proton. And in each case the assignment is based on the outcomes of a simultaneous measurement of three observables: the atom's energy (specified by the value of the principal quantum number n), its total angular momentum⁵ l (specified by a letter, here p , d , or f), and the vertical component of its angular momentum m .

1.3 Fuzzy observables

We say that an observable Q with a finite or countable number of possible values q_k is **fuzzy** (or that it has a fuzzy value) if and only if at least one of the propositions "The value of Q is

⁴ <http://en.wikipedia.org/wiki/Truth%20value>

⁵ http://en.wikipedia.org/wiki/Angular_momentum%23Angular_momentum_in_quantum_mechanics

q_k " lacks a truth value. This is equivalent to the following necessary and sufficient condition: the probability assigned to at least one of the values q_k is neither 0 nor 1.

What about observables that are generally described as continuous, like a position?

The description of an observable as "continuous" is potentially misleading. For one thing, we cannot separate an observable and its possible values from a measurement and its possible outcomes, and a measurement with an uncountable set of possible outcomes is not even in principle possible. For another, there is not a single observable called "position". Different partitions of space define different position measurements with different sets of possible outcomes.

- Corollary: The possible outcomes of a position measurement (or the possible values of a position observable) are defined by a partition of space. They make up a finite or countable set of *regions* of space. An exact position is therefore neither a possible measurement outcome nor a possible value of a position observable.

So how do those cloud-like blurs represent the electron's fuzzy position relative to the proton? Strictly speaking, they graphically represent probability densities in the mathematical space of exact relative positions, rather than fuzzy positions. It is these probability densities that represent fuzzy positions by allowing us to calculate the probability of every possible value of every position observable.

It should now be clear why we cannot describe the atom's internal relative position *as it is*. To describe a fuzzy observable is to assign probabilities to the possible outcomes of a measurement. But a description that rests on the assumption that a measurement is made, does not describe an observable *as it is* (by itself, *regardless of measurements*).

2 Serious illnesses require drastic remedies

2.1 Planck

Quantum mechanics began as a desperate measure to get around some spectacular failures of what subsequently came to be known as classical physics¹.

In 1900 Max Planck² discovered a law that perfectly describes the spectrum of a glowing hot object. Planck's radiation formula³ turned out to be irreconcilable with the physics of his time. (If classical physics were right, you would be blinded by ultraviolet light if you looked at the burner of a stove, aka the UV catastrophe⁴.) At first, it was just a fit to the data, "a fortuitous guess at an interpolation formula" as Planck himself called it. Only weeks later did it turn out to imply the quantization of energy for the emission of electromagnetic radiation⁵: the energy E of a quantum⁶ of radiation is proportional to the frequency ν of the radiation, the constant of proportionality being Planck's constant⁷ h :

$$E = h\nu$$

We can of course use the angular frequency⁸ $\omega = 2\pi\nu$ instead of ν . Introducing the reduced Planck constant $\hbar = h/2\pi$, we then have

$$E = \hbar\omega$$

2.2 Rutherford

In 1911 Ernest Rutherford⁹ proposed a model of the atom¹⁰ based on experiments by Geiger and Marsden¹¹. Geiger and Marsden had directed a beam of alpha particles¹² at a thin

1 [http://en.wikipedia.org/wiki/Classical%20physics](http://en.wikipedia.org/wiki/Classical_physics)
2 [http://en.wikipedia.org/wiki/Max%20Planck](http://en.wikipedia.org/wiki/Max_Planck)
3 [http://en.wikipedia.org/wiki/Planck%27s%20law%20of%20black%20body%20radiation](http://en.wikipedia.org/wiki/Planck%27s_20law%20of%20black%20body%20radiation)
4 [http://en.wikipedia.org/wiki/UV%20catastrophe](http://en.wikipedia.org/wiki/UV_catastrophe)
5 [http://en.wikipedia.org/wiki/Electromagnetic%20radiation](http://en.wikipedia.org/wiki/Electromagnetic_radiation)
6 <http://en.wikipedia.org/wiki/Quantum>
7 <http://en.wikipedia.org/wiki/Planck%27s%20constant>
8 [http://en.wikipedia.org/wiki/Angular%20frequency](http://en.wikipedia.org/wiki/Angular_frequency)
9 [http://en.wikipedia.org/wiki/Ernest%20Rutherford](http://en.wikipedia.org/wiki/Ernest_Rutherford)
10 [http://en.wikipedia.org/wiki/Rutherford%20model](http://en.wikipedia.org/wiki/Rutherford_model)
11 [http://en.wikipedia.org/wiki/Geiger-Marsden%20experiment](http://en.wikipedia.org/wiki/Geiger-Marsden_experiment)
12 [http://en.wikipedia.org/wiki/Alpha%20particle](http://en.wikipedia.org/wiki/Alpha_particle)

gold foil. Most of the particles passed the foil more or less as expected, but about one in 8000 bounced back as if it had encountered a much heavier object. In Rutherford's own words this was as incredible as if you fired a 15 inch cannon ball at a piece of tissue paper and it came back and hit you. After analysing the data collected by Geiger and Marsden, Rutherford concluded that the diameter of the atomic nucleus (which contains over 99.9% of the atom's mass) was less than 0.01% of the diameter of the entire atom, and he suggested that atomic electrons orbit the nucleus much like planets orbit a star.

The problem of having electrons orbit the nucleus the same way that a planet orbits a star is that classical electromagnetic theory demands that an orbiting electron will radiate away its energy and spiral into the nucleus in about 0.5×10^{-10} seconds. This was the worst quantitative failure in the history of physics, under-predicting the lifetime of hydrogen by at least forty orders of magnitude! (This figure is based on the experimentally established lower bound on the proton's lifetime.)

2.3 Bohr

In 1913 Niels Bohr¹³ postulated that the angular momentum L of an orbiting atomic electron was quantized: its "allowed" values are integral multiples of \hbar :

$$L = n\hbar \text{ where } n = 1, 2, 3, \dots$$

Why quantize angular momentum, rather than any other quantity?

- Radiation energy of a given frequency is quantized in multiples of Planck's constant.
- Planck's constant is measured in the same units as angular momentum¹⁴.

Bohr's postulate explained not only the stability of atoms but also why the emission and absorption of electromagnetic radiation by atoms is discrete. In addition it enabled him to calculate with remarkable accuracy the spectrum of atomic hydrogen — the frequencies at which it is able to emit and absorb light (visible as well as infrared and ultraviolet). The following image shows the visible emission spectrum of atomic hydrogen, which contains four lines of the Balmer series¹⁵.

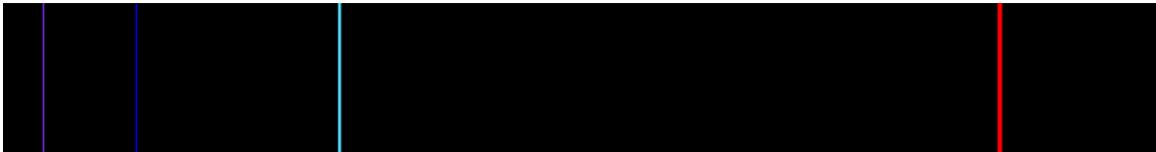


Figure 21 Visible emission spectrum of atomic hydrogen, containing four lines of the Balmer series^a.

^a <http://en.wikipedia.org/wiki/Balmer%20series>

13 <http://en.wikipedia.org/wiki/Niels%20Bohr>

14 <http://en.wikipedia.org/wiki/Angular%20momentum>

15 <http://en.wikipedia.org/wiki/Balmer%20series>

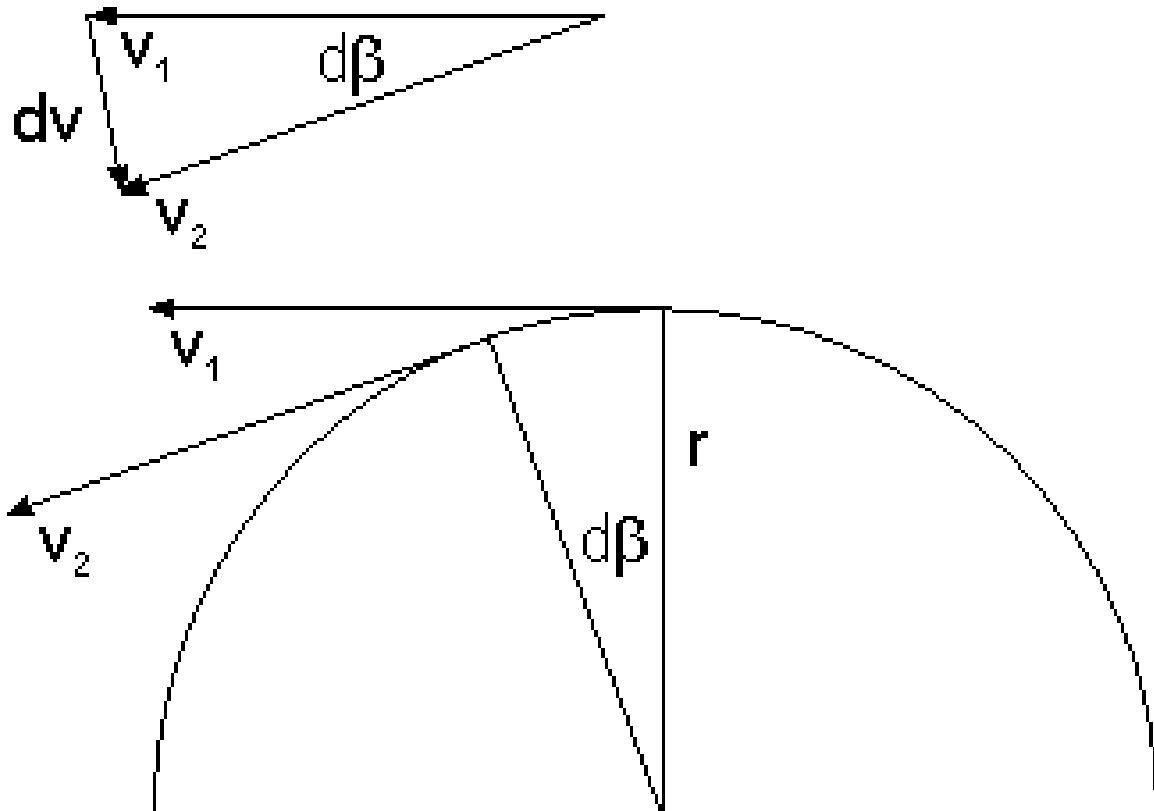


Figure 22 frame

Apart from his quantization postulate, Bohr's reasoning at this point remained completely classical. Let's assume with Bohr that the electron's orbit is a circle of radius r . The speed of the electron is then given by $v = r d\beta/dt$, and the magnitude of its acceleration by $a = dv/dt = v d\beta/dt$. Eliminating $d\beta/dt$ yields $a = v^2/r$. In the cgs system of units¹⁶, the magnitude of the Coulomb force¹⁷ is simply $F = e^2/r^2$, where e is the magnitude of the charge of both the electron and the proton. Via Newton's¹⁸ $F = ma$ the last two equations yield $m_e v^2 = e^2/r$, where m_e is the electron's mass. If we take the proton to be at rest, we obtain $T = m_e v^2/2 = e^2/2r$ for the electron's kinetic energy.

If the electron's potential energy at infinity is set to 0, then its potential energy V at a distance r from the proton is minus the work¹⁹ required to move it from r to infinity,

$$V = - \int_r^\infty F(r') dr' = - \int_r^\infty \frac{e^2}{(r')^2} dr' = + \left[\frac{e^2}{r'} \right]_r^\infty = 0 - \frac{e^2}{r}.$$

The total energy of the electron thus is

16 <http://en.wikipedia.org/wiki/Centimeter%20gram%20second%20system%20of%20units>

17 <http://en.wikipedia.org/wiki/Coulomb%27s%20law>

18 <http://en.wikipedia.org/wiki/Newton%27s%20laws%20of%20motion>

19 <http://en.wikipedia.org/wiki/Mechanical%20work>

$$E = T + V = e^2/2r - e^2/r = -e^2/2r.$$

We want to express this in terms of the electron's angular momentum $L = m_e v r$. Remembering that $m_e v^2 = e^2/r$, and hence $r m_e^2 v^2 = m_e e^2$, and multiplying the numerator e^2 by $m_e e^2$ and the denominator $2r$ by $r m_e^2 v^2$, we obtain

$$E = -\frac{e^2}{2r} = -\frac{m_e e^4}{2m_e^2 v^2 r^2} = -\frac{m_e e^4}{2L^2}.$$

Now comes Bohr's break with classical physics: he simply replaced L by $n\hbar$. The "allowed" values for the angular momentum define a series of allowed values for the atom's energy:

$$E_n = -\frac{1}{n^2} \left(\frac{m_e e^4}{2\hbar^2} \right), \quad n = 1, 2, 3, \dots$$

As a result, the atom can emit or absorb energy only by amounts equal to the absolute values of the differences

$$\Delta E_{nm} = E_n - E_m = \left(\frac{1}{n^2} - \frac{1}{m^2} \right) \text{Ry},$$

one Rydberg²⁰ (Ry) being equal to $m_e e^4 / 2\hbar^2 = 13.6056923(12)$ eV. This is also the ionization energy²¹ $\Delta E_{1\infty}$ of atomic hydrogen — the energy needed to completely remove the electron from the proton. Bohr's predicted value was found to be in excellent agreement with the measured value.

Using two of the above expressions for the atom's energy and solving for r , we obtain $r = n^2 \hbar^2 / m_e e^2$. For the ground state ($n = 1$) this is the Bohr radius of the hydrogen atom²², which equals $\hbar^2 / m_e e^2 = 5.291772108(18) \times 10^{-11} m$. The mature theory yields the same figure but interprets it as the *most likely* distance from the proton at which the electron would be found if its distance from the proton were measured.

2.4 de Broglie

In 1923, ten years after Bohr had derived the spectrum of atomic hydrogen by postulating the quantization of angular momentum, Louis de Broglie²³ hit on an explanation of why the atom's angular momentum comes in multiples of \hbar . Since 1905, Einstein²⁴ had argued that electromagnetic radiation itself was quantized (and not merely its emission and absorption,

²⁰ <http://en.wikipedia.org/wiki/Rydberg>

²¹ <http://en.wikipedia.org/wiki/Ionization%20potential>

²² <http://en.wikipedia.org/wiki/Bohr%20radius>

²³ <http://en.wikipedia.org/wiki/Louis%20de%20Broglie>

²⁴ <http://en.wikipedia.org/wiki/Albert%20Einstein>

as Planck held). If electromagnetic waves can behave like particles (now known as photons²⁵), de Broglie reasoned, why cannot electrons behave like waves?

Suppose that the electron in a hydrogen atom is a standing wave²⁶ on what has so far been thought of as the electron's circular orbit. (The crests, troughs²⁷, and nodes²⁸ of a standing wave are stationary.) For such a wave to exist on a circle, the circumference of the latter must be an integral multiple of the wavelength²⁹ λ of the former: $2\pi r = n\lambda$.

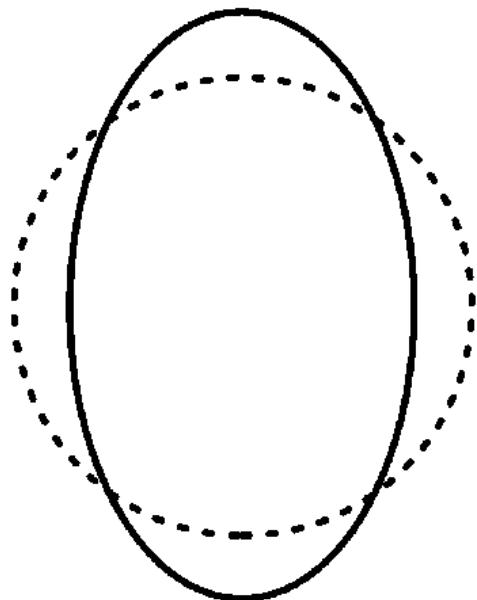


Figure 23

25 <http://en.wikipedia.org/wiki/Photon>
26 <http://en.wikipedia.org/wiki/Standing%20wave>
27 <http://en.wikipedia.org/wiki/Crest%20%28physics%29>
28 <http://en.wikipedia.org/wiki/Node%20%28physics%29>
29 <http://en.wikipedia.org/wiki/Wavelength>

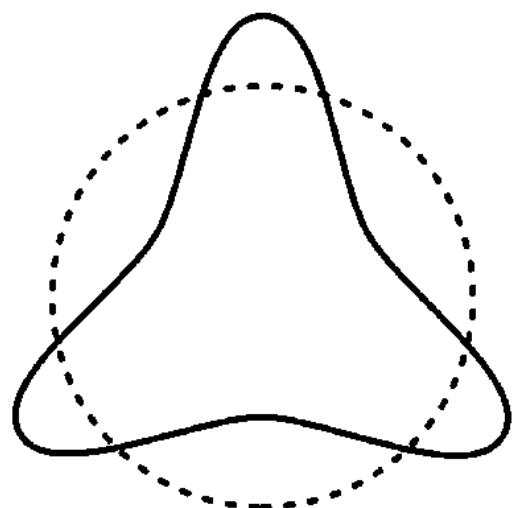


Figure 24

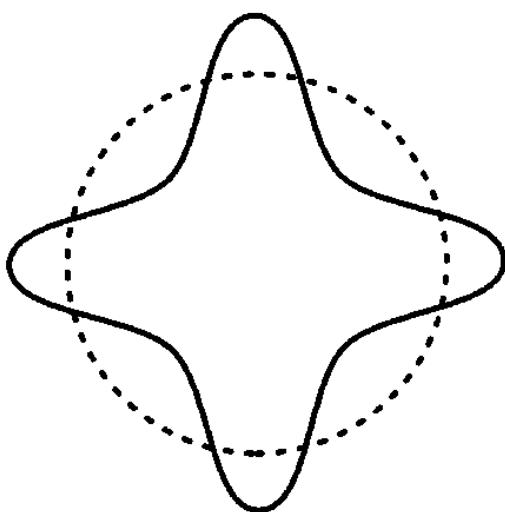


Figure 25

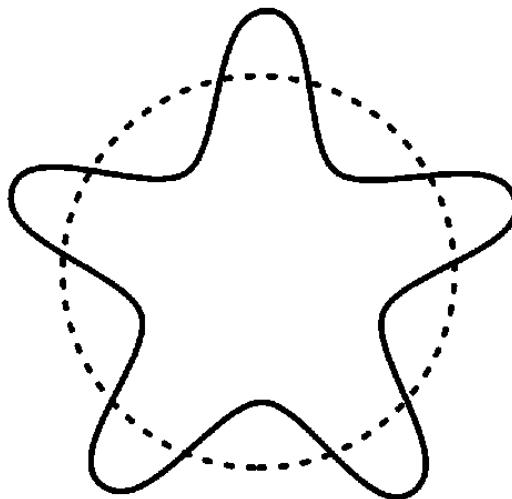


Figure 26

Einstein had established not only that electromagnetic radiation of frequency ν comes in quanta of energy $E = h\nu$ but also that these quanta carry a momentum $p = h/\lambda$. Using this formula to eliminate λ from the condition $2\pi r = n\lambda$, one obtains $pr = n\hbar$. But $pr = mvr$ is just the angular momentum L of a classical electron with an orbit of radius r . In this way de Broglie derived the condition $L = n\hbar$ that Bohr had simply postulated.

2.5 Schrödinger

If the electron is a standing wave, why should it be confined to a circle? After de Broglie's crucial insight that particles are waves of some sort, it took less than three years for the mature quantum theory to be found, not once but twice, by Werner Heisenberg³⁰ in 1925 and by Erwin Schrödinger³¹ in 1926. If we let the electron be a standing wave in three dimensions, we have all it takes to arrive at the Schrödinger equation, which is at the heart of the mature theory.

Let's keep to one spatial dimension. The simplest mathematical description of a wave of angular wavenumber³² $k = 2\pi/\lambda$ and angular frequency³³ $\omega = 2\pi/T = 2\pi\nu$ (at any rate, if you are familiar with complex numbers³⁴) is the function

$$\psi(x, t) = e^{i(kx - \omega t)}.$$

Let's express the phase³⁵ $\phi(x, t) = kx - \omega t$ in terms of the electron's energy $E = h\nu = \hbar\omega$ and momentum $p = h/\lambda = \hbar k$:

30 <http://en.wikipedia.org/wiki/Werner%20Heisenberg>

31 <http://en.wikipedia.org/wiki/Erwin%20Schr%F6dinger>

32 <http://en.wikipedia.org/wiki/wavenumber>

33 <http://en.wikipedia.org/wiki/Angular%20frequency>

34 <http://en.wikipedia.org/wiki/Complex%20number>

35 <http://en.wikipedia.org/wiki/Phase%20%28waves%29>

$$\psi(x, t) = e^{i(px - Et)/\hbar}.$$

The partial derivatives³⁶ with respect to x and t are

$$\frac{\partial \psi}{\partial x} = \frac{i}{\hbar} p \psi \quad \text{and} \quad \frac{\partial \psi}{\partial t} = -\frac{i}{\hbar} E \psi.$$

We also need the second partial derivative of ψ with respect to x :

$$\frac{\partial^2 \psi}{\partial x^2} = \left(\frac{ip}{\hbar}\right)^2 \psi.$$

We thus have

$$E\psi = i\hbar \frac{\partial \psi}{\partial t}, \quad p\psi = -i\hbar \frac{\partial \psi}{\partial x}, \quad \text{and} \quad p^2\psi = -\hbar^2 \frac{\partial^2 \psi}{\partial x^2}.$$

In non-relativistic classical physics³⁷ the *kinetic* energy³⁸ and the *kinetic* momentum p of a free particle³⁹ are related via the dispersion relation⁴⁰

$$E = p^2/2m.$$

This relation also holds in non-relativistic quantum physics. Later you will learn why.

In three spatial dimensions, p is the magnitude of a vector \mathbf{p} . If the particle also has a potential energy $V(\mathbf{r}, t)$ and a potential momentum $\mathbf{A}(\mathbf{r}, t)$ (in which case it is *not* free), and if E and \mathbf{p} stand for the particle's *total* energy and *total* momentum, respectively, then the dispersion relation is

$$E - V = (\mathbf{p} - \mathbf{A})^2/2m.$$

By the square of a vector \mathbf{v} we mean the dot (or scalar) product⁴¹ $\mathbf{v} \cdot \mathbf{v}$. Later you will learn why we represent possible influences on the motion of a particle by such fields⁴² as $V(\mathbf{r}, t)$ and $\mathbf{A}(\mathbf{r}, t)$.

Returning to our fictitious world with only one spatial dimension, allowing for a potential energy $V(x, t)$, substituting the differential operators⁴³ $i\hbar \frac{\partial}{\partial t}$ and $-\hbar^2 \frac{\partial^2}{\partial x^2}$ for E and p^2 in the resulting dispersion relation, and applying both sides of the resulting operator equation to ψ , we arrive at the one-dimensional (time-dependent) **Schrödinger equation**:

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial x^2} + V\psi$$

³⁶ <http://en.wikipedia.org/wiki/Partial%20derivative>

³⁷ <http://en.wikipedia.org/wiki/Classical%20physics>

³⁸ <http://en.wikipedia.org/wiki/Kinetic%20energy>

³⁹ <http://en.wikipedia.org/wiki/Free%20particle>

⁴⁰ <http://en.wikipedia.org/wiki/Dispersion%20relation>

⁴¹ <http://en.wikipedia.org/wiki/Dot%20product>

⁴² <http://en.wikipedia.org/wiki/Field%20%28physics%29>

⁴³ <http://en.wikipedia.org/wiki/Differential%20operator>

In three spatial dimensions and with both potential energy $V(\mathbf{r}, t)$ and potential momentum $\mathbf{A}(\mathbf{r}, t)$ present, we proceed from the relation $E - V = (\mathbf{p} - \mathbf{A})^2/2m$, substituting $i\hbar\frac{\partial}{\partial t}$ for E and $-i\hbar\frac{\partial}{\partial \mathbf{r}}$ for \mathbf{p} . The differential operator $\frac{\partial}{\partial \mathbf{r}}$ is a vector whose components are the differential operators $\left(\frac{\partial\psi}{\partial x}, \frac{\partial\psi}{\partial y}, \frac{\partial\psi}{\partial z}\right)$. The result:

$$i\hbar\frac{\partial\psi}{\partial t} = \frac{1}{2m} \left(-i\hbar\frac{\partial}{\partial \mathbf{r}} - \mathbf{A} \right)^2 \psi + V\psi,$$

where ψ is now a function of $\mathbf{r} = (x, y, z)$ and t . This is the three-dimensional **Schrödinger equation**. In non-relativistic investigations (to which the Schrödinger equation is confined) the potential momentum can generally be ignored, which is why the Schrödinger equation⁴⁴ is often given this form:

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m} \left(\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} \right) + V\psi$$

The *free* Schrödinger equation (without even the potential energy term) is satisfied by $\psi(x, t) = e^{i(kx - \omega t)}$ (in one dimension) or $\psi(\mathbf{r}, t) = e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)}$ (in three dimensions) *provided that* $E = \hbar\omega$ equals $p^2/2m = (\hbar k)^2/2m$, which is to say: $\omega(k) = \hbar k^2/2m$. However, since we are dealing with a homogeneous linear differential equation⁴⁵ — which tells us that solutions may be added and/or multiplied by an arbitrary constant to yield additional solutions — any function of the form

$$\psi(x, t) = \frac{1}{\sqrt{2\pi}} \int \bar{\psi}(k) e^{i[kx - \omega(k)t]} dk = \frac{1}{\sqrt{2\pi}} \int \bar{\psi}(k, t) e^{ikx} dk$$

with $\bar{\psi}(k, t) = \bar{\psi}(k) e^{-i\omega(k)t}$ solves the (one-dimensional) Schrödinger equation. If no integration boundaries are specified, then we integrate over the real line⁴⁶, i.e., the integral is defined as the limit $\lim_{L \rightarrow \infty} \int_{-L}^{+L}$. The converse also holds: every solution is of this form. The factor in front of the integral is present for purely cosmetic reasons, as you will realize presently. $\bar{\psi}(k, t)$ is the Fourier transform⁴⁷ of $\psi(x, t)$, which means that

$$\bar{\psi}(k, t) = \frac{1}{\sqrt{2\pi}} \int \psi(x, t) e^{-ikx} dx.$$

The Fourier transform of $\psi(x, t)$ exists because the integral $\int |\psi(x, t)| dx$ is finite. In the next section⁴⁸ we will come to know the physical reason why this integral is finite.

So now we have a condition that every electron "wave function" must satisfy in order to satisfy the appropriate dispersion relation. If this (and hence the Schrödinger equation) contains either or both of the potentials⁴⁹ \mathbf{V} and \mathbf{A} , then finding solutions can be tough. As a budding quantum mechanician, you will spend a considerable amount of time learning to solve the Schrödinger equation with various potentials.

⁴⁴ <http://en.wikipedia.org/wiki/Schr%F6dinger%20equation>

⁴⁵ <http://en.wikipedia.org/wiki/Differential%20equation>

⁴⁶ <http://en.wikipedia.org/wiki/Real%20line>

⁴⁷ <http://en.wikipedia.org/wiki/Fourier%20transform>

⁴⁸ Chapter 2.5 on page 26

⁴⁹ <http://en.wikipedia.org/wiki/Potential>

2.6 Born

In the same year that Erwin Schrödinger published the equation that now bears his name, the nonrelativistic theory was completed by Max Born's⁵¹ insight that the Schrödinger wave function⁵² $\psi(\mathbf{r}, t)$ is actually nothing but a tool for calculating probabilities, and that the probability of detecting a particle "described by" $\psi(\mathbf{r}, t)$ in a region of space R is given by the volume integral⁵³

$$\int_R |\psi(t, \mathbf{r})|^2 d^3r = \int_R \psi^* \psi d^3r$$

— provided that the appropriate measurement is made, in this case a test for the particle's presence in R . Since the probability of finding the particle *somewhere* (no matter where) has to be 1, only a square integrable⁵⁴ function can "describe" a particle. This rules out $\psi(\mathbf{r}) = e^{i\mathbf{k} \cdot \mathbf{r}}$, which is not square integrable. In other words, no particle can have a momentum so sharp as to be given by \hbar times a wave vector⁵⁵ \mathbf{k} , rather than by a genuine probability distribution over different momenta.

Given a probability density function $|\psi(x)|^2$, we can define the expected value⁵⁶

$$\langle x \rangle = \int |\psi(x)|^2 x dx = \int \psi^* x \psi dx$$

and the standard deviation⁵⁷ $\Delta x = \sqrt{\int |\psi|^2 (x - \langle x \rangle)^2}$

as well as higher moments⁵⁸ of $|\psi(x)|^2$. By the same token,

$$\langle k \rangle = \int \bar{\psi}^* k \bar{\psi} dk \text{ and } \Delta k = \sqrt{\int |\bar{\psi}|^2 (k - \langle k \rangle)^2}.$$

Here is another expression for $\langle k \rangle$:

$$\langle k \rangle = \int \psi^*(x) \left(-i \frac{d}{dx} \right) \psi(x) dx.$$

To check that the two expressions are in fact equal, we plug $\psi(x) = (2\pi)^{-1/2} \int \bar{\psi}(k) e^{ikx} dk$ into the latter expression:

$$\langle k \rangle = \frac{1}{\sqrt{2\pi}} \int \psi^*(x) \left(-i \frac{d}{dx} \right) \int \bar{\psi}(k) e^{ikx} dk dx = \frac{1}{\sqrt{2\pi}} \int \psi^*(x) \int \bar{\psi}(k) k e^{ikx} dk dx.$$

50 <http://en.wikibooks.org/wiki/Category%3A>

51 <http://en.wikipedia.org/wiki/Max%20Born>

52 <http://en.wikipedia.org/wiki/Wavefunction>

53 <http://en.wikipedia.org/wiki/Volume%20integral>

54 <http://en.wikipedia.org/wiki/Integrable%20function>

55 <http://en.wikipedia.org/wiki/Wave%20vector>

56 <http://en.wikipedia.org/wiki/Expected%20value>

57 <http://en.wikipedia.org/wiki/Standard%20deviation>

58 <http://en.wikipedia.org/wiki/Moment%20%28mathematics%29>

Next we replace $\psi^*(x)$ by $(2\pi)^{-1/2} \int \bar{\psi}^*(k') e^{-ik'x} dk'$ and shuffle the integrals with the mathematical nonchalance that is common in physics:

$$\langle k \rangle = \int \int \bar{\psi}^*(k') k \bar{\psi}(k) \left[\frac{1}{2\pi} \int e^{i(k-k')x} dx \right] dk dk'.$$

The expression in square brackets is a representation of Dirac's delta distribution⁵⁹ $\delta(k - k')$, the defining characteristic of which is $\int_{-\infty}^{+\infty} f(x) \delta(x) dx = f(0)$ for any continuous function $f(x)$. (In case you didn't notice, this proves what was to be proved.)

2.7 Heisenberg

In the same *annus mirabilis* of quantum mechanics, 1926, Werner Heisenberg⁶⁰ proved the so-called "uncertainty" relation⁶¹

$$\Delta x \Delta p \geq \hbar/2.$$

Heisenberg spoke of *Unschärfe*, the literal translation of which is "fuzziness" rather than "uncertainty". Since the relation $\Delta x \Delta k \geq 1/2$ is a consequence of the fact that $\psi(x)$ and $\bar{\psi}(k)$ are related to each other via a Fourier transformation⁶², we leave the proof to the mathematicians. The fuzziness relation for position and momentum follows via $p = \hbar k$. It says that the fuzziness of a position (as measured by Δx) and the fuzziness of the corresponding momentum (as measured by $\Delta p = \hbar \Delta k$) must be such that their product equals at least $\hbar/2$.

⁶³

59 <http://en.wikipedia.org/wiki/Dirac%20delta%20function>

60 <http://en.wikipedia.org/wiki/Werner%20Heisenberg>

61 <http://en.wikipedia.org/wiki/Uncertainty%20principle>

62 <http://en.wikipedia.org/wiki/Fourier%20transform>

63 <http://en.wikibooks.org/wiki/Category%3A>

3 The Feynman route to Schrödinger

The probabilities of the possible outcomes of measurements performed at a time t_2 are determined by the Schrödinger wave function $\psi(\mathbf{r}, t_2)$. The wave function $\psi(\mathbf{r}, t_2)$ is determined via the Schrödinger equation¹ by $\psi(\mathbf{r}, t_1)$. What determines $\psi(\mathbf{r}, t_1)$? Why, the outcome of a measurement performed at t_1 — what else? Actual measurement outcomes determine the probabilities of possible measurement outcomes.

3.1 Two rules

In this chapter we develop the quantum-mechanical probability algorithm from two fundamental rules. To begin with, two definitions:

- **Alternatives** are possible sequences of measurement outcomes.
- With each alternative is associated a complex number² called **amplitude**.

Suppose that you want to calculate the probability of a possible outcome of a measurement given the actual outcome of an earlier measurement. Here is what you have to do:

- Choose any sequence of measurements that may be made in the meantime.
- Assign an amplitude to each alternative.
- Apply either of the following rules:

Rule A: If the intermediate measurements are made (or if it is possible to infer from other measurements what their outcomes would have been if they had been made), first square the absolute values of the amplitudes of the alternatives and then add the results.

Rule B: If the intermediate measurements are not made (and if it is not possible to infer from other measurements what their outcomes would have been), first add the amplitudes of the alternatives and then square the absolute value of the result.

In subsequent sections we will explore the consequences of these rules for a variety of setups, and we will think about their origin — their *raison d'être*. Here we shall use Rule B to determine the interpretation of $\bar{\psi}(k)$ given Born's probabilistic interpretation of $\psi(x)$.

In the so-called "continuum normalization", the unphysical limit of a particle with a sharp momentum $\hbar k'$ is associated with the wave function

$$\psi_{k'}(x, t) = \frac{1}{\sqrt{2\pi}} \int \delta(k - k') e^{i[kx - \omega(k)t]} dk = \frac{1}{\sqrt{2\pi}} e^{i[k'x - \omega(k')t]}.$$

¹ Chapter 2.4 on page 23

² <http://en.wikipedia.org/wiki/Complex%20number>

Hence we may write $\psi(x, t) = \int \bar{\psi}(k) \psi_k(x, t) dk$.

$\bar{\psi}(k)$ is the amplitude for the outcome $\hbar k$ of an infinitely precise momentum measurement. $\psi_k(x, t)$ is the amplitude for the outcome x of an infinitely precise position measurement performed (at time t) subsequent to an infinitely precise momentum measurement with outcome $\hbar k$. And $\psi(x, t)$ is the amplitude for obtaining x by an infinitely precise position measurement performed at time t .

The preceding equation therefore tells us that the *amplitude* for finding x at t is the product of

1. the *amplitude* for the outcome $\hbar k$ and
2. the *amplitude* for the outcome x (at time t) subsequent to a momentum measurement with outcome $\hbar k$,

summed over all values of k .

Under the conditions stipulated by Rule A, we would have instead that the *probability* for finding x at t is the product of

1. the *probability* for the outcome $\hbar k$ and
2. the *probability* for the outcome x (at time t) subsequent to a momentum measurement with outcome $\hbar k$,

summed over all values of k .

The latter is what we expect on the basis of standard probability theory. But if this holds under the conditions stipulated by Rule A, then the same holds with "amplitude" substituted from "probability" under the conditions stipulated by Rule B. Hence, given that $\psi_k(x, t)$ and $\psi(x, t)$ are amplitudes for obtaining the outcome x in an infinitely precise position measurement, $\bar{\psi}(k)$ is the amplitude for obtaining the outcome $\hbar k$ in an infinitely precise momentum measurement.

Notes:

1. Since Rule B stipulates that the momentum measurement is not actually made, we need not worry about the impossibility of making an infinitely precise momentum measurement.
2. If we refer to $|\psi(x)|^2$ as "the probability of obtaining the outcome x ," what we mean is that $|\psi(x)|^2$ integrated over any interval or subset of the real line³ is the probability of finding our particle in this interval or subset.

³ <http://en.wikipedia.org/wiki/Real%20line>

3.2 An experiment with two slits

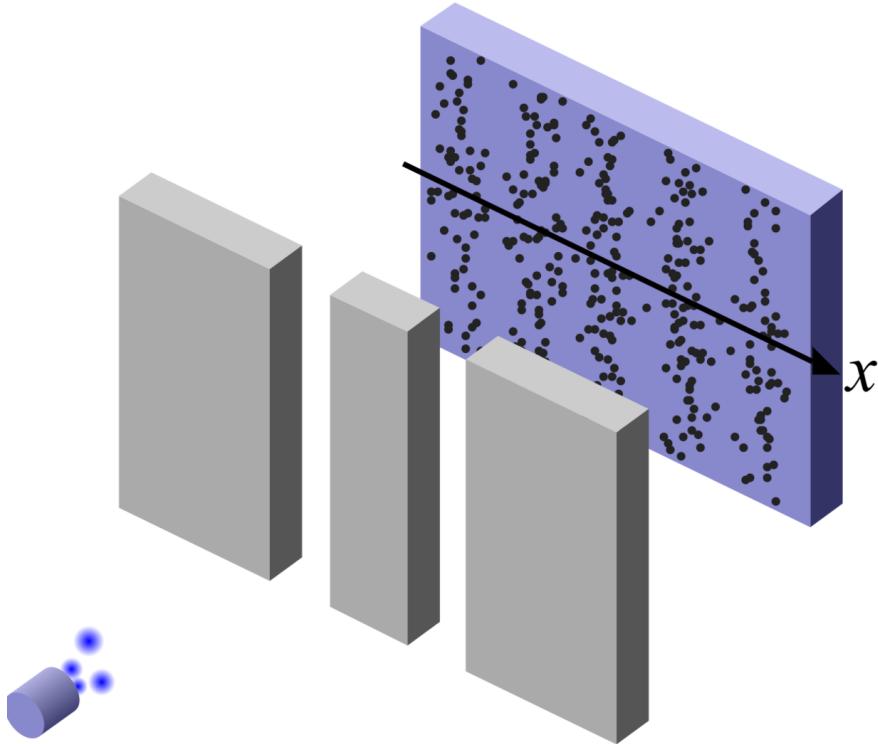


Figure 27 The setup

In this experiment, the final measurement (to the possible outcomes of which probabilities are assigned) is the detection of an electron at the backdrop, by a detector situated at **D** (**D** being a particular value of **x**). The initial measurement outcome, on the basis of which probabilities are assigned, is the launch of an electron by an electron gun **G**. (Since we assume that **G** is the only source of free electrons, the detection of an electron behind the slit plate also indicates the launch of an electron in front of the slit plate.) The alternatives or possible intermediate outcomes are

- the electron went through the left slit (**L**),
- the electron went through the right slit (**R**).

The corresponding amplitudes are A_L and A_R .

Here is what we need to know in order to calculate them:

- A_L is the product of two complex numbers, for which we shall use the symbols $\langle D|L \rangle$ and $\langle L|G \rangle$.
- By the same token, $A_R = \langle D|R \rangle \langle R|G \rangle$.

- The absolute value of $\langle B|A \rangle$ is inverse proportional to the distance $d(BA)$ between **A** and **B**.
- The phase of $\langle B|A \rangle$ is proportional to $d(BA)$.

For obvious reasons $\langle B|A \rangle$ is known as a *propagator*.

3.2.1 Why product?

Recall the fuzziness ("uncertainty") relation⁴, which implies that $\Delta p \rightarrow \infty$ as $\Delta x \rightarrow 0$. In this limit the particle's momentum is completely indefinite or, what comes to the same, has no value at all. As a consequence, the probability of finding a particle at **B**, given that it was last "seen" at **A**, depends on the initial position **A** but not on any initial momentum, inasmuch as there is none. Hence whatever the particle does after its detection at **A** is independent of what it did before then. In probability-theoretic terms this means that the particle's propagation from **G** to **L** and its propagation from **L** to **D** are independent events. So the probability of propagation from **G** to **D** via **L** is the product of the corresponding probabilities, and so the amplitude of propagation from **G** to **D** via **L** is the product $\langle D|L \rangle \langle L|G \rangle$ of the corresponding amplitudes.

3.2.2 Why is the absolute value inverse proportional to the distance?

Imagine (i) a sphere of radius r whose center is **A** and (ii) a detector monitoring a unit area of the surface of this sphere. Since the total surface area is proportional to r^2 , and since for a free particle the probability of detection per unit area is constant over the entire surface (explain why!), the probability of detection per unit area is inverse proportional to r^2 . The absolute value of the *amplitude* of detection per unit area, being the square root of the probability, is therefore inverse proportional to r .

3.2.3 Why is the phase proportional to the distance?

The multiplicativity of successive propagators implies the additivity of their phases. Together with the fact that, in the case of a free particle, the propagator $\langle B|A \rangle$ (and hence its phase) can only depend on the distance between **A** and **B**, it implies the proportionality of the phase of $\langle B|A \rangle$ to $d(BA)$.

3.2.4 Calculating the interference pattern

According to Rule A, the probability of detecting at **D** an electron launched at **G** is

$$p_A(D) = |\langle D|L \rangle \langle L|G \rangle|^2 + |\langle D|R \rangle \langle R|G \rangle|^2.$$

If the slits are equidistant from **G**, then $\langle L|G \rangle$ and $\langle R|G \rangle$ are equal and $p_A(D)$ is proportional to

⁴ Chapter 2.5 on page 26

$$|\langle D|L \rangle|^2 + |\langle D|R \rangle|^2 = 1/d^2(DL) + 1/d^2(DR).$$

Here is the resulting plot of p_A against the position x of the detector:

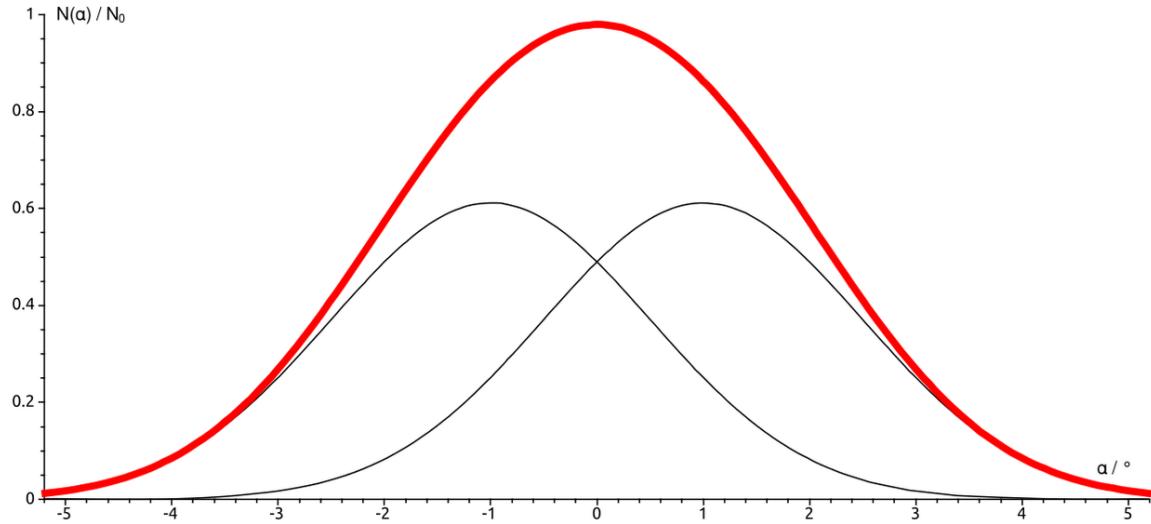


Figure 28 Predicted relative frequency of detection according to Rule A

$p_A(x)$ (solid line) is the sum of two distributions (dotted lines), one for the electrons that went through **L** and one for the electrons that went through **R**.

According to Rule B, the probability $p_B(D)$ of detecting at **D** an electron launched at **G** is proportional to

$$|\langle D|L \rangle + \langle D|R \rangle|^2 = 1/d^2(DL) + 1/d^2(DR) + 2 \cos(k\Delta)/[d(DL)d(DR)],$$

where Δ is the difference $d(DR) - d(DL)$ and $k = p/\hbar$ is the wavenumber, which is sufficiently sharp to be approximated by a number. (And it goes without saying that you should check this result.)

Here is the plot of p_B against x for a particular set of values for the wavenumber, the distance between the slits, and the distance between the slit plate and the backdrop:

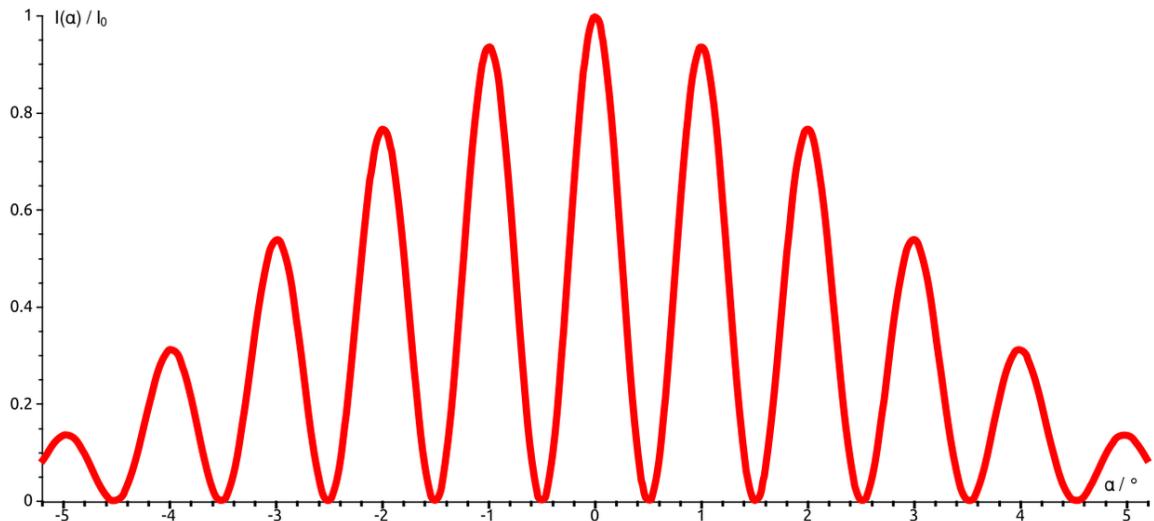


Figure 29 Predicted relative frequency of detection according to Rule B

Observe that near the minima the probability of detection is *less* if both slits are open than it is if one slit is shut. It is customary to say that destructive interference occurs at the minima and that constructive interference occurs at the maxima, but *do not* think of this as the description of a physical process. All we mean by "constructive interference" is that a probability calculated according to Rule B is greater than the same probability calculated according to Rule A, and all we mean by "destructive interference" is that a probability calculated according to Rule B is less than the same probability calculated according to Rule A.

Here is how an interference pattern builds up over time⁵:

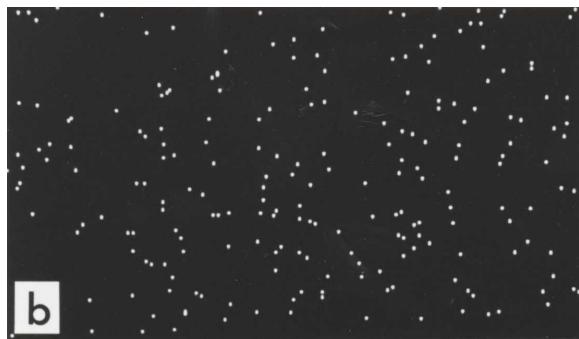


Figure 30 100 electrons

⁵ A. Tonomura, J. Endo, T. Matsuda, T. Kawasaki, & H. Ezawa, "Demonstration of single-electron buildup of an interference pattern", American Journal of Physics **57**, 117-120, 1989.

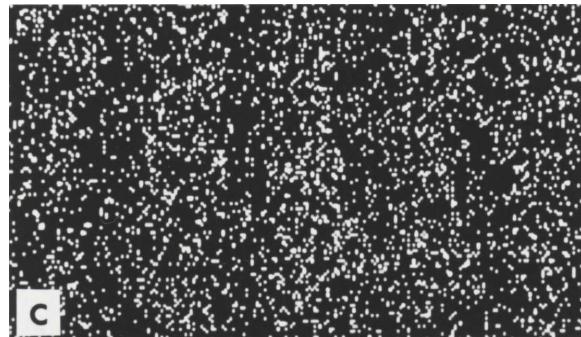


Figure 31 3000 electrons

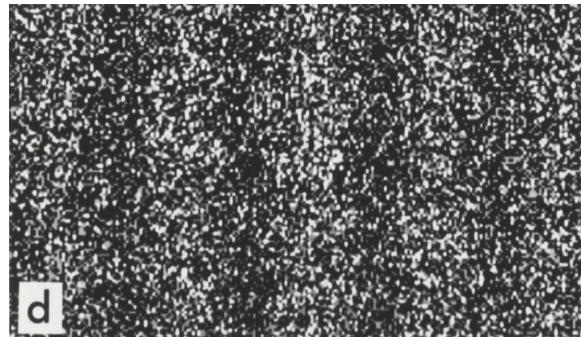


Figure 32 20000 electrons

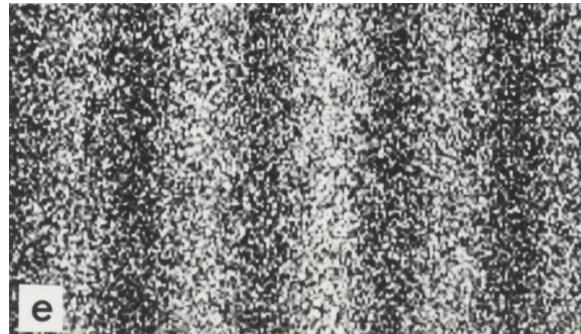


Figure 33 70000 electrons

6

3.3 Bohm's story

3.3.1 Hidden Variables

Suppose that the conditions stipulated by Rule B⁷ are met: there is nothing — no event, no state of affairs, anywhere, anytime — from which the slit taken by an electron can be inferred. Can it be true, in this case,

6 http://en.wikibooks.org/wiki/Category%3A_Quantum_mechanics

7 Chapter 2.7 on page 27

- that each electron goes through a single slit — either **L** or **R** — and
- that the behavior of an electron that goes through one slit does not depend on whether the other slit is open or shut?

To keep the language simple, we will say that an electron leaves a mark where it is detected at the backdrop. If each electron goes through a single slit, then the observed distribution of marks when both slits are open is the sum of two distributions, one from electrons that went through **L** and one from electrons that went through **R**:

$$p_B(x) = p_L(x) + p_R(x)$$

If in addition the behavior of an electron that goes through one slit does not depend on whether the other slit is open or shut, then we can observe $p_L(x)$ by keeping **R** shut, and we can observe $p_R(x)$ by keeping **L** shut. What we observe if **R** is shut is the left dashed hump, and what we observed if **L** is shut is the right dashed hump:

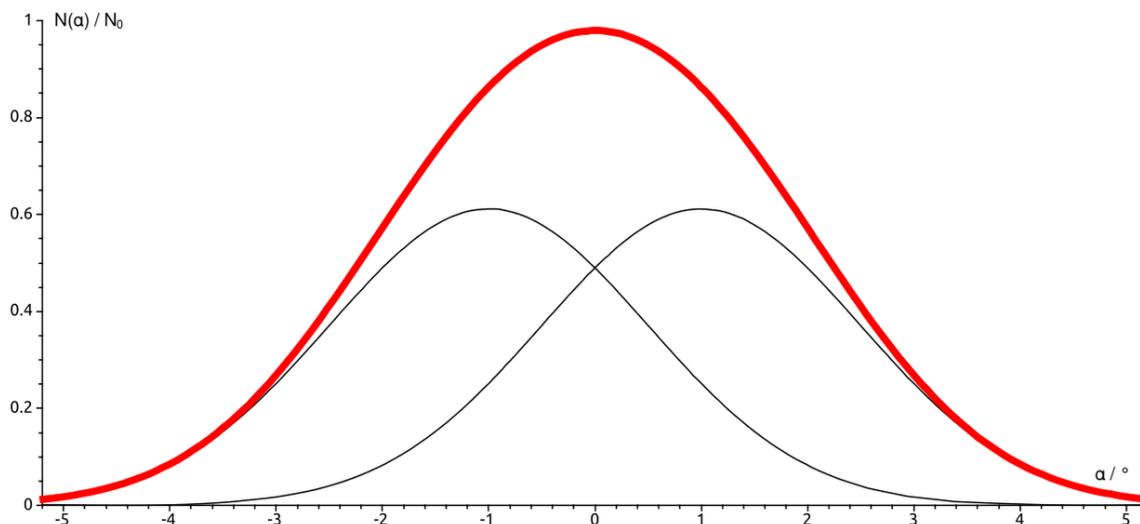
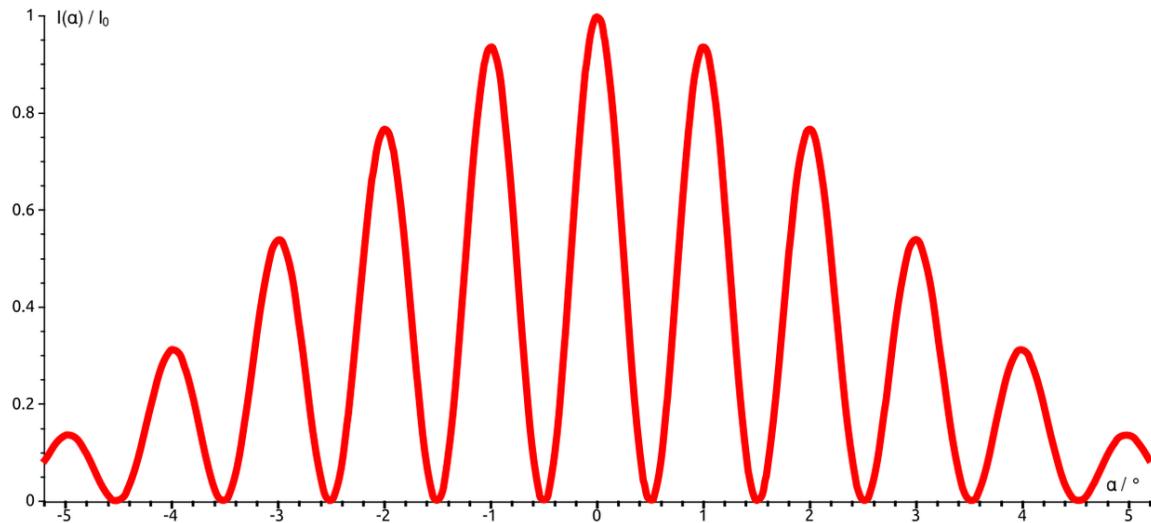


Figure 34

Hence if the above two conditions (as well as those stipulated by Rule B) are satisfied, we will see the sum of these two humps. In reality what we see is this:

**Figure 35**

Thus all of those conditions cannot be simultaneously satisfied. If Rule B applies, then either it is false that each electron goes through a single slit or the behavior of an electron that goes through one slit does depend on whether the other slit is open or shut.

Which is it?

According to one attempt to make physical sense of the mathematical formalism of quantum mechanics, due to Louis de Broglie⁸ and David Bohm⁹, each electron goes through a single slit, and the behavior of an electron that goes through one slit depends on whether the other slit is open or shut.

So how does the state of, say, the right slit (open or shut) affect the behavior of an electron that goes through the left slit? In both de Broglie's pilot wave theory and Bohmian mechanics¹⁰, the electron is assumed to be a well-behaved particle in the sense that it follows a precise path — its position at any moment is given by three coordinates — and in addition there exists a wave that guides the electron by exerting on it a force. If only one slit is open, this passes through one slit. If both slits are open, this passes through both slits and interferes with itself (in the "classical" sense of interference). As a result, it guides the electrons along wiggly paths that cluster at the backdrop so as to produce the observed interference pattern:

⁸ <http://en.wikipedia.org/wiki/Louis%2C%207th%20duc%20de%20Broglie>

⁹ <http://en.wikipedia.org/wiki/David%20Bohm>

¹⁰ <http://en.wikipedia.org/wiki/Bohm%20interpretation>

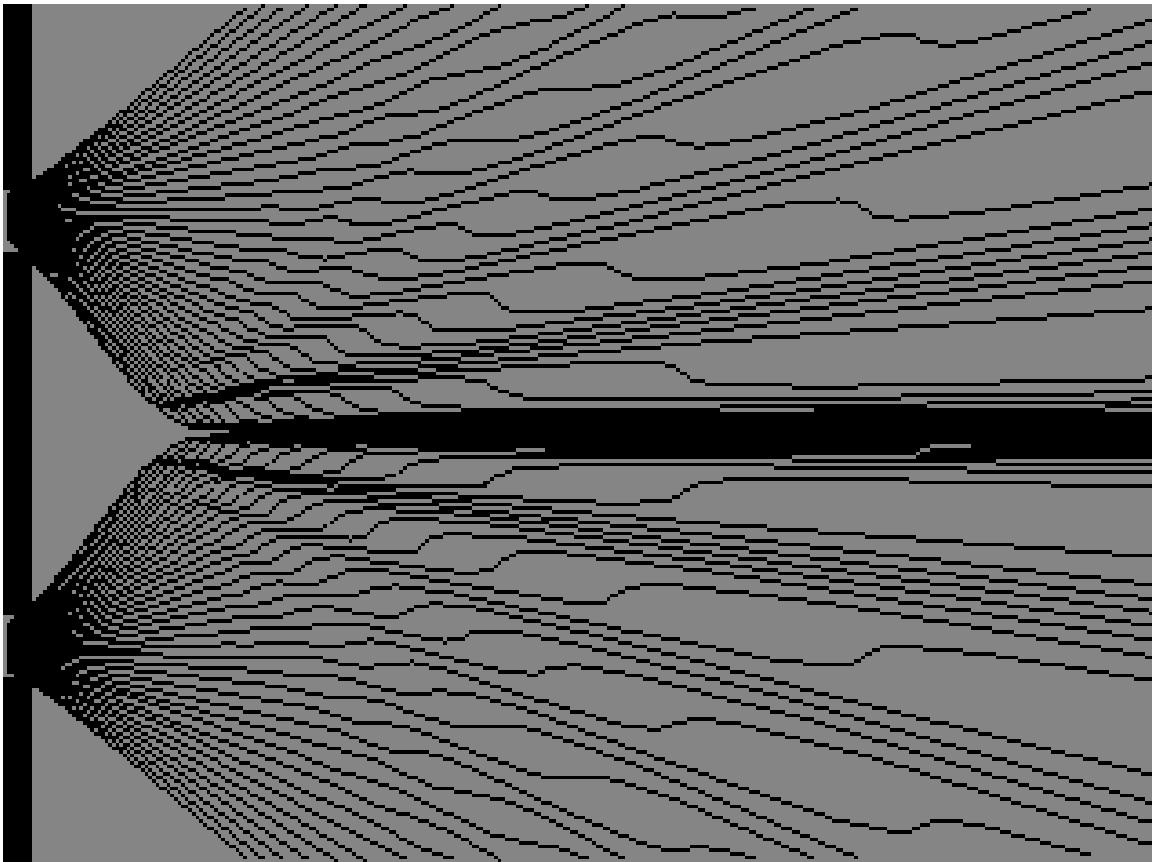


Figure 36 none

According to this story, the reason why electrons coming from the same source or slit arrive in different places, is that they start out in slightly different directions and/or with slightly different speeds. If we had exact knowledge of their initial positions and momenta, we could make an exact prediction of each electron's subsequent motion. This, however, is impossible. The uncertainty principle¹¹ prevents us from making exact predictions of a particle's motion. Hence even though according to Bohm the initial positions and momenta are in possession of precise values, we can never know them.

If positions and momenta have precise values, then why can we not measure them? It used to be said that this is because a measurement exerts an uncontrollable influence on the value of the observable being measured. Yet this merely raises another question: why do measurements exert uncontrollable influences? This may be true for all practical purposes, but the uncertainty principle does not say that $\Delta x \Delta p \geq \hbar/2$ merely holds for all practical purposes. Moreover, it isn't the case that measurements necessarily "disturb" the systems on which they are performed.

The statistical element of quantum mechanics is an essential feature of the theory. The postulate of an underlying determinism, which in order to be consistent with the theory has to be a *crypto*¹²-determinism, not only adds nothing to our understanding of the theory but

¹¹ Chapter 2.7 on page 27

¹² <http://en.wikipedia.org/wiki/Crypto>

also precludes any proper understanding of this essential feature of the theory. There is, in fact, a simple and obvious reason why hidden variables¹³ are hidden: the reason why they are strictly (rather than merely for all practical purposes) unobservable is that *they do not exist*.

At one time Einstein insisted that theories ought to be formulated without reference to unobservable quantities. When Heisenberg later mentioned to Einstein that this maxim had guided him in his discovery of the uncertainty principle, Einstein replied something to this effect: "Even if I once said so, it is nonsense." His point was that before one has a theory, one cannot know what is observable and what is not. Our situation here is different. We have a theory, and this tells in no uncertain terms what is observable and what is not.

3.4 Propagator for a free and stable particle

3.4.1 The propagator as a path integral

Suppose that we make \mathbf{m} intermediate position measurements at fixed intervals of duration Δt . Each of these measurements is made with the help of an array of detectors monitoring \mathbf{n} mutually disjoint regions R_k , $k = 1, \dots, n$. Under the conditions stipulated by **Rule B**, the propagator¹⁴ $\langle B | A \rangle$ now equals the sum of amplitudes

$$\sum_{k_1=1}^n \cdots \sum_{k_m=1}^n \langle B | R_{k_m} \rangle \cdots \langle R_{k_2} | R_{k_1} \rangle \langle R_{k_1} | A \rangle.$$

It is not hard to see what happens in the double limit $\Delta t \rightarrow 0$ (which implies that $m \rightarrow \infty$) and $n \rightarrow \infty$. The multiple sum $\sum_{k_1=1}^n \cdots \sum_{k_m=1}^n$ becomes an integral $\int \mathcal{D}\mathcal{C}$ over continuous spacetime paths from **A** to **B**, and the amplitude $\langle B | R_{k_m} \rangle \cdots \langle R_{k_1} | A \rangle$ becomes a complex-valued functional $Z[\mathcal{C} : A \rightarrow B]$ — a complex function of continuous functions representing continuous spacetime paths from **A** to **B**:

$$\langle B | A \rangle = \int \mathcal{D}\mathcal{C} Z[\mathcal{C} : A \rightarrow B]$$

The integral $\int \mathcal{D}\mathcal{C}$ is not your standard Riemann integral¹⁵ $\int_a^b dx f(x)$, to which each infinitesimal interval dx makes a contribution proportional to the value that $f(x)$ takes inside the interval, but a functional or path integral¹⁶, to which each "bundle" of paths of infinitesimal width $\mathcal{D}\mathcal{C}$ makes a contribution proportional to the value that $Z[\mathcal{C}]$ takes inside the bundle.

As it stands, the path integral $\int \mathcal{D}\mathcal{C}$ is just the idea of an idea. Appropriate evaluation methods have to be devised on a more or less case-by-case basis.

¹³ [http://en.wikipedia.org/wiki/Hidden%20variable%20theory](http://en.wikipedia.org/wiki/Hidden_variable_theory)

¹⁴ Chapter 3.1 on page 30

¹⁵ [http://en.wikipedia.org/wiki/Riemann%20integral](http://en.wikipedia.org/wiki/Riemann_integral)

¹⁶ <http://en.wikipedia.org/wiki/Functional%20integration>

3.4.2 A free particle

Now pick any path \mathcal{C} from **A** to **B**, and then pick any infinitesimal segment $d\mathcal{C}$ of \mathcal{C} . Label the start and end points of $d\mathcal{C}$ by inertial coordinates¹⁷ t, x, y, z and $t + dt, x + dx, y + dy, z + dz$, respectively. In the general case, the amplitude $Z(d\mathcal{C})$ will be a function of t, x, y, z and dt, dx, dy, dz . In the case of a free particle, $Z(d\mathcal{C})$ depends neither on the position of $d\mathcal{C}$ in spacetime (given by t, x, y, z) nor on the spacetime orientation of $d\mathcal{C}$ (given by the four-velocity¹⁸ $(cdt/ds, dx/ds, dy/ds, dz/ds)$) but only on the proper time¹⁹ interval $ds = \sqrt{dt^2 - (dx^2 + dy^2 + dz^2)/c^2}$.

(Because its norm equals the speed of light, the four-velocity depends on three rather than four independent parameters. Together with ds , they contain the same information as the four independent numbers dt, dx, dy, dz .)

Thus for a free particle $Z(d\mathcal{C}) = Z(ds)$. With this, the multiplicativity of successive propagators²⁰ tells us that

$$\prod_j Z(ds_j) = Z\left(\sum_j ds_j\right) \longrightarrow Z\left(\int_{\mathcal{C}} ds\right)$$

It follows that there is a complex number z such that $Z[\mathcal{C}] = e^{zs[\mathcal{C}:A \rightarrow B]}$, where the line integral²¹ $s[\mathcal{C} : A \rightarrow B] = \int_{\mathcal{C}} ds$ gives the time that passes on a clock as it travels from **A** to **B** via \mathcal{C} .

3.4.3 A free and stable particle

By integrating $|\langle B|A \rangle|^2$ (as a function of \mathbf{r}_B) over the whole of space, we obtain the probability of finding that a particle launched at the spacetime point t_A, \mathbf{r}_A still exists at the time t_B . For a stable particle this probability equals 1:

$${}^3 r_B |\langle t_B, \mathbf{r}_B | t_A, \mathbf{r}_A \rangle|^2 = {}^3 r_B \left| \int \mathcal{D}\mathcal{C} e^{zs[\mathcal{C}:A \rightarrow B]} \right|^2 = 1$$

If you contemplate this equation with a calm heart and an open mind, you will notice that if the complex number $z = a + ib$ had a real part $a \neq 0$, then the integral between the two equal signs would either blow up ($a > 0$) or drop off ($a < 0$) exponentially as a function of t_B , due to the exponential factor $e^{as[\mathcal{C}]}$.

3.4.4 Meaning of mass

The propagator for a free and stable particle thus has a single "degree of freedom": it depends solely on the value of b . If proper time is measured in seconds, then b is measured in radians

¹⁷ <http://en.wikipedia.org/wiki/Inertial%20frame%20of%20reference>

¹⁸ <http://en.wikipedia.org/wiki/Four-velocity>

¹⁹ <http://en.wikipedia.org/wiki/Proper%20time>

²⁰ Chapter 3.2.1 on page 32

²¹ <http://en.wikipedia.org/wiki/Line%20integral>

per second. We may think of $e^{ib s}$, with s a proper-time parametrization of \mathcal{C} , as a *clock* carried by a particle that travels from **A** to **B** via \mathcal{C} , provided we keep in mind that we are thinking of an aspect of the mathematical formalism of quantum mechanics rather than an aspect of the real world.

It is customary

- to insert a minus (so the clock actually turns clockwise!): $Z = e^{-ib s[\mathcal{C}]}$,
- to multiply by 2π (so that we may think of b as the rate at which the clock "ticks" — the number of cycles it completes each second): $Z = e^{-i2\pi b s[\mathcal{C}]}$,
- to divide by Planck's constant \hbar (so that b is measured in energy units and called the rest energy²² of the particle): $Z = e^{-i(2\pi/\hbar)b s[\mathcal{C}]} = e^{-(i/\hbar)b s[\mathcal{C}]}$,
- and to multiply by c^2 (so that b is measured in mass units and called the particle's rest mass²³): $Z = e^{-(i/\hbar)bc^2 s[\mathcal{C}]}$.

The purpose of using the same letter b everywhere is to emphasize that it denotes the same physical quantity, merely measured in different units. If we use natural units²⁴ in which $\hbar = c = 1$, rather than conventional ones, the identity of the various b 's is immediately obvious.

3.5 From quantum to classical

3.5.1 Action

Let's go back to the propagator

$$\langle B|A \rangle = \int \mathcal{D}\mathcal{C} Z[\mathcal{C} : A \rightarrow B].$$

For a free and stable particle we found that

$$Z[\mathcal{C}] = e^{-(i/\hbar)mc^2 s[\mathcal{C}]}, \quad s[\mathcal{C}] = \int_{\mathcal{C}} ds,$$

where $ds = \sqrt{dt^2 - (dx^2 + dy^2 + dz^2)/c^2}$ is the proper-time interval associated with the path element $d\mathcal{C}$. For the general case we found that the amplitude $Z(d\mathcal{C})$ is a function of t, x, y, z and dt, dx, dy, dz or, equivalently, of the coordinates t, x, y, z , the components $c dt/ds, dx/ds, dy/ds, dz/ds$ of the 4-velocity, as well as ds . For a particle that is stable but not free, we obtain, by the same argument that led to the above amplitude,

$$Z[\mathcal{C}] = e^{(i/\hbar)S[\mathcal{C}]},$$

²² <http://en.wikipedia.org/wiki/Rest%20energy>

²³ <http://en.wikipedia.org/wiki/Invariant%20mass>

²⁴ <http://en.wikipedia.org/wiki/Natural%20units>

where we have introduced the functional $S[\mathcal{C}] = \int_{\mathcal{C}} dS$, which goes by the name *action*.

For a free and stable particle, $S[\mathcal{C}]$ is the proper time (or proper duration) $s[\mathcal{C}] = \int_{\mathcal{C}} ds$ multiplied by $-mc^2$, and the infinitesimal action $dS[d\mathcal{C}]$ is proportional to ds :

$$S[\mathcal{C}] = -mc^2 s[\mathcal{C}], \quad dS[d\mathcal{C}] = -mc^2 ds.$$

Let's recap. We know all about the motion of a stable particle if we know how to calculate the probability $p(B|A)$ (in all circumstances). We know this if we know the amplitude $\langle B|A \rangle$. We know the latter if we know the functional $Z[\mathcal{C}]$. And we know this functional if we know the infinitesimal action $dS(t, x, y, z, dt, dx, dy, dz)$ or $dS(t, \mathbf{r}, dt, d\mathbf{r})$ (in all circumstances).

What do we know about dS ?

The multiplicativity of successive propagators implies the additivity of actions associated with neighboring infinitesimal path segments $d\mathcal{C}_1$ and $d\mathcal{C}_2$. In other words,

$$e^{(i/\hbar)dS(d\mathcal{C}_1+d\mathcal{C}_2)} = e^{(i/\hbar)dS(d\mathcal{C}_2)} e^{(i/\hbar)dS(d\mathcal{C}_1)}$$

implies

$$dS(d\mathcal{C}_1 + d\mathcal{C}_2) = dS(d\mathcal{C}_1) + dS(d\mathcal{C}_2).$$

It follows that the differential dS is homogeneous (of degree 1) in the differentials $dt, d\mathbf{r}$:

$$dS(t, \mathbf{r}, \lambda dt, \lambda d\mathbf{r}) = \lambda dS(t, \mathbf{r}, dt, d\mathbf{r}).$$

This property of dS makes it possible to think of the action $S[\mathcal{C}]$ as a (particle-specific) length associated with \mathcal{C} , and of dS as defining a (particle-specific) spacetime geometry. By substituting $1/dt$ for λ we get:

$$dS(t, \mathbf{r}, \mathbf{v}) = \frac{ds}{dt}.$$

Something is wrong, isn't it? Since the right-hand side is now a finite quantity, we shouldn't use the symbol dS for the left-hand side. What we have actually found is that there is a function $L(t, \mathbf{r}, \mathbf{v})$, which goes by the name *Lagrange function*, such that $dS = L dt$.

3.5.2 Geodesic equations

Consider a spacetime path \mathcal{C} from A to B . Let's change ("vary") it in such a way that every point (t, \mathbf{r}) of \mathcal{C} gets shifted by an infinitesimal amount to a corresponding point $(t + \delta t, \mathbf{r} + \delta \mathbf{r})$, except the end points, which are held fixed: $\delta t = 0$ and $\delta \mathbf{r} = 0$ at both A and B .

If $t \rightarrow t + \delta t$, then $dt = t_2 - t_1 \rightarrow t_2 + \delta t_2 - (t_1 + \delta t_1) = (t_2 - t_1) + (\delta t_2 - \delta t_1) = dt + d\delta t$.

By the same token, $d\mathbf{r} \rightarrow d\mathbf{r} + d\delta \mathbf{r}$.

In general, the change $\mathcal{C} \rightarrow \mathcal{C}'$ will cause a corresponding change in the action: $S[\mathcal{C}] \rightarrow S[\mathcal{C}'] \neq S[\mathcal{C}]$. If the action does not change (that is, if it is *stationary* at \mathcal{C}),

$$\delta S = \int_{\mathcal{C}'} dS - \int_{\mathcal{C}} dS = 0,$$

then \mathcal{C} is a *geodesic* of the geometry defined by dS . (A function $f(x)$ is stationary at those values of x at which its value does not change if x changes infinitesimally. By the same token we call a functional $S[\mathcal{C}]$ stationary if its value does not change if \mathcal{C} changes infinitesimally.)

To obtain a handier way to characterize geodesics, we begin by expanding

$$dS(\mathcal{C}') = dS(t + \delta t, \mathbf{r} + \delta \mathbf{r}, dt + d\delta t, d\mathbf{r} + d\delta \mathbf{r})$$

$$= dS(t, \mathbf{r}, dt, d\mathbf{r}) + \frac{\partial dS}{\partial t} \delta t + \frac{\partial dS}{\partial \mathbf{r}} \cdot \delta \mathbf{r} + \frac{\partial dS}{\partial dt} d\delta t + \frac{\partial dS}{\partial d\mathbf{r}} \cdot d\delta \mathbf{r}.$$

This gives us

$$(*) \quad \int_{\mathcal{C}'} dS - \int_{\mathcal{C}} dS = \int_{\mathcal{C}} \left[\frac{\partial dS}{\partial t} \delta t + \frac{\partial dS}{\partial \mathbf{r}} \cdot \delta \mathbf{r} + \frac{\partial dS}{\partial dt} d\delta t + \frac{\partial dS}{\partial d\mathbf{r}} \cdot d\delta \mathbf{r} \right].$$

Next we use the product rule for derivatives,

$$d \left(\frac{\partial dS}{\partial dt} \delta t \right) = \left(d \frac{\partial dS}{\partial dt} \right) \delta t + \frac{\partial dS}{\partial dt} d\delta t,$$

$$d \left(\frac{\partial dS}{\partial d\mathbf{r}} \cdot \delta \mathbf{r} \right) = \left(d \frac{\partial dS}{\partial d\mathbf{r}} \right) \cdot \delta \mathbf{r} + \frac{\partial dS}{\partial d\mathbf{r}} \cdot d\delta \mathbf{r},$$

to replace the last two terms of (*), which takes us to

$$\delta S = \int \left[\left(\frac{\partial dS}{\partial t} - d \frac{\partial dS}{\partial dt} \right) \delta t + \left(\frac{\partial dS}{\partial \mathbf{r}} - d \frac{\partial dS}{\partial d\mathbf{r}} \right) \cdot \delta \mathbf{r} \right] + \int d \left(\frac{\partial dS}{\partial dt} \delta t + \frac{\partial dS}{\partial d\mathbf{r}} \cdot \delta \mathbf{r} \right).$$

The second integral vanishes because it is equal to the difference between the values of the expression in brackets at the end points A and B , where $\delta t = 0$ and $\delta \mathbf{r} = 0$. If \mathcal{C} is a geodesic, then the first integral vanishes, too. In fact, in this case $\delta S = 0$ must hold for all possible (infinitesimal) variations δt and $\delta \mathbf{r}$, whence it follows that the *integrand* of the first integral vanishes. The bottom line is that the geodesics defined by dS satisfy the *geodesic equations*

$$\frac{\partial dS}{\partial t} = d \frac{\partial dS}{\partial dt}, \quad \frac{\partial dS}{\partial \mathbf{r}} = d \frac{\partial dS}{\partial d\mathbf{r}}.$$

3.5.3 Principle of least action

If an object travels from A to B , it travels along all paths from A to B , in the same sense in which an electron goes through both slits. Then how is it that a big thing (such as a planet, a tennis ball, or a mosquito) appears to move along a single well-defined path?

There are at least two reasons. One of them is that the bigger an object is, the harder it is to satisfy the conditions stipulated by Rule B . Another reason is that even if these conditions are satisfied, the likelihood of finding an object of mass m where according to the laws of classical physics it should not be, decreases as m increases.

To see this, we need to take account of the fact that it is strictly impossible to check whether an object that has travelled from A to B , has done so along a mathematically precise path \mathcal{C} . Let us make the half realistic assumption that what we *can* check is whether an object has travelled from A to B within a narrow bundle of paths — the paths contained in a narrow tube \mathcal{T} . The probability of finding that it has, is the absolute square of the path integral $I(\mathcal{T}) = \int_{\mathcal{T}} \mathcal{D}\mathcal{C} e^{(i/\hbar)S[\mathcal{C}]}$, which sums over the paths contained in \mathcal{T} .

Let us assume that there is exactly one path from A to B for which $S[\mathcal{C}]$ is stationary: its length does not change if we vary the path ever so slightly, no matter how. In other words, we assume that there is exactly one geodesic. Let's call it \mathcal{G} , and let's assume it lies in \mathcal{T} .

No matter how rapidly the phase $S[\mathcal{C}]/\hbar$ changes under variation of a generic path \mathcal{C} , it will be stationary at \mathcal{G} . This means, loosely speaking, that a large number of paths near \mathcal{G} contribute to $I(\mathcal{T})$ with almost equal phases. As a consequence, the magnitude of the sum of the corresponding phase factors $e^{(i/\hbar)S[\mathcal{C}]}$ is large.

If $S[\mathcal{C}]/\hbar$ is not stationary at \mathcal{C} , all depends on how rapidly it changes under variation of \mathcal{C} . If it changes sufficiently rapidly, the phases associated with paths near \mathcal{C} are more or less equally distributed over the interval $[0, 2\pi]$, so that the corresponding phase factors add up to a complex number of comparatively small magnitude. In the limit $S[\mathcal{C}]/\hbar \rightarrow \infty$, the only significant contributions to $I(\mathcal{T})$ come from paths in the infinitesimal neighborhood of \mathcal{G} .

We have assumed that \mathcal{G} lies in \mathcal{T} . If it does not, and if $S[\mathcal{C}]/\hbar$ changes sufficiently rapidly, the phases associated with paths near *any* path in \mathcal{T} are more or less equally distributed over the interval $[0, 2\pi]$, so that in the limit $S[\mathcal{C}]/\hbar \rightarrow \infty$ there are no significant contributions to $I(\mathcal{T})$.

For a free particle, as you will remember, $S[\mathcal{C}] = -mc^2 s[\mathcal{C}]$. From this we gather that the likelihood of finding a freely moving object where according to the laws of classical physics it should not be, decreases as its mass increases. Since for sufficiently massive objects the contributions to the action due to influences on their motion are small compared to $|-mc^2 s[\mathcal{C}]|$, this is equally true of objects that are not moving freely.

What, then, *are* the laws of classical physics?

They are what the laws of quantum physics degenerate into in the limit $\hbar \rightarrow 0$. In this limit, as you will gather from the above, the probability of finding that a particle has traveled within a tube (however narrow) containing a geodesic, is 1, and the probability of finding that a particle has traveled within a tube (however wide) not containing a geodesic, is 0. Thus we may state the laws of classical physics (for a single "point mass", to begin with) by saying that it follows a geodesic of the geometry defined by dS .

This is readily generalized. The propagator for a system with n degrees of freedom — such as an m -particle system with $n = 3m$ degrees of freedom — is

$$\langle \mathcal{P}_f, t_f | \mathcal{P}_i, t_i \rangle = \int \mathcal{D}\mathcal{C} e^{(i/\hbar)S[\mathcal{C}]},$$

where \mathcal{P}_i and \mathcal{P}_f are the system's respective configurations at the initial time t_i and the final time t_f , and the integral sums over all paths in the system's $n+1$ -dimensional configuration spacetime leading from (\mathcal{P}_i, t_i) to (\mathcal{P}_f, t_f) . In this case, too, the corresponding classical system follows a geodesic of the geometry defined by the action differential dS , which now depends on n spatial coordinates, one time coordinate, and the corresponding $n+1$ differentials.

The statement that a classical system follows a geodesic of the geometry defined by its action, is often referred to as the *principle of least action*. A more appropriate name is *principle of stationary action*.

3.5.4 Energy and momentum

Observe that if dS does not depend on t (that is, $\partial dS / \partial t = 0$) then

$$E = -\frac{\partial dS}{\partial dt}$$

is constant along geodesics. (We'll discover the reason for the negative sign in a moment.)

Likewise, if dS does not depend on \mathbf{r} (that is, $\partial dS / \partial \mathbf{r} = 0$) then

$$\mathbf{p} = \frac{\partial dS}{\partial d\mathbf{r}}$$

is constant along geodesics.

E tells us how much the projection dt of a segment $d\mathcal{C}$ of a path \mathcal{C} onto the time axis contributes to the action of \mathcal{C} . \mathbf{p} tells us how much the projection $d\mathbf{r}$ of $d\mathcal{C}$ onto space contributes to $S[\mathcal{C}]$. If dS has no explicit time dependence, then equal intervals of the time axis make equal contributions to $S[\mathcal{C}]$, and if dS has no explicit space dependence, then equal intervals of any spatial axis make equal contributions to $S[\mathcal{C}]$. In the former case, equal time intervals are *physically equivalent*: they represent *equal durations*. In the latter case, equal space intervals are *physically equivalent*: they represent *equal distances*.

If equal intervals of the time coordinate or equal intervals of a space coordinate are *not* physically equivalent, this is so for either of two reasons. The first is that non-inertial coordinates are used. For if inertial coordinates are used, then *every* freely moving point mass moves by equal intervals of the space coordinates in equal intervals of the time coordinate, which means that equal coordinate intervals *are* physically equivalent. The second is that whatever it is that is moving is not moving *freely*: something, no matter what, influences its motion, no matter how. This is because one way of incorporating effects on the motion of an object into the mathematical formalism of quantum physics, is to make inertial coordinate intervals physically *inequivalent*, by letting dS depend on t and/or \mathbf{r} .

Thus for a freely moving classical object, both E and \mathbf{p} are constant. Since the constancy of E follows from the physical equivalence of equal intervals of coordinate time (a.k.a. the "homogeneity" of time), and since (classically) energy is defined as the quantity whose constancy is implied by the homogeneity of time, E is the object's *energy*.

By the same token, since the constancy of \mathbf{p} follows from the physical equivalence of equal intervals of any spatial coordinate axis (a.k.a. the "homogeneity" of space), and since (classically) momentum is defined as the quantity whose constancy is implied by the homogeneity of space, \mathbf{p} is the object's *momentum*.

Let us differentiate a former result²⁵,

$$dS(t, \mathbf{r}, \lambda dt, \lambda d\mathbf{r}) = \lambda dS(t, \mathbf{r}, dt, d\mathbf{r}),$$

with respect to λ . The left-hand side becomes

$$\frac{d(dS)}{d\lambda} = \frac{\partial dS}{\partial(\lambda dt)} \frac{\partial(\lambda dt)}{\partial \lambda} + \frac{\partial dS}{\partial(\lambda d\mathbf{r})} \cdot \frac{\partial(\lambda d\mathbf{r})}{\partial \lambda} = \frac{\partial dS}{\partial(\lambda dt)} dt + \frac{\partial dS}{\partial(\lambda d\mathbf{r})} \cdot d\mathbf{r},$$

while the right-hand side becomes just dS . Setting $\lambda = 1$ and using the above definitions of E and \mathbf{p} , we obtain

$$-E dt + \mathbf{p} \cdot d\mathbf{r} = dS.$$

$dS = -mc^2 ds$ is a 4-scalar. Since $(cdt, d\mathbf{r})$ are the components of a 4-vector, the left-hand side, $-E dt + \mathbf{p} \cdot d\mathbf{r}$, is a 4-scalar if and only if $(E/c, \mathbf{p})$ are the components of another 4-vector.

(If we had defined E without the minus, this 4-vector would have the components $(-E/c, \mathbf{p})$.)

In the rest frame \mathcal{F}' of a free point mass, $dt' = ds$ and $dS = -mc^2 dt'$. Using the Lorentz transformations²⁶, we find that this equals

$$dS = -mc^2 \frac{dt - v dx/c^2}{\sqrt{1-v^2/c^2}} = -\frac{mc^2}{\sqrt{1-v^2/c^2}} dt + \frac{m\mathbf{v}}{\sqrt{1-v^2/c^2}} \cdot d\mathbf{r},$$

where $\mathbf{v} = (v, 0, 0)$ is the velocity of the point mass in \mathcal{F} . Compare with the above framed equation to find that for a free point mass,

$$E = \frac{mc^2}{\sqrt{1-v^2/c^2}} \quad \mathbf{p} = \frac{m\mathbf{v}}{\sqrt{1-v^2/c^2}}.$$

²⁵ http://en.wikibooks.org/wiki/This_quantum_world/Feynman_route/From_quantum_to_classical#Action

²⁶ http://en.wikibooks.org/wiki/This_quantum_world/Appendix/Relativity/Lorentz#The_actual_Lorentz_transformations

3.5.5 Lorentz force law

To incorporate effects on the motion of a particle (regardless of their causes), we must modify the action differential $dS = -mc^2 dt \sqrt{1-v^2/c^2}$ that a free particle associates with a path segment dC . In doing so we must take care that the modified dS (i) remains homogeneous in the differentials²⁷ and (ii) remains a 4-scalar. The most straightforward way to do this is to add a term that is not just homogeneous but linear in the coordinate differentials:

$$(*) \quad dS = -mc^2 dt \sqrt{1-v^2/c^2} - qV(t, \mathbf{r}) dt + (q/c)\mathbf{A}(t, \mathbf{r}) \cdot d\mathbf{r}.$$

Believe it or not, all classical electromagnetic effects (as against their causes) are accounted for by this expression. $V(t, \mathbf{r})$ is a scalar field (that is, a function of time and space coordinates that is invariant under rotations of the space coordinates), $\mathbf{A}(t, \mathbf{r})$ is a 3-vector field, and (V, \mathbf{A}) is a 4-vector field. We call V and \mathbf{A} the *scalar potential* and the *vector potential*, respectively. The particle-specific constant q is the electric charge, which determines how strongly a particle of a given species is affected by influences of the electromagnetic kind.

If a point mass is not free, the expressions at the end of the previous section²⁸ give its *kinetic energy* E_k and its *kinetic momentum* \mathbf{p}_k . Casting $(*)$ into the form

$$dS = -(E_k + qV) dt + [\mathbf{p}_k + (q/c)\mathbf{A}] \cdot d\mathbf{r}$$

and plugging it into the definitions²⁹

$$(**) \quad E = -\frac{\partial dS}{\partial dt}, \quad \mathbf{p} = \frac{\partial dS}{\partial d\mathbf{r}},$$

we obtain

$$E = E_k + qV, \quad \mathbf{p} = \mathbf{p}_k + (q/c)\mathbf{A}.$$

qV and $(q/c)\mathbf{A}$ are the particle's *potential energy* and *potential momentum*, respectively.

Now we plug $(**)$ into the geodesic equation

$$\frac{\partial dS}{\partial \mathbf{r}} = d \frac{\partial dS}{\partial d\mathbf{r}}.$$

For the right-hand side we obtain

$$d\mathbf{p}_k + \frac{q}{c}d\mathbf{A} = d\mathbf{p}_k + \frac{q}{c} \left[dt \frac{\partial \mathbf{A}}{\partial t} + \left(d\mathbf{r} \cdot \frac{\partial}{\partial \mathbf{r}} \right) \mathbf{A} \right],$$

27 http://en.wikibooks.org/wiki/This_quantum_world/Feynman_route/From_quantum_to_classical#Action

28 Chapter 3.5.4 on page 45

29 Chapter 3.5.4 on page 45

while the left-hand side works out at

$$-q \frac{\partial V}{\partial \mathbf{r}} dt + \frac{q}{c} \frac{\partial(\mathbf{A} \cdot d\mathbf{r})}{\partial \mathbf{r}} = -q \frac{\partial V}{\partial \mathbf{r}} dt + \frac{q}{c} \left[\left(d\mathbf{r} \cdot \frac{\partial}{\partial \mathbf{r}} \right) \mathbf{A} + d\mathbf{r} \times \left(\frac{\partial}{\partial \mathbf{r}} \times \mathbf{A} \right) \right].$$

Two terms cancel out, and the final result is

$$d\mathbf{p}_k = q \underbrace{\left(-\frac{\partial V}{\partial \mathbf{r}} - \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} \right)}_{\equiv \mathbf{E}} dt + d\mathbf{r} \times \underbrace{\frac{q}{c} \left(\frac{\partial}{\partial \mathbf{r}} \times \mathbf{A} \right)}_{\equiv \mathbf{B}} = q \mathbf{E} dt + d\mathbf{r} \times \frac{q}{c} \mathbf{B}.$$

As a classical object travels along the segment $d\mathcal{G}$ of a geodesic, its kinetic momentum changes by the sum of two terms, one linear in the temporal component dt of $d\mathcal{G}$ and one linear in the spatial component $d\mathbf{r}$. How much dt contributes to the change of \mathbf{p}_k depends on the *electric field* \mathbf{E} , and how much $d\mathbf{r}$ contributes depends on the *magnetic field* \mathbf{B} . The last equation is usually written in the form

$$\frac{d\mathbf{p}_k}{dt} = q \mathbf{E} + \frac{q}{c} \mathbf{v} \times \mathbf{B},$$

called the *Lorentz force law*, and accompanied by the following story: there is a physical entity known as the electromagnetic field, which is present everywhere, and which exerts on a charge q an electric force $q\mathbf{E}$ and a magnetic force $(q/c)\mathbf{v} \times \mathbf{B}$.

(Note: This form of the Lorentz force law holds in the Gaussian system of units³⁰. In the MKSA system of units³¹ the c is missing.)

3.5.6 Whence the classical story?

Imagine a small rectangle in spacetime with corners

$$A = (0, 0, 0, 0), B = (dt, 0, 0, 0), C = (0, dx, 0, 0), D = (dt, dx, 0, 0).$$

Let's calculate the electromagnetic contribution to the action of the path from A to D via B for a unit charge ($q = 1$) in natural units ($c = 1$):

$$\begin{aligned} S_{ABD} &= -V(dt/2, 0, 0, 0) dt + A_x(dt, dx/2, 0, 0) dx \\ &= -V(dt/2, 0, 0, 0) dt + \left[A_x(0, dx/2, 0, 0) + \frac{\partial A_x}{\partial t} dt \right] dx. \end{aligned}$$

Next, the contribution to the action of the path from A to D via C :

$$S_{ACD} = A_x(0, dx/2, 0, 0) dx - V(dt/2, dx, 0, 0) dt$$

30 <http://en.wikipedia.org/wiki/Centimeter%20gram%20second%20system%20of%20units>

31 <http://en.wikipedia.org/wiki/International%20System%20of%20Units>

$$= A_x(0, dx/2, 0, 0) dx - \left[V(dt/2, 0, 0, 0) + \frac{\partial V}{\partial x} dx \right] dt.$$

Look at the difference:

$$\Delta S = S_{ACD} - S_{ABD} = \left(-\frac{\partial V}{\partial x} - \frac{\partial A_x}{\partial t} \right) dt dx = E_x dt dx.$$

Alternatively, you may think of ΔS as the electromagnetic contribution to the action of the loop $A \rightarrow B \rightarrow D \rightarrow C \rightarrow A$.

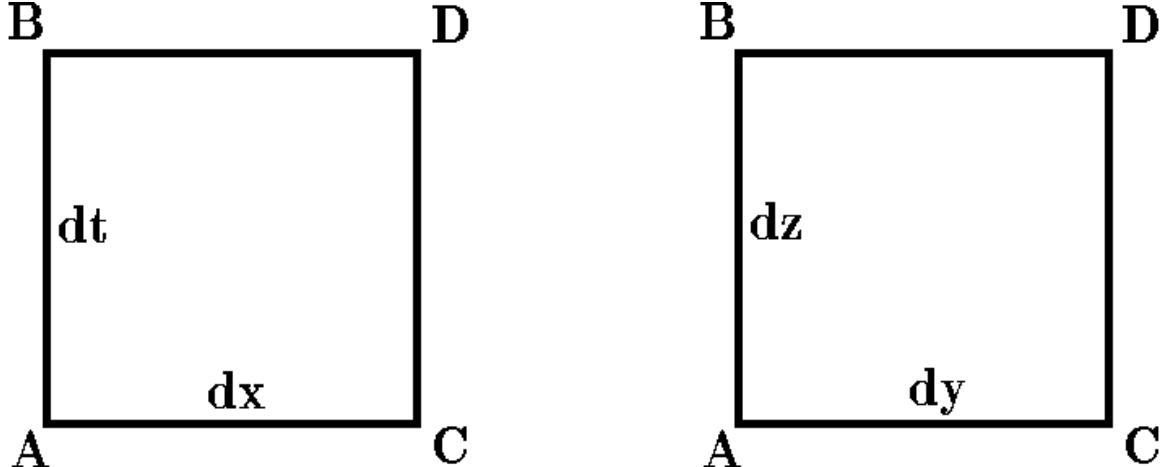


Figure 37

Let's repeat the calculation for a small rectangle with corners

$$A = (0, 0, 0, 0), B = (0, 0, dy, 0), C = (0, 0, 0, dz), D = (0, 0, dy, dz).$$

$$S_{ABD} = A_z(0, 0, 0, dz/2) dz + A_y(0, 0, dy/2, dz) dy$$

$$= A_z(0, 0, 0, dz/2) dz + \left[A_y(0, 0, dy/2, 0) + \frac{\partial A_y}{\partial z} dz \right] dy,$$

$$S_{ACD} = A_y(0, 0, dy/2, 0) dy + A_z(0, 0, dy, dz/2) dz$$

$$= A_y(0, 0, dy/2, 0) dy + \left[A_z(0, 0, 0, dz/2) + \frac{\partial A_z}{\partial y} dy \right] dz,$$

$$\Delta S = S_{ACD} - S_{ABD} = \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \right) dy dz = B_x dy dz.$$

Thus the electromagnetic contribution to the action of *this* loop equals the flux of **B** through the loop.

Remembering (i) Stokes' theorem³² and (ii) the definition³³ of \mathbf{B} in terms of \mathbf{A} , we find that

$$\oint_{\partial\Sigma} \mathbf{A} \cdot d\mathbf{r} = \int_{\Sigma} \text{curl } \mathbf{A} \cdot d\Sigma = \int_{\Sigma} \mathbf{B} \cdot d\Sigma.$$

In (other) words, the magnetic flux through a loop $\partial\Sigma$ (or through any surface Σ bounded by $\partial\Sigma$) equals the circulation of \mathbf{A} around the loop (or around any surface bounded by the loop).

The effect of a circulation $\oint_{\partial\Sigma} \mathbf{A} \cdot d\mathbf{r}$ around the *finite* rectangle $A \rightarrow B \rightarrow D \rightarrow C \rightarrow A$ is to increase (or decrease) the action associated with the segment $A \rightarrow B \rightarrow D$ relative to the action associated with the segment $A \rightarrow C \rightarrow D$. If the actions of the two segments are equal, then we can expect the path of least action from A to D to be a straight line. If one segment has a greater action than the other, then we can expect the path of least action from A to D to curve away from the segment with the larger action.

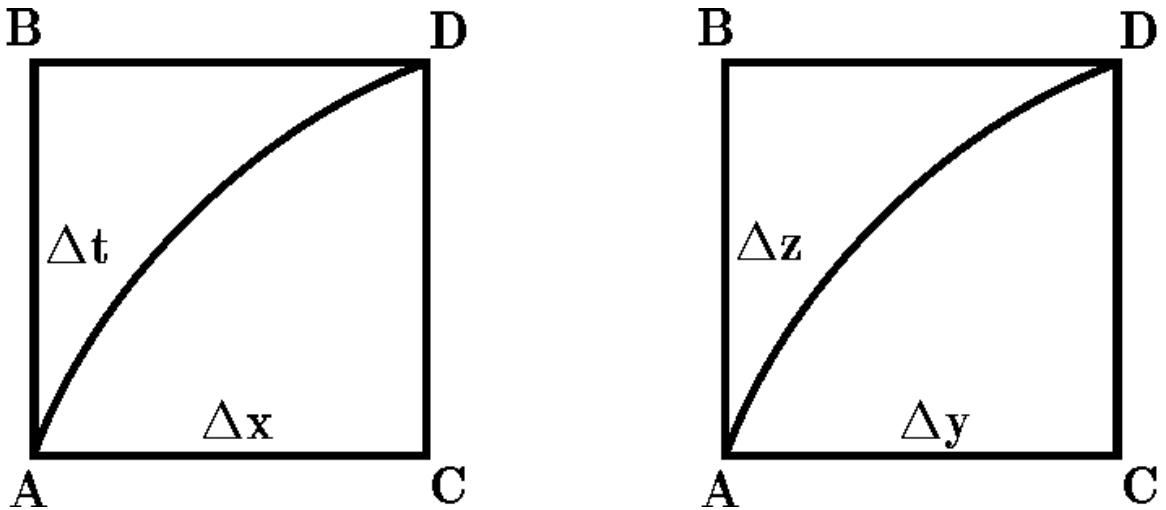


Figure 38

Compare this with the classical story, which explains the curvature of the path of a charged particle in a magnetic field by invoking a force that acts at right angles to both the magnetic field and the particle's direction of motion³⁴. The quantum-mechanical treatment of the same effect offers no such explanation. *Quantum mechanics invokes no mechanism of any kind.* It simply tells us that for a sufficiently massive charge traveling from A to D , the probability of finding that it has done so within any bundle of paths *not* containing the action-geodesic connecting A with D , is virtually 0.

Much the same goes for the classical story according to which the curvature of the path of a charged particle in a *spacetime* plane is due to a force that acts in the direction of the electric field. (Observe that curvature in a spacetime plane is equivalent to acceleration or deceleration. In particular, curvature in a spacetime plane containing the x axis is equivalent

32 Chapter 7.2.3 on page 115

33 Chapter 3.5.5 on page 47

34 Chapter 3.5.5 on page 47

to acceleration in a direction parallel to the x axis.) In this case the corresponding circulation is that of the 4-vector potential (cV, \mathbf{A}) around a spacetime loop.

35

3.6 Schrödinger at last

The Schrödinger equation is non-relativistic. We obtain the non-relativistic version of the electromagnetic action differential,

$$dS = -mc^2 dt \sqrt{1 - v^2/c^2} - qV(t, \mathbf{r}) dt + (q/c)\mathbf{A}(t, \mathbf{r}) \cdot d\mathbf{r},$$

by expanding the root and ignoring all but the first two terms:

$$\sqrt{1 - v^2/c^2} = 1 - \frac{1}{2} \frac{v^2}{c^2} - \frac{1}{8} \frac{v^4}{c^4} - \dots \approx 1 - \frac{1}{2} \frac{v^2}{c^2}.$$

This is obviously justified if $v \ll c$, which defines the non-relativistic regime.

Writing the potential part of dS as $q[-V + \mathbf{A}(t, \mathbf{r}) \cdot (\mathbf{v}/c)]dt$ makes it clear that in most non-relativistic situations the effects represented by the vector potential \mathbf{A} are small compared to those represented by the scalar potential V . If we ignore them (or assume that \mathbf{A} vanishes), and if we include the charge q in the definition of V (or assume that $q = 1$), we obtain

$$S[\mathcal{C}] = -mc^2(t_B - t_A) + \int_{\mathcal{C}} dt \left[\frac{m}{2}v^2 - V(t, \mathbf{r}) \right]$$

for the action associated with a spacetime path \mathcal{C} .

Because the first term is the same for all paths from A to B , it has no effect on the *differences* between the phases of the amplitudes associated with different paths. By dropping it we change neither the classical phenomena (inasmuch as the extremal path remains the same) nor the quantum phenomena (inasmuch as interference effects only depend on those differences). Thus

$$\langle B | A \rangle = \int \mathcal{D}\mathcal{C} e^{(i/\hbar) \int_{\mathcal{C}} dt [(m/2)v^2 - V]}.$$

We now introduce the so-called *wave function* $\psi(t, \mathbf{r})$ as the *amplitude* of finding our particle at \mathbf{r} if the appropriate measurement is made at time t . $\langle t, \mathbf{r} | t', \mathbf{r}' \rangle \psi(t', \mathbf{r}')$, accordingly, is the amplitude of finding the particle first at \mathbf{r}' (at time t') and then at \mathbf{r} (at time t). Integrating over \mathbf{r} , we obtain the amplitude of finding the particle at \mathbf{r} (at time t), provided that Rule B applies. The wave function thus satisfies the equation

$$\psi(t, \mathbf{r}) = {}^3 r' \langle t, \mathbf{r} | t', \mathbf{r}' \rangle \psi(t', \mathbf{r}').$$

35 <http://en.wikibooks.org/wiki/Category%3A>

We again simplify our task by pretending that space is one-dimensional. We further assume that t and t' differ by an infinitesimal interval ϵ . Since ϵ is infinitesimal, there is only one path leading from x' to x . We can therefore forget about the path integral except for a normalization factor \mathcal{A} implicit in the integration measure \mathcal{DC} , and make the following substitutions:

$$dt = \epsilon, \quad v = \frac{x - x'}{\epsilon}, \quad V = V\left(t + \frac{\epsilon}{2}, \frac{x+x'}{2}\right).$$

This gives us

$$\psi(t+\epsilon, x) = \mathcal{A}' e^{im(x-x')^2/2\hbar\epsilon} e^{-(i\epsilon/\hbar)V(t+\epsilon/2, (x+x')/2)} \psi(t, x').$$

We obtain a further simplification if we introduce $\eta = x' - x$ and integrate over η instead of x' . (The integration "boundaries" $-\infty$ and $+\infty$ are the same for both x' and η .) We now have that

$$\psi(t+\epsilon, x) = \mathcal{A}\eta e^{im\eta^2/2\hbar\epsilon} e^{-(i\epsilon/\hbar)V(t+\epsilon/2, x+\eta/2)} \psi(t, x+\eta).$$

Since we are interested in the limit $\epsilon \rightarrow 0$, we expand all terms to first order in ϵ . To which power in η should we expand? As η increases, the phase $m\eta^2/2\hbar\epsilon$ increases at an infinite rate (in the limit $\epsilon \rightarrow 0$) unless η^2 is of the same order as ϵ . In this limit, higher-order contributions to the integral cancel out. Thus the left-hand side expands to

$$\psi(t+\epsilon, x) \approx \psi(t, x) + \frac{\partial\psi}{\partial t}\epsilon,$$

while $e^{-(i\epsilon/\hbar)V(t+\epsilon/2, x+\eta/2)} \psi(t, x+\eta)$ expands to

$$\left[1 - \frac{i\epsilon}{\hbar}V(t, x)\right] \left[\psi(t, x) + \frac{\partial\psi}{\partial x}\eta + \frac{1}{2}\frac{\partial^2\psi}{\partial x^2}\eta^2\right] = \left[1 - \frac{i\epsilon}{\hbar}V(t, x)\right] \psi(t, x) + \frac{\partial\psi}{\partial x}\eta + \frac{\partial^2\psi}{\partial x^2}\frac{\eta^2}{2}.$$

The following integrals need to be evaluated:

$$I_1 = \eta e^{im\eta^2/2\hbar\epsilon}, \quad I_2 = \eta e^{im\eta^2/2\hbar\epsilon}\eta, \quad I_3 = \eta e^{im\eta^2/2\hbar\epsilon}\eta^2.$$

The results are

$$I_1 = \sqrt{2\pi i\hbar\epsilon/m}, \quad I_2 = 0, \quad I_3 = \sqrt{2\pi\hbar^3\epsilon^3/im^3}.$$

Putting Humpty Dumpty back together again yields

$$\psi(t, x) + \frac{\partial\psi}{\partial t}\epsilon = \mathcal{A}\sqrt{\frac{2\pi i\hbar\epsilon}{m}} \left(1 - \frac{i\epsilon}{\hbar}V(t, x)\right) \psi(t, x) + \frac{\mathcal{A}}{2}\sqrt{\frac{2\pi\hbar^3\epsilon^3}{im^3}} \frac{\partial^2\psi}{\partial x^2}.$$

The factor of $\psi(t,x)$ must be the same on both sides, so $\mathcal{A} = \sqrt{m/2\pi i\hbar\epsilon}$, which reduces Humpty Dumpty to

$$\frac{\partial\psi}{\partial t}\epsilon = -\frac{i\epsilon}{\hbar}V\psi + \frac{i\hbar\epsilon}{2m}\frac{\partial^2\psi}{\partial x^2}.$$

Multiplying by $i\hbar/\epsilon$ and taking the limit $\epsilon \rightarrow 0$ (which is trivial since ϵ has dropped out), we arrive at the Schrödinger equation for a particle with one degree of freedom subject to a potential $V(t,x)$:

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2} + V\psi.$$

Trumpets please! The transition to three dimensions is straightforward:

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\left(\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2}\right) + V\psi.$$

4 The Schrödinger equation: implications and applications

In this chapter we take a look at some of the implications of the Schrödinger equation

$$i\hbar \frac{\partial \psi}{\partial t} = \frac{1}{2m} \left(\frac{\hbar}{i} \frac{\partial}{\partial \mathbf{r}} - \mathbf{A} \right)^2 \psi + V\psi.$$

4.1 How fuzzy positions get fuzzier

We will calculate the rate at which the fuzziness of a position probability distribution increases, in consequence of the fuzziness of the corresponding momentum, when there is no counterbalancing attraction (like that between the nucleus and the electron in atomic hydrogen).

Because it is easy to handle, we choose a Gaussian function

$$\psi(0, x) = Ne^{-x^2/2\sigma^2},$$

which has a bell-shaped graph. It defines a position probability distribution

$$|\psi(0, x)|^2 = N^2 e^{-x^2/\sigma^2}.$$

If we normalize this distribution so that $\int dx |\psi(0, x)|^2 = 1$, then $N^2 = 1/\sigma\sqrt{\pi}$, and

$$|\psi(0, x)|^2 = e^{-x^2/\sigma^2}/\sigma\sqrt{\pi}.$$

We also have that

- $\Delta x(0) = \sigma/\sqrt{2}$,
- the Fourier transform of $\psi(0, x)$ is $\bar{\psi}(0, k) = \sqrt{\sigma/\sqrt{\pi}} e^{-\sigma^2 k^2/2}$,
- this defines the momentum probability distribution $|\bar{\psi}(0, k)|^2 = \sigma e^{-\sigma^2 k^2}/\sqrt{\pi}$,
- and $\Delta k(0) = 1/\sigma\sqrt{2}$.

The fuzziness of the position and of the momentum of a particle associated with $\psi(0, x)$ is therefore the minimum allowed by the "uncertainty" relation¹: $\Delta x(0)\Delta k(0) = 1/2$.

Now recall that

¹ Chapter 2.7 on page 27

$$\bar{\psi}(t, k) = \phi(0, k)e^{-i\omega t},$$

where $\omega = \hbar k^2/2m$. This has the Fourier transform

$$\psi(t, x) = \sqrt{\frac{\sigma}{\sqrt{\pi}}} \frac{1}{\sqrt{\sigma^2 + i(\hbar/m)t}} e^{-x^2/2[\sigma^2 + i(\hbar/m)t]},$$

and this defines the position probability distribution

$$|\psi(t, x)|^2 = \frac{1}{\sqrt{\pi} \sqrt{\sigma^2 + (\hbar^2/m^2\sigma^2)t^2}} e^{-x^2/[\sigma^2 + (\hbar^2/m^2\sigma^2)t^2]}.$$

Comparison with $|\psi(0, x)|^2$ reveals that $\sigma(t) = \sqrt{\sigma^2 + (\hbar^2/m^2\sigma^2)t^2}$. Therefore,

$$\Delta x(t) = \frac{\sigma(t)}{\sqrt{2}} = \sqrt{\frac{\sigma^2}{2} + \frac{\hbar^2 t^2}{2m^2\sigma^2}} = \sqrt{[\Delta x(0)]^2 + \frac{\hbar^2 t^2}{4m^2[\Delta x(0)]^2}}.$$

The graphs below illustrate how rapidly the fuzziness of a particle the mass of an electron grows, when compared to an object the mass of a C_{60} molecule or a peanut. Here we see one reason, though by no means the only one, why for all intents and purposes "once sharp, always sharp" is true of the positions of macroscopic objects.

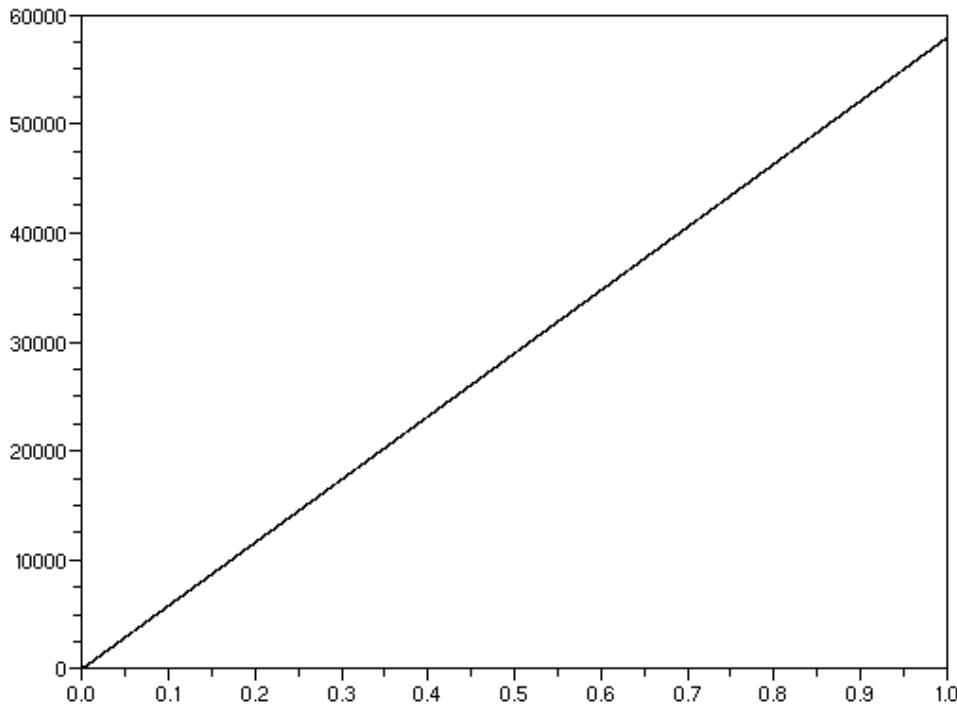


Figure 39

Above: an electron with $\Delta x(0) = 1$ nanometer. In a second, $\Delta x(t)$ grows to nearly 60 km.
Below: an electron with $\Delta x(0) = 1$ centimeter. $\Delta x(t)$ grows only 16% in a second.

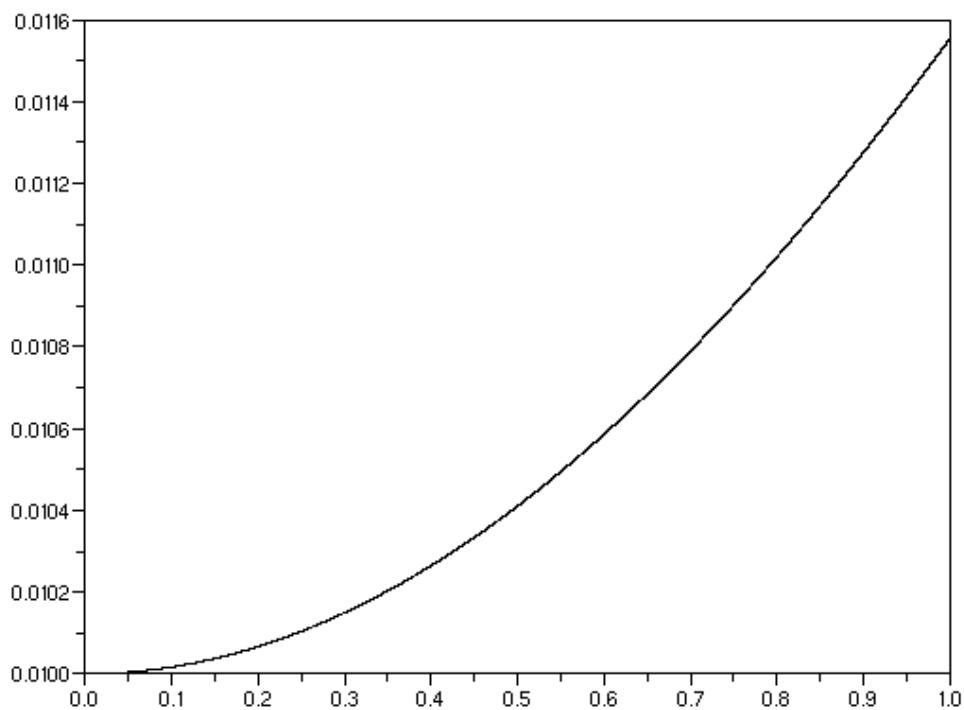


Figure 40

Next, a C_{60} molecule with $\Delta x(0) = 1$ nanometer. In a second, $\Delta x(t)$ grows to 4.4 centimeters.

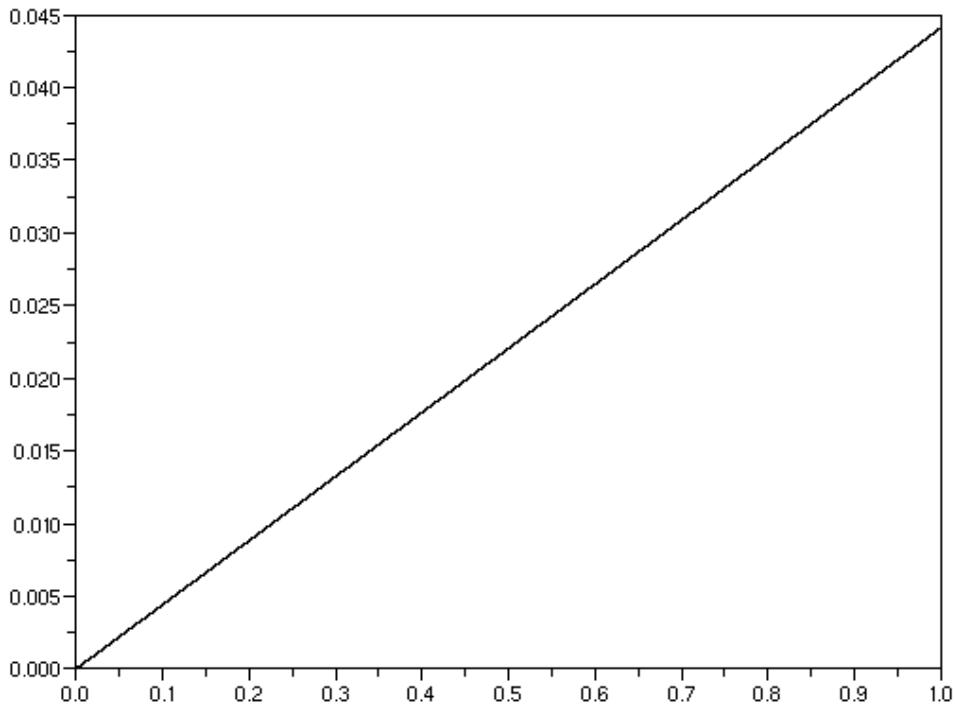


Figure 41

Finally, a peanut (2.8 g) with $\Delta x(0) = 1$ nanometer. $\Delta x(t)$ takes the present age of the universe to grow to 7.5 micrometers.

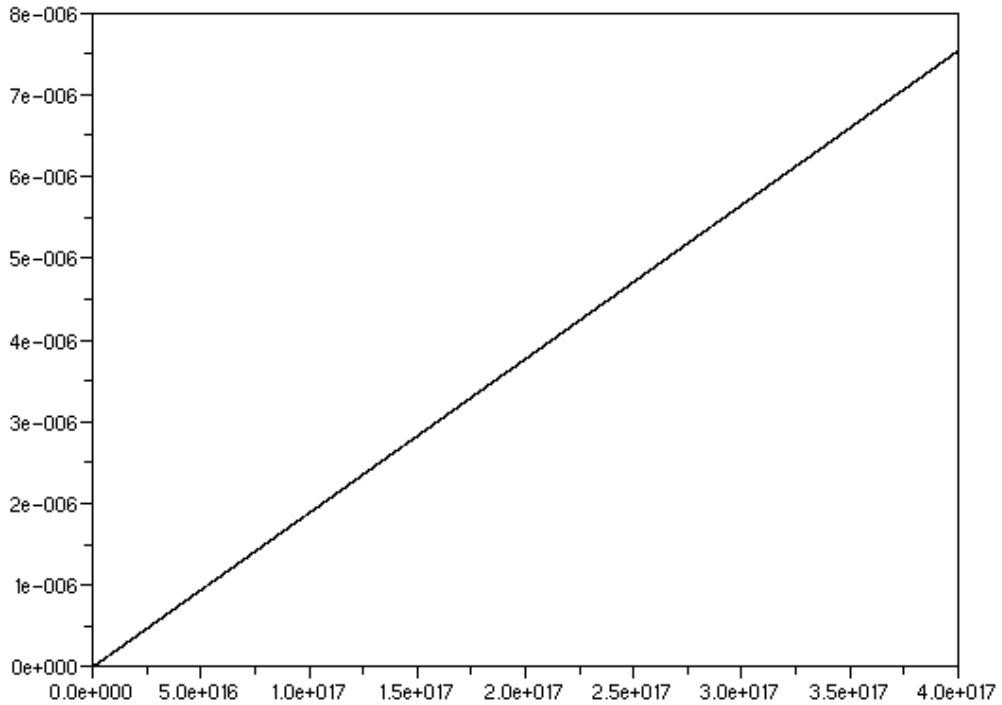


Figure 42

2

4.2 Time-independent Schrödinger equation

If the potential V does not depend on time, then the Schrödinger equation has solutions that are products of a time-independent function $\psi(\mathbf{r})$ and a time-dependent phase factor $e^{-(i/\hbar)Et}$:

$$\psi(t, \mathbf{r}) = \psi(\mathbf{r}) e^{-(i/\hbar)Et}.$$

Because the probability density $|\psi(t, \mathbf{r})|^2$ is independent of time, these solutions are called *stationary*.

Plug $\psi(\mathbf{r}) e^{-(i/\hbar)Et}$ into

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial}{\partial \mathbf{r}} \cdot \frac{\partial}{\partial \mathbf{r}} \psi + V\psi$$

² <http://en.wikibooks.org/wiki/Category%3A>

to find that $\psi(\mathbf{r})$ satisfies the time-independent Schrödinger equation

$$E\psi(\mathbf{r}) = -\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \psi(\mathbf{r}) + V(\mathbf{r}) \psi(\mathbf{r}).$$

³

4.3 Why energy is quantized

Limiting ourselves again to one spatial dimension, we write the time independent Schrödinger equation in this form:

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

Since this equation contains no complex numbers except possibly ψ itself, it has real solutions, and these are the ones in which we are interested. You will notice that if $V > E$, then A is positive and $\psi(x)$ has the same sign as its second derivative. This means that the graph of $\psi(x)$ curves upward above the x axis and downward below it. Thus it cannot cross the axis. On the other hand, if $V < E$, then A is negative and $\psi(x)$ and its second derivative have opposite signs. In this case the graph of $\psi(x)$ curves downward above the x axis and upward below it. As a result, the graph of $\psi(x)$ keeps crossing the axis — it is a wave. Moreover, the larger the difference $E - V$, the larger the curvature of the graph; and the larger the curvature, the smaller the wavelength. In particle terms, the higher the kinetic energy, the higher the momentum.

Let us now find the solutions that describe a particle "trapped" in a potential well — a bound state. Consider this potential:

³ <http://en.wikibooks.org/wiki/Category%3A>

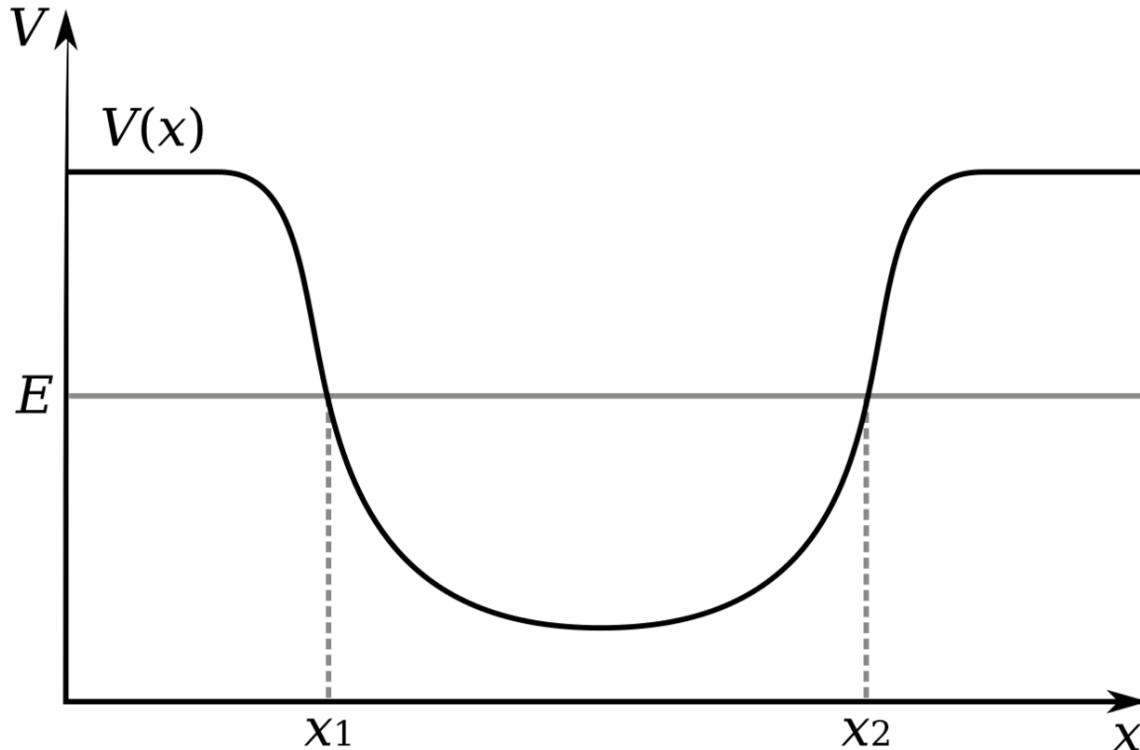


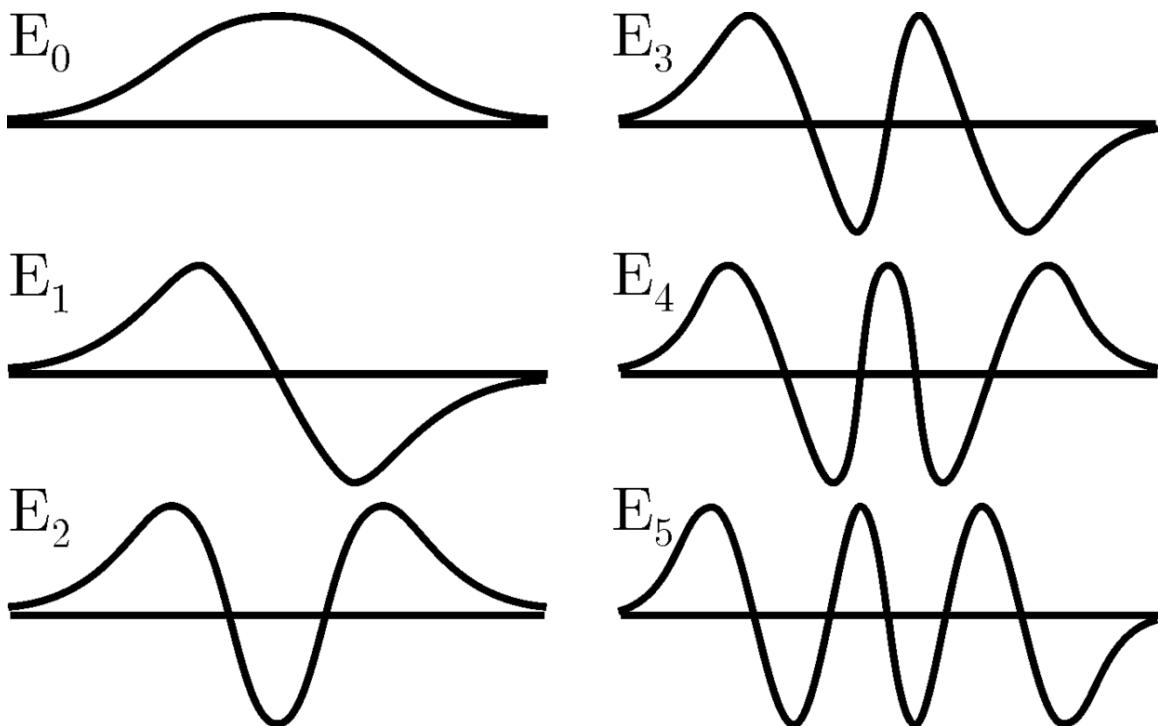
Figure 43

Observe, to begin with, that at x_1 and x_2 , where $E = V$, the slope of $\psi(x)$ does not change since $d^2\psi(x)/dx^2 = 0$ at these points. This tells us that the probability of finding the particle cannot suddenly drop to zero at these points. It will therefore be possible to find the particle to the left of x_1 or to the right of x_2 , where classically it could not be. (A classical particle would oscillate back and forth between these points.)

Next, take into account that the probability distributions defined by $\psi(x)$ must be normalizable. For the graph of $\psi(x)$ this means that it must approach the x axis asymptotically as $x \rightarrow \pm\infty$.

Suppose that we have a normalized solution for a particular value E . If we increase or decrease the value of E , the curvature of the graph of $\psi(x)$ between x_1 and x_2 increases or decreases. A small increase or decrease won't give us another solution: $\psi(x)$ won't vanish asymptotically for both positive and negative x . To obtain another solution, we must increase E by just the right amount to increase or decrease by one the number of wave nodes between the "classical" turning points x_1 and x_2 and to make $\psi(x)$ again vanish asymptotically in both directions.

The bottom line is that the energy of a bound particle — a particle "trapped" in a potential well — is *quantized*: only certain values E_k yield solutions $\psi_k(x)$ of the time-independent Schrödinger equation:

**Figure 44**

4

4.4 A quantum bouncing ball

As a specific example, consider the following potential:

$$V(z) = mgz \quad \text{if } z > 0 \quad \text{and} \quad V(z) = \infty \quad \text{if } z < 0.$$

g is the gravitational acceleration at the floor. For $z < 0$, the Schrödinger equation as given in the previous section⁵ tells us that $d^2\psi(z)/dz^2 = \infty$ unless $\psi(z) = 0$. The only sensible solution for negative z is therefore $\psi(z) = 0$. The requirement that $V(z) = \infty$ for $z < 0$ ensures that our perfectly elastic, frictionless quantum bouncer won't be found below the floor.

Since a picture is worth more than a thousand words, we won't solve the time-independent Schrödinger equation for this particular potential but merely plot its first eight solutions:

⁴ <http://en.wikibooks.org/wiki/Category%3A>

⁵ Chapter 4.2 on page 61

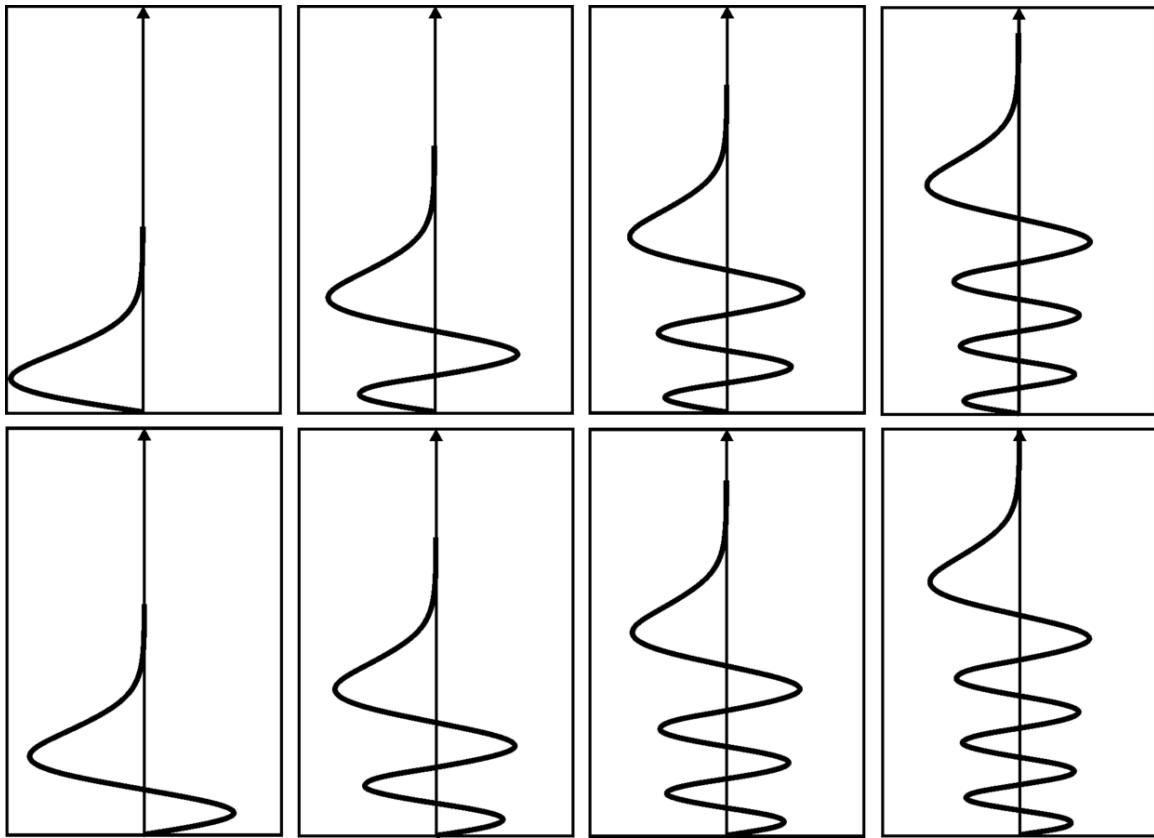


Figure 45

Where would a classical bouncing ball subject to the same potential reverse its direction of motion? Observe the correlation between position and momentum (wavenumber).

All of these states are stationary; the probability of finding the quantum bouncer in any particular interval of the z axis is independent of time. So how do we get it to move?

Recall that any linear combination of solutions of the Schrödinger equation is another solution. Consider this linear combination of two stationary states:

$$\psi(t, x) = A\psi_1(x)e^{-i\omega_1 t} + B\psi_2(x)e^{-i\omega_2 t}.$$

Assuming that the coefficients A, B and the wave functions $\psi_1(x), \psi_2(x)$ are real, we calculate the mean position of a particle associated with $\psi(t, x)$:

$$\begin{aligned} \psi^* x \psi &= (A\psi_1 e^{i\omega_1 t} + B\psi_2 e^{i\omega_2 t}) x (A\psi_1 e^{-i\omega_1 t} + B\psi_2 e^{-i\omega_2 t}) \\ &= A^2 \psi_1^2 x + B^2 \psi_2^2 x + AB(e^{i(\omega_1 - \omega_2)t} + e^{i(\omega_2 - \omega_1)t}) \psi_1 x \psi_2. \end{aligned}$$

The first two integrals are the (time-independent) mean positions of a particle associated with $\psi_1(x) e^{i\omega_1 t}$ and $\psi_2(x) e^{i\omega_2 t}$, respectively. The last term equals

$$2AB \cos(\Delta\omega t) \psi_1 x \psi_2,$$

and this tells us that the particle's mean position oscillates with frequency $\Delta\omega = \omega_2 - \omega_1$ and amplitude $2AB\psi_1x\psi_2$ about the sum of the first two terms.

Visit this site⁶ to watch the time-dependence of the probability distribution associated with a quantum bouncer that is initially associated with a Gaussian distribution.

7

4.5 Atomic hydrogen

While de Broglie's theory of 1923 featured circular electron waves, Schrödinger's "wave mechanics" of 1926 features standing waves in three dimensions. Finding them means finding the solutions of the time-independent Schrödinger equation

$$E\psi(\mathbf{r}) = -\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \psi(\mathbf{r}) + V(\mathbf{r})\psi(\mathbf{r}).$$

with $V(\mathbf{r}) = -e^2/r$, the potential energy of a classical electron at a distance $r = |\mathbf{r}|$ from the proton. (Only when we come to the relativistic theory will we be able to shed the last vestige of classical thinking.)

$$E\psi(\mathbf{r}) = -\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \psi(\mathbf{r}) - \frac{e^2}{r} V(\mathbf{r})\psi(\mathbf{r}).$$

In using this equation, we ignore (i) the influence of the electron on the proton, whose mass is some 1836 times larger than that of the electron, and (ii) the electron's spin. Since relativistic and spin effects on the measurable properties of atomic hydrogen are rather small, this non-relativistic approximation nevertheless gives excellent results.

For bound states the total energy E is negative, and the Schrödinger equation has a discrete set of solutions. As it turns out, the "allowed" values of E are precisely the values that Bohr obtained in 1913:

$$E_n = -\frac{1}{n^2} \frac{\mu e^4}{2\hbar^2}, \quad n = 1, 2, 3, \dots$$

However, for each n there are now n^2 linearly independent solutions. (If ψ_1, \dots, ψ_k are independent solutions, then none of them can be written as a linear combination $\sum a_i\psi_i$ of the others.)

Solutions with different n correspond to different energies. What physical differences correspond to linearly independent solutions with the same n ?

⁶ http://www.uark.edu/misc/julio/bouncing_ball/bouncing_ball.html

⁷ <http://en.wikibooks.org/wiki/Category%3A>

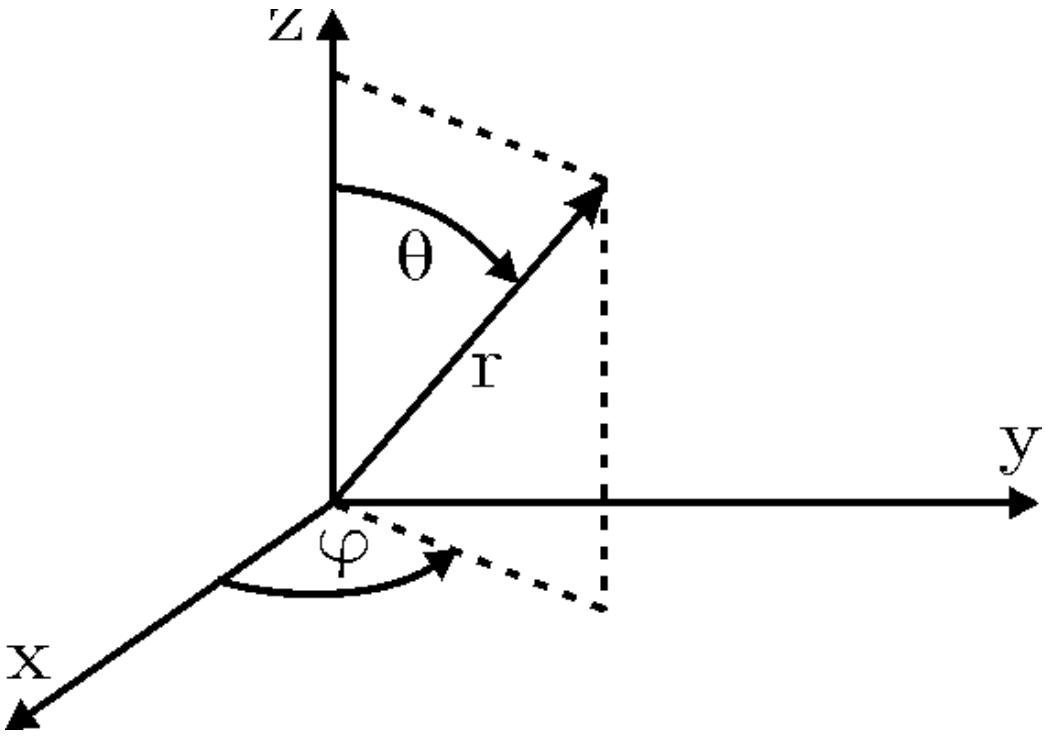


Figure 46

Using polar coordinates, one finds that all solutions for a particular value E_n are linear combinations of solutions that have the form

$$\psi(r, \phi, \theta) = e^{(i/\hbar)l_z\phi}\psi(r, \theta).$$

l_z turns out to be another *quantized* variable, for $e^{(i/\hbar)l_z\phi} = e^{(i/\hbar)l_z(\phi \pm 2\pi)}$ implies that $l_z = m\hbar$ with $m = 0, \pm 1, \pm 2, \dots$. In addition, $|m|$ has an upper bound, as we shall see in a moment.

Just as the factorization of $\psi(t, \mathbf{r})$ into $e^{-(i/\hbar)Et}\psi(\mathbf{r})$ made it possible to obtain a t -independent Schrödinger equation, so the factorization of $\psi(r, \phi, \theta)$ into $e^{(i/\hbar)l_z\phi}\psi(r, \theta)$ makes it possible to obtain a ϕ -independent Schrödinger equation. This contains another real parameter Λ , over and above m , whose "allowed" values are given by $l(l+1)\hbar^2$, with l an integer satisfying $0 \leq l \leq n-1$. The range of possible values for m is bounded by the inequality $|m| \leq l$. The possible values of the *principal quantum number* n , the *angular momentum quantum number* l , and the so-called *magnetic quantum number* m thus are:

$n = 1$	$l = 0$	$m = 0$
$n = 2$	$l = 0$	$m = 0$
	$l = 1$	$m = 0, \pm 1$
$n = 3$	$l = 0$	$m = 0$
	$l = 1$	$m = 0, \pm 1$
	$l = 2$	$m = 0, \pm 1, \pm 2$
$n = 4$

Each possible set of quantum numbers n, l, m defines a unique wave function $\psi_{nlm}(t, \mathbf{r})$, and together these make up a complete set of bound-state solutions ($E < 0$) of the Schrödinger equation with $V(\mathbf{r}) = -e^2/r$. The following images give an idea of the position probability distributions of the first three $l = 0$ states (not to scale). Below them are the probability densities plotted against r . Observe that these states have $n - 1$ nodes, all of which are spherical, that is, surfaces of constant r . (The nodes of a wave in three dimensions are two-dimensional surfaces. The nodes of a "probability wave" are the surfaces at which the sign of ψ changes and, consequently, the probability density $|\psi|^2$ vanishes.)

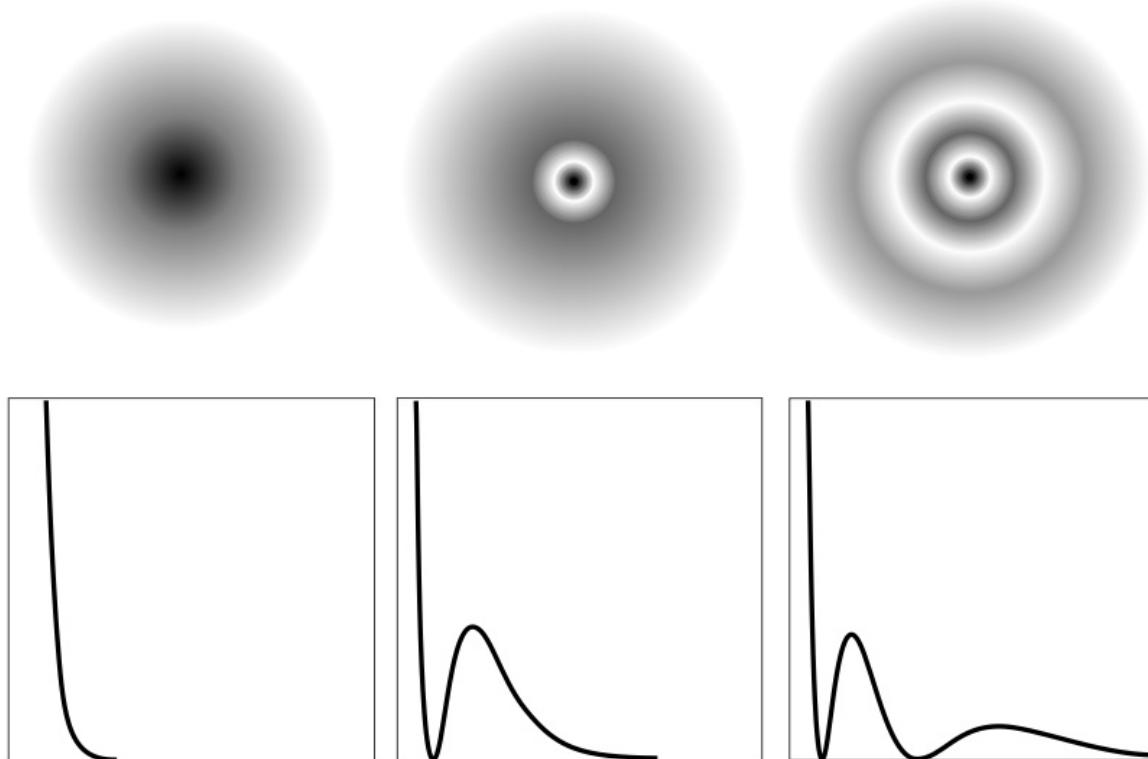


Figure 47

Take another look at these images:

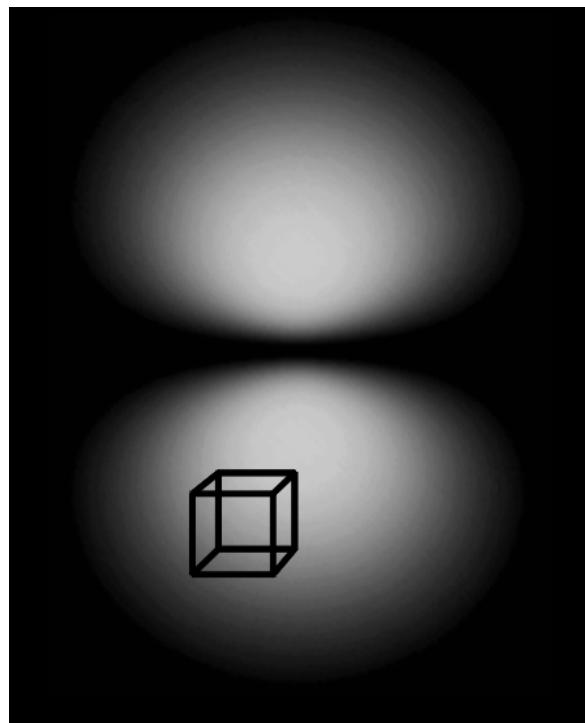


Figure 48 *Image : Orbitals2.png|*

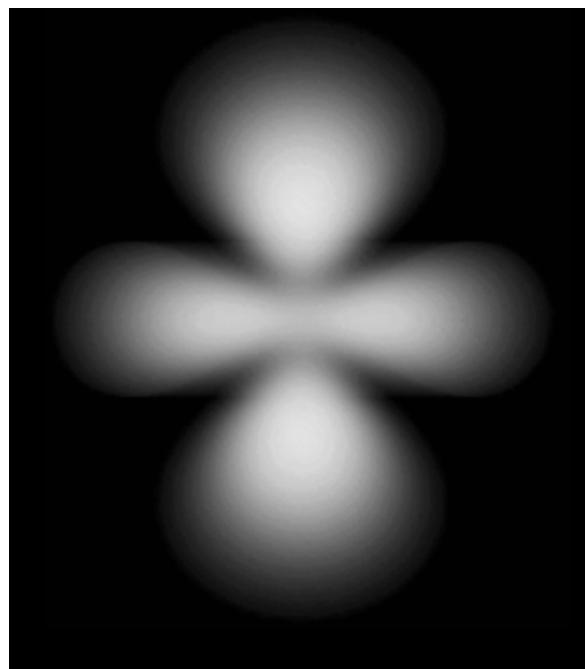


Figure 49 *Image : Orbitals4.png|*

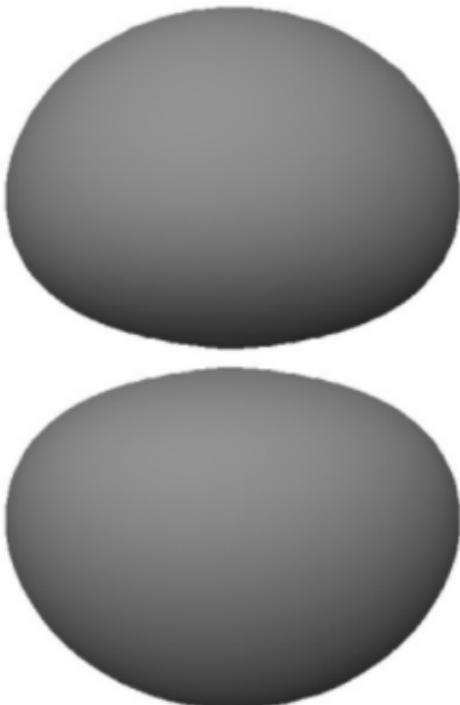


Figure 50 ρ_{2p0}

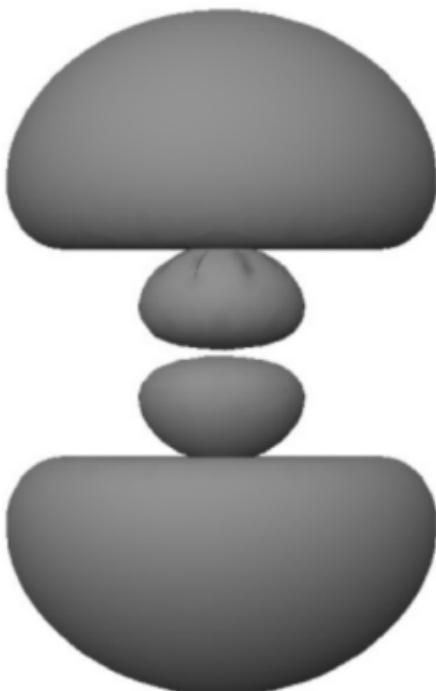


Figure 51 ρ_{3p0}

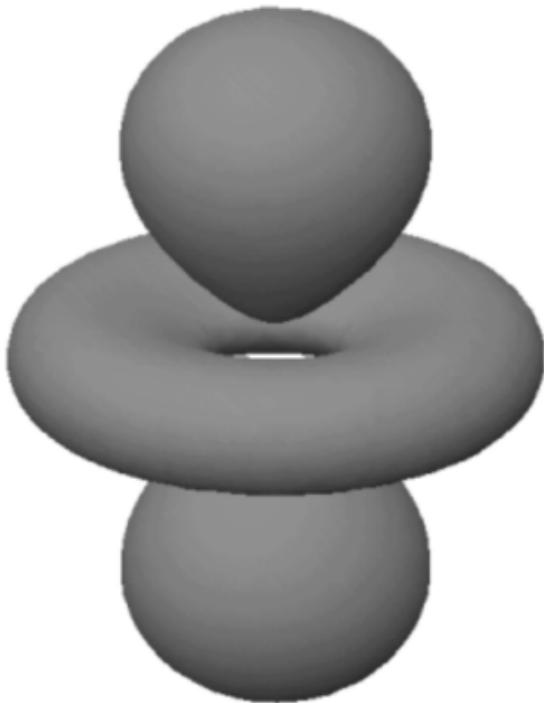


Figure 52 ρ_{3d0}

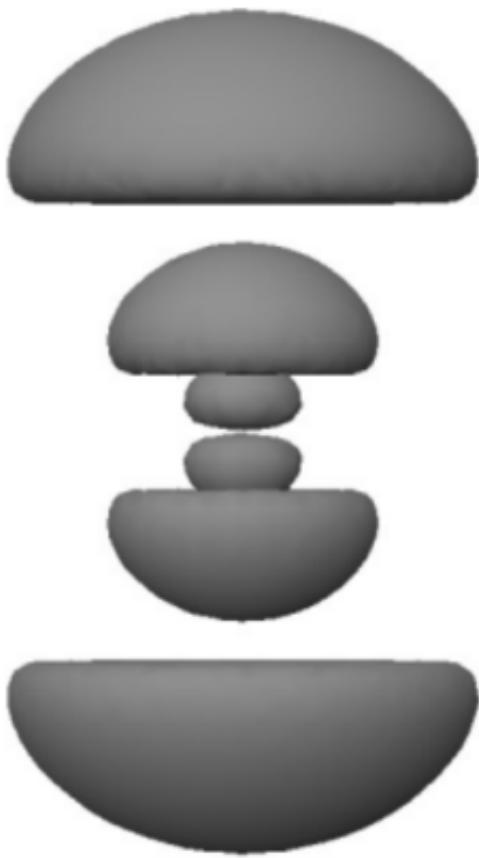


Figure 53 ρ_{4p0}

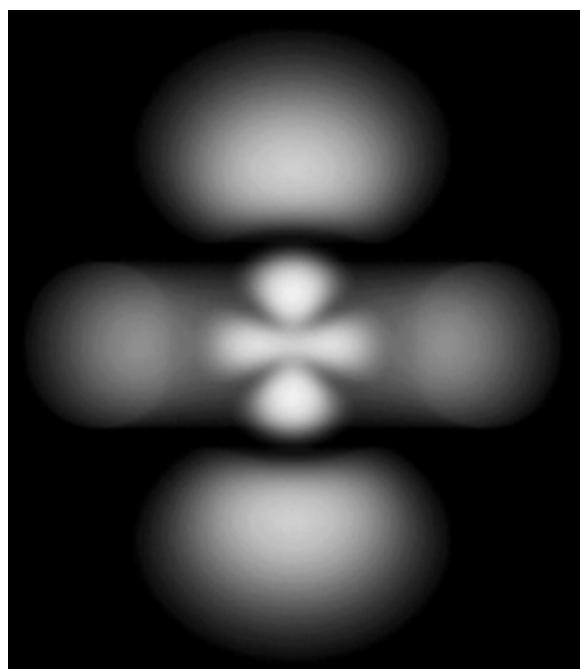


Figure 54 *Image : Orbitals6.png|*

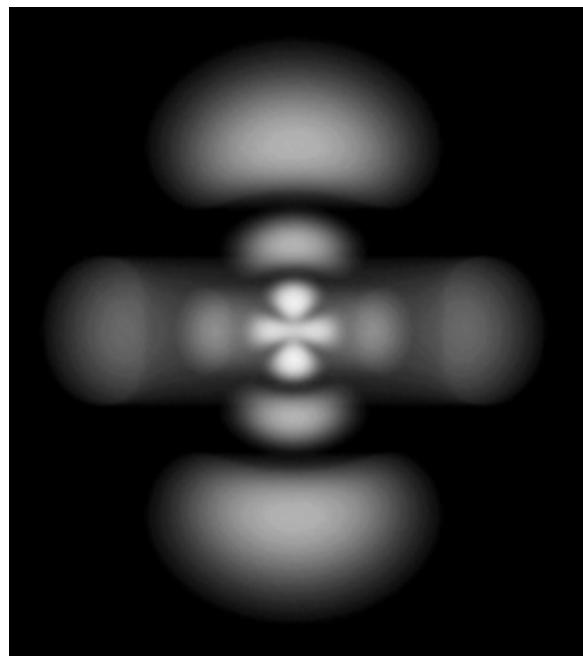


Figure 55 *Image : Orbitals8.png|*

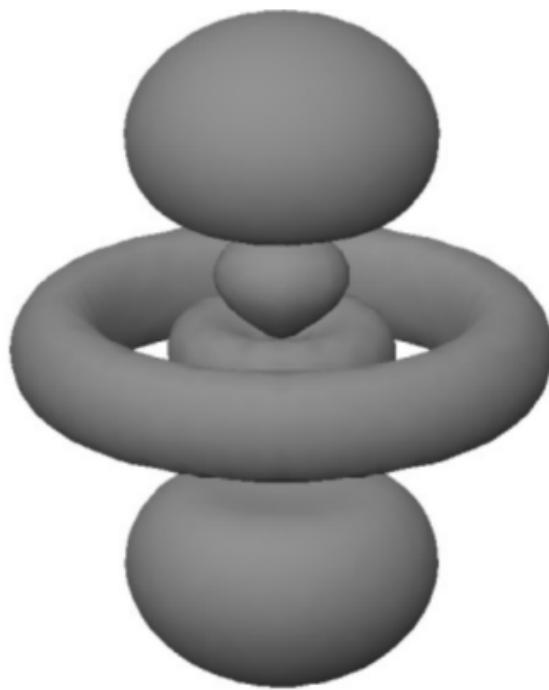


Figure 56 ρ_{4d0}

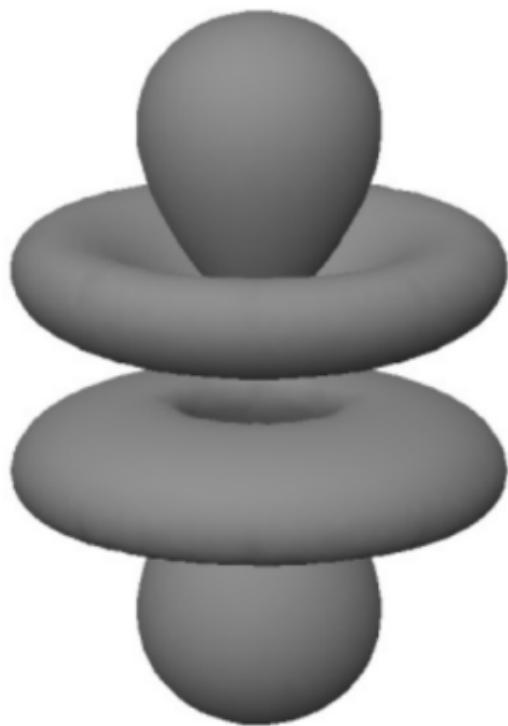


Figure 57 ρ_{4f0}

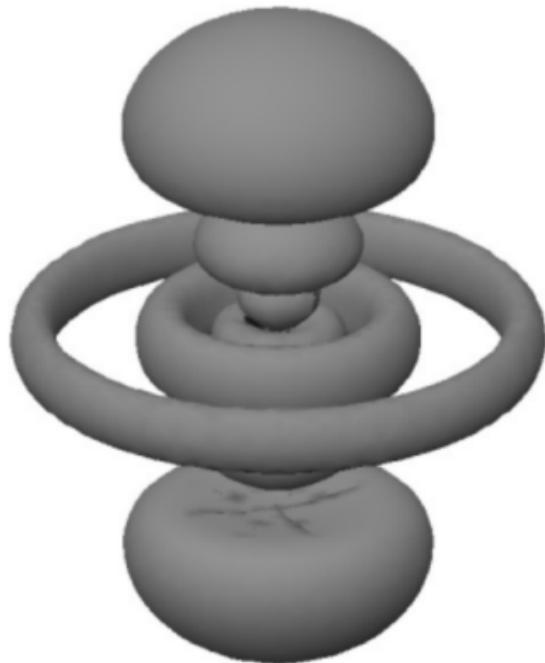


Figure 58 ρ_{5d0}

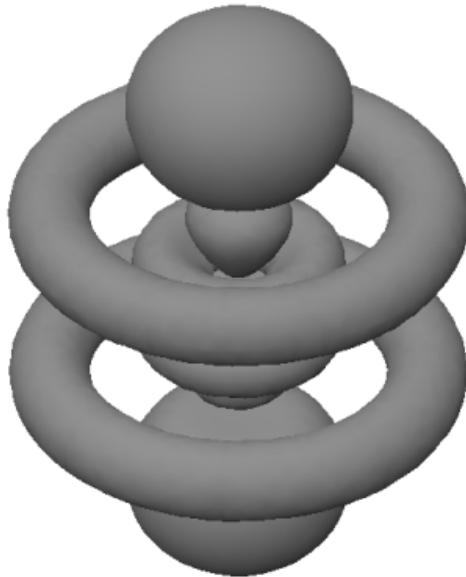


Figure 59 ρ_{5f_0}

The letters s,p,d,f stand for $l=0,1,2,3$, respectively. (Before the quantum-mechanical origin of atomic spectral lines was understood, a distinction was made between "sharp," "principal," "diffuse," and "fundamental" lines. These terms were subsequently found to correspond to the first four values that l can take. From $l = 3$ onward the labels follows the alphabet: f,g,h,...) Observe that these states display both spherical and conical nodes, the latter being surfaces of constant θ . (The "conical" node with $\theta = 0$ is a horizontal plane.) These states, too, have a total of $n - 1$ nodes, l of which are conical.

Because the "waviness" in ϕ is contained in a phase factor $e^{im\phi}$, it does not show up in representations of $|\psi|^2$. To make it visible, it is customary to replace $e^{im\phi}$ by its real part $\cos(m\phi)$, as in the following images, which do *not* represent probability distributions.

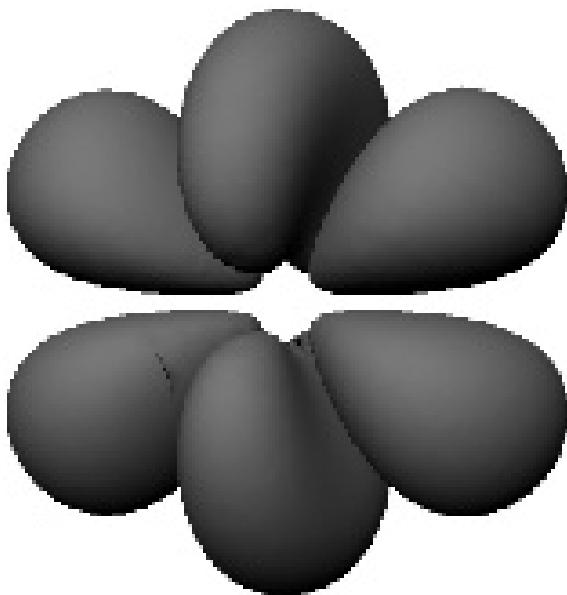


Figure 60 ρ_{4f1}

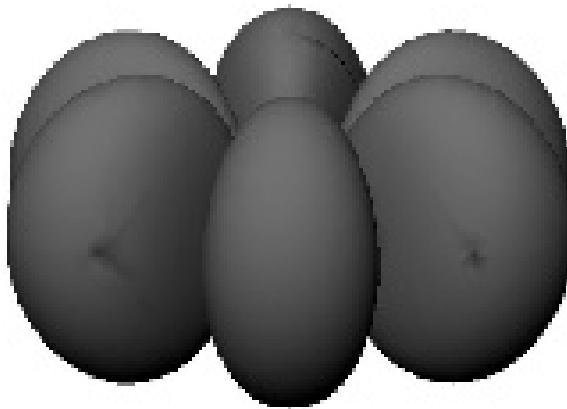


Figure 61 ρ_{4f3}

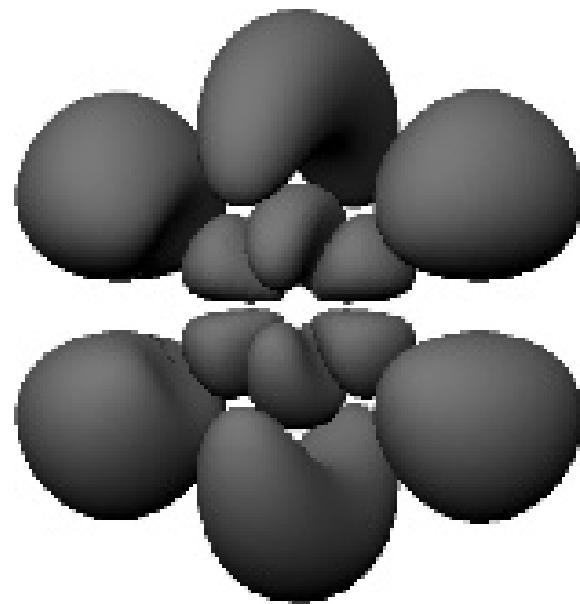


Figure 62 ρ_{5f1}

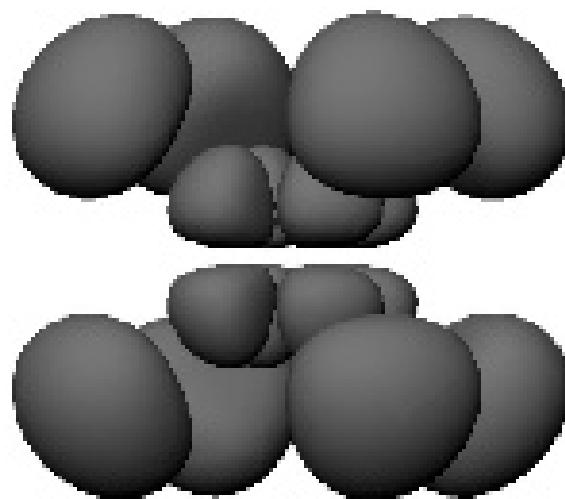


Figure 63 ρ_{5f2}

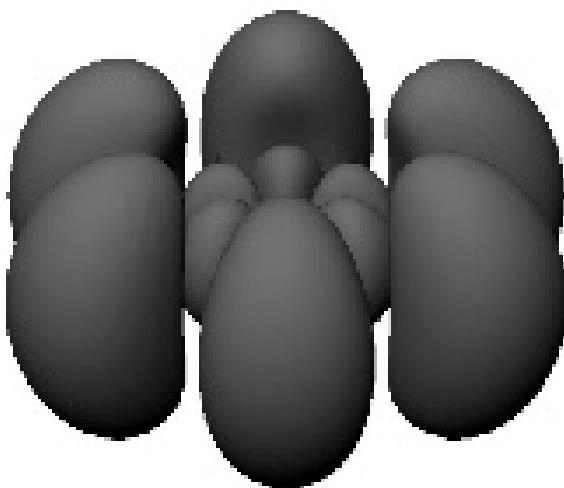


Figure 64 ρ_{5f3}

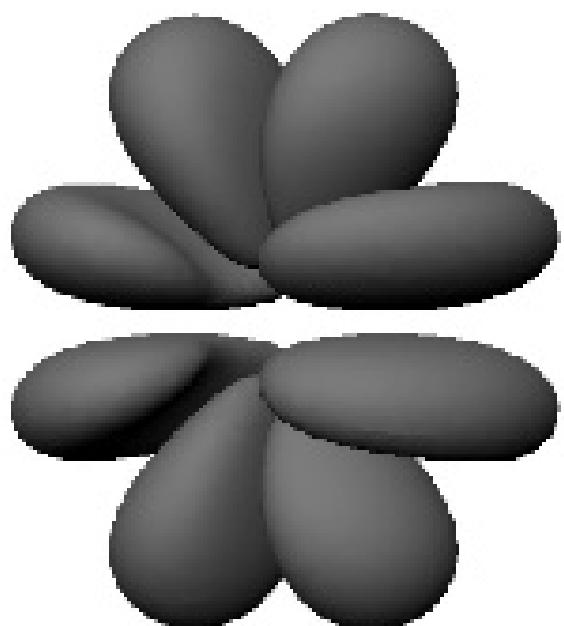


Figure 65 ρ_{5g1}

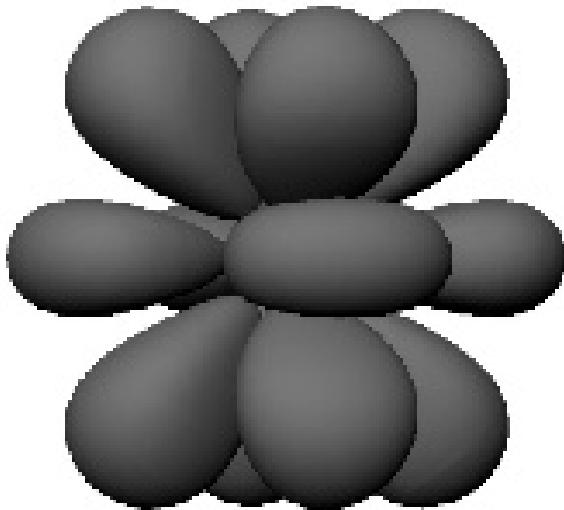


Figure 66 ρ_{5g2}

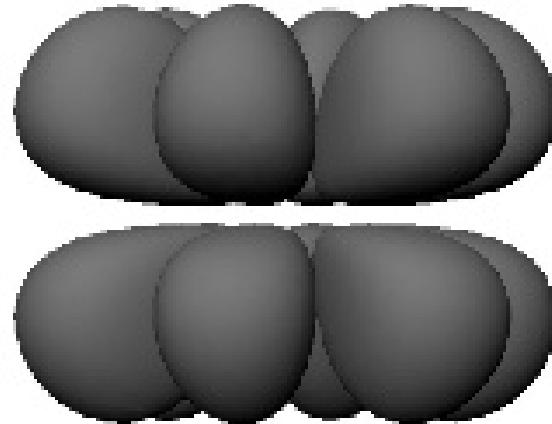


Figure 67 ρ_{5g3}

The total number of nodes is again $n - 1$, the total number of non-spherical nodes is again l , but now there are m plane nodes containing the z axis and $l - m$ conical nodes.

What is so special about the z axis? Absolutely nothing, for the wave functions ψ'_{nlm} , which are defined with respect to a different axis, make up another complete set of bound-state solutions. This means that every wave function ψ'_{nlm} can be written as a linear combination of the functions ψ_{nlm} , and vice versa.

8

4.6 Observables and operators

Remember the mean values

8 <http://en.wikibooks.org/wiki/Category%3A>

$$\langle x \rangle = \int |\psi|^2 x dx \quad \text{and} \quad \langle p \rangle = \hbar \langle k \rangle = \int |\bar{\psi}|^2 \hbar k dk.$$

As noted already, if we define the operators

$$\hat{x} = x \text{ ("multiply with } x\text{") and } \hat{p} = -i\hbar \frac{\partial}{\partial x},$$

then we can write

$$\langle x \rangle = \int \psi^* \hat{x} \psi dx \quad \text{and} \quad \langle p \rangle = \int \psi^* \hat{p} \psi dx.$$

By the same token,

$$\langle E \rangle = \int \psi^* \hat{E} \psi dx \quad \text{with} \quad \hat{E} = i\hbar \frac{\partial}{\partial t}.$$

Which observable is associated with the differential operator $\partial/\partial\phi$? If r and θ are constant (as the partial derivative with respect to ϕ requires), then z is constant, and

$$\frac{\partial \psi}{\partial \phi} = \frac{\partial y}{\partial \phi} \frac{\partial \psi}{\partial y} + \frac{\partial x}{\partial \phi} \frac{\partial \psi}{\partial x}.$$

Given that $x = r \sin \theta \cos \phi$ and $y = r \sin \theta \sin \phi$, this works out at $x \frac{\partial \psi}{\partial y} - y \frac{\partial \psi}{\partial x}$ or

$$-i\hbar \frac{\partial}{\partial \phi} = \hat{x} \hat{p}_y - \hat{y} \hat{p}_x.$$

Since, classically, orbital angular momentum is given by $\mathbf{L} = \mathbf{r} \times \mathbf{p}$, so that $L_z = x p_y - y p_x$, it seems obvious that we should consider $\hat{x} \hat{p}_y - \hat{y} \hat{p}_x$ as the operator \hat{l}_z associated with the z component of the atom's angular momentum.

Yet we need to be wary of basing quantum-mechanical definitions on classical ones. Here are the quantum-mechanical definitions:

Consider the wave function $\psi(q_k, t)$ of a closed system \mathcal{S} with K degrees of freedom. Suppose that the probability distribution $|\psi(q_k, t)|^2$ (which is short for $|\psi(q_1, \dots, q_K, t)|^2$) is invariant under translations in time: waiting for any amount of time τ makes no difference to it:

$$|\psi(q_k, t)|^2 = |\psi(q_k, t + \tau)|^2.$$

Then the time dependence of ψ is confined to a phase factor $e^{i\alpha(q_k, t)}$.

Further suppose that the time coordinate t and the space coordinates q_k are homogeneous — equal intervals are physically equivalent. Since \mathcal{S} is closed, the phase factor $e^{i\alpha(q_k, t)}$ cannot then depend on q_k , and its phase can at most linearly depend on t : waiting for 2τ should have the same effect as twice waiting for τ . In other words, multiplying the wave function by $e^{i\alpha(2\tau)}$ should have same effect as multiplying it twice by $e^{i\alpha(\tau)}$:

$$e^{i\alpha(2\tau)} = [e^{i\alpha(\tau)}]^2 = e^{i2\alpha(\tau)}.$$

Thus

$$\psi(q_k, t) = \psi(q_k) e^{-i\omega t} = \psi(q_k) e^{-(i/\hbar)E t}.$$

So the existence of a constant ("conserved") quantity ω or (in conventional units) E is implied for a closed system, and this is what we mean by the *energy* of the system.

Now suppose that $|\psi(q_k, t)|^2$ is invariant under translations in the direction of one of the spatial coordinates q_k , say q_j :

$$|\psi(q_j, q_{k \neq j}, t)|^2 = |\psi(q_j + \kappa, q_{k \neq j}, t)|^2.$$

Then the dependence of ψ on q_j is confined to a phase factor $e^{i\beta(q_k, t)}$.

And suppose again that the time coordinates t and q_k are homogeneous. Since \mathcal{S} is closed, the phase factor $e^{i\beta(q_k, t)}$ cannot then depend on $q_{k \neq j}$ or t , and its phase can at most linearly depend on q_j : translating \mathcal{S} by 2κ should have the same effect as twice translating it by κ . In other words, multiplying the wave function by $e^{i\beta(2\kappa)}$ should have same effect as multiplying it twice by $e^{i\beta(\kappa)}$:

$$e^{i\beta(2\kappa)} = [e^{i\beta(\kappa)}]^2 = e^{i2\beta(\kappa)}.$$

Thus

$$\psi(q_k, t) = \psi(q_{k \neq j}, t) e^{ik_j q_k} = \psi(q_{k \neq j}, t) e^{(i/\hbar)p_j q_k}.$$

So the existence of a constant ("conserved") quantity k_j or (in conventional units) p_j is implied for a closed system, and this is what we mean by the j -component of the system's *momentum*.

You get the picture. Moreover, the spatial coordinates might as well be the spherical coordinates r, θ, ϕ . If $|\psi(r, \theta, \phi, t)|^2$ is invariant under rotations about the z axis, and if the longitudinal coordinate ϕ is homogeneous, then

$$\psi(r, \theta, \phi, t) = \psi(r, \theta, t) e^{im\phi} = \psi(r, \theta, t) e^{(i/\hbar)l_z\phi}.$$

In this case we call the conserved quantity the z component of the system's *angular momentum*.

Now suppose that O is an observable, that \hat{O} is the corresponding operator, and that $\psi_{\hat{O}, v}$ satisfies

$$\hat{O} \psi_{\hat{O}, v} = v \psi_{\hat{O}, v}.$$

We say that $\psi_{\hat{O},v}$ is an *eigenfunction* or *eigenstate* of the operator \hat{O} , and that it has the *eigenvalue* v . Let's calculate the mean and the standard deviation of O for $\psi_{\hat{O},v}$. We obviously have that

$$\langle O \rangle = \int \psi_{\hat{O},v}^* \hat{O} \psi_{\hat{O},v} dx = \int \psi_{\hat{O},v}^* v \psi_{\hat{O},v} dx = v \int |\psi_{\hat{O},v}|^2 dx = v.$$

Hence

$$\Delta O = \sqrt{\int \psi_{\hat{O},v}^* (\hat{O} - v)(\hat{O} - v) \psi_{\hat{O},v} dx} = 0,$$

since $(\hat{O} - v)\psi_{\hat{O},v} = 0$. For a system associated with $\psi_{\hat{O},v}$, O is dispersion-free. Hence the probability of finding that the value of O lies in an interval containing v , is 1. But we have that

$$\hat{E}\psi(q_k)e^{-(i/\hbar)Et} = E\psi(q_k)e^{-(i/\hbar)Et}$$

$$\hat{p}_j\psi(q_{k \neq j},t)e^{(i/\hbar)p_j q_k} = p_j\psi(q_{k \neq j},t)e^{(i/\hbar)p_j q_k}$$

$$\hat{l}_z\psi(r,\theta,t)e^{(i/\hbar)l_z\phi} = l_z\psi(r,\theta,t)e^{(i/\hbar)l_z\phi}.$$

So, indeed, \hat{l}_z is the operator associated with the z component of the atom's angular momentum.

Observe that the eigenfunctions of any of these operators are associated with systems for which the corresponding observable is "sharp": the standard deviation measuring its fuzziness vanishes.

For obvious reasons we also have

$$\hat{l}_x = -i\hbar \left(y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right) \quad \text{and} \quad \hat{l}_y = -i\hbar \left(z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z} \right).$$

If we define the *commutator* $[\hat{A}, \hat{B}] \equiv \hat{A}\hat{B} - \hat{B}\hat{A}$, then saying that the operators \hat{A} and \hat{B} commute is the same as saying that their commutator vanishes. Later we will prove that two observables are *compatible* (can be simultaneously measured) if and only if their operators *commute*.

Exercise: Show that $[\hat{l}_x, \hat{l}_y] = i\hbar\hat{l}_z$.

One similarly finds that $[\hat{l}_y, \hat{l}_z] = i\hbar\hat{l}_x$ and $[\hat{l}_z, \hat{l}_x] = i\hbar\hat{l}_y$. The upshot: different components of a system's angular momentum are *incompatible*.

Exercise: Using the above commutators, show that the operator $\hat{\mathbf{L}}^2 \equiv \hat{l}_x^2 + \hat{l}_y^2 + \hat{l}_z^2$ commutes with \hat{l}_x , \hat{l}_y , and \hat{l}_z .

4.7 Beyond hydrogen: the Periodic Table

If we again assume that the nucleus is fixed at the center and ignore relativistic and spin effects, then the stationary states of helium are the solutions of the following equation:

$$E \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \left[\frac{\partial^2 \psi}{\partial x_1^2} + \frac{\partial^2 \psi}{\partial y_1^2} + \frac{\partial^2 \psi}{\partial z_1^2} + \frac{\partial^2 \psi}{\partial x_2^2} + \frac{\partial^2 \psi}{\partial y_2^2} + \frac{\partial^2 \psi}{\partial z_2^2} \right] + \left[-\frac{2e^2}{r_1} - \frac{2e^2}{r_2} + \frac{e^2}{r_{12}} \right] \psi.$$

The wave function now depends on six coordinates, and the potential energy V is made up of three terms. $r_1 = \sqrt{x_1^2 + y_1^2 + z_1^2}$ and $r_2 = \sqrt{x_2^2 + y_2^2 + z_2^2}$ are associated with the respective distances of the electrons from the nucleus, and $r_{12} = \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2 + (z_2-z_1)^2}$ is associated with the distance between the electrons. Think of e^2/r_{12} as the value the potential energy associated with the two electrons would have if they were at r_1 and r_2 , respectively.

Why are there no separate wave functions for the two electrons? The joint probability of finding the first electron in a region A and the second in a region B (relative to the nucleus) is given by

$$p(A, B) = \int_A d^3 r_1 \int_B d^3 r_2 |\psi(\mathbf{r}_1, \mathbf{r}_2)|^2.$$

If the probability of finding the first electron in A were independent of the whereabouts of the second electron, then we could assign to it a wave function $\psi_1(\mathbf{r}_1)$, and if the probability of finding the second electron in B were independent of the whereabouts of the first electron, we could assign to it a wave function $\psi_2(\mathbf{r}_2)$. In this case $\psi(\mathbf{r}_1, \mathbf{r}_2)$ would be given by the product $\psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2)$ of the two wave functions, and $p(A, B)$ would be the product of $p(A) = \int_A d^3 r_1 |\psi_1(\mathbf{r}_1)|^2$ and $p(B) = \int_B d^3 r_2 |\psi_2(\mathbf{r}_2)|^2$. But in general, and especially inside a helium atom, the positional probability distribution for the first electron is conditional on the whereabouts of the second electron, and vice versa, given that the two electrons repel each other (to use the language of classical physics).

For the lowest energy levels, the above equation has been solved by numerical methods. With three or more electrons it is hopeless to look for exact solutions of the corresponding Schrödinger equation. Nevertheless, the Periodic Table¹⁰ and many properties of the chemical elements can be understood by using the following approximate theory.

First, we disregard the details of the interactions between the electrons. Next, since the chemical properties of atoms depend on their outermost electrons, we consider each of these atoms subject to a potential that is due to (i) the nucleus and (ii) a continuous, spherically symmetric, charge distribution doing duty for the other electrons. We again neglect spin effects *except* that we take account of the Pauli exclusion principle¹¹, according to which

9 <http://en.wikibooks.org/wiki/Category%3A>

10 <http://en.wikipedia.org/wiki/Periodic%20table>

11 <http://en.wikipedia.org/wiki/Pauli%20exclusion%20principle>

the probability of finding two electrons (more generally, two fermions¹²) having exactly the same properties is 0. Thus two electrons can be associated with exactly the same wave function provided that their spin states differ in the following way: whenever the spins of the two electrons are measured with respect to a given axis, the outcomes are perfectly anticorrelated; one will be "up" and the other will be "down". Since there are only two possible outcomes, a third electron cannot be associated with the same wave function.

This approximate theory yields stationary wave functions $\psi_{nlm}(\mathbf{r})$ called orbitals¹³ for individual electrons. These are quite similar to the stationary wave functions one obtains for the single electron of hydrogen, except that their dependence on the radial coordinate is modified by the negative charge distribution representing the remaining electrons. As a consequence of this modification, the energies associated with orbitals with the same quantum number n but different quantum numbers l are no longer equal. For any given $n \geq 1$, orbitals with higher l yield a larger mean distance between the electron and the nucleus, and the larger this distance, the more the negative charge of the remaining electrons screens the positive charge of the nucleus. As a result, an electron with higher l is less strongly bound (given the same n), so its ionization energy¹⁴ is lower.

Chemists group orbitals into shells¹⁵ according to their principal quantum number. As we have seen, the n -th shell can "accommodate" up to $n^2 \times 2$ electrons. Helium has the first shell completely "filled" and the second shell "empty." Because the helium nucleus has twice the charge of the hydrogen nucleus, the two electrons are, on average, much nearer the nucleus than the single electron of hydrogen. The ionization energy of helium is therefore much larger, 2372.3 J/mol¹⁶ as compared to 1312.0 J/mol for hydrogen. On the other hand, if you tried to add an electron to create a negative helium ion, it would have to go into the second shell, which is almost completely screened from the nucleus by the electrons in the first shell. Helium is therefore neither prone to give up an electron nor able to hold an extra electron. It is chemically inert, as are all elements in the rightmost column of the Periodic Table.

In the second row of the Periodic Table the second shell gets filled. Since the energies of the 2p orbitals are higher than that of the 2s orbital, the latter gets "filled" first. With each added electron (and proton!) the entire electron distribution gets pulled in, and the ionization energy goes up, from 520.2 J/mol for lithium (atomic number Z=3) to 2080.8 J/mol for neon (Z=10). While lithium readily parts with an electron, fluorine (Z=9) with a single empty "slot" in the second shell is prone to grab one. Both are therefore quite active chemically. The progression from sodium (Z=11) to argon (Z=18) parallels that from lithium to neon.

There is a noteworthy peculiarity in the corresponding sequences of ionization energies¹⁷: The ionization energy of oxygen (Z=8, 1313.9 J/mol) is lower than that of nitrogen (Z=7, 1402.3 J/mol), and that of sulfur (Z=16, 999.6 J/mol) is lower than that of phosphorus (Z=15, 1011.8 J/mol). To understand why this is so, we must take account of certain details of the inter-electronic forces that we have so far ignored.

12 <http://en.wikipedia.org/wiki/Fermion>

13 <http://en.wikipedia.org/wiki/Atomic%20orbital>

14 <http://en.wikipedia.org/wiki/Ionization%20potential>

15 <http://en.wikipedia.org/wiki/Electron%20shell>

16 <http://en.wikipedia.org/wiki/Joule%20per%20mole>

17 <http://en.wikipedia.org/wiki/Ionization%20energies%20of%20the%20elements>

Suppose that one of the two 2p electrons of carbon ($Z=6$) goes into the $m=0$ orbital with respect to the z axis. Where will the other 2p electron go? It will go into any vacant orbital that minimizes the repulsion between the two electrons, by maximizing their mean distance. This is neither of the orbitals with $|m|=1$ with respect to the z axis but an orbital with $m=0$ with respect to some axis perpendicular to the z axis. If we call this the x axis, then the third 2p electron of nitrogen goes into the orbital with $m=0$ relative to y axis. The fourth 2p electron of oxygen then has no choice but to go — with opposite spin — into an already occupied 2p orbital. This raises its energy significantly and accounts for the drop in ionization from nitrogen to oxygen.

By the time the 3p orbitals are "filled," the energies of the 3d states are pushed up so high (as a result of screening) that the 4s state is energetically lower. The "filling up" of the 3d orbitals therefore begins only after the 4s orbitals are "occupied," with scandium ($Z=21$).

Thus even this simplified and approximate version of the quantum theory of atoms has the power to predict the qualitative and many of the quantitative features of the Period Table.

18

4.8 Probability flux

The time rate of change of the probability density $\rho(t, \mathbf{r}) = |\psi(t, \mathbf{r})|^2$ (at a fixed location \mathbf{r}) is given by

$$\frac{\partial \rho}{\partial t} = \psi^* \frac{\partial \psi}{\partial t} + \psi \frac{\partial \psi^*}{\partial t}.$$

With the help of the Schrödinger equation and its complex conjugate,

$$i\hbar \frac{\partial \psi}{\partial t} = \frac{1}{2m} \left(\frac{\hbar}{i} \frac{\partial}{\partial \mathbf{r}} - \mathbf{A} \right) \cdot \left(\frac{\hbar}{i} \frac{\partial}{\partial \mathbf{r}} - \mathbf{A} \right) \psi + V\psi,$$

$$\frac{\hbar}{i} \frac{\partial \psi^*}{\partial t} = \frac{1}{2m} \left(i\hbar \frac{\partial}{\partial \mathbf{r}} - \mathbf{A} \right) \cdot \left(i\hbar \frac{\partial}{\partial \mathbf{r}} - \mathbf{A} \right) \psi^* + V\psi^*,$$

one obtains

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= -\frac{i}{\hbar} \psi^* \left[\frac{1}{2m} \left(\frac{\hbar}{i} \frac{\partial}{\partial \mathbf{r}} - \mathbf{A} \right) \cdot \left(\frac{\hbar}{i} \frac{\partial}{\partial \mathbf{r}} - \mathbf{A} \right) \psi + V\psi \right] \\ &\quad + \frac{i}{\hbar} \psi \left[\frac{1}{2m} \left(i\hbar \frac{\partial}{\partial \mathbf{r}} - \mathbf{A} \right) \cdot \left(i\hbar \frac{\partial}{\partial \mathbf{r}} - \mathbf{A} \right) \psi^* + V\psi^* \right]. \end{aligned}$$

The terms containing V cancel out, so we are left with

$$\frac{\partial \rho}{\partial t} = -\frac{i}{2m\hbar} \left[\psi^* \left(i\hbar \frac{\partial}{\partial \mathbf{r}} + \mathbf{A} \right) \cdot \left(i\hbar \frac{\partial}{\partial \mathbf{r}} + \mathbf{A} \right) \psi - \psi \left(i\hbar \frac{\partial}{\partial \mathbf{r}} - \mathbf{A} \right) \cdot \left(i\hbar \frac{\partial}{\partial \mathbf{r}} - \mathbf{A} \right) \psi^* \right]$$

18 http://en.wikibooks.org/wiki/Category%3A_Schrodinger_Equation

$$= \dots = -\frac{\hbar}{2mi} \left(\frac{\partial^2 \psi}{\partial \mathbf{r}^2} \psi^* - \psi \frac{\partial^2 \psi^*}{\partial \mathbf{r}^2} \right) + \frac{1}{m} \left(\psi \psi^* \frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{A} + \mathbf{A} \frac{\partial \psi}{\partial \mathbf{r}} \psi^* + \mathbf{A} \psi \frac{\partial \psi^*}{\partial \mathbf{r}} \right).$$

Next, we calculate the divergence of $\mathbf{j} = \frac{\hbar}{2mi} \left(\psi^* \frac{\partial \psi}{\partial \mathbf{r}} - \frac{\partial \psi^*}{\partial \mathbf{r}} \psi \right) - \frac{1}{m} \mathbf{A} \psi^* \psi$:

$$\frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{j} = \frac{\hbar}{2mi} \left(\frac{\partial^2 \psi}{\partial \mathbf{r}^2} \psi^* - \psi \frac{\partial^2 \psi^*}{\partial \mathbf{r}^2} \right) - \frac{1}{m} \left(\psi \psi^* \frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{A} + \mathbf{A} \frac{\partial \psi}{\partial \mathbf{r}} \psi^* + \mathbf{A} \psi \frac{\partial \psi^*}{\partial \mathbf{r}} \right).$$

The upshot:

$$\frac{\partial \rho}{\partial t} = -\frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{j}.$$

Integrated over a spatial region R with unchanging boundary ∂R :

$$\frac{\partial}{\partial t} \int_R \rho d^3 r = - \int_R \frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{j} d^3 r.$$

According to Gauss's law¹⁹, the outward flux of \mathbf{j} through ∂R equals the integral of the divergence²⁰ of \mathbf{j} over R :

$$\oint_{\partial R} \mathbf{j} \cdot d\Sigma = \int_R \frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{j} d^3 r.$$

We thus have that

$$\frac{\partial}{\partial t} \int_R \rho d^3 r = - \oint_{\partial R} \mathbf{j} \cdot d\Sigma.$$

If ρ is the continuous density of some kind of stuff (stuff per unit volume) and \mathbf{j} is its flux (stuff per unit area per unit time), then on the left-hand side we have the rate at which the stuff inside R increases, and on the right-hand side we have the rate at which stuff enters through the surface of R . So if some stuff moves from place A to place B, it crosses the boundary of any region that contains either A or B. This is why the framed equation is known as a continuity equation²¹.

In the quantum world, however, there is no such thing as continuously distributed and/or continuously moving stuff. ρ and \mathbf{j} , respectively, are a density (something per unit volume) and a flux (something per unit area per unit time) only in a formal sense. If ψ is the wave function associated with a particle, then the integral $\int_R \rho d^3 r = \int_R |\psi|^2 d^3 r$ gives the probability of finding the particle in R *if the appropriate measurement is made*, and the framed equation tells us this: if the probability of finding the particle inside R , as a function of the time at which the measurement is made, increases, then the probability of finding the particle outside R , as a function of the same time, decreases by the same amount. (Much

19 <http://en.wikipedia.org/wiki/Gauss%27s%20law>

20 <http://en.wikipedia.org/wiki/Divergence>

21 <http://en.wikipedia.org/wiki/Continuity%20equation>

the same holds if ψ is associated with a system having n degrees of freedom and R is a region of the system's configuration space.) This is sometimes expressed by saying that "probability is (locally) conserved." When you hear this, then remember that the probability for something to happen in a given place at a given time isn't anything that is situated at that place or that exists at that time.

22

22 <http://en.wikibooks.org/wiki/Category%3A>

5 Entanglement (a preview)

5.1 Bell's theorem: the simplest version

Quantum mechanics permits us to create the following scenario.

- Pairs of particles are launched in opposite directions.
- Each particle is subjected to one of three possible measurements (**1**, **2**, or **3**).
- Each time the two measurements are chosen at random.
- Each measurement has two possible results, indicated by a red or green light.

Here is what we find:

- If both particles are subjected to the same measurement, identical results are never obtained.
- The two sequences of recorded outcomes are completely random. In particular, half of the time both lights are the same color.

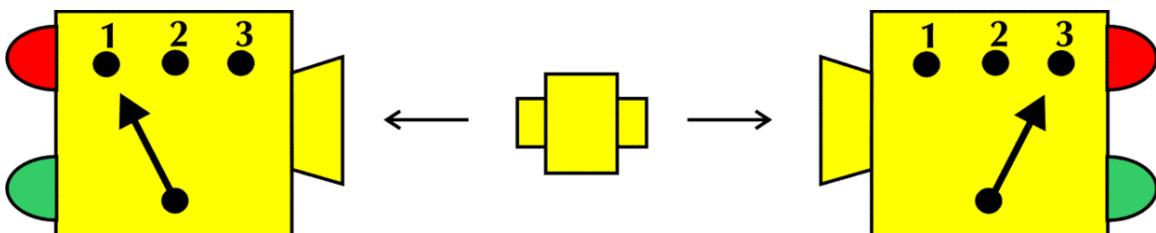


Figure 68

If this doesn't bother you, then please explain how it is that the colors differ whenever identical measurements are performed!

The obvious explanation would be that each particle arrives with an "instruction set" — some property that pre-determines the outcome of every possible measurement. Let's see what this entails.

Each particle arrives with one of the following $23 = 8$ instruction sets:

RRR, RRG, RGR, GRR, RGG, GRG, GGR, or GGG.

(If a particle arrives with, say, **RGG**, then the apparatus flashes red if it is set to **1** and green if it is set to **2** or **3**.) In order to explain why the outcomes differ whenever both particles are subjected to the same measurement, we have to assume that particles launched together arrive with opposite instruction sets. If one carries the instruction (or arrives with the property denoted by) **RRG**, then the other carries the instruction **GGR**.

Suppose that the instruction sets are **RRG** and **GGR**. In this case we observe different colors with the following five of the $32 = 9$ possible combinations of apparatus settings:

1—1,2—2,3—3,1—2, and **2—1**,

and we observe equal colors with the following four:

1—3,2—3,3—1, and **3—2**.

Because the settings are chosen at random, this particular pair of instruction sets thus results in different colors $5/9$ of the time. The same is true for the other pairs of instruction sets *except* the pair **RRR**, **GGG**. If the two particles carry these respective instruction sets, we see different colors *every* time. It follows that we see different colors *at least* $5/9$ of the time.

But different colors are observed half of the time! In reality the probability of observing different colors is $1/2$. Conclusion: the statistical predictions of quantum mechanics cannot be explained with the help of instruction sets. In other words, these measurements do not reveal *pre-existent* properties. They *create* the properties the possession of which they indicate.

Then how is it that the colors differ whenever identical measurements are made? How does one apparatus "know" which measurement is performed and which outcome is obtained *by the other apparatus*?

Whenever the joint probability $p(\mathbf{A},\mathbf{B})$ of the respective outcomes **A** and **B** of two measurements does not equal the product $p(\mathbf{A}) p(\mathbf{B})$ of the individual probabilities, the outcomes — or their probabilities — are said to be *correlated*. With equal apparatus settings we have $p(\mathbf{R},\mathbf{R}) = p(\mathbf{G},\mathbf{G}) = 0$, and this obviously differs from the products $p(\mathbf{R}) p(\mathbf{R})$ and $p(\mathbf{G}) p(\mathbf{G})$, which equal $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. What kind of mechanism is responsible for the correlations between the measurement outcomes?

You understand this as much as anybody else!

The conclusion that we see different colors at least $5/9$ of the time is *Bell's theorem* (or *Bell's inequality*) for this particular setup. The fact that the universe violates the logic of Bell's Theorem is evidence that particles do not carry instruction sets embedded within them and instead have instantaneous knowledge of other particles at a great distance. Here is a comment by a distinguished Princeton physicist as quoted by David Mermin¹

Anybody who's not bothered by Bell's theorem has to have rocks in his head.

And here is why Einstein wasn't happy with quantum mechanics:

I cannot seriously believe in it because it cannot be reconciled with the idea that physics should represent a reality in time and space, free from spooky actions at a distance.²

Sadly, Einstein (1879 - 1955) did not know Bell's theorem of 1964. We know now that

¹ N. David Mermin, "Is the Moon there when nobody looks? Reality and the quantum theory," *Physics Today*, April 1985. The version of Bell's theorem discussed in this section first appeared in this article.

² Albert Einstein, *The Born-Einstein Letters*, with comments by Max Born (New York: Walker, 1971).

there must be a mechanism whereby the setting of one measurement device can influence the reading of another instrument, however remote.³

Spooky actions at a distance are here to stay!

4

5.2 A quantum game

Here are the rules:⁵

- Two teams play against each other: Andy, Bob, and Charles (the "players") versus the "interrogators".
- Each player is asked either "What is the value of **X**?" or "What is the value of **Y**?"
- Only two answers are allowed: +1 or -1.
- Either each player is asked the **X** question, or one player is asked the **X** question and the two other players are asked the **Y** question.
- The players win if the product of their answers is -1 in case only **X** questions are asked, and if the product of their answers is +1 in case **Y** questions are asked. Otherwise they lose.
- The players are not allowed to communicate with each other once the questions are asked. Before that, they are permitted to work out a strategy.

Is there a failsafe strategy? Can they make sure that they will win? Stop to ponder the question.

Let us try pre-agreed answers, which we will call **XA**, **XB**, **XC** and **YA**, **YB**, **YC**. The winning combinations satisfy the following equations:

$$X_A Y_B Y_C = 1, \quad Y_A X_B Y_C = 1, \quad Y_A Y_B X_C = 1, \quad X_A X_B X_C = -1.$$

Consider the first three equations. The product of their right-hand sides equals +1. The product of their left-hand sides equals **XAXBXC**, implying that **XAXBXC = 1**. (Remember that the possible values are ± 1 .) But if **XAXBXC = 1**, then the fourth equation **XAXBXC = -1** obviously cannot be satisfied.

The bottom line: There is no failsafe strategy with pre-agreed answers.

6

3 John S. Bell, "On the Einstein Podolsky Rosen paradox," *Physics* 1, pp. 195-200, 1964.

4 <http://en.wikibooks.org/wiki/Category%3A>

5 Lev Vaidman, "Variations on the theme of the Greenberger-Horne-Zeilinger proof," *Foundations of Physics* 29, pp. 615-30, 1999.

6 <http://en.wikibooks.org/wiki/Category%3A>

5.3 The experiment of Greenberger, Horne, and Zeilinger

And yet there is a failsafe strategy.⁷

Here goes:

- Andy, Bob, and Charles prepare three particles (for instance, electrons) in a particular way. As a result, they are able to predict the probabilities of the possible outcomes of any spin measurement to which the three particles may subsequently be subjected. In principle these probabilities do not depend on how far the particles are apart.
- Each player takes one particle with him.
- Whoever is asked the **X** question measures the x component of the spin of his particle and answers with his outcome, and whoever is asked the **Y** question measures the y component of the spin of his particle and answers likewise. (All you need to know at this point about the spin of a particle is that its component with respect to any one axis can be measured, and that for the type of particle used by the players there are two possible outcomes, namely +1 and -1.)

Proceeding in this way, the team of players is sure to win every time.

Is it possible for the x and y components of the spins of the three particles to be in possession of values before their values are actually measured?

Suppose that the y components of the three spins have been measured. The three equations

$$X_A Y_B Y_C = 1, \quad Y_A X_B Y_C = 1, \quad Y_A Y_B X_C = 1$$

of the previous section⁸ tell us what we would have found if the x component of any one of the three particles had been measured instead of the y component. If we assume that the x components are in possession of values even though they are *not* measured, then their values can be inferred from the measured values of the three y components.

Try to fill in the following table in such a way that

- each cell contains either +1 or -1,
- the product of the three X values equals -1, and
- the product of every pair of Y values equals the remaining X value.

Can it be done?

	A	B	C
X			
Y			

The answer is negative, for the same reason that the four equations

⁷ D. M. Greenberger, M. A. Horne, and A. Zeilinger, "Going beyond Bell's theorem," in *Bell's theorem, Quantum Theory, and Conception of the Universe*, edited by M. Kafatos (Dordrecht: Kluwer Academic, 1989), pp. 69-72.

⁸ Chapter 5.1 on page 89

$$X_A Y_B Y_C = 1, \quad Y_A X_B Y_C = 1, \quad Y_A Y_B X_C = 1, \quad X_A X_B X_C = -1$$

cannot all be satisfied. Just as there can be no strategy with pre-agreed answers, there can be no pre-existent values. We seem to have no choice but to conclude that these spin components are in possession of values *only if* (and only when) they are actually measured.

Any two outcomes suffice to predict a third outcome. If two x components are measured, the third x component can be predicted, if two y components are measured, the x component of the third spin can be predicted, and if one x and one y component are measurement, the y component of the third spin can be predicted. How can we understand this given that

- the values of the spin components are created as and when they are measured,
- the relative times of the measurements are irrelevant,
- in principle the three particles can be millions of miles apart.

How does the third spin "know" which components of the other spins are measured and which outcomes are obtained? What mechanism correlates the outcomes?

You understand this as much as anybody else!

9

9 <http://en.wikibooks.org/wiki/Category%3A>

6 Authors

List authors here.

7 Appendix

7.1 Probability

7.1.1 Basic Concepts

Probability is a numerical measure of likelihood. If an event has a probability equal to 1 (or 100%), then it is certain to occur. If it has a probability equal to 0, then it will definitely not occur. And if it has a probability equal to 1/2 (or 50%), then it is as likely as not to occur.

You will know that tossing a fair coin has probability 1/2 to yield heads, and that casting a fair die has probability 1/6 to yield a 1. How do we know this?

There is a principle known as the *principle of indifference*, which states: if there are n mutually exclusive and jointly exhaustive possibilities, and if, as far as we know, there are no differences between the n possibilities apart from their names (such as "heads" or "tails"), then each possibility should be assigned a probability equal to $1/n$. (*Mutually exclusive*: only one possibility can be realized in a single trial. *Jointly exhaustive*: at least one possibility is realized in a single trial. *Mutually exclusive and jointly exhaustive*: exactly one possibility is realized in a single trial.)

Since this principle appeals to what we *know*, it concerns *epistemic* probabilities (a.k.a. *subjective* probabilities) or *degrees of belief*. If you are certain of the truth of a proposition, then you assign to it a probability equal to 1. If you are certain that a proposition is false, then you assign to it a probability equal to 0. And if you have no information that makes you believe that the truth of a proposition is more likely (or less likely) than its falsity, then you assign to it probability 1/2. Subjective probabilities are therefore also known as *ignorance probabilities*: if you are ignorant of any differences between the possibilities, you assign to them equal probabilities.

If we assign probability 1 to a proposition because we *believe* that it is true, we assign a subjective probability, and if we assign probability 1 to an event because it *is* certain that it will occur, we assign an objective probability. Until the advent of quantum mechanics, the only objective probabilities known were *relative frequencies*.

The advantage of the frequentist definition of probability is that it allows us to measure probabilities, at least approximately. The trouble with it is that it refers to *ensembles*. You can't measure the probability of heads by tossing a single coin. You get better and better approximations to the probability of heads by tossing a larger and larger number N of coins and dividing the number N_H of heads by N . The exact probability of heads is the limit

$$p(H) = \lim_{N \rightarrow \infty} \frac{N_H}{N}.$$

The meaning of this formula is that for any positive number ϵ , however small, you can find a (sufficiently large but finite) number N such that

$$\left| p(H) - \frac{N_H}{N} \right| < \epsilon.$$

The probability that m events from a mutually exclusive and jointly exhaustive set of n possible events happen is the sum of the probabilities of the m events. Suppose, for example, you win if you cast either a 1 or a 6. The probability of winning is

$$p(1 \text{ or } 6) = p(1) + p(6) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

In frequentist terms, this is virtually self-evident. $N(1)/N$ approximates $p(1)$, $N(6)/N$ approximates $p(6)$, and $[N(1) + N(6)]/N$ approximates $p(1 \text{ or } 6)$.

The probability that two *independent* events happen is the product of the probabilities of the individual events. Suppose, for example, you cast two dice and you win if the total is 12. Then

$$p(6 \text{ and } 6) = p(6) \times p(6) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}.$$

By the principle of indifference, there are now $6 \times 6 = 36$ equiprobable possibilities, and casting a total of 12 with two dice is one of them.

It is important to remember that the *joint probability* $p(A, B) = p(A \text{ and } B)$ of two events A, B equals the product of the individual probabilities $p(A)$ and $p(B)$ *only if* the two events are independent, meaning that the probability of one does not depend on whether or not the other happens. In terms of propositions: the probability that the conjunction P_1 and P_2 is true is the probability that P_1 is true times the probability that P_2 is true *only if* the probability that either proposition is true does not depend on whether the other is true or false. Ignoring this can have the most tragic consequences¹.

The general rule for the joint probability of two events is

$$p(A, B) = p(B|A)p(A) = p(A|B)p(B).$$

$p(B|A)$ is a *conditional probability*: the probability of B given that A .

To see this, let $N(A, B)$ be the number of trials in which both A and B happen or are true. $N(A, B)/N$ approximates $p(A, B)$, $N(A, B)/N(A)$ approximates $p(B|A)$, and $N(A)/N$ approximates $p(A)$. But

$$p(A, B) \xleftarrow{N \rightarrow \infty} \frac{N(A, B)}{N} = \frac{N(A, B)}{N(A)} \times \frac{N(A)}{N} \xrightarrow{N \rightarrow \infty} p(B|A)p(A).$$

¹ <http://en.wikipedia.org/wiki/Sally%20Clark>

An immediate consequence of this is *Bayes' theorem*:

$$p(B|A) = \frac{p(A|B)}{p(A)} p(B).$$

The following is just as readily established:

$$p(X) = p(X|Y)p(Y) + p(X|\bar{Y})p(\bar{Y}),$$

where \bar{Y} happens or is true whenever Y does not happen or is false. The generalization to $n > 2$ mutually exclusive and jointly exhaustive possibilities should be obvious.

Given a *random variable*, which is a set $X = \{x_1, \dots, x_n\}$ of random numbers, we may want to know the arithmetic mean

$$\langle X \rangle = \frac{1}{n} \sum_{k=1}^n x_k = \frac{x_1 + \dots + x_n}{n}$$

as well as the *standard deviation*, which is the root-mean-square deviation from the arithmetic mean,

$$\sigma(X) = \sqrt{\frac{1}{n} \sum_{k=1}^n (x_k - \langle X \rangle)^2}.$$

The standard deviation is an important measure of *statistical dispersion*.

Given n possible measurement outcomes v_1, \dots, v_n with probabilities $p_k = p(v_k)$, we have a *probability distribution* $\{p_1, \dots, p_n\}$, and we may want to know the *expected value* of X , defined by

$$\langle X \rangle = \sum_{k=1}^n p_k x_k$$

as well as the corresponding standard deviation

$$\sigma(X) = \sqrt{\sum_{k=1}^n p_k (x_k - \langle X \rangle)^2},$$

which is a handy measure of the fuzziness of X .

We have defined probability as a numerical measure of likelihood. So what is likelihood? What is probability apart from being a numerical measure? The frequentist definition covers some cases, the epistemic definition covers others, but which definition would cover all cases? It seems that probability is one of those concepts that are intuitively meaningful to us, but — just like time or the experience of purple — cannot be explained in terms of other concepts.

7.1.2 Some Problems

Problem 1 (Monty Hall). A player in a game show is given the choice of three doors. Behind one door is the Grand Prize (say, a car); behind the other two doors are booby prizes (say, goats). The player picks a door, and the show host peeks behind the doors and opens one of the remaining doors. There is a booby prize behind the door he opened. The host then offers the player either to stay with the door that was chosen at the beginning, or to switch to the other closed door. What gives the player the better chance of winning: to switch doors or to stay with the original choice? Or are the chances equal?

Problem 2. Imagine you toss a coin successively and wait till the first time the pattern HTT appears. For example, if the sequence of tosses was

H H T H H T H H T T H H T T T H T H

then the pattern HTT would appear after the 10th toss. Let $A(\text{HTT})$ be the average number of tosses until HTT occurs, and let $A(\text{HTH})$ be the average number of tosses until HTH occurs. Which of the following is true?

- (a) $A(\text{HTH}) < A(\text{HTT})$, (b) $A(\text{HTH}) = A(\text{HTT})$, or (c) $A(\text{HTH}) > A(\text{HTT})$.

Problem 3. Imagine a test for a certain disease (say, HIV) that is 99% accurate. And suppose a person picked at random tests positive. What is the probability that the person actually has the disease?

Solutions

Problem 1. Let $p(C1)$ be the probability that the car is behind door 1, $p(O3)$ the probability that the host opens door 3, and $p(O3|C1)$ the probability that the host opens door 3 given that the car is behind door 1. We have

$$p(O3) = p(O3|C1)p(C1) + p(O3|C2)p(C2) + p(O3|C3)p(C3)$$

as well as

$$p(O3|C2)p(C2) = p(C2|O3)p(O3).$$

If the first choice is door 1, then $p(O3|C1) = 1/2$, $p(O3|C2) = 1$, and $p(O3|C3) = 0$. Hence

$$p(O3) = \frac{1}{2} \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3} = \frac{1}{2}$$

and thus

² <http://en.wikibooks.org/wiki/Category%3A>

$$p(C2|O3) = \frac{p(O3|C2)p(C2)}{p(O3)} = \frac{1 \times \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}.$$

In words: If the player's first choice is door 1 and the host opens door 3, then the probability that the car is behind door 2 is $2/3$, whereas the probability that it is behind door 1 is $1 - 2/3 = 1/3$. A quicker way to see that switching *doubles* the chances of winning is to compare this game with another one, in which the show host offers the choice of either opening the originally chosen door or opening *both* other doors (and winning regardless of which, if any, has the car).

Note: This result depends on the show host *deliberately* opening only a door with a goat behind it. If she doesn't know - or doesn't care (!) - which door the car is behind, and opens a remaining door at random, then $1/3$ of the outcomes that were initially possible have been removed by her having opened a door with a goat. In this case the player gains no advantage (or disadvantage) by switching. So the answer depends on the rules of the game, not just the sequence of events. Of course the player may not know what the 'rules' are in this respect, in which case he should still switch doors because there can be no disadvantage in doing so.

Problem 2. The average number of tosses until HTT occurs, $A(\text{HTT})$, equals 8, whereas $A(\text{HTH}) = 10$. To see why the latter is greater, imagine you have tossed HT. If you are looking for HTH and the next toss gives you HTT, then your next chance to see HTH is after a total of 6 tosses, whereas if you are looking for HTT and the next toss gives you HTH, then your next chance to see HTT is after a total of 5 tosses.

Problem 3. The answer depends on how rare the disease is. Suppose that one in 10,000 has it. This means 100 in a million. If a million are tested, there will be 99 true positives and one false negative. 99% of the remaining 999,900 — that is, 989,901 — will yield true negatives and 1% — that is, 9,999 — will yield false positives. The probability that a randomly picked person testing positive actually has the disease is the number of true positives divided by the number of positives, which in this particular example is $99/(9999+99) = 0.0098$ — less than 1%!

7.1.3 Moral

Be it scientific data or evidence in court — there are usually competing explanations, and usually each explanation has a likely bit and an unlikely bit. For example, having the disease is unlikely, but the test is likely to be correct; not having the disease is likely, but a false test result is unlikely. You can see the importance of accurate assessments of the likelihood of competing explanations, and if you have tried the problems, you have seen that we aren't very good at such assessments.

³

³ http://en.wikibooks.org/wiki/Category%3A_Probability

7.2 Mathematical tools

7.2.1 Elements of calculus

A definite integral

Imagine an object \mathcal{O} that is free to move in one dimension — say, along the x axis. Like every physical object, it has a more or less fuzzy position (relative to whatever reference object we choose). For the purpose of describing its fuzzy position, quantum mechanics provides us with a probability density $\rho(x)$. This depends on actual measurement outcomes, and it allows us to calculate the probability of finding the particle in any given interval of the x axis, provided that an appropriate measurement is made. (Remember our mantra: the mathematical formalism of quantum mechanics serves to assign probabilities to possible measurement outcomes on the basis of actual outcomes.)

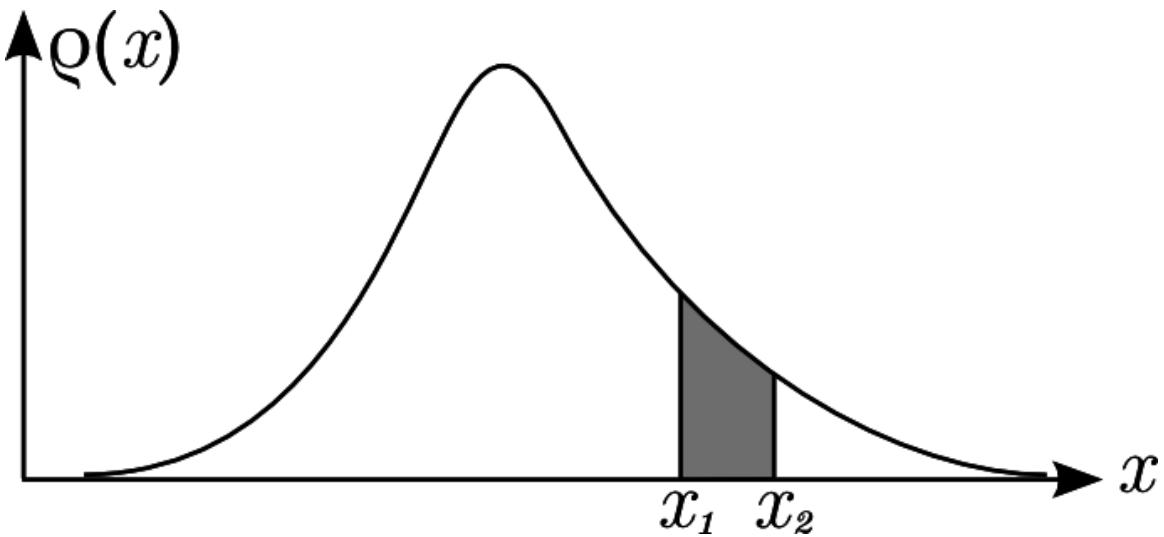


Figure 69

We call $\rho(x)$ a probability *density* because it represents a probability per unit length. The probability of finding \mathcal{O} in the interval between x_1 and x_2 is given by the area A between the graph of $\rho(x)$, the x axis, and the vertical lines at x_1 and x_2 , respectively. How do we calculate this area? The trick is to cover it with narrow rectangles of width Δx .

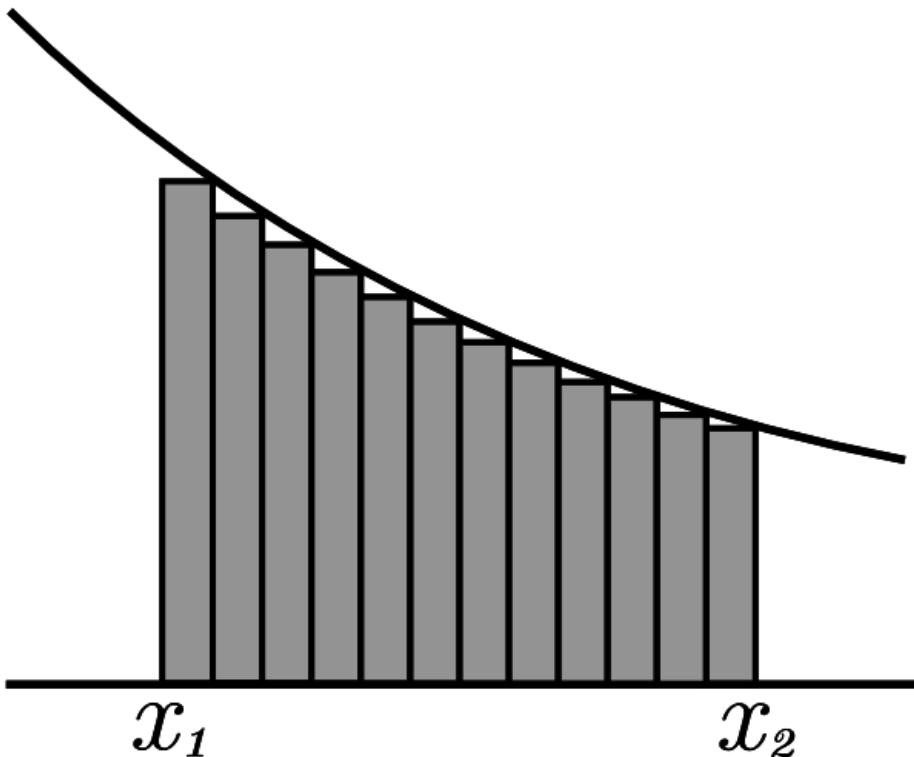


Figure 70

The area of the first rectangle from the left is $\rho(x_1 + \Delta x)\Delta x$, the area of the second is $\rho(x_1 + 2\Delta x)\Delta x$, and the area of the last is $\rho(x_1 + 12\Delta x)\Delta x$. For the sum of these areas we have the shorthand notation

$$\sum_{k=1}^{12} \rho(x + k\Delta x)\Delta x.$$

It is not hard to visualize that if we increase the number N of rectangles and at the same time decrease the width Δx of each rectangle, then the sum of the areas of all rectangles fitting under the graph of $\rho(x)$ between x_1 and x_2 gives us a better and better approximation to the area A and thus to the probability of finding \mathcal{O} in the interval between x_1 and x_2 . As Δx tends toward 0 and N tends toward infinity (∞), the above sum tends toward the *integral*

$$\int_{x_1}^{x_2} \rho(x) dx.$$

We sometimes call this a *definite* integral to emphasize that it's just a number. (As you can guess, there are also *indefinite* integrals, about which more later.) The uppercase delta has turned into a *d* indicating that dx is an infinitely small (or *infinitesimal*) width, and the summation symbol (the uppercase sigma) has turned into an elongated S indicating that we are adding infinitely many infinitesimal areas.

Don't let the term "infinitesimal" scare you. An infinitesimal quantity means nothing by itself. It is the *combination* of the integration symbol \int with the infinitesimal quantity dx that makes sense as a *limit*, in which N grows above any number however large, dx (and hence the area of each rectangle) shrinks below any (positive) number however small, while the sum of the areas tends toward a well-defined, finite number.

4

Differential calculus: a very brief introduction

Another method by which we can obtain a well-defined, finite number from infinitesimal quantities is to divide one such quantity by another.

We shall assume throughout that we are dealing with well-behaved functions, which means that you can plot the graph of such a function without lifting up your pencil, and you can do the same with each of the function's derivatives. So what is a function, and what is the derivative of a function?

A *function* $f(x)$ is a machine with an input and an output. Insert a number x and out pops the number $f(x)$. Rather confusingly, we sometimes think of $f(x)$ not as a machine that churns out numbers but as the number churned out when x is inserted.

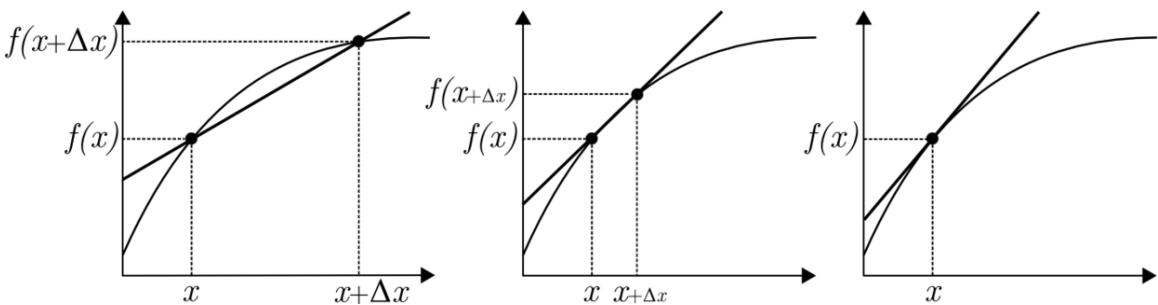


Figure 71

The (first) *derivative* $f'(x)$ of $f(x)$ is a function that tells us how much $f(x)$ increases as x increases (starting from a given value of x , say x_0) in the limit in which both the increase Δx in x and the corresponding increase $\Delta f = f(x + \Delta x) - f(x)$ in $f(x)$ (which of course may be negative) tend toward 0:

$$f'(x_0) = \lim_{\Delta x \rightarrow 0} \frac{\Delta f}{\Delta x} = \frac{df}{dx}(x_0).$$

The above diagrams illustrate this limit. The ratio $\Delta f / \Delta x$ is the *slope* of the straight line through the black circles (that is, the tan of the angle between the positive x axis and the straight line, measured counterclockwise from the positive x axis). As Δx decreases, the black circle at $x + \Delta x$ slides along the graph of $f(x)$ towards the black circle at x , and the slope of the straight line through the circles increases. In the limit $\Delta x \rightarrow 0$, the straight line

4 <http://en.wikibooks.org/wiki/Category%3A>

becomes a tangent on the graph of $f(x)$, touching it at x . The slope of the tangent on $f(x)$ at x_0 is what we mean by *the slope of $f(x)$ at x_0* .

So the first derivative $f'(x)$ of $f(x)$ is the function that equals the slope of $f(x)$ for every x . To *differentiate* a function f is to obtain its first derivative f' . By differentiating f' , we obtain the second derivative $f'' = \frac{d^2f}{dx^2}$ of f , by differentiating f'' we obtain the third derivative $f''' = \frac{d^3f}{dx^3}$, and so on.

It is readily shown that if a is a number and f and g are functions of x , then

$$\frac{d(af)}{dx} = a \frac{df}{dx} \quad \text{and} \quad \frac{d(f+g)}{dx} = \frac{df}{dx} + \frac{dg}{dx}.$$

A slightly more difficult problem is to differentiate the product $e = fg$ of two functions of x . Think of f and g as the vertical and horizontal sides of a rectangle of area e . As x increases by Δx , the product fg increases by the sum of the areas of the three white rectangles in this diagram:

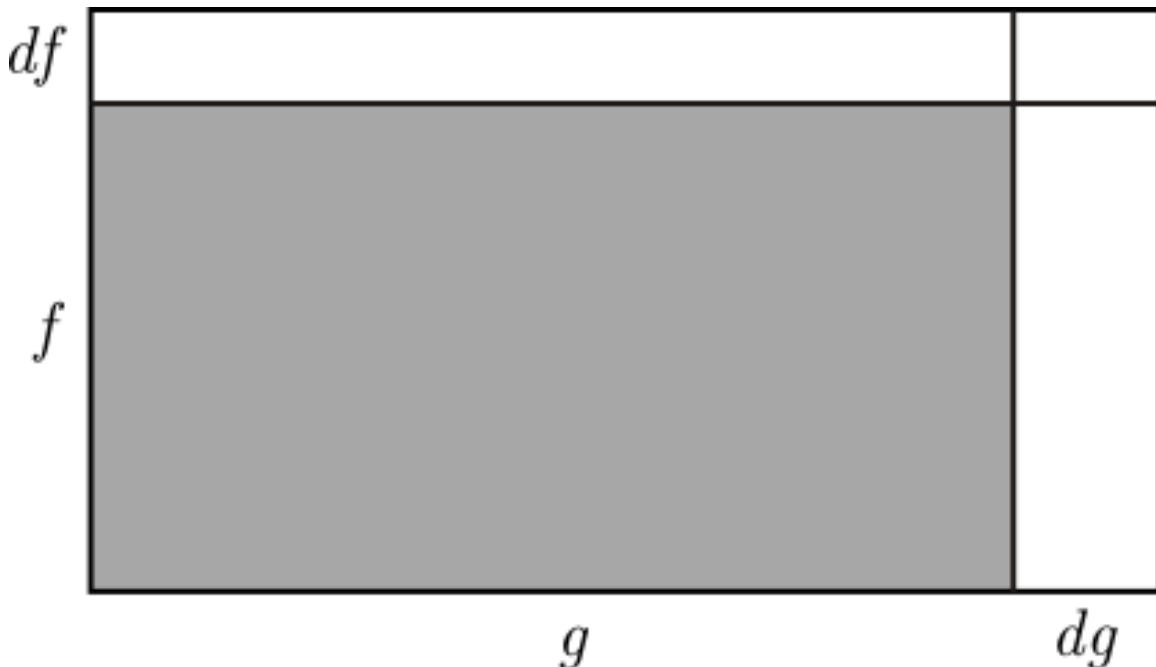


Figure 72

In other "words",

$$\Delta e = f(\Delta g) + (\Delta f)g + (\Delta f)(\Delta g)$$

and thus

$$\frac{\Delta e}{\Delta x} = f \frac{\Delta g}{\Delta x} + \frac{\Delta f}{\Delta x} g + \frac{\Delta f \Delta g}{\Delta x}.$$

If we now take the limit in which Δx and, hence, Δf and Δg tend toward 0, the first two terms on the right-hand side tend toward $fg' + f'g$. What about the third term? Because it

is the product of an expression (either Δf or Δg) that tends toward 0 and an expression (either $\Delta g/\Delta x$ or $\Delta f/\Delta x$) that tends toward a finite number, it tends toward 0. The bottom line:

$$e' = (fg)' = fg' + f'g.$$

This is readily generalized to products of n functions. Here is a special case:

$$(f^n)' = f^{n-1} f' + f^{n-2} f' f + f^{n-3} f' f^2 + \cdots + f' f^{n-1} = n f^{n-1} f'.$$

Observe that there are n equal terms between the two equal signs. If the function f returns whatever you insert, this boils down to

$$(x^n)' = nx^{n-1}.$$

Now suppose that g is a function of f and f is a function of x . An increase in x by Δx causes an increase in f by $\Delta f \approx \frac{df}{dx} \Delta x$, and this in turn causes an increase in g by $\Delta g \approx \frac{dg}{df} \Delta f$. Thus $\frac{\Delta g}{\Delta x} \approx \frac{dg}{df} \frac{df}{dx}$. In the limit $\Delta x \rightarrow 0$ the \approx becomes a $=$:

$$\frac{dg}{dx} = \frac{dg}{df} \frac{df}{dx}.$$

We obtained $(x^n)' = nx^{n-1}$ for integers $n \geq 2$. Obviously it also holds for $n = 0$ and $n = 1$.

1. Show that it also holds for negative integers n . Hint: Use the product rule to calculate $(x^n x^{-n})'$.
2. Show that $(\sqrt{x})' = 1/(2\sqrt{x})$. Hint: Use the product rule to calculate $(\sqrt{x}\sqrt{x})'$.
3. Show that $(x^n)' = nx^{n-1}$ also holds for $n = 1/m$ where m is a natural number.
4. Show that this equation also holds if n is a rational number. Use $\frac{dg}{dx} = \frac{dg}{df} \frac{df}{dx}$.

Since every real number is the limit of a sequence of rational numbers, we may now confidently proceed on the assumption that $(x^n)' = nx^{n-1}$ holds for all real numbers n .

5

Taylor series

A well-behaved function can be expanded into a power series. This means that for all non-negative integers k there are real numbers a_k such that

$$f(x) = \sum_{k=0}^{\infty} a_k x^k = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + \dots$$

Let us calculate the first four derivatives using $(x^n)' = nx^{n-1}$:

5 <http://en.wikibooks.org/wiki/Category%3A>

$$f'(x) = a_1 + 2a_2x + 3a_3x^2 + 4a_4x^3 + 5a_5x^4 + \dots$$

$$f''(x) = 2a_2 + 2 \cdot 3a_3x + 3 \cdot 4a_4x^2 + 4 \cdot 5a_5x^3 + \dots$$

$$f'''(x) = 2 \cdot 3a_3 + 2 \cdot 3 \cdot 4a_4x + 3 \cdot 4 \cdot 5a_5x^2 + \dots$$

$$f''''(x) = 2 \cdot 3 \cdot 4a_4 + 2 \cdot 3 \cdot 4 \cdot 5a_5x + \dots$$

Setting x equal to zero, we obtain

$$f(0) = a_0, \quad f'(0) = a_1, \quad f''(0) = 2a_2, \quad f'''(0) = 2 \times 3a_3, \quad f''''(0) = 2 \times 3 \times 4a_4.$$

Let us write $f^{(n)}(x)$ for the n -th derivative of $f(x)$. We also write $f^{(0)}(x) = f(x)$ — think of $f(x)$ as the "zeroth derivative" of $f(x)$. We thus arrive at the general result $f^{(k)}(0) = k! a_k$, where the *factorial* $k!$ is defined as equal to 1 for $k = 0$ and $k = 1$ and as the product of all natural numbers $n \leq k$ for $k > 1$. Expressing the coefficients a_k in terms of the derivatives of $f(x)$ at $x = 0$, we obtain

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k = f(0) + f'(0)x + f''(0)\frac{x^2}{2!} + f'''(0)\frac{x^3}{3!} + \dots$$

This is the *Taylor series* for $f(x)$.

A remarkable result: if you know the value of a well-behaved function $f(x)$ and the values of all of its derivatives *at the single point* $x = 0$ then you know $f(x)$ at *all* points x . Besides, there is nothing special about $x = 0$, so $f(x)$ is also determined by its value and the values of its derivatives at any other point x_0 :

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k.$$

6

The exponential function

We define the function $\exp(x)$ by requiring that

$$\exp'(x) = \exp(x) \text{ and } \exp(0) = 1.$$

The value of this function is everywhere equal to its slope. Differentiating the first defining equation repeatedly we find that

$$\exp^{(n)}(x) = \exp^{(n-1)}(x) = \dots = \exp(x).$$

6 <http://en.wikibooks.org/wiki/Category%3A>

The second defining equation now tells us that $\exp^{(k)}(0) = 1$ for all k . The result is a particularly simple Taylor series:

$$\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \dots$$

Let us check that a well-behaved function satisfies the equation

$$f(a)f(b) = f(a+b)$$

if and only if

$$f^{(i+k)}(0) = f^{(i)}(0)f^{(k)}(0).$$

We will do this by expanding the f 's in powers of a and b and compare coefficients. We have

$$f(a)f(b) = \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} \frac{f^{(i)}(0)f^{(k)}(0)}{i!k!} a^i b^k,$$

and using the binomial expansion

$$(a+b)^i = \sum_{l=0}^i \frac{i!}{(i-l)!l!} a^{i-l} b^l,$$

we also have that

$$f(a+b) = \sum_{i=0}^{\infty} \frac{f^{(i)}(0)}{i!} (a+b)^i = \sum_{i=0}^{\infty} \sum_{l=0}^i \frac{f^{(i)}(0)}{(i-l)!l!} a^{i-l} b^l = \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} \frac{f^{(i+k)}(0)}{i!k!} a^i b^k.$$

Voilà.

The function $\exp(x)$ obviously satisfies $f^{(i+k)}(0) = f^{(i)}(0)f^{(k)}(0)$ and hence $f(a)f(b) = f(a+b)$.

So does the function $f(x) = \exp(ux)$.

Moreover, $f^{(i+k)}(0) = f^{(i)}(0)f^{(k)}(0)$ implies $f^{(n)}(0) = [f'(0)]^n$.

We gather from this

- that the functions satisfying $f(a)f(b) = f(a+b)$ form a one-parameter family, the parameter being the real number $f'(0)$, and
- that the one-parameter family of functions $\exp(ux)$ satisfies $f(a)f(b) = f(a+b)$, the parameter being the real number u .

But $f(x) = v^x$ also defines a one-parameter family of functions that satisfies $f(a)f(b) = f(a+b)$, the parameter being the positive number v .

Conclusion: for every real number u there is a positive number v (and vice versa) such that $v^x = \exp(ux)$.

One of the most important numbers is e , defined as the number v for which $u = 1$, that is: $e^x = \exp(x)$:

$$e = \exp(1) = \sum_{n=0}^{\infty} \frac{1}{n!} = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \dots = 2.7182818284590452353602874713526\dots$$

The *natural logarithm* $\ln(x)$ is defined as the inverse of $\exp(x)$, so $\exp[\ln(x)] = \ln[\exp(x)] = x$. Show that

$$\frac{d\ln f(x)}{dx} = \frac{1}{f(x)} \frac{df}{dx}.$$

Hint: differentiate $\exp\{\ln[f(x)]\}$.

7

The indefinite integral

How do we add up infinitely many infinitesimal areas? This is elementary if we know a function $F(x)$ of which $f(x)$ is the first derivative. If $f(x) = \frac{dF}{dx}$ then $dF(x) = f(x)dx$ and

$$\int_a^b f(x) dx = \int_a^b dF(x) = F(b) - F(a).$$

All we have to do is to add up the infinitesimal amounts dF by which $F(x)$ increases as x increases from a to b , and this is simply the difference between $F(b)$ and $F(a)$.

A function $F(x)$ of which $f(x)$ is the first derivative is called an *integral* or *antiderivative* of $f(x)$. Because the integral of $f(x)$ is determined only up to a constant, it is also known as *indefinite integral* of $f(x)$. Note that wherever $f(x)$ is negative, the area between its graph and the x axis counts as negative.

How do we calculate the integral $I = \int_a^b dx f(x)$ if we don't know any antiderivative of the *integrand* $f(x)$? Generally we look up a table of integrals. Doing it ourselves calls for a significant amount of skill. As an illustration, let us do the Gaussian integral

$$I = \int_{-\infty}^{+\infty} dx e^{-x^2/2}.$$

7 <http://en.wikibooks.org/wiki/Category%3A>

For this integral someone has discovered the following trick. (The trouble is that different integrals generally require different tricks.) Start with the square of I :

$$I^2 = \int_{-\infty}^{+\infty} dx e^{-x^2/2} \int_{-\infty}^{+\infty} dy e^{-y^2/2} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx dy e^{-(x^2+y^2)/2}.$$

This is an integral over the $x-y$ plane. Instead of dividing this plane into infinitesimal rectangles $dx dy$, we may divide it into concentric rings of radius r and infinitesimal width dr . Since the area of such a ring is $2\pi r dr$, we have that

$$I^2 = 2\pi \int_0^{+\infty} dr r e^{-r^2/2}.$$

Now there is only one integration to be done. Next we make use of the fact that $\frac{dr^2}{dr} = 2r$, hence $dr r = d(r^2/2)$, and we introduce the variable $w = r^2/2$:

$$I^2 = 2\pi \int_0^{+\infty} d(r^2/2) e^{-r^2/2} = 2\pi \int_0^{+\infty} dw e^{-w}.$$

Since we know that the antiderivative of e^{-w} is $-e^{-w}$, we also know that

$$\int_0^{+\infty} dw e^{-w} = (-e^{-\infty}) - (-e^{-0}) = 0 + 1 = 1.$$

Therefore $I^2 = 2\pi$ and

$$\int_{-\infty}^{+\infty} dx e^{-x^2/2} = \sqrt{2\pi}.$$

Believe it or not, a significant fraction of the literature in theoretical physics concerns variations and elaborations of this basic *Gaussian* integral.

One variation is obtained by substituting $\sqrt{a}x$ for x :

$$\int_{-\infty}^{+\infty} dx e^{-ax^2/2} = \sqrt{2\pi/a}.$$

Another variation is obtained by thinking of both sides of this equation as functions of a and differentiating them with respect to a . The result is

$$\int_{-\infty}^{+\infty} dx e^{-ax^2/2} x^2 = \sqrt{2\pi/a^3}.$$

⁸ <http://en.wikibooks.org/wiki/Category%3A>

Sine and cosine

We define the function $\cos(x)$ by requiring that

$$\cos''(x) = -\cos(x), \quad \cos(0) = 1 \quad \text{and} \quad \cos'(0) = 0.$$

If you sketch the graph of this function using only this information, you will notice that wherever $\cos(x)$ is positive, its slope decreases as x increases (that is, its graph curves downward), and wherever $\cos(x)$ is negative, its slope increases as x increases (that is, its graph curves upward).

Differentiating the first defining equation repeatedly yields

$$\cos^{(n+2)}(x) = -\cos^{(n)}(x)$$

for all natural numbers n . Using the remaining defining equations, we find that $\cos^{(k)}(0)$ equals 1 for $k = 0, 4, 8, 12, \dots$, -1 for $k = 2, 6, 10, 14, \dots$, and 0 for odd k . This leads to the following Taylor series:

$$\cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots$$

The function $\sin(x)$ is similarly defined by requiring that

$$\sin''(x) = -\sin(x), \quad \sin(0) = 0, \quad \text{and} \quad \sin'(0) = 1.$$

This leads to the Taylor series

$$\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots$$

9

7.2.2 Complex numbers

The *natural numbers*¹⁰ are used for counting. By subtracting natural numbers from natural numbers, we can create *integers*¹¹ that are not natural numbers. By dividing integers by integers (other than zero) we can create *rational numbers*¹² that are not integers. By taking the square roots of positive rational numbers we can create *real numbers*¹³ that are

9 <http://en.wikibooks.org/wiki/Category%3A>

10 <http://en.wikipedia.org/wiki/Natural%20number>

11 <http://en.wikipedia.org/wiki/Integers>

12 <http://en.wikipedia.org/wiki/Rational%20number>

13 <http://en.wikipedia.org/wiki/Real%20number>

irrational¹⁴. And by taking the square roots of negative numbers we can create *complex numbers*¹⁵ that are imaginary¹⁶.

Any imaginary number is a real number multiplied by the positive square root of -1 , for which we have the symbol $i = +\sqrt{-1}$.

Every complex number z is the sum of a real number a (the *real part*¹⁷ of z) and an imaginary number ib . Somewhat confusingly, the *imaginary part*¹⁸ of z is the *real* number b .

Because real numbers can be visualized as points on a line, they are also referred to as (or thought of as constituting) the *real line*¹⁹. Because complex numbers can be visualized as points in a plane, they are also referred to as (or thought of as constituting) the *complex plane*²⁰. This plane contains two axes, one horizontal (the *real axis* constituted by the real numbers) and one vertical (the *imaginary axis* constituted by the imaginary numbers).

Do not be misled by the whimsical tags "real" and "imaginary". No number is real in the sense in which, say, apples are real. The real numbers are no less imaginary in the ordinary sense than the imaginary numbers, and the imaginary numbers are no less real in the mathematical sense than the real numbers. If you are not yet familiar with complex numbers, it is because you don't need them for counting or measuring. You need them for calculating the probabilities of measurement outcomes.

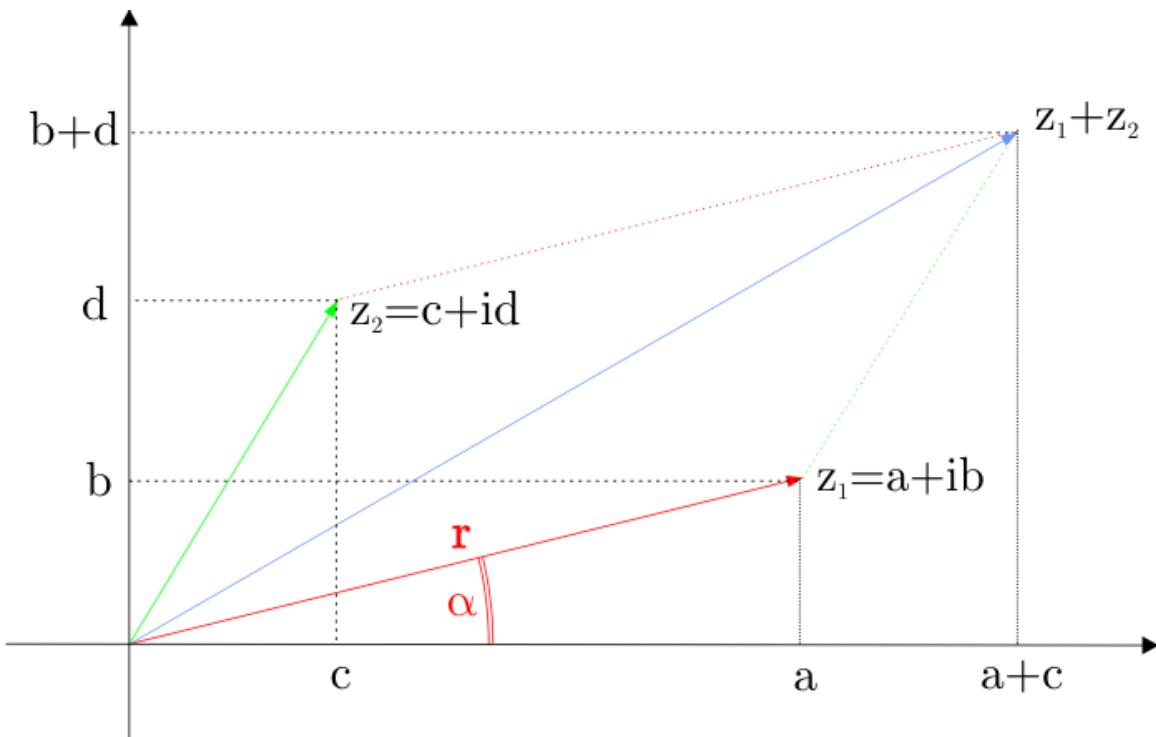


Figure 73

14 <http://en.wikipedia.org/wiki/Irrational%20number>

15 <http://en.wikipedia.org/wiki/Complex%20number>

16 <http://en.wikipedia.org/wiki/Imaginary%20number>

17 <http://en.wikipedia.org/wiki/Real%20part>

18 <http://en.wikipedia.org/wiki/Imaginary%20part>

19 <http://en.wikipedia.org/wiki/Real%20line>

20 <http://en.wikipedia.org/wiki/Complex%20plane>

This diagram illustrates, among other things, the addition of complex numbers:

$$z_1 + z_2 = (a_1 + ib_1) + (a_2 + ib_2) = (a_1 + a_2) + i(b_1 + b_2).$$

As you can see, adding two complex numbers is done in the same way as adding two vectors²¹ (a, b) and (c, d) in a plane.

Instead of using rectangular coordinates specifying the real and imaginary parts of a complex number, we may use polar coordinates specifying the *absolute value* or *modulus* $r = |z|$ and the *complex argument* or *phase*²² α , which is an angle measured in radians²³. Here is how these coordinates are related:

$$a = r \cos \alpha, \quad b = r \sin \alpha, \quad r = \sqrt{a^2 + b^2},$$

(Remember Pythagoras²⁴?)

$$\alpha = \begin{cases} \arctan\left(\frac{b}{a}\right) & \text{if } a > 0 \\ \arctan\left(\frac{b}{a}\right) + \pi & \text{if } a < 0 \text{ and } b \geq 0 \\ \arctan\left(\frac{b}{a}\right) - \pi & \text{if } a < 0 \text{ and } b < 0 \\ +\frac{\pi}{2} & \text{if } a = 0 \text{ and } b > 0 \\ -\frac{\pi}{2} & \text{if } a = 0 \text{ and } b < 0 \end{cases} \quad \text{or} \quad \alpha = \begin{cases} +\arccos\left(\frac{a}{r}\right) & \text{if } b \geq 0 \\ -\arccos\left(\frac{a}{r}\right) & \text{if } b < 0 \end{cases}$$

All you need to know to be able to multiply complex numbers is that $i^2 = -1$:

$$z_1 z_2 = (a_1 + ib_1)(a_2 + ib_2) = (a_1 a_2 - b_1 b_2) + i(a_1 b_2 + b_1 a_2).$$

There is, however, an easier way to multiply complex numbers. Plugging the power series²⁵ (or Taylor series²⁶) for cos and sin,

$$\cos x = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} x^{2k} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots$$

$$\sin x = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} x^{2k+1} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots,$$

into the expression $\cos \alpha + i \sin \alpha$ and rearranging terms, we obtain

21 <http://en.wikipedia.org/wiki/Vector%20%28spatial%29#Vector%20addition%20and%20subtraction>

22 <http://mathworld.wolfram.com/ComplexNumber.html>

23 <http://en.wikipedia.org/wiki/Radian>

24 <http://en.wikipedia.org/wiki/Pythagorean%20theorem>

25 <http://en.wikipedia.org/wiki/Power%20series>

26 <http://en.wikipedia.org/wiki/Taylor%20series>

$$\sum_{k=0}^{\infty} \frac{(ix)^k}{k!} = 1 + ix + \frac{(ix)^2}{2!} + \frac{(ix)^3}{3!} + \frac{(ix)^4}{4!} + \frac{(ix)^5}{5!} + \frac{(ix)^6}{6!} + \frac{(ix)^7}{7!} + \dots$$

But this is the power/Taylor series for the exponential function²⁷ e^y with $y = ix$! Hence Euler's formula²⁸

$$e^{i\alpha} = \cos \alpha + i \sin \alpha,$$

and this reduces multiplying two complex numbers to multiplying their absolute values and adding their phases:

$$(z_1)(z_2) = r_1 e^{i\alpha_1} r_2 e^{i\alpha_2} = (r_1 r_2) e^{i(\alpha_1 + \alpha_2)}.$$

An extremely useful definition is the complex conjugate²⁹ $z^* = a - ib$ of $z = a + ib$. Among other things, it allows us to calculate the *absolute square* $|z|^2$ by calculating the product

$$zz^* = (a + ib)(a - ib) = a^2 + b^2.$$

1. Show that

$$\cos x = \frac{e^{ix} + e^{-ix}}{2} \quad \text{and} \quad \sin x = \frac{e^{ix} - e^{-ix}}{2i}.$$

2. Arguably the five most important numbers are $0, 1, i, \pi, e$. Write down an equation containing each of these numbers just once. (Answer?)³⁰

31

7.2.3 Vectors (spatial)

A **vector**³² is a quantity that has both a magnitude and a direction. Vectors can be visualized as arrows. The following figure shows what we mean by the *components* (a_x, a_y, a_z) of a vector \mathbf{a} .

²⁷ <http://en.wikipedia.org/wiki/Exponential%20function>

²⁸ <http://en.wikipedia.org/wiki/Euler%27s%20formula>

²⁹ <http://en.wikipedia.org/wiki/Complex%20conjugate>

³⁰ <http://en.wikipedia.org/wiki/Euler%27s%20identity>

³¹ <http://en.wikibooks.org/wiki/Category%3A>

³² <http://en.wikipedia.org/wiki/Vector%20%28spatial%29>

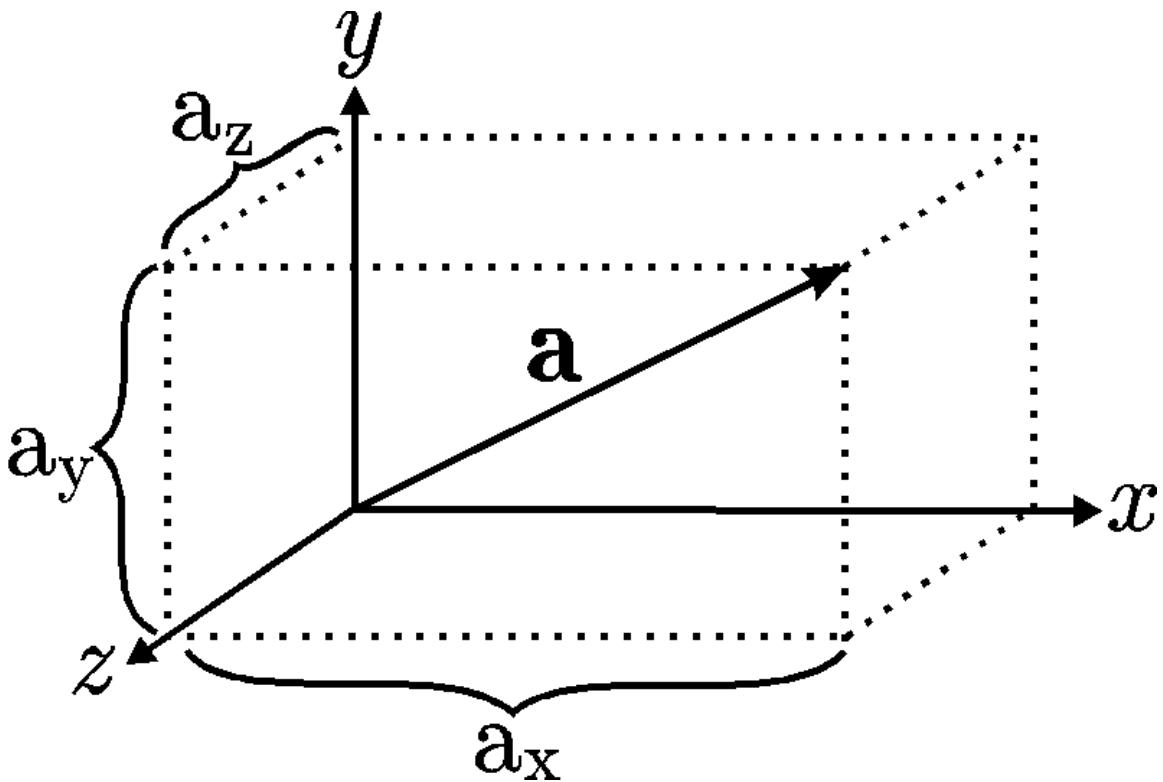


Figure 74

The *sum* $\mathbf{a} + \mathbf{b}$ of two vectors has the components $(a_x + b_x, a_y + b_y, a_z + b_z)$.

- Explain the addition of vectors in terms of arrows.

The *dot product*³³ of two vectors is the number

$$\mathbf{a} \cdot \mathbf{b} = a_x b_x + a_y b_y + a_z b_z.$$

Its importance arises from the fact that it is invariant under rotations³⁴. To see this, we calculate

$$(\mathbf{a} + \mathbf{b}) \cdot (\mathbf{a} + \mathbf{b}) = (a_x + b_x)^2 + (a_y + b_y)^2 + (a_z + b_z)^2 =$$

$$a_x^2 + a_y^2 + a_z^2 + b_x^2 + b_y^2 + b_z^2 + 2(a_x b_x + a_y b_y + a_z b_z) = \mathbf{a} \cdot \mathbf{a} + \mathbf{b} \cdot \mathbf{b} + 2\mathbf{a} \cdot \mathbf{b}.$$

According to Pythagoras³⁵, the magnitude of \mathbf{a} is $a = \sqrt{a_x^2 + a_y^2 + a_z^2}$. If we use a different coordinate system, the components of \mathbf{a} will be different: $(a_x, a_y, a_z) \rightarrow (a'_x, a'_y, a'_z)$. But

33 http://en.wikipedia.org/wiki/Dot_product

34 <http://en.wikipedia.org/wiki/Rotation%20%28mathematics%29>

35 <http://en.wikipedia.org/wiki/Pythagorean%20theorem>

if the new system of axes differs only by a rotation and/or translation³⁶ of the axes, the magnitude of \mathbf{a} will remain the same:

$$\sqrt{a_x^2 + a_y^2 + a_z^2} = \sqrt{(a'_x)^2 + (a'_y)^2 + (a'_z)^2}.$$

The squared magnitudes $\mathbf{a} \cdot \mathbf{a}$, $\mathbf{b} \cdot \mathbf{b}$, and $(\mathbf{a} + \mathbf{b}) \cdot (\mathbf{a} + \mathbf{b})$ are invariant under rotations, and so, therefore, is the product $\mathbf{a} \cdot \mathbf{b}$.

- Show that the dot product is also invariant under translations.

Since by a *scalar* we mean a number that is invariant under certain transformations (in this case rotations and/or translations of the coordinate axes), the dot product is also known as (a) *scalar product*. Let us prove that

$$\mathbf{a} \cdot \mathbf{b} = ab \cos \theta,$$

where θ is the angle between \mathbf{a} and \mathbf{b} . To do so, we pick a coordinate system \mathcal{F} in which $\mathbf{a} = (a, 0, 0)$. In this coordinate system $\mathbf{a} \cdot \mathbf{b} = ab_x$ with $b_x = b \cos \theta$. Since $\mathbf{a} \cdot \mathbf{b}$ is a scalar, and since scalars are invariant under rotations and translations, the result $\mathbf{a} \cdot \mathbf{b} = ab \cos \theta$ (which makes no reference to any particular frame) holds in all frames that are rotated and/or translated relative to \mathcal{F} .

We now introduce the *unit vectors* $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$, whose directions are defined by the coordinate axes. They are said to form an *orthonormal basis*. *Ortho* because they are mutually orthogonal:

$$\hat{\mathbf{x}} \cdot \hat{\mathbf{y}} = \hat{\mathbf{x}} \cdot \hat{\mathbf{z}} = \hat{\mathbf{y}} \cdot \hat{\mathbf{z}} = 0.$$

Normal because they are unit vectors:

$$\hat{\mathbf{x}} \cdot \hat{\mathbf{x}} = \hat{\mathbf{y}} \cdot \hat{\mathbf{y}} = \hat{\mathbf{z}} \cdot \hat{\mathbf{z}} = 1.$$

And *basis* because every vector \mathbf{v} can be written as a *linear combination*³⁷ of these three vectors — that is, a sum in which each basis vector appears once, multiplied by the corresponding component of \mathbf{v} (which may be 0):

$$\mathbf{v} = v_x \hat{\mathbf{x}} + v_y \hat{\mathbf{y}} + v_z \hat{\mathbf{z}}.$$

It is readily seen that $v_x = \hat{\mathbf{x}} \cdot \mathbf{v}$, $v_y = \hat{\mathbf{y}} \cdot \mathbf{v}$, $v_z = \hat{\mathbf{z}} \cdot \mathbf{v}$, which is why we have that

$$\mathbf{v} = \hat{\mathbf{x}}(\hat{\mathbf{x}} \cdot \mathbf{v}) + \hat{\mathbf{y}}(\hat{\mathbf{y}} \cdot \mathbf{v}) + \hat{\mathbf{z}}(\hat{\mathbf{z}} \cdot \mathbf{v}).$$

Another definition that is useful (albeit only in a 3-dimensional space) is the *cross product*³⁸ of two vectors:

36 <http://en.wikipedia.org/wiki/Translation%20%28geometry%29>

37 <http://en.wikipedia.org/wiki/Linear%20combination>

38 <http://en.wikipedia.org/wiki/Cross%20product>

$$\mathbf{a} \times \mathbf{b} = (a_y b_z - a_z b_y) \hat{\mathbf{x}} + (a_z b_x - a_x b_z) \hat{\mathbf{y}} + (a_x b_y - a_y b_x) \hat{\mathbf{z}}.$$

- Show that the cross product is antisymmetric: $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$.

As a consequence, $\mathbf{a} \times \mathbf{a} = 0$.

- Show that $\mathbf{a} \cdot (\mathbf{a} \times \mathbf{b}) = \mathbf{b} \cdot (\mathbf{a} \times \mathbf{b}) = 0$.

Thus $\mathbf{a} \times \mathbf{b}$ is perpendicular to both \mathbf{a} and \mathbf{b} .

- Show that the magnitude of $\mathbf{a} \times \mathbf{b}$ equals $a b \sin \alpha$, where α is the angle between \mathbf{a} and \mathbf{b} . Hint: use a coordinate system in which $\mathbf{a} = (a, 0, 0)$ and $\mathbf{b} = (b \cos \alpha, b \sin \alpha, 0)$.

Since $a b \sin \alpha$ is also the area A of the parallelogram P spanned by \mathbf{a} and \mathbf{b} , we can think of $\mathbf{a} \times \mathbf{b}$ as a vector of magnitude A perpendicular to P . Since the cross product yields a vector, it is also known as *vector product*.

(We save ourselves the trouble of showing that the cross product is invariant under translations and rotations of the coordinate axes, as is required of a vector. Let us however note in passing that if \mathbf{a} and \mathbf{b} are polar vectors, then $\mathbf{a} \times \mathbf{b}$ is an axial vector. Under a reflection (for instance, the inversion of a coordinate axis) an ordinary (or *polar*) vector is invariant, whereas an *axial vector*³⁹ changes its sign.)

Here is a useful relation involving both scalar and vector products:

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{c} \cdot \mathbf{a}) - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}.$$

40

7.2.4 Fields

As you will remember, a function is a machine that accepts a number and returns a number. A *field* is a function that accepts the three coordinates of a point or the four coordinates of a spacetime point and returns a scalar, a vector, or a tensor (either of the spatial variety or of the 4-dimensional spacetime variety).

Gradient

Imagine a curve \mathcal{C} in 3-dimensional space. If we label the points of this curve by some parameter λ , then \mathcal{C} can be represented by a 3-vector function $\mathbf{r}(\lambda)$. We are interested in how much the value of a scalar field $f(x, y, z)$ changes as we go from a point $\mathbf{r}(\lambda)$ of \mathcal{C} to the point $\mathbf{r}(\lambda + d\lambda)$ of \mathcal{C} . By how much f changes will depend on how much the coordinates (x, y, z) of \mathbf{r} change, which are themselves functions of λ . The changes in the coordinates are evidently given by

39 <http://en.wikipedia.org/wiki/Pseudovector>

40 <http://en.wikibooks.org/wiki/Category%3A>

$$(*) \quad dx = \frac{dx}{d\lambda} d\lambda, \quad dy = \frac{dy}{d\lambda} d\lambda, \quad dz = \frac{dz}{d\lambda} d\lambda,$$

while the change in f is a compound of three changes, one due to the change in x , one due to the change in y , and one due to the change in z :

$$(**) \quad df = \frac{df}{dx} dx + \frac{df}{dy} dy + \frac{df}{dz} dz.$$

The first term tells us by how much f changes as we go from (x, y, z) to $(x+dx, y, z)$, the second tells us by how much f changes as we go from (x, y, z) to $(x, y+dy, z)$, and the third tells us by how much f changes as we go from (x, y, z) to $(x, y, z+dz)$.

Shouldn't we add the changes in f that occur as we go first from (x, y, z) to $(x+dx, y, z)$, then from $(x+dx, y, z)$ to $(x+dx, y+dy, z)$, and then from $(x+dx, y+dy, z)$ to $(x+dx, y+dy, z+dz)$? Let's calculate.

$$\frac{\partial f(x+dx, y, z)}{\partial y} = \frac{\partial \left[f(x, y, z) + \frac{\partial f}{\partial x} dx \right]}{\partial y} = \frac{\partial f(x, y, z)}{\partial y} + \frac{\partial^2 f}{\partial y \partial x} dx.$$

If we take the limit $dx \rightarrow 0$ (as we mean to whenever we use dx), the last term vanishes. Hence we may as well use $\frac{\partial f(x, y, z)}{\partial y}$ in place of $\frac{\partial f(x+dx, y, z)}{\partial y}$. Plugging $(*)$ into $(**)$, we obtain

$$df = \left(\frac{\partial f}{\partial x} \frac{dx}{d\lambda} + \frac{\partial f}{\partial y} \frac{dy}{d\lambda} + \frac{\partial f}{\partial z} \frac{dz}{d\lambda} \right) d\lambda.$$

Think of the expression in brackets as the dot product of two vectors:

- the *gradient* $\frac{\partial f}{\partial \mathbf{r}}$ of the scalar field f , which is a vector field with components $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$,
- the vector $\frac{d\mathbf{r}}{d\lambda}$, which is tangent on \mathcal{C} .

If we think of λ as the time at which an object moving along \mathcal{C} is at $\mathbf{r}(\lambda)$, then the magnitude of $\frac{d\mathbf{r}}{d\lambda}$ is this object's speed.

$\frac{\partial}{\partial \mathbf{r}}$ is a differential operator that accepts a function $f(\mathbf{r})$ and returns its gradient $\frac{\partial f}{\partial \mathbf{r}}$.

The gradient of f is another input-output device: pop in $d\mathbf{r}$, and get the difference

$$\frac{\partial f}{\partial \mathbf{r}} \cdot d\mathbf{r} = df = f(\mathbf{r} + d\mathbf{r}) - f(\mathbf{r}).$$

The differential operator $\frac{\partial}{\partial \mathbf{r}}$ is also used in conjunction with the dot and cross products.

Curl

The *curl* of a vector field \mathbf{A} is defined by

$$\operatorname{curl} \mathbf{A} = \frac{\partial}{\partial \mathbf{r}} \times \mathbf{A} = \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \right) \hat{\mathbf{x}} + \left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \right) \hat{\mathbf{y}} + \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \right) \hat{\mathbf{z}}.$$

To see what this definition is good for, let us calculate the integral $\oint \mathbf{A} \cdot d\mathbf{r}$ over a closed curve \mathcal{C} . (An integral over a curve is called a *line integral*, and if the curve is closed it is called a *loop integral*.) This integral is called the *circulation* of \mathbf{A} along \mathcal{C} (or around the surface enclosed by \mathcal{C}). Let's start with the boundary of an infinitesimal rectangle with corners $A = (0, 0, 0)$, $B = (0, dy, 0)$, $C = (0, dy, dz)$, and $D = (0, 0, dz)$.

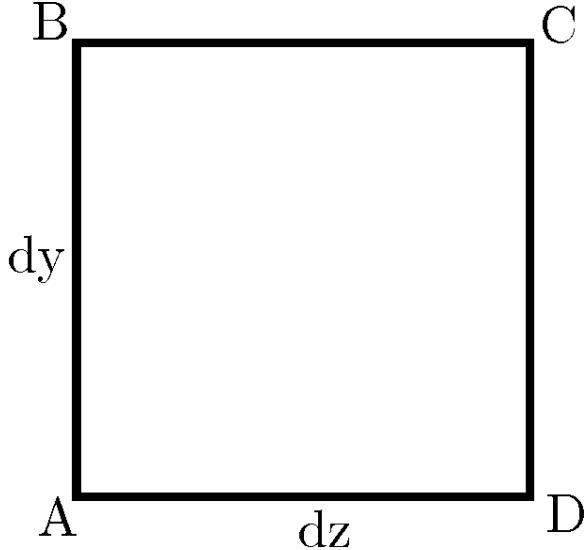


Figure 75

The contributions from the four sides are, respectively,

- \overline{AB} : $A_y(0, dy/2, 0) dy$,
- \overline{BC} : $A_z(0, dy, dz/2) dz = \left[A_z(0, 0, dz/2) + \frac{\partial A_z}{\partial y} dy \right] dz$,
- \overline{CD} : $-A_y(0, dy/2, dz) dy = - \left[A_y(0, dy/2, 0) + \frac{\partial A_y}{\partial z} dz \right] dy$,
- \overline{DA} : $-A_z(0, 0, dz/2) dz$.

These add up to

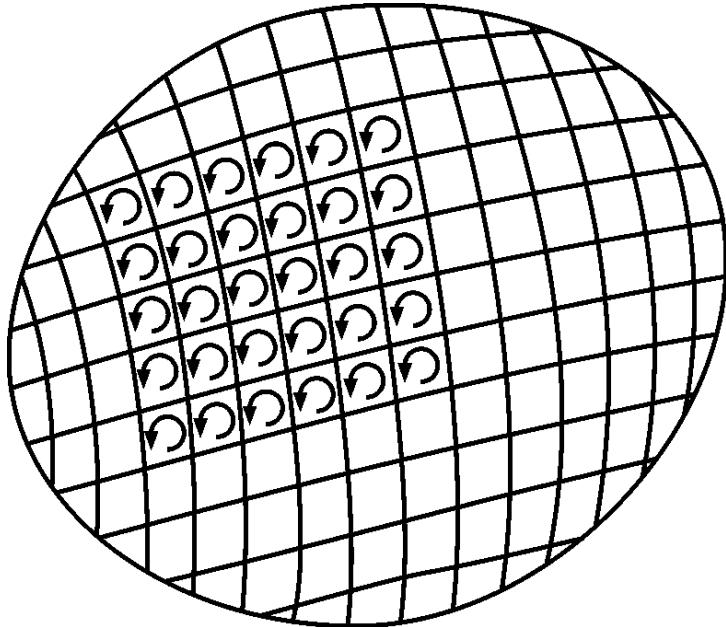
$$(\text{***}) \quad \left[\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \right] dy dz = (\operatorname{curl} \mathbf{A})_x dy dz.$$

**Figure 76**

Let us represent this infinitesimal rectangle of area $dy dz$ (lying in the y - z plane) by a vector $d\mathbf{\Sigma}$ whose magnitude equals $d\mathbf{\Sigma} = dy dz$, and which is perpendicular to the rectangle. (There are two possible directions. The right-hand rule illustrated on the right indicates how the direction of $d\mathbf{\Sigma}$ is related to the direction of circulation.) This allows us to write (****) as a scalar (product) $\text{curl } \mathbf{A} \cdot d\mathbf{\Sigma}$. Being a scalar, it is invariant under rotations either of the coordinate axes or of the infinitesimal rectangle. Hence if we cover a surface Σ with infinitesimal rectangles and add up their circulations, we get $\int_{\Sigma} \text{curl } \mathbf{A} \cdot d\mathbf{\Sigma}$.

Observe that the common sides of all neighboring rectangles are integrated over twice in opposite directions. Their contributions cancel out and only the contributions from the boundary $\partial\Sigma$ of Σ survive.

The bottom line: $\oint_{\partial\Sigma} \mathbf{A} \cdot d\mathbf{r} = \int_{\Sigma} \text{curl } \mathbf{A} \cdot d\mathbf{\Sigma}$.

**Figure 77**

This is *Stokes' theorem*. Note that the left-hand side depends solely on the boundary $\partial\Sigma$ of Σ . So, therefore, does the right-hand side. The value of the surface integral of the curl of a vector field depends solely on the values of the vector field at the boundary of the surface integrated over.

If the vector field \mathbf{A} is the gradient of a scalar field f , and if \mathcal{C} is a curve from \mathbf{A} to \mathbf{b} , then

$$\int_{\mathcal{C}} \mathbf{A} \cdot d\mathbf{r} = \int_{\mathcal{C}} df = f(\mathbf{b}) - f(\mathbf{A}).$$

The line integral of a gradient thus is the same for all curves having identical end points. If $\mathbf{b} = \mathbf{A}$ then \mathcal{C} is a loop and $\int_{\mathcal{C}} \mathbf{A} \cdot d\mathbf{r}$ vanishes. By Stokes' theorem it follows that the curl of a gradient vanishes identically:

$$\int_{\Sigma} \left(\operatorname{curl} \frac{\partial f}{\partial \mathbf{r}} \right) \cdot d\Sigma = \oint_{\partial\Sigma} \frac{\partial f}{\partial \mathbf{r}} \cdot d\mathbf{r} = 0.$$

Divergence

The *divergence* of a vector field \mathbf{A} is defined by

$$\operatorname{div} \mathbf{A} = \frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z}.$$

To see what *this* definition is good for, consider an infinitesimal volume element d^3r with sides dx, dy, dz . Let us calculate the net (outward) flux of a vector field \mathbf{A} through the

surface of d^3r . There are three pairs of opposite sides. The net flux through the surfaces perpendicular to the x axis is

$$A_x(x+dx, y, z) dy dz - A_x(x, y, z) dy dz = \frac{\partial A_x}{\partial x} dx dy dz.$$

It is obvious what the net flux through the remaining surfaces will be. The net flux of \mathbf{A} out of d^3r thus equals

$$\left[\frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \right] dx dy dz = \operatorname{div} \mathbf{A} d^3r.$$

If we fill up a region R with infinitesimal parallelepipeds and add up their net outward fluxes, we get $\int_R \operatorname{div} \mathbf{A} d^3r$. Observe that the common sides of all neighboring parallelepipeds are integrated over twice with opposite signs — the flux out of one equals the flux into the other. Hence their contributions cancel out and only the contributions from the surface ∂R of R survive. The bottom line:

$$\int_{\partial R} \mathbf{A} \cdot d\mathbf{\Sigma} = \int_R \operatorname{div} \mathbf{A} d^3r.$$

This is *Gauss' law*. Note that the left-hand side depends solely on the boundary ∂R of R . So, therefore, does the right-hand side. The value of the volume integral of the divergence of a vector field depends solely on the values of the vector field at the boundary of the region integrated over.

If Σ is a closed surface — and thus the boundary ∂R or a region of space R — then Σ itself has no boundary (symbolically, $\partial\Sigma = 0$). Combining Stokes' theorem with Gauss' law we have that

$$\oint_{\partial\partial R} \mathbf{A} \cdot d\mathbf{r} = \int_{\partial R} \operatorname{curl} \mathbf{A} \cdot d\mathbf{\Sigma} = \int_R \operatorname{div} \operatorname{curl} \mathbf{A} d^3r.$$

The left-hand side is an integral over the boundary of a boundary. But a boundary has no boundary! The boundary of a boundary is zero: $\partial\partial = 0$. It follows, in particular, that the right-hand side is zero. Thus not only the curl of a gradient but also the divergence of a curl vanishes identically:

$$\frac{\partial}{\partial \mathbf{r}} \times \frac{\partial f}{\partial \mathbf{r}} = 0, \quad \frac{\partial}{\partial \mathbf{r}} \cdot \frac{\partial}{\partial \mathbf{r}} \times \mathbf{A} = 0.$$

Some useful identities

$$d\mathbf{r} \times \left(\frac{\partial}{\partial \mathbf{r}} \times \mathbf{A} \right) \frac{\partial}{\partial \mathbf{r}} (\mathbf{A} \cdot d\mathbf{r}) - \left(d\mathbf{r} \cdot \frac{\partial}{\partial \mathbf{r}} \right) \mathbf{A}$$

$$\frac{\partial}{\partial \mathbf{r}} \times \left(\frac{\partial}{\partial \mathbf{r}} \times \mathbf{A} \right) = \frac{\partial}{\partial \mathbf{r}} \left(\frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{A} \right) - \left(\frac{\partial}{\partial \mathbf{r}} \cdot \frac{\partial}{\partial \mathbf{r}} \right) \mathbf{A}.$$

7.3 The ABCs of relativity

See also the Wikibook Special relativity⁴² that contains an in-depth text on this subject.

7.3.1 The principle of relativity

If we use an *inertial system*⁴³ (a.k.a. inertial coordinate system, inertial frame of reference, or inertial reference frame), then the components x, y, z of the position of any freely moving classical object ("point mass") change by equal amounts $\Delta x, \Delta y, \Delta z$ in equal time intervals Δt . Evidently, if \mathcal{F}_1 is an inertial frame then so is a reference frame \mathcal{F}_2 that is, relative to \mathcal{F}_1 ,

1. shifted ("translated") in space by any distance and/or in any direction,
2. translated in time by any interval,
3. rotated by any angle about any axis, and/or
4. moving with any constant velocity.

The *principle of relativity* states that all inertial systems are "created equal": the laws of physics are the same as long as they are formulated with respect to an inertial frame — no matter which. (Describing the same physical event or state of affairs using different inertial systems is like saying the same thing in different languages.) The first three items tell us that one inertial frame is as good as any other frame as long as the other frame differs by a shift of the coordinate origin in space and/or time and/or by a rotation of the spatial coordinate axes. What matters in physics are *relative* positions (the positions of objects relative to each other), *relative* times (the times of events relative to each other), and *relative* orientations (the orientations of objects relative to each other), inasmuch as these are unaffected by translations in space and/or time and by rotations of the spatial axes. In the physical world, there are no absolute positions, absolute times, or absolute orientations.

The fourth item tells us, in addition, that one inertial frame is as good as any other frame as long as the two frames move with a constant velocity relative to each other. What matters are relative velocities (the velocities of objects relative to each other), inasmuch as these are unaffected by a coordinate *boost* — the switch from an inertial frame \mathcal{F} to a frame moving with a constant velocity relative to \mathcal{F} . In the physical world, there are no absolute velocities and, in particular, *there is no absolute rest*.

It stands to reason. For one thing, positions are properties of objects, not things that exist even when they are not "occupied" or possessed. For another, the positions of objects are defined relative to the positions of other objects. In a universe containing a single object, there is no position that one could attribute to that object. By the same token, all physically meaningful times are the times of physical events, and they too are relatively defined, as the times *between* events. In a universe containing a single event, there is not time that one

41 <http://en.wikibooks.org/wiki/Category%3A>

42 <http://en.wikibooks.org/wiki/Special%20relativity>

43 <http://en.wikipedia.org/wiki/Inertial%20frame%20of%20reference>

could attribute to that event. But if positions and times are relatively defined, then so are velocities.

That there is no such thing as absolute rest has not always been as obvious as it should have been. Two ideas were responsible for the erroneous notion that there is a special class of inertial frames defining "rest" in an absolute sense: the idea that electromagnetic effects are transmitted by waves, and the idea that these waves require a physical medium (dubbed "ether") for their propagation. If there were such a medium, one could define absolute rest as equivalent to being at rest with respect to it.

44

7.3.2 Lorentz transformations (general form)

We want to express the coordinates t and $\mathbf{r} = (x, y, z)$ of an inertial frame \mathcal{F}_1 in terms of the coordinates t' and $\mathbf{r}' = (x', y', z')$ of another inertial frame \mathcal{F}_2 . We will assume that the two frames meet the following conditions:

1. their spacetime coordinate origins coincide ($t'=0, \mathbf{r}'=0$ mark the same spacetime location as $t=0, \mathbf{r}=0$),
2. their space axes are parallel, and
3. \mathcal{F}_2 moves with a constant velocity \mathbf{w} relative to \mathcal{F}_1 .

What we know at this point is that whatever moves with a constant velocity in \mathcal{F}_1 will do so in \mathcal{F}_2 . It follows that the transformation $t, \mathbf{r} \rightarrow t', \mathbf{r}'$ maps straight lines in \mathcal{F}_1 onto straight lines in \mathcal{F}_2 . Coordinate lines of \mathcal{F}_1 , in particular, will be mapped onto straight lines in \mathcal{F}_2 . This tells us that the dashed coordinates are linear combinations of the undashed ones,

$$t' = At + \mathbf{B} \cdot \mathbf{r}, \quad \mathbf{r}' = C\mathbf{r} + (\mathbf{D} \cdot \mathbf{r})\mathbf{w} + t.$$

We also know that the transformation from \mathcal{F}_1 to \mathcal{F}_2 can only depend on \mathbf{w} , so $A, \mathbf{B}, C, \mathbf{D}$, and

are functions of

w. Our task is to find these functions. The real-valued functions A and C actually can depend only on $w = |\mathbf{w}| = \sqrt{\mathbf{w} \cdot \mathbf{w}}$, so $A = a(w)$ and $C = c(w)$. A vector function depending only on \mathbf{w} must be parallel (or antiparallel) to \mathbf{w} , and its magnitude must be a function of w . We can therefore write $\mathbf{B} = b(w)\mathbf{w}$, $\mathbf{D} = [d(w)/w^2]\mathbf{w}$, and $= e(w)\mathbf{w}$. (It will become clear in a moment why the factor w^{-2} is included in the definition of \mathbf{D} .) So,

$$t' = a(w)t + b(w)\mathbf{w} \cdot \mathbf{r}, \quad \mathbf{r}' = c(w)\mathbf{r} + d(w)\frac{\mathbf{w} \cdot \mathbf{r}}{w^2}\mathbf{w} + e(w)\mathbf{w}t.$$

Let's set \mathbf{r} equal to $\mathbf{w}t$. This implies that $\mathbf{r}' = (c + d + e)\mathbf{w}t$. As we are looking at the trajectory of an object at rest in \mathcal{F}_2 , \mathbf{r}' must be constant. Hence,

$$c + d + e = 0.$$

44 <http://en.wikibooks.org/wiki/Category%3A>

Let's write down the inverse transformation. Since \mathcal{F}_1 moves with velocity $-\mathbf{w}$ relative to \mathcal{F}_2 , it is

$$t = a(w)t' - b(w)\mathbf{w} \cdot \mathbf{r}', \quad \mathbf{r} = c(w)\mathbf{r}' + d(w)\frac{\mathbf{w} \cdot \mathbf{r}'}{w^2}\mathbf{w} - e(w)\mathbf{w}t'.$$

To make life easier for us, we now chose the space axes so that $\mathbf{w} = (w, 0, 0)$. Then the above two (mutually inverse) transformations simplify to

$$t' = at + bwx, \quad x' = cx + dx + ewt, \quad y' = cy, \quad z' = cz,$$

$$t = at' - bwx', \quad x = cx' + dx' - ewt', \quad y = cy', \quad z = cz'.$$

Plugging the first transformation into the second, we obtain

$$t = a(at + bwx) - bw(cx + dx + ewt) = (a^2 - bew^2)t + (abw - bcw - bdw)x,$$

$$x = c(cx + dx + ewt) + d(cx + dx + ewt) - ew(at + bwx)$$

$$= (c^2 + 2cd + d^2 - bew^2)x + (cew + dew - aew)t,$$

$$y = c^2y,$$

$$z = c^2z.$$

The first of these equations tells us that

$$a^2 - bew^2 = 1 \quad \text{and} \quad abw - bcw - bdw = 0.$$

The second tells us that

$$c^2 + 2cd + d^2 - bew^2 = 1 \quad \text{and} \quad cew + dew - aew = 0.$$

Combining $abw - bcw - bdw = 0$ with $c + d + e = 0$ (and taking into account that $w \neq 0$), we obtain $b(a + e) = 0$.

Using $c + d + e = 0$ to eliminate d , we obtain $e^2 - bew^2 = 1$ and $e(a + e) = 0$.

Since the first of the last two equations implies that $e \neq 0$, we gather from the second that $e = -a$.

$y = c^2y$ tells us that $c^2 = 1$. c must, in fact, be equal to 1, since we have assumed that the space axes of the two frames are parallel (rather than antiparallel).

With $c = 1$ and $e = -a$, $c + d + e = 0$ yields $d = a - 1$. Upon solving $e^2 - bew^2 = 1$ for b , we are left with expressions for b, c, d , and e depending solely on a :

$$b = \frac{1 - a^2}{aw^2}, \quad c = 1, \quad d = a - 1, \quad e = -a.$$

Quite an improvement!

To find the remaining function $a(w)$, we consider a third inertial frame \mathcal{F}_3 , which moves with velocity $\mathbf{v} = (v, 0, 0)$ relative to \mathcal{F}_2 . Combining the transformation from \mathcal{F}_1 to \mathcal{F}_2 ,

$$t' = a(w)t + \frac{1-a^2(w)}{a(w)w}x, \quad x' = a(w)x - a(w)wt,$$

with the transformation from \mathcal{F}_2 to \mathcal{F}_3 ,

$$t'' = a(v)t' + \frac{1-a^2(v)}{a(v)v}x', \quad x'' = a(v)x' - a(v)vt',$$

we obtain the transformation from \mathcal{F}_1 to \mathcal{F}_3 :

$$t'' = a(v) \left[a(w)t + \frac{1-a^2(w)}{a(w)w}x \right] + \frac{1-a^2(v)}{a(v)v} \left[a(w)x - a(w)wt \right]$$

$$= \underbrace{\left[a(v)a(w) - \frac{1-a^2(v)}{a(v)v}a(w)w \right]}_{\star} t + [\dots]x,$$

$$x'' = a(v) \left[a(w)x - a(w)wt \right] - a(v)v \left[a(w)t + \frac{1-a^2(w)}{a(w)w}x \right]$$

$$= \underbrace{\left[a(v)a(w) - a(v)v \frac{1-a^2(w)}{a(w)w} \right]}_{\star\star} x - [\dots]t.$$

The direct transformation from \mathcal{F}_1 to \mathcal{F}_3 must have the same form as the transformations from \mathcal{F}_1 to \mathcal{F}_2 and from \mathcal{F}_2 to \mathcal{F}_3 , namely

$$t'' = \underbrace{a(u)t}_{\star} + \frac{1-a^2(u)}{a(u)u}x, \quad x'' = \underbrace{a(u)x - a(u)ut}_{\star\star},$$

where u is the speed of \mathcal{F}_3 relative to \mathcal{F}_1 . Comparison of the coefficients marked with stars yields two expressions for $a(u)$, which of course must be equal:

$$a(v)a(w) - \frac{1-a^2(v)}{a(v)v}a(w)w = a(v)a(w) - a(v)v \frac{1-a^2(w)}{a(w)w}.$$

It follows that $[1-a^2(v)]a^2(w)w^2 = [1-a^2(w)]a^2(v)v^2$, and this tells us that

$$K = \frac{1-a^2(w)}{a^2(w)w^2} = \frac{1-a^2(v)}{a^2(v)v^2}$$

is a universal constant. Solving the first equality for $a(w)$, we obtain

$$a(w) = 1/\sqrt{1+Kw^2}.$$

This allows us to cast the transformation

$$t' = at + bwx, \quad x' = cx + dx + ewt, \quad y' = cy, \quad z' = cz,$$

into the form

$$t' = \frac{t + Kwx}{\sqrt{1 + Kw^2}}, \quad x' = \frac{x - wt}{\sqrt{1 + Kw^2}}, \quad y' = y, \quad z' = z.$$

Trumpets, please! We have managed to reduce five unknown functions to a single constant.

45

7.3.3 Composition of velocities

In fact, there are only three physically distinct possibilities. (If $K \neq 0$, the magnitude of K depends on the choice of units, and this tells us something about us rather than anything about the physical world.)

The possibility $K = 0$ yields the *Galilean transformations* of Newtonian ("non-relativistic") mechanics:

$$t' = t, \quad \mathbf{r}' = \mathbf{r} - \mathbf{w}t, \quad u = v + w, \quad ds = dt.$$

(The common practice of calling theories with this transformation law "non-relativistic" is inappropriate, inasmuch as they too satisfy the principle of relativity.) In the remainder of this section we assume that $K \neq 0$.

Suppose that object C moves with speed v relative to object B , and that this moves with speed w relative to object A . If B and C move in the same direction, what is the speed u of C relative to A ? In the previous section we found that

$$a(u) = a(v)a(w) - \frac{1 - a^2(v)}{a(v)v}a(w)w,$$

and that

$$K = \frac{1 - a^2(v)}{a^2(v)v^2}.$$

This allows us to write

$$a(u) = a(v)a(w) - \frac{1 - a^2(v)}{a^2(v)v^2}a(v)v a(w)w = a(v)a(w)(1 - Kv w).$$

Expressing a in terms of K and the respective velocities, we obtain

$$\frac{1}{\sqrt{1 + Ku^2}} = \frac{1 - Kv w}{\sqrt{1 + Kv^2}\sqrt{1 + Kw^2}},$$

45 <http://en.wikibooks.org/wiki/Category%3A>

which implies that

$$1 + Ku^2 = \frac{(1 + Kv^2)(1 + Kw^2)}{(1 - Kvw)^2}.$$

We massage this into

$$Ku^2 = \frac{(1 + Kv^2)(1 + Kw^2) - (1 - Kvw)^2}{(1 - Kvw)^2} = \frac{K(v + w)^2}{(1 - Kvw)^2},$$

divide by K , and end up with:

$$u = \frac{v + w}{1 - Kvw}.$$

Thus, unless $K = 0$, we don't get the speed of C relative to A by simply adding the speed of C relative to B to the speed of B relative to A .

7.3.4 Proper time

Consider an infinitesimal segment $d\mathcal{C}$ of a spacetime path \mathcal{C} . In \mathcal{F}_1 it has the components (dt, dx, dy, dz) , in \mathcal{F}_2 it has the components (dt', dx', dy', dz') . Using the Lorentz transformation in its general form,

$$t' = \frac{t + Kwx}{\sqrt{1 + Kw^2}}, \quad x' = \frac{x - wt}{\sqrt{1 + Kw^2}}, \quad y' = y, \quad z' = z,$$

it is readily shown that

$$(dt')^2 + K d\mathbf{r}' \cdot d\mathbf{r}' = dt^2 + K d\mathbf{r} \cdot d\mathbf{r}.$$

We conclude that the expression

$$ds^2 = dt^2 + K d\mathbf{r} \cdot d\mathbf{r} = dt^2 + K(dx^2 + dy^2 + dz^2)$$

is invariant under this transformation. It is also invariant under rotations of the spatial axes (why?) and translations of the spacetime coordinate origin. This makes ds a 4-scalar.

What is the physical significance of ds ?

A clock that travels along $d\mathcal{C}$ is at rest in any frame in which $d\mathcal{C}$ lacks spatial components. In such a frame, $ds^2 = dt^2$. Hence ds is the time it takes to travel along $d\mathcal{C}$ as measured by a clock that travels along $d\mathcal{C}$. ds is the *proper time* (or *proper duration*) of $d\mathcal{C}$. The proper time (or proper duration) of a finite spacetime path \mathcal{C} , accordingly, is

$$\int_{\mathcal{C}} ds = \int_{\mathcal{C}} \sqrt{dt^2 + K d\mathbf{r} \cdot d\mathbf{r}} = \int_{\mathcal{C}} dt \sqrt{1 + Kv^2}.$$

7.3.5 An invariant speed

If $K < 0$, then there is a universal constant $c \equiv 1/\sqrt{-K}$ with the dimension of a velocity, and we can cast $u = v + w/(1 - Kvw)$ into the form

$$u = \frac{v + w}{1 + vw/c^2}.$$

If we plug in $v = w = c/2$, then instead of the Galilean $u = v + w = c$, we have $u = \frac{4}{5}c < c$. More intriguingly, if object O moves with speed c relative to \mathcal{F}_2 , and if \mathcal{F}_2 moves with speed w relative to \mathcal{F}_1 , then O moves with the same speed c relative to \mathcal{F}_1 : $(w + c)/(1 + wc/c^2) = c$. The *speed of light* c thus is an *invariant speed*: whatever travels with it in one inertial frame, travels with the same speed in *every* inertial frame.

Starting from

$$ds^2 = (dt')^2 - d\mathbf{r}' \cdot d\mathbf{r}'/c^2 = dt^2 - d\mathbf{r} \cdot d\mathbf{r}/c^2,$$

we arrive at the same conclusion: if O travels with c relative to \mathcal{F}_1 , then it travels the distance $d\mathbf{r} = cdt$ in the time dt . Therefore $ds^2 = dt^2 - dr^2/c^2 = 0$. But then $(dt')^2 - (dr')^2/c^2 = 0$, and this implies $dr' = cdt'$. It follows that O travels with the same speed c relative to \mathcal{F}_2 .

An invariant speed also exists if $K = 0$, but in this case it is infinite: whatever travels with infinite speed in *one* inertial frame — it takes no time to get from one place to another — does so in *every* inertial frame.

The existence of an invariant speed prevents objects from making U-turns in spacetime. If $K = 0$, it obviously takes an infinite amount of energy to reach $v = \infty$. Since an infinite amount of energy isn't at our disposal, we cannot start vertically in a spacetime diagram and then make a U-turn (that is, we cannot reach, let alone "exceed", a horizontal slope. ("Exceeding" a horizontal slope here means changing from a positive to a negative slope, or from going forward to going backward in time.)

If $K < 0$, it takes an infinite amount of energy to reach even the finite speed of light. Imagine you spent a finite amount of fuel accelerating from 0 to $0.1c$. In the frame in which you are now at rest, your speed is not a whit closer to the speed of light. And this remains true no matter how many times you repeat the procedure. Thus no finite amount of energy can make you reach, let alone "exceed", a slope equal to $1/c$. ("Exceeding" a slope equal to $1/c$ means attaining a smaller slope. As we will see, if we were to travel faster than light in any one frame, then there would be frames in which we travel backward in time.)

46

7.3.6 The case against $K > 0$

In a hypothetical world with $K > 0$ we can define $k \equiv 1/\sqrt{K}$ (a universal constant with the dimension of a velocity), and we can cast $u = v + w/(1 - Kvw)$ into the form

46 <http://en.wikibooks.org/wiki/Category%3A>

$$u = \frac{v + w}{1 - vw/k^2}.$$

If we plug in $v = w = k/2$, then instead of the Galilean $u = v + w = k$ we have $u = \frac{4}{3}k > k$. Worse, if we plug in $v = w = k$, we obtain $u = \infty$: if object O travels with speed k relative to \mathcal{F}_2 , and if \mathcal{F}_2 travels with speed k relative to \mathcal{F}_1 (in the same direction), then O travels with an infinite speed relative to \mathcal{F}_1 ! And if O travels with $2k$ relative to \mathcal{F}_2 and \mathcal{F}_2 travels with $2k$ relative to \mathcal{F}_1 , O 's speed relative to \mathcal{F}_1 is negative: $u = -\frac{4}{3}k$.

If we use units in which $K = k = 1$, then the invariant proper time associated with an infinitesimal path segment is related to the segment's inertial components via

$$ds^2 = dt^2 + dx^2 + dy^2 + dz^2.$$

This is the 4-dimensional version of the 3-scalar $dx^2 + dy^2 + dz^2$, which is invariant under rotations in space. Hence if K is positive, the transformations between inertial systems are rotations in spacetime. I guess you now see why in this hypothetical world the composition of two positive speeds can be a negative speed.

Let us confirm this conclusion by deriving the composition theorem (for $k=1$) from the assumption that the x' and t' axes are rotated relative to the x and t axes.

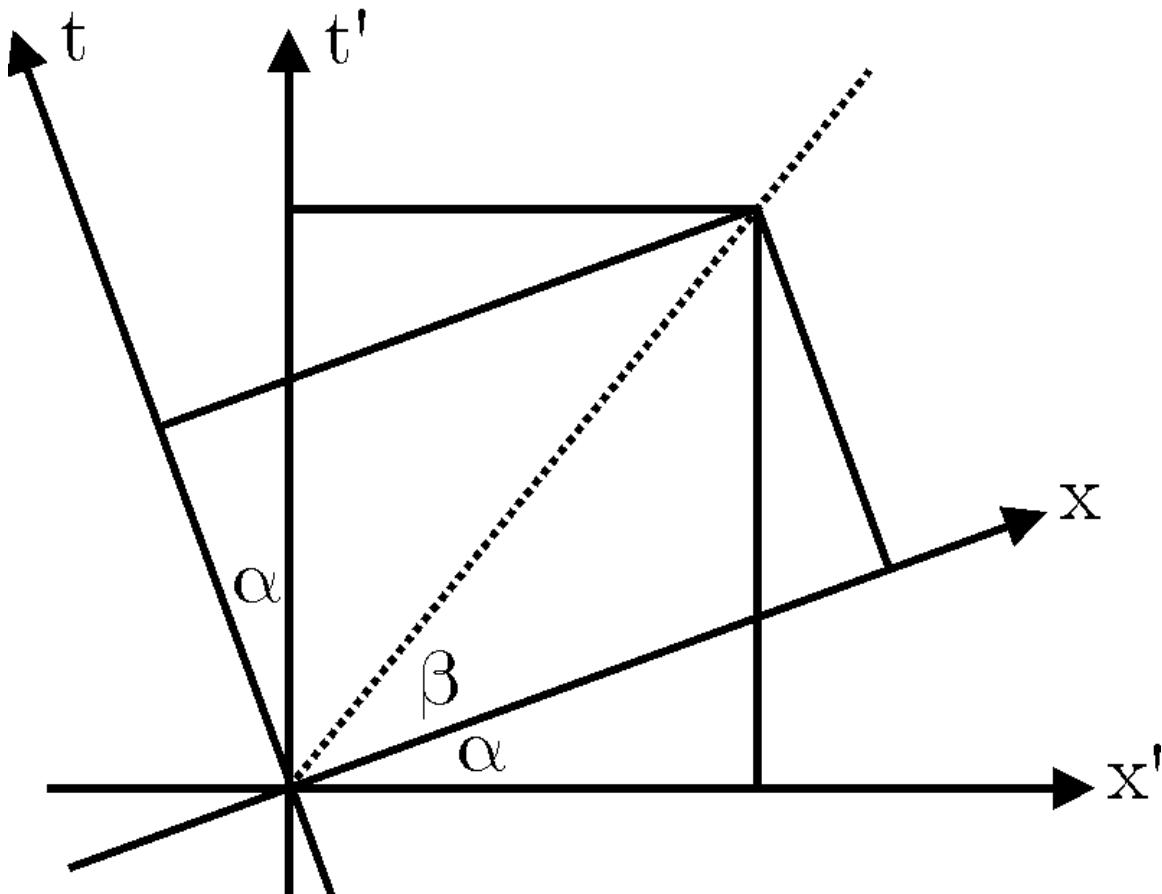


Figure 78

The speed of an object O following the dotted line is $w = \cot(\alpha + ta)$ relative to \mathcal{F}' , the speed of \mathcal{F}' relative to \mathcal{F} is $v = \tan \alpha$, and the speed of O relative to \mathcal{F} is $u = \cot \beta$. Invoking the trigonometric relation

$$\tan(\alpha + \beta) = \frac{\tan \alpha + \tan \beta}{1 - \tan \alpha \tan \beta},$$

we conclude that $\frac{1}{w} = \frac{v+1/u}{1-v/u}$. Solving for u , we obtain $u = \frac{v+w}{1-vw}$.

How can we rule out the *a priori* possibility that $K > 0$? As shown in the body of the book, the stability of matter — to be precise, the existence of stable objects that (i) have spatial extent (they "occupy" space) and (ii) are composed of a finite number of objects that lack spatial extent (they don't "occupy" space) — rests on the existence of relative positions that are (a) more or less fuzzy and (b) independent of time. Such relative positions are described by probability distributions that are (a) *inhomogeneous* in space and (b) *homogeneous* in time. Their objective existence thus requires an objective difference between spacetime's temporal dimension and its spatial dimensions. This rules out the possibility that $K > 0$.

How? If $K < 0$, and if we use natural units, in which $c = 1$, we have that

$$ds^2 = +dt^2 - dx^2 - dy^2 - dz^2.$$

As far as physics is concerned, the difference between the positive sign in front of dt and the negative signs in front of dx , dy , and dz is the *only* objective difference between time and the spatial dimensions of spacetime. If K were positive, not even this difference would exist.

7.3.7 The case against zero K

And what argues against the possibility that $K = 0$?

Recall the propagator for a free and stable particle:

$$\langle B|A \rangle = \int \mathcal{D}\mathcal{C} e^{-i b s[\mathcal{C}]}.$$

If K were to vanish, we would have $ds^2 = dt^2$. There would be no difference between inertial time and proper time, and every spacetime path leading from A to B would contribute the same amplitude $e^{-ib(t_B-t_A)}$ to the propagator $\langle B|A \rangle$, which would be hopelessly divergent as a result. Worse, $\langle B|A \rangle$ would be independent of the distance between A and B . To obtain well-defined, finite probabilities, cancellations ("destructive interference") must occur, and this rules out that $K = 0$.

7.3.8 The actual Lorentz transformations

In the real world, therefore, the Lorentz transformations take the form

$$t' = \frac{t - wx/c^2}{\sqrt{1 - w^2/c^2}}, \quad x' = \frac{x - wt}{\sqrt{1 - w^2/c^2}}, \quad y' = y, \quad z' = z.$$

Let's explore them diagrammatically, using natural units ($c = 1$). Setting $t' = 0$, we have $t = wx$. This tells us that the slope of the x' axis relative to the undashed frame is $w = \tan \alpha$. Setting $x' = 0$, we have $t = x/w$. This tells us that the slope of the t' axis is $1/w$. The dashed axes are thus rotated by the same angle in *opposite* directions; if the t' axis is rotated clockwise relative to the t axis, then the x' axis is rotated counterclockwise relative to the x axis.

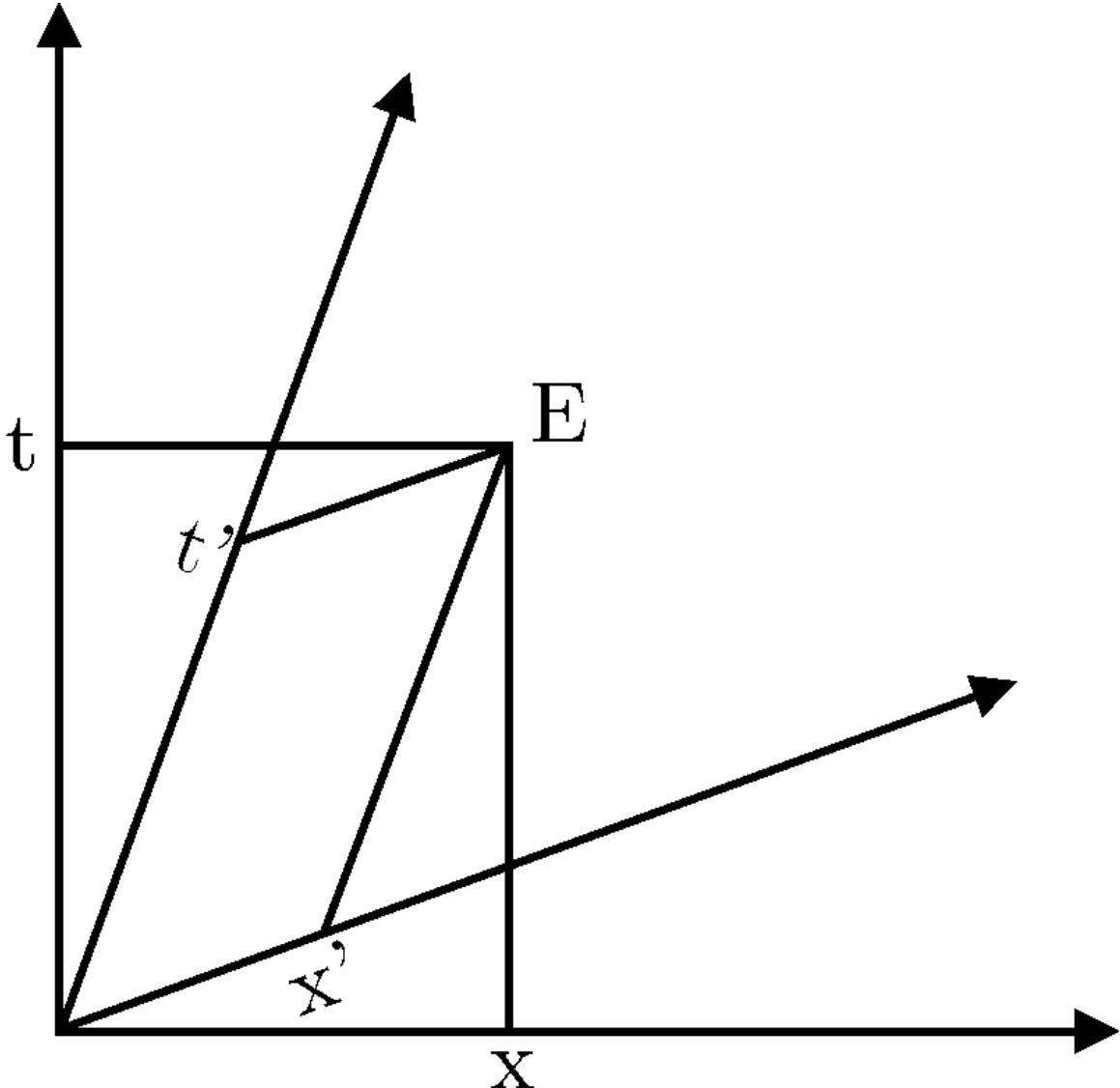


Figure 79

We arrive at the same conclusion if we think about the synchronization of clocks in motion. Consider three clocks (1,2,3) that travel with the same speed $w = \tan \alpha$ relative to \mathcal{F} . To synchronize them, we must send signals from one clock to another. What kind of signals? If

we want our synchronization procedure to be independent of the language we use (that is, independent of the reference frame), then we must use signals that travel with the invariant speed c .

Here is how it's done:

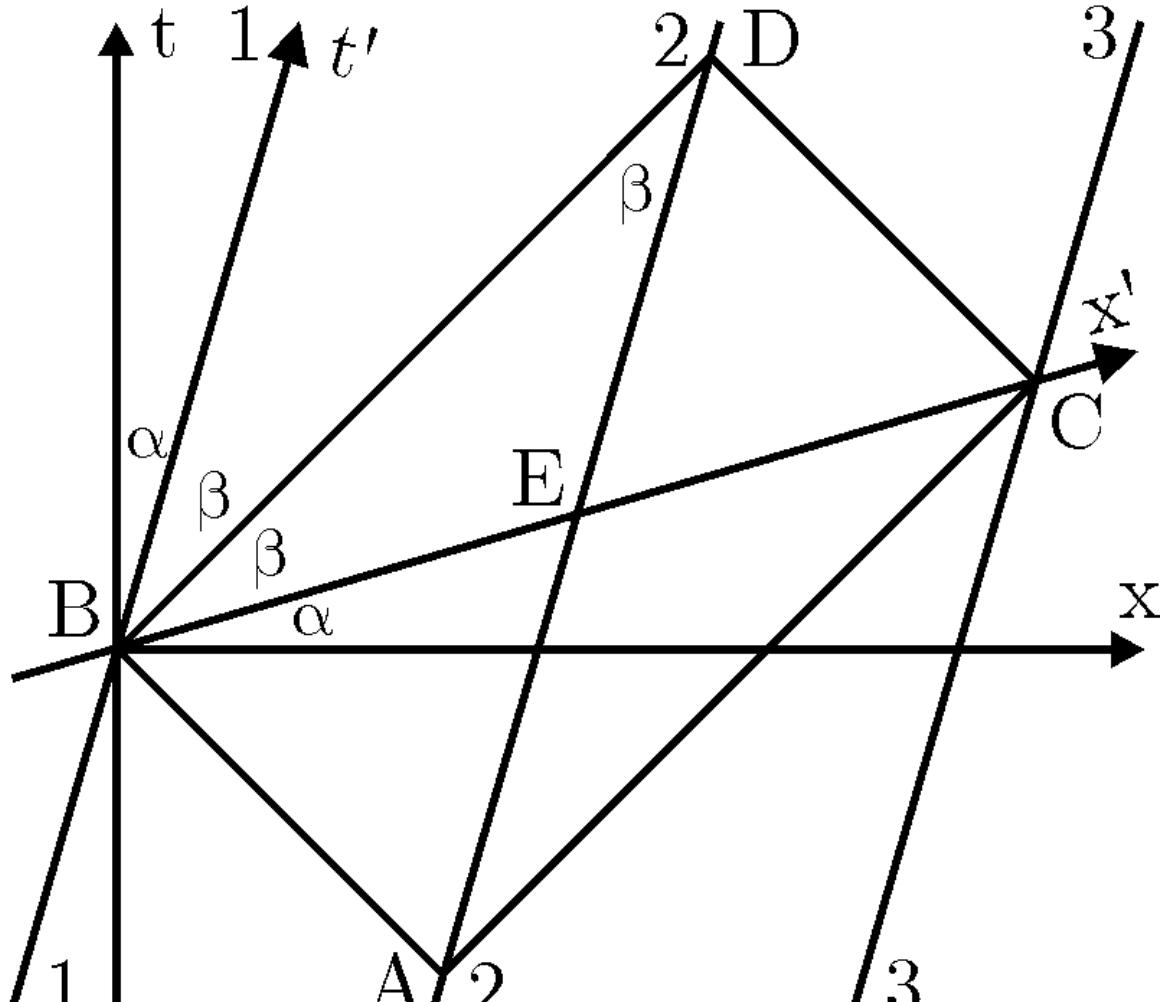


Figure 80

Light signals are sent from clock 2 (event A) and are reflected by clocks 1 and 3 (events B and C , respectively). The distances between the clocks are adjusted so that the reflected signals arrive simultaneously at clock 2 (event D). This ensures that the distance between clocks 1 and 2 equals the distance between clocks 2 and 3, regardless of the inertial frame in which they are compared. In \mathcal{F}' , where the clocks are at rest, the signals from A have traveled equal distances when they reach the first and the third clock, respectively. Since they also have traveled with the same speed c , they have traveled for equal times. Therefore the clocks must be synchronized so that B and C are simultaneous. We may use the *worldline* of clock 1 as the t' axis and the straight line through B and C as the x' axis. It is readily seen that the three angles α in the above diagram are equal. From this and the fact that the slope of the signal from B to D equals 1 (given that $c=1$), the equality of the two angles α follows.

Simultaneity thus depends on the language — the inertial frame — that we use to describe a physical situation. If two events E_1, E_2 are simultaneous in one frame, then there are frames in which E_1 happens after E_2 as well as frames in which E_1 happens before E_2 .

Where do we place the unit points on the space and time axes? The unit point of the time axis of \mathcal{F}' has the coordinates $t' = 1, x' = 0$ and satisfies $t^2 - x^2 = 1$, as we gather from the version $(t')^2 - (x')^2 = t^2 - x^2$ of ($\text{ref}\{\text{ds2}\}$). The unit point of the x' axis has the coordinates $t' = 0, x' = 1$ and satisfies $x^2 - t^2 = 1$. The loci of the unit points of the space and time axes are the hyperbolas that are defined by these equations:

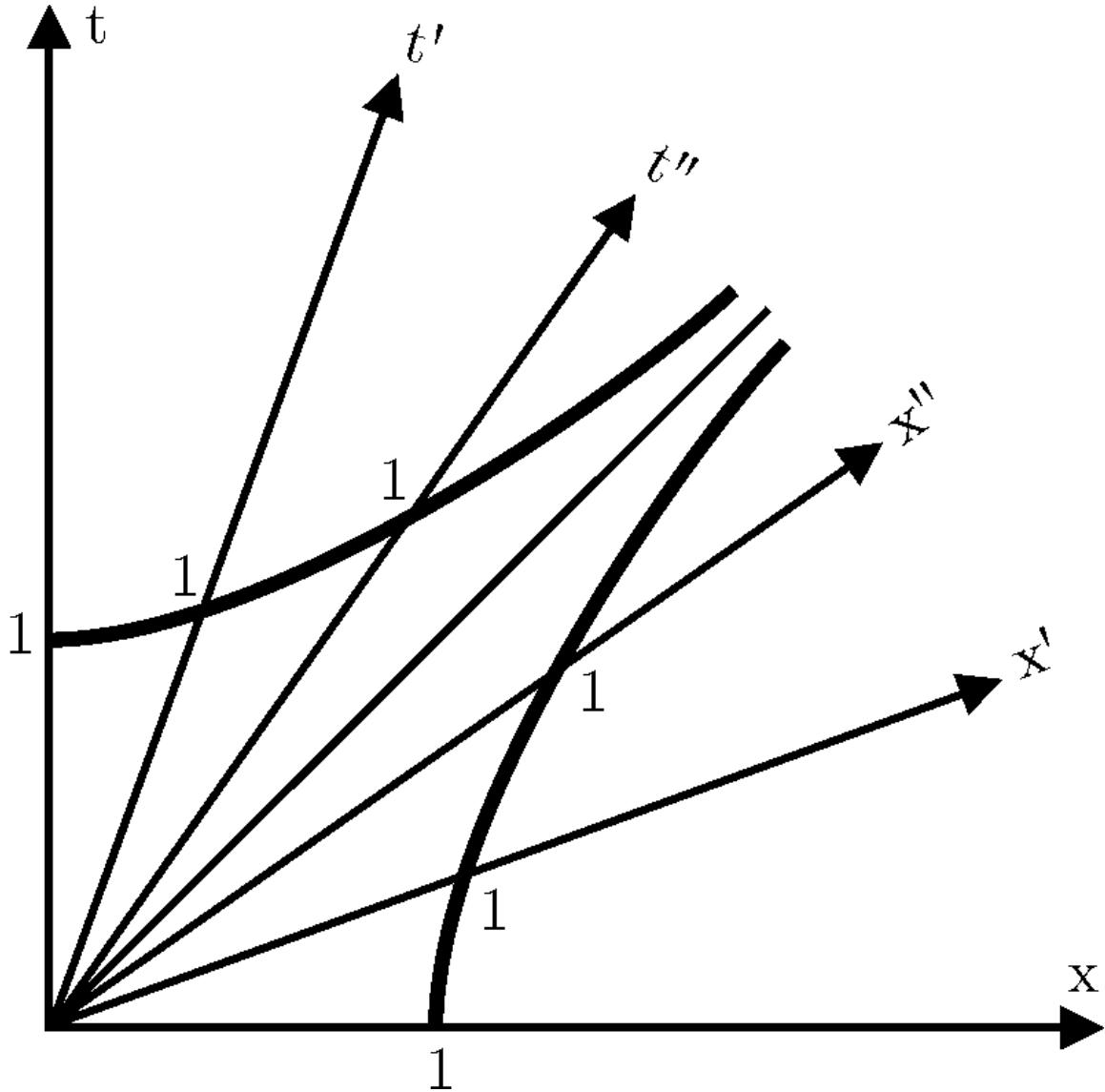


Figure 81

7.3.9 Lorentz contraction, time dilatation

Imagine a meter stick at rest in \mathcal{F}' . At the time $t' = 0$, its ends are situated at the points O and C . At the time $t = 0$, they are situated at the points O and A , which are less than a meter apart. Now imagine a stick (not a *meter* stick) at rest in \mathcal{F} , whose end points at the time $t' = 0$ are O and C . In \mathcal{F}' they are a meter apart, but in the stick's rest-frame they are at O and B and thus more than a meter apart. The bottom line: a moving object is contracted (shortened) in the direction in which it is moving.

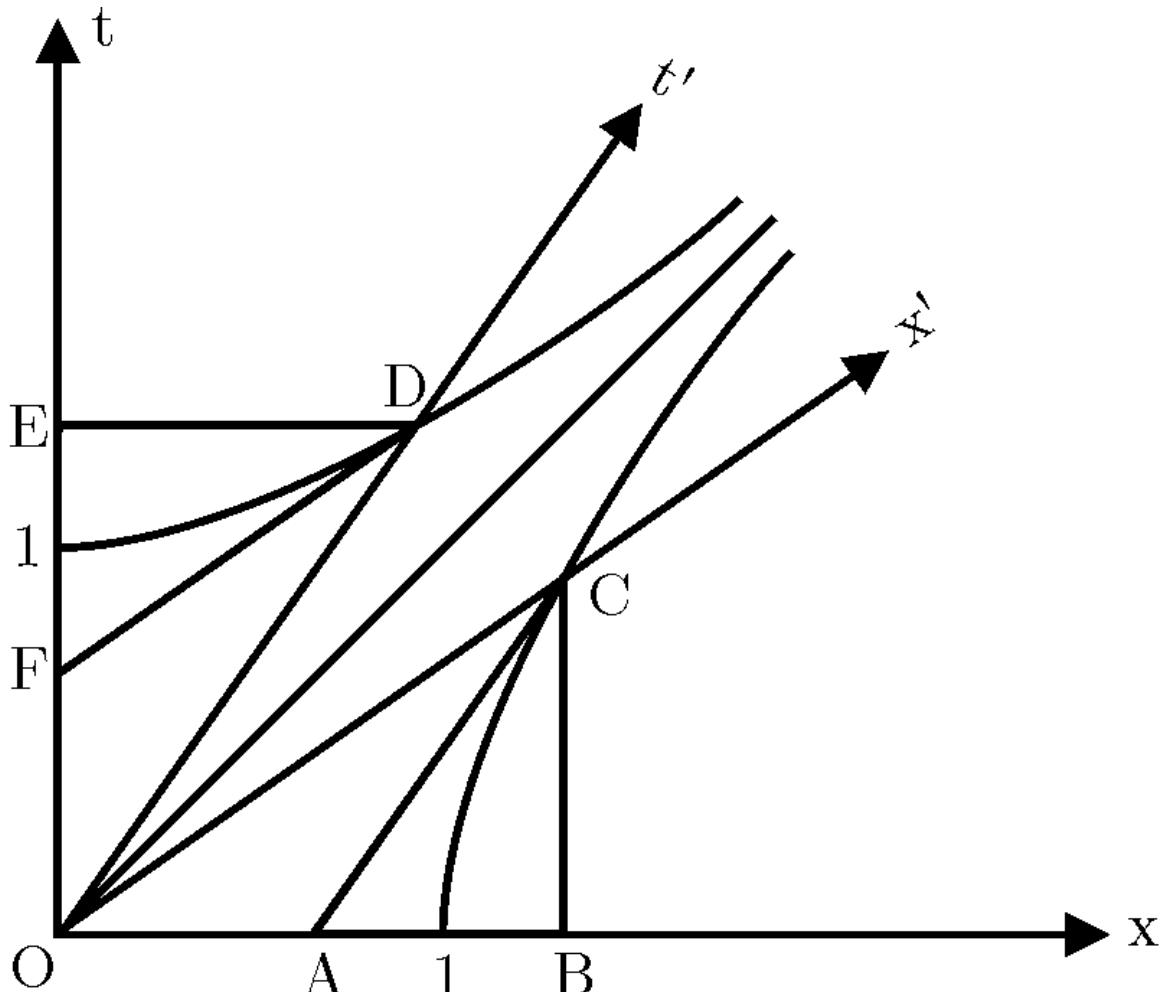


Figure 82

Next imagine two clocks, one (\mathcal{C}) at rest in \mathcal{F} and located at $x = 0$, and one (\mathcal{C}') at rest in \mathcal{F}' and located at $x' = 0$. At D , \mathcal{C}' indicates that one second has passed, while at E (which in \mathcal{F} is simultaneous with D), \mathcal{C} indicates that more than a second has passed. On the other hand, at F (which in \mathcal{F}' is simultaneous with D), \mathcal{C} indicates that less than a second has passed. The bottom line: a moving clock runs slower than a clock at rest.

Example: Muons (μ particles) are created near the top of the atmosphere, some ten kilometers up, when high-energy particles of cosmic origin hit the atmosphere. Since muons decay

spontaneously after an average lifetime of 2.2 microseconds, they don't travel much farther than 600 meters. Yet many are found at sea level. How do they get that far?

The answer lies in the fact that most of them travel at close to the speed of light. While from its own point of view (that is, relative to the inertial system in which it is at rest), a muon only lives for about 2 microseconds, from our point of view (that is, relative to an inertial system in which it travels close to the speed of light), it lives much longer and has enough time to reach the Earth's surface.

48

7.3.10 4-vectors

3-vectors are triplets of real numbers that transform under rotations like the coordinates x, y, z . 4-vectors are quadruplets of real numbers that transform under Lorentz transformations like the coordinates of $\vec{x} = (ct, x, y, z)$.

You will remember that the scalar product of two 3-vectors is invariant under rotations of the (spatial) coordinate axes; after all, this is why we call it a scalar. Similarly, the scalar product of two 4-vectors $\vec{a} = (a_t, \mathbf{a}) = (a_0, a_1, a_2, a_3)$ and $\vec{b} = (b_t, \mathbf{b}) = (b_0, b_1, b_2, b_3)$, defined by

$$(\vec{a}, \vec{b}) = a_0 b_0 - a_1 b_1 - a_2 b_2 - a_3 b_3,$$

is invariant under Lorentz transformations (as well as translations of the coordinate origin and rotations of the spatial axes). To demonstrate this, we consider the sum of two 4-vectors $\vec{c} = \vec{a} + \vec{b}$ and calculate

$$(\vec{c}, \vec{c}) = (\vec{a} + \vec{b}, \vec{a} + \vec{b}) = (\vec{a}, \vec{a}) + (\vec{b}, \vec{b}) + 2(\vec{a}, \vec{b}).$$

The products (\vec{a}, \vec{a}) , (\vec{b}, \vec{b}) , and (\vec{c}, \vec{c}) are invariant 4-scalars. But if *they* are invariant under Lorentz transformations, then so is the scalar product (\vec{a}, \vec{b}) .

One important 4-vector, apart from \vec{x} , is the *4-velocity* $\vec{u} = \frac{d\vec{x}}{ds}$, which is tangent on the worldline $\vec{x}(s)$. \vec{u} is a 4-vector because \vec{x} is one and because ds is a scalar (to be precise, a 4-scalar).

The norm or "magnitude" of a 4-vector \vec{a} is defined as $\sqrt{|(\vec{a}, \vec{a})|}$. It is readily shown that the norm of \vec{u} equals c (exercise!).

Thus if we use natural units, the 4-velocity is a unit vector.

49

48 <http://en.wikibooks.org/wiki/Category%3A>

49 <http://en.wikibooks.org/wiki/Category%3A>

8 Contributors

Edits	User
1	Adicarlo ¹
94	Adrignola ²
1	Bdesham ³
2	BlackMeph ⁴
1	Bookaneer ⁵
1	CommonsDelinker ⁶
1	Dirk H��nniger ⁷
1	Donaldamccartor ⁸
109	Glich ⁹
4	Herbythyme ¹⁰
2	JasonWoof ¹¹
2	Jomegat ¹²
1	Jsilve1 ¹³
2	Kdkeller ¹⁴
75	Koantum ¹⁵
2	Mattpickman ¹⁶
9	Mike.lifeguard ¹⁷
1	Ok123jump ¹⁸
4	QuiteUnusual ¹⁹
1	Ramac ²⁰
1	Recent Runes ²¹

-
- 1 <http://en.wikibooks.org/w/index.php?title=User:Adicarlo>
 - 2 <http://en.wikibooks.org/w/index.php?title=User:Adrignola>
 - 3 <http://en.wikibooks.org/w/index.php?title=User:Bdesham>
 - 4 <http://en.wikibooks.org/w/index.php?title=User:BlackMeph>
 - 5 <http://en.wikibooks.org/w/index.php?title=User:Bookaneer>
 - 6 <http://en.wikibooks.org/w/index.php?title=User:CommonsDelinker>
 - 7 http://en.wikibooks.org/w/index.php?title=User:Dirk_H  nniger
 - 8 <http://en.wikibooks.org/w/index.php?title=User:Donaldamccartor>
 - 9 <http://en.wikibooks.org/w/index.php?title=User:Glich>
 - 10 <http://en.wikibooks.org/w/index.php?title=User:Herbythyme>
 - 11 <http://en.wikibooks.org/w/index.php?title=User:JasonWoof>
 - 12 <http://en.wikibooks.org/w/index.php?title=User:Jomegat>
 - 13 <http://en.wikibooks.org/w/index.php?title=User:Jsilve1>
 - 14 <http://en.wikibooks.org/w/index.php?title=User:Kdkeller>
 - 15 <http://en.wikibooks.org/w/index.php?title=User:Koantum>
 - 16 <http://en.wikibooks.org/w/index.php?title=User:Mattpickman>
 - 17 <http://en.wikibooks.org/w/index.php?title=User:Mike.lifeguard>
 - 18 <http://en.wikibooks.org/w/index.php?title=User:Ok123jump>
 - 19 <http://en.wikibooks.org/w/index.php?title=User:QuiteUnusual>
 - 20 <http://en.wikibooks.org/w/index.php?title=User:Ramac>
 - 21 http://en.wikibooks.org/w/index.php?title=User:Recent_Runes

- 1 Richardlj²²
- 1 RobinH²³
- 1 Rowan²⁴
- 1 Spica61²⁵
- 1 Taober²⁶
- 1 Thenub314²⁷
- 1 Tom Bachmann²⁸
- 4 Whiteknight²⁹
- 3 Yohomes³⁰

22 <http://en.wikibooks.org/w/index.php?title=User:Richardlj>
23 <http://en.wikibooks.org/w/index.php?title=User:RobinH>
24 <http://en.wikibooks.org/w/index.php?title=User:Rowan>
25 <http://en.wikibooks.org/w/index.php?title=User:Spica61>
26 <http://en.wikibooks.org/w/index.php?title=User:Taober>
27 <http://en.wikibooks.org/w/index.php?title=User:Thenub314>
28 http://en.wikibooks.org/w/index.php?title=User:Tom_Bachmann
29 <http://en.wikibooks.org/w/index.php?title=User:Whiteknight>
30 <http://en.wikibooks.org/w/index.php?title=User:Yohomes>

List of Figures

- GFDL: Gnu Free Documentation License. <http://www.gnu.org/licenses/fdl.html>
- cc-by-sa-3.0: Creative Commons Attribution ShareAlike 3.0 License. <http://creativecommons.org/licenses/by-sa/3.0/>
- cc-by-sa-2.5: Creative Commons Attribution ShareAlike 2.5 License. <http://creativecommons.org/licenses/by-sa/2.5/>
- cc-by-sa-2.0: Creative Commons Attribution ShareAlike 2.0 License. <http://creativecommons.org/licenses/by-sa/2.0/>
- cc-by-sa-1.0: Creative Commons Attribution ShareAlike 1.0 License. <http://creativecommons.org/licenses/by-sa/1.0/>
- cc-by-2.0: Creative Commons Attribution 2.0 License. <http://creativecommons.org/licenses/by/2.0/>
- cc-by-2.0: Creative Commons Attribution 2.0 License. <http://creativecommons.org/licenses/by/2.0/deed.en>
- cc-by-2.5: Creative Commons Attribution 2.5 License. <http://creativecommons.org/licenses/by/2.5/deed.en>
- cc-by-3.0: Creative Commons Attribution 3.0 License. <http://creativecommons.org/licenses/by/3.0/deed.en>
- GPL: GNU General Public License. <http://www.gnu.org/licenses/gpl-2.0.txt>
- LGPL: GNU Lesser General Public License. <http://www.gnu.org/licenses/lgpl.html>
- PD: This image is in the public domain.
- ATTR: The copyright holder of this file allows anyone to use it for any purpose, provided that the copyright holder is properly attributed. Redistribution, derivative work, commercial use, and all other use is permitted.
- EURO: This is the common (reverse) face of a euro coin. The copyright on the design of the common face of the euro coins belongs to the European Commission. Authorised is reproduction in a format without relief (drawings, paintings, films) provided they are not detrimental to the image of the euro.
- LFK: Lizenz Freie Kunst. <http://artlibre.org/licence/lal/de>
- CFR: Copyright free use.

- EPL: Eclipse Public License. <http://www.eclipse.org/org/documents/epl-v10.php>

Copies of the GPL, the LGPL as well as a GFDL are included in chapter Licenses³¹. Please note that images in the public domain do not require attribution. You may click on the image numbers in the following table to open the webpage of the images in your webbrowser.

31 Chapter 9 on page 143

1		PD
2		GFDL
3		GFDL
4	en:User:Cburnett ³²	GFDL
5	Ulrich Mohrhoff	GFDL
6	Ulrich Mohrhoff	GFDL
7	Ulrich Mohrhoff	GFDL
8	Ulrich Mohrhoff	GFDL
9	Ulrich Mohrhoff	GFDL
10	Ulrich Mohrhoff	GFDL
11	Ulrich Mohrhoff	GFDL
12	Ulrich Mohrhoff	GFDL
13	Ulrich Mohrhoff	GFDL
14	Ulrich Mohrhoff	GFDL
15	Ulrich Mohrhoff	GFDL
16	Ulrich Mohrhoff	GFDL
17	Ulrich Mohrhoff	GFDL
18	Ulrich Mohrhoff	GFDL
19	Ulrich Mohrhoff	GFDL
20	Ulrich Mohrhoff	GFDL
21	Merikanto ³³ , Adrignola ³⁴	
22	Ulrich Mohrhoff	GFDL
23	Ulrich Mohrhoff	GFDL
24	Ulrich Mohrhoff	GFDL
25	Ulrich Mohrhoff	GFDL
26	Ulrich Mohrhoff	GFDL
27	Koantum ³⁵ , svg version by Trutz Behn ³⁶	GFDL
28	Klaus-Dieter Keller ³⁷	cc-by-sa-3.0
29	Klaus-Dieter Keller ³⁸	cc-by-sa-3.0
30	Dr. Tonomura	GFDL
31	Dr. Tonomura	GFDL
32	Dr. Tonomura	GFDL
33	Dr. Tonomura	GFDL
34	Klaus-Dieter Keller ³⁹	cc-by-sa-3.0
35	Klaus-Dieter Keller ⁴⁰	cc-by-sa-3.0
36	self	GFDL
37	Ulrich Mohrhoff	GFDL
38	Ulrich Mohrhoff	GFDL
39	Koantum ⁴¹	GFDL
40	Koantum ⁴²	GFDL

32 <http://en.wikibooks.org/wiki/%3Aen%3AUser%3ACburnett>
 33 <http://en.wikibooks.org/wiki/User%3AMerikanto>
 34 <http://en.wikibooks.org/wiki/User%3AAdrignola>
 35 <http://en.wikibooks.org/wiki/User%3AKoantum>
 36 <http://en.wikibooks.org/wiki/%3Ade%3AUser%3ATrutz%20Behn>
 37 <http://en.wikibooks.org/wiki/User%3AKdkeller>
 38 <http://en.wikibooks.org/wiki/User%3AKdkeller>
 39 <http://en.wikibooks.org/wiki/User%3AKdkeller>
 40 <http://en.wikibooks.org/wiki/User%3AKdkeller>
 41 <http://en.wikibooks.org/wiki/User%3AKoantum>
 42 <http://en.wikibooks.org/wiki/User%3AKoantum>

41	Koantum ⁴³	GFDL
42	Koantum ⁴⁴	GFDL
43		PD
44	Koantum ⁴⁵	GFDL
45	Koantum ⁴⁶	GFDL
46	Koantum ⁴⁷	GFDL
47	Koantum ⁴⁸	GFDL
48	Ulrich Mohrhoff	GFDL
49	Ulrich Mohrhoff	GFDL
50	Ulrich Mohrhoff	GFDL
51	Ulrich Mohrhoff	GFDL
52	Ulrich Mohrhoff	GFDL
53	Ulrich Mohrhoff	GFDL
54	Ulrich Mohrhoff	GFDL
55	Ulrich Mohrhoff	GFDL
56	Ulrich Mohrhoff	GFDL
57	Ulrich Mohrhoff	GFDL
58	Ulrich Mohrhoff	GFDL
59	Ulrich Mohrhoff	GFDL
60	Koantum ⁴⁹	GFDL
61	Koantum ⁵⁰	GFDL
62	Koantum ⁵¹	GFDL
63	Koantum ⁵²	GFDL
64	Koantum ⁵³	GFDL
65	Koantum ⁵⁴	GFDL
66	Koantum ⁵⁵	GFDL
67	Koantum ⁵⁶	GFDL
68	Ulrich Mohrhoff	GFDL
69	Ulrich Mohrhoff	GFDL
70	Ulrich Mohrhoff	GFDL
71	Ulrich Mohrhoff	GFDL
72	Ulrich Mohrhoff	GFDL
73	self	GFDL
74	Koantum ⁵⁷	GFDL
75	Ulrich Mohrhoff	GFDL

-
- 43 <http://en.wikibooks.org/wiki/User%3AKoantum>
 44 <http://en.wikibooks.org/wiki/User%3AKoantum>
 45 <http://en.wikibooks.org/wiki/User%3AKoantum>
 46 <http://en.wikibooks.org/wiki/User%3AKoantum>
 47 <http://en.wikibooks.org/wiki/User%3AKoantum>
 48 <http://en.wikibooks.org/wiki/User%3AKoantum>
 49 <http://en.wikibooks.org/wiki/User%3AKoantum>
 50 <http://en.wikibooks.org/wiki/User%3AKoantum>
 51 <http://en.wikibooks.org/wiki/User%3AKoantum>
 52 <http://en.wikibooks.org/wiki/User%3AKoantum>
 53 <http://en.wikibooks.org/wiki/User%3AKoantum>
 54 <http://en.wikibooks.org/wiki/User%3AKoantum>
 55 <http://en.wikibooks.org/wiki/User%3AKoantum>
 56 <http://en.wikibooks.org/wiki/User%3AKoantum>
 57 <http://en.wikibooks.org/wiki/User%3AKoantum>

76	Original uploader was Schorschi2 ⁵⁸ at de.wikipedia ⁵⁹	PD
77	Ulrich Mohrhoff	GFDL
78	Ulrich Mohrhoff	GFDL
79	Ulrich Mohrhoff	GFDL
80	Ulrich Mohrhoff	GFDL
81	Ulrich Mohrhoff	GFDL
82	Ulrich Mohrhoff	GFDL

58 <http://en.wikibooks.org/wiki/%3Ade%3AUser%3ASchorschi2>
59 <http://de.wikipedia.org>

9 Licenses

9.1 GNU GENERAL PUBLIC LICENSE

Version 3, 29 June 2007

Copyright © 2007 Free Software Foundation, Inc.
<http://fsf.org/>

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed. Preamble

The GNU General Public License is a free, copyleft license for software and other kinds of works.

The licenses for most software and other practical works are designed to take away your freedom to share and change the works. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change all versions of a program—to make sure it remains free software for all its users. We, the Free Software Foundation, use the GNU General Public License for most of our software; it applies also to any other work released this way by its authors. You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for them if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs, and that you know you can do these things.

To protect your rights, we need to prevent others from denying you these rights or asking you to surrender the rights. Therefore, you have certain responsibilities if you distribute copies of the software, or if you modify it: responsibilities to respect the freedom of others.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must pass on to the recipients the same freedoms that you received. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

Developers that use the GNU GPL protect your rights with two steps: (1) assert copyright on the software, and (2) offer you this License giving you legal permission to copy, distribute and/or modify it.

For the developers' and authors' protection, the GPL clearly explains that there is no warranty for this free software. For both users' and authors' sake, the GPL requires that modified versions be marked as changed, so that their problems will not be attributed erroneously to authors of previous versions.

Some devices are designed to deny users access to install or run modified versions of the software inside them, although the manufacturer can do so. This is fundamentally incompatible with the aim of protecting users' freedom to change the software. The systematic pattern of such abuse occurs in the area of products for individuals to use, which is precisely where it is most unacceptable. Therefore, we have designed this version of the GPL to prohibit the practice for those products. If such problems arise substantially in other domains, we stand ready to extend this provision to those domains in future versions of the GPL, as needed to protect the freedom of users.

Finally, every program is threatened constantly by software patents. States should not allow patents to restrict development and use of software on general-purpose computers, but in those that do, we wish to avoid the special danger that patents applied to a free program could make it effectively proprietary. To prevent this, the GPL assures that patents cannot be used to render the program non-free.

The precise terms and conditions for copying, distribution and modification follow. TERMS AND CONDITIONS 0. Definitions.

"This License" refers to version 3 of the GNU General Public License.

"Copyright" also means copyright-like laws that apply to other kinds of works, such as semiconductor masks.

"The Program" refers to any copyrighted work licensed under this License. Each licensee is addressed as "you". "Licensees" and "recipients" may be individuals or organizations.

To "modify" a work means to copy from or adapt all or part of the work in a fashion requiring copyright permission, other than the making of an exact copy. The resulting work is called a "modified version" of the earlier work or a work "based on" the earlier work.

A "covered work" means either the unmodified Program or a work based on the Program.

To "propagate" a work means to do anything with it that, without permission, would make you directly or secondarily liable for infringement under applicable copyright law, except executing it on a computer or modifying a private copy. Propagation includes copying, distribution (with or without modification), making available to the public, and in some countries other activities as well.

To "convey" a work means any kind of propagation that enables other parties to make or receive copies. Mere interaction with a user through a computer

network, with no transfer of a copy, is not conveying.

An interactive user interface displays "Appropriate Legal Notices" to the extent that it includes a convenient and prominently visible feature that (1) displays an appropriate copyright notice, and (2) tells the user that there is no warranty for the work (except to the extent that warranties are provided), that licensees may convey the work under this License, and how to view a copy of this License. If the interface presents a list of user commands or options, such as a menu, a prominent item in the list meets this criterion. 1. Source Code.

The "source code" for a work means the preferred form of the work for making modifications to it. "Object code" means any non-source form of a work.

A "Standard Interface" means an interface that either is an official standard defined by a recognized standards body, or, in the case of interfaces specified for a particular programming language, one that is widely used among developers working in that language.

The "System Libraries" of an executable work include anything, other than the work as a whole, that (a) is included in the normal form of packaging a Major Component, but which is not part of that Major Component, and (b) serves only to enable use of the work with that Major Component, or to implement a Standard Interface for which an implementation is available to the public in source code form. A "Major Component", in this context, means a major essential component (kernel, window system, and so on) of the specific operating system (if any) on which the executable work runs, or a compiler used to produce the work, or an object code interpreter used to run it.

The "Corresponding Source" for a work in object code form means all the source code needed to generate, install, and (for an executable work) run the object code and to modify the work, including scripts to control those activities. However, it does not include the work's System Libraries, or general-purpose tools or generally available free programs which are used unmodified in performing those activities but which are not part of the work. For example, Corresponding Source includes interface definition files associated with source files for the work, and the source code for shared libraries and dynamically linked subprograms that the work is specifically designed to require, such as by intimate data communication or control flow between those subprograms and other parts of the work.

The Corresponding Source need not include anything that users can regenerate automatically from other parts of the Corresponding Source.

The Corresponding Source for a work in source code form is that same work. 2. Basic Permissions.

All rights granted under this License are granted for the term of copyright on the Program, and are irrevocable provided the stated conditions are met. This License explicitly affirms your unlimited permission to run the unmodified Program. The output from running a covered work is covered by this License only if the output, given its content, constitutes a covered work. This License acknowledges your rights of fair use or other equivalent, as provided by copyright law.

You may make, run and propagate covered works that you do not convey, without conditions so long as your license otherwise remains in force. You may convey covered works to others for the sole purpose of having them make modifications exclusively for you, or provide you with facilities for running those works, provided that you comply with the terms of this License in conveying all material for which you do not control copyright. Those thus making or running the covered works for you must do so exclusively on your behalf, under your direction and control, on terms that prohibit them from making any copies of your copyrighted material outside their relationship with you.

Conveying under any other circumstances is permitted solely under the conditions stated below. Sublicensing is not allowed; section 10 makes it unnecessary. 3. Protecting Users' Legal Rights From Anti-Circumvention Law.

No covered work shall be deemed part of an effective technological measure under any applicable law fulfilling obligations under article 11 of the WIPO copyright treaty adopted on 20 December 1996, or similar laws prohibiting or restricting circumvention of such measures.

When you convey a covered work, you waive any legal power to forbid circumvention of technological measures to the extent such circumvention is effected by exercising rights under this License with respect to the covered work, and you disclaim any intention to limit operation or modification of the work as a means of enforcing, against the work's users, your or third parties' legal rights to forbid circumvention of technological measures. 4. Conveying Verbatim Copies.

You may convey copyrighted copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice; keep intact all notices stating that this License and any non-permissive terms added in accord with section 7 apply to the code; keep intact all notices of the absence of any warranty; and give all recipients a copy of this License along with the Program.

You may charge any price or no price for each copy that you convey, and you may offer support or warranty protection for a fee. 5. Conveying Modified Source Versions.

You may convey a work based on the Program, or the modifications to produce it from the Program, in the form of source code under the terms of section 4, provided that you also meet all of these conditions:

- * a) The work must carry prominent notices stating that you modified it, and giving a relevant date.
- * b) The work must carry prominent notices stating that it is released under this License and any conditions added under section 7. This requirement modifies the requirement in section 4 to "keep intact all notices".
- * c) You must license the entire work, as a whole, under this License to anyone who comes into possession of a copy. This License will therefore apply, along with any applicable section 7 additional terms, to the whole of the work, and all its parts, regardless of how they are packaged. This License gives no permission to license the work in any other way, but it does not invalidate such permission if you have separately received it.
- * d) If the work has interactive user interfaces, each must display Appropriate Legal Notices; however, if the Program has interactive interfaces that do not display Appropriate Legal Notices, your work need not make them do so.

A compilation of a covered work with other separate and independent works, which are not by their nature extensions of the covered work, and which are not combined with it such as to form a larger program, in or on a volume of a storage or distribution medium, is called an "aggregate" if the compilation and its resulting copyright are not used to limit the access or legal rights of the compilation's users beyond what the individual works permit. Inclusion of a covered work in an aggregate does not cause this License to apply to the other parts of the aggregate. 6. Conveying Non-Source Forms.

You may convey a covered work in object code form under the terms of sections 4 and 5, provided that you also convey the machine-readable Corresponding Source under the terms of this License, in one of these ways:

- * a) Convey the object code in, or embodied in, a physical product (including a physical distribution medium), accompanied by the Corresponding Source fixed on a durable physical medium customarily used for software interchange.
- * b) Convey the object code in, or embodied in, a physical product (including a physical distribution medium), accompanied by a written offer, valid for at least three years and valid for as long as you offer spare parts or customer support for that product model, to give anyone who possesses the object code either (1) a copy of the Corresponding Source for all the software in the product that is covered by this License, on a durable physical medium customarily used for software interchange, for a price no more than your reasonable cost of physically performing this conveying of source, or (2) access to copy the Corresponding Source from a network server at no charge.
- * c) Convey individual copies of the object code with a copy of the written offer to provide the Corresponding Source. This alternative is allowed only occasionally and noncommercially, and only if you received the object code with such an offer, in accord with subsection 6b.
- * d) Convey the object code by offering access from a designated place (gratis or for a charge), and offer equivalent access to the Corresponding Source in the same way through the same place at no further charge. You need not require recipients to copy the Corresponding Source along with the object code. If the place to copy the object code is a network server, the Corresponding Source may be on a different server (operated by you or a third party) that supports equivalent copying facilities, provided you maintain clear directions next to the object code saying where to find the Corresponding Source. Regardless of what server hosts the Corresponding Source, you remain obligated to ensure that it is available for as long as needed to satisfy these requirements.
- * e) Convey the object code using peer-to-peer transmission, provided you inform other peers where the object code and Corresponding Source of the work are being offered to the same place at no further charge.

You need not require recipients to copy the Corresponding Source along with the object code. If the place to copy the object code is a network server, the Corresponding Source may be on a different server (operated by you or a third party) that supports equivalent copying facilities, provided you maintain clear directions next to the object code saying where to find the Corresponding Source. Regardless of what server hosts the Corresponding Source, you remain obligated to ensure that it is available for as long as needed to satisfy these requirements.

A separable portion of the object code, whose source code is excluded from the Corresponding Source as a System Library, need not be included in conveying the object code work.

A "User Product" is either (1) a "consumer product", which means any tangible personal property which is normally used for personal, family, or household purposes, or (2) anything designed or sold for incorporation into a dwelling. In determining whether a product is a consumer product, doubtful cases shall be resolved in favor of coverage. For a particular product received by a particular user, "normally used" refers to a typical or common use of that class of product, regardless of the status of the particular user or of the way in which the particular user actually uses, or expects or is expected to use, the product. A product is a consumer product regardless of whether the product has substantial commercial, industrial or non-consumer uses, unless such uses represent the only significant mode of use of the product.

"Installation Information" for a User Product means any methods, procedures, authorization keys, or other information required to install and execute modified versions of a covered work in that User Product from a modified version of its Corresponding Source. The information must suffice to ensure that the continued functioning of the modified object code is in no case prevented or interfered with solely because modification has been made.

If you convey an object code work under this section in, or with, or specifically for use in, a User Product, and the conveying occurs as part of a transaction in which the right of possession and use of the User Product is transferred to the recipient in perpetuity or for a fixed term (regardless of how the transaction is characterized), the Corresponding Source conveyed under this section must be accompanied by the Installation Information. But this requirement does not apply if neither you nor any third party retains the ability to install modified object code on the User Product (for example, the work has been installed in ROM).

The requirement to provide Installation Information does not include a requirement to continue to provide support service, warranty, or updates for a work that has been modified or installed by the recipient, or for the User Product in which it has been modified or installed. Access to a network may be denied when the modification itself materially and adversely affects the operation of the network or violates the rules and protocols for communication across the network.

Corresponding Source conveyed, and Installation Information provided, in accord with this section must be in a format that is publicly documented (and with an implementation available to the public in source code form), and must require no special password or key for unpacking, reading or copying. 7. Additional Terms.

"Additional permissions" are terms that supplement the terms of this License by making exceptions from one or more of its conditions. Additional permissions that are applicable to the entire Program shall be treated as though they were included in this License, to the extent that they are valid under applicable law. If additional permissions apply only to part of the Program, that part may be used separately under those permissions, but the entire Program remains governed by this License without regard to the additional permissions.

Wher you convey a copy of a covered work, you may at your option remove any additional permissions from that copy, or from any part of it. (Additional permissions may be written to require their own removal in certain cases when you modify the work.) You may place additional permissions on material, added by you to a covered work, for which you have or can give appropriate copyright permission.

Notwithstanding any other provision of this License, for material you add to a covered work, you may (if authorized by the copyright holders of that material) supplement the terms of this License with terms:

- * a) Disclaiming warranty or limiting liability differently from the terms of sections 15 and 16 of this License; or
- * b) Requiring preservation of specified reasonable legal notices or author attributions in that material or in the Appropriate Legal Notices displayed by works containing it; or
- * c) Prohibiting misrepresentation of the origin of that material, or requiring that modified versions of such material be marked in reasonable ways as different from the original version; or
- * d) Limiting the use for publicity purposes of names of licensors or authors of the material; or
- * e) Declining to grant rights under trademark law for use of some trade names, trademarks, or service marks; or
- * f) Requiring indemnification of licensors and authors of that material by anyone who conveys the material (or modified versions of it) with contractual assumptions of liability to the recipient, for any liability that these contractual assumptions directly impose on those licensors and authors.

All other non-permissive additional terms are considered "further restrictions" within the meaning of section 10. If the Program as you received it, or any part of it, contains a notice stating that it is governed by this License along with a term that is a further restriction, you may remove that term. If a license document contains a further restriction but permits relicensing or conveying under this License, you may add to a covered work material governed by the terms of that license document, provided that the further restriction does not survive such relicensing or conveying.

If you add terms to a covered work in accord with this section, you must place, in the relevant source files, a statement of the additional terms that apply to those files, or a notice indicating where to find the applicable terms.

Additional terms, permissive or non-permissive, may be stated in the form of a separately written license, or stated as exceptions; the above requirements apply either way. 8. Termination.

You may not propagate or modify a covered work except as expressly provided under this License. Any attempt otherwise to propagate or modify it is void, and will automatically terminate your rights under this License (including any patent licenses granted under the third paragraph of section 11).

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work)

from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, you do not qualify to receive new licenses for the same material under section 10. 9. Acceptance Not Required for Having Copies.

You are not required to accept this License in order to receive or run a copy of the Program. Ancillary propagation of a covered work occurring solely as a consequence of using peer-to-peer transmission to receive a copy likewise does not require acceptance. However, nothing other than this License grants you permission to propagate or modify any covered work. These actions infringe copyright if you do not accept this License. Therefore, by modifying or propagating a covered work, you indicate your acceptance of this License to do so. 10. Automatic Licensing of Downstream Recipients.

Each time you convey a covered work, the recipient automatically receives a license from the original licensors, to run, modify and propagate that work, subject to this License. You are not responsible for enforcing compliance by third parties with this License.

An "entity transaction" is a transaction transferring control of an organization, or substantially all assets of one, or subdividing an organization, or merging organizations. If propagation of a covered work results from an entity transaction, each party to that transaction who receives a copy of the work also receives whatever licenses to the work the party's predecessor in interest had or could give under the previous paragraph, plus a right to possession of the Corresponding Source of the work from the predecessor in interest, if the predecessor has it or can get it with reasonable efforts.

You may not impose any further restrictions on the exercise of the rights granted or affirmed under this License. For example, you may not impose a license fee, royalty, or other charge for exercise of rights granted under this License, and you may not initiate litigation (including a cross-claim or counter-claim in a lawsuit) alleging that any patent claim is infringed by making, using, selling, offering for sale, or importing the Program or any portion of it. 11. Patents.

A "contributor" is a copyright holder who authorizes use under this License of the Program or a work on which the Program is based. The work thus licensed is called the contributor's "contributor version".

A contributor's "essential patent claims" are all patent claims owned or controlled by the contributor, whether already acquired or hereafter acquired, that would be infringed by some manner, permitted by this License, of making, using, or selling its contributor version, but do not include claims that would be infringed only as a consequence of further modification of the contributor version. For purposes of this definition, "control" includes the right to grant patent sublicenses in a manner consistent with the requirements of this License.

Each contributor grants you a non-exclusive, worldwide, royalty-free patent license under the contributor's essential patent claims, to make, use, sell, offer for sale, import and otherwise run, modify and propagate the contents of its contributor version.

In the following three paragraphs, a "patent license" is any express agreement or commitment, however denominated, not to enforce a patent (such as an express permission to practice a patent or covenant not to sue for patent infringement). To "grant" such a patent license to a party means to make such an agreement or commitment not to enforce a patent against the party.

If you convey a covered work, knowingly relying on a patent license, and the Corresponding Source of the work is not available for anyone to copy, free of charge and under the terms of this License, through a publicly available network server or other readily accessible means, then you must either (1) cause the Corresponding Source to be so available, or (2) arrange to deprive yourself of the benefit of the patent license for this particular work, or (3) arrange, in a manner consistent with the requirements of this License, to extend the patent license to downstream recipients. "Knowingly relying" means you have actual knowledge that, but for the patent license, you conveying the covered work in a country, or your recipient's use of the covered work in a country, would infringe one or more identifiable patents in that country that you have reason to believe are valid.

If, pursuant to or in connection with a single transaction or arrangement, you convey, or propagate by procuring conveyance of, a covered work, and granting a patent license to some of the parties receiving the covered work authorizing them to use, propagate, modify or convey a specific copy of the covered work, then the patent license you grant is automatically extended to all recipients of the covered work and works based on it.

A patent license is "discriminatory" if it does not include within the scope of its coverage, prohibits the exercise of, or is conditioned on the non-exercise of one or more of the rights that are specifically granted under this License. You may not convey a covered work if you are a party to an arrangement with a third party that is in the business of distributing software, under which you make payment to the third party based on the extent of your activity of conveying the work, and under which the third party grants, to any of the parties who would receive the covered work from you, a discriminatory patent license (a) in connection with copies of the covered work conveyed by you (or copies made from those copies), or (b) primarily for and in connection with specific products or compilations that contain the covered work, unless you entered into that arrangement, or that patent license was granted, prior to 28 March 2007.

Nothing in this License shall be construed as excluding or limiting any implied license or other defenses to infringement that may otherwise be available to you under applicable patent law. 12. No Surrender of Others' Freedom.

If conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot convey a covered work so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not convey it at all. For example, if you agree to terms that obligate you to collect a royalty for further conveying from those to whom you convey the Program, the only way you could satisfy both those terms and this License would be to refrain entirely from conveying the Program. 13. Use with the GNU Afferro General Public License.

9.2 GNU Free Documentation License

Version 1.3, 3 November 2008

Copyright © 2000, 2001, 2002, 2007, 2008 Free Software Foundation, Inc. <http://fsf.org/>

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed. 0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference. 1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or

Notwithstanding any other provision of this License, you have permission to link or combine any covered work with a work licensed under version 3 of the GNU Afferro General Public License into a single combined work, and to convey the resulting work. The terms of this License will continue to apply to the part which is the covered work, but the special requirements of the GNU Afferro General Public License, section 13, concerning interaction through a network will apply to the combination as such. 14. Revised Versions of this License.

The Free Software Foundation may publish revised and/or new versions of the GNU General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Program specifies that a certain numbered version of the GNU General Public License "or any later version" applies to it, you have the option of following the terms and conditions either of that numbered version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of the GNU General Public License, you may choose any version ever published by the Free Software Foundation.

If the Program specifies that a proxy can decide which future versions of the GNU General Public License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Program.

Later license versions may give you additional or different permissions. However, no additional obligations are imposed on any author or copyright holder as a result of your choosing to follow a later version. 15. Disclaimer of Warranty.

THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION. 16. Limitation of Liability.

IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MODIFIES AND/OR CONVEYS THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. 17. Interpretation of Sections 15 and 16.

If the disclaimer of warranty and limitation of liability provided above cannot be given local legal effect

according to their terms, reviewing courts shall apply local law that most closely approximates an absolute waiver of all civil liability in connection with the Program, unless a warranty or assumption of liability accompanies a copy of the Program in return for a fee.

END OF TERMS AND CONDITIONS How to Apply These Terms to Your New Programs

If you develop a new program, and you want it to be of the greatest possible use to the public, the best way to achieve this is to make it free software which everyone can redistribute and change under these terms.

To do so, attach the following notices to the program. It is safest to attach them to the start of each source file to most effectively state the exclusion of warranty; and each file should have at least the "copyright" line and a pointer to where the full notice is found.

<one line to give the program's name and a brief idea of what it does.> Copyright (C) <year> <name of author>

This program is free software: you can redistribute it or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see <http://www.gnu.org/licenses/>.

Also add information on how to contact you by electronic and paper mail.

If the program does terminal interaction, make it output a short notice like this when it starts in an interactive mode:

<program> Copyright (C) <year> <name of author> This program comes with ABSOLUTELY NO WARRANTY; for details type 'show w'. This is free software, and you are welcome to redistribute it under certain conditions; type 'show c' for details.

The hypothetical commands 'show w' and 'show c' should show the appropriate parts of the General Public License. Of course, your program's commands might be different; for a GUI interface, you would use an "about box".

You should also get your employer (if you work as a programmer) or school, if any, to sign a "copyright disclaimer" for the program, if necessary. For more information on this, and how to apply and follow the GNU GPL, see <http://www.gnu.org/licenses/>.

The GNU General Public License does not permit incorporating your program into proprietary programs. If your program is a subroutine library, you may consider it more useful to permit linking proprietary applications with the library. If this is what you want to do, use the GNU Lesser General Public License instead of this License. But first, please read <http://www.gnu.org/philosophy/why-not-lGPL.html>.

PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

The "publisher" means any person or entity that distributes copies of the Document to the public.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License. 2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies. 3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you

must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document. 4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

* A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission. * B. List on the Title

Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement. * C. State on the Title page the name of the publisher of the Modified Version, as the publisher. * D. Preserve all the copyright notices of the Document. * E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices. * F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below. * G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice. * H. Include an unaltered copy of this License. * I. Preserve the section Entitled "History". Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence. * J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, if the original publisher of the version it refers to gives permission. * K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein. * L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles. * M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version. * N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section. * O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add an-

other; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version. 5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements". 6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document. 7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate. 8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title. 9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it. 10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of

this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document. 11. RELICENSING

"Massive Multiauthor Collaboration Site" (or "MMC Site") means any World Wide Web server that publishes copyrighted works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A "Massive Multiauthor Collaboration" (or "MMC") contained in the site means any set of copyrighted works thus published on the MMC site.

"CC-BY-SA" means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

"Incorporate" means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is "eligible for relicensing" if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing. ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

Copyright (C) YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with ... Texts." line with this:

with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.

4. Combined Works.

You may convey a Combined Work under terms of your choice that, taken together, effectively do not restrict modification of the portions of the Library contained in the Combined Work and reverse engineering for debugging such modifications, if you also do each of the following:

* a) Give prominent notice with each copy of the Combined Work that the Library is used in it and that the Library and its use are covered by this License. * b) Accompany the Combined Work with a copy of the GNU GPL and this license document. * c) For a Combined Work that displays copyright notices during execution, include the copyright notice for the Library among these notices, as well as a reference directing the user to the copy of the GNU GPL and this license document. * d) Do one of the following: o 0) Convey the Minimal Corresponding Source under the terms of this License, and the Corresponding Application Code in a form suitable for, and under terms that permit, the user to recombine or relink the Application with a modified version of the Linked Version to produce a modified Combined Work, in the manner specified by section 6 of the GNU GPL for conveying Corresponding Source. o 1) Use a suitable shared library mechanism for linking with the Library. A suitable mechanism is one that (a) uses at run time a copy of the Library already present on the user's computer system, and (b) will operate properly with a modified version of the Library that is interface-compatible with the Linked Version. * e) Provide Installation Information, but only if you would otherwise be required to provide such information under section 6 of the GNU GPL, and only to the extent that such information is necessary to install and execute a modified version of the Combined Work produced by recombining or relinking the Application with a modified version of the Linked Version. (If you use option 4d0, the Installation Information must accompany the Minimal Corresponding Source and Corresponding Application Code. If you use option 4d1, you must provide the Installation Information in the manner specified by section 6 of the GNU GPL for conveying Corresponding Source.)

Each version is given a distinguishing version number. If the Library as you received it specifies that a certain numbered version of the GNU Lesser General Public License "or any later version" applies to it, you have the option of following the terms and conditions either of that published version or of any later version published by the Free Software Foundation. If the Library as you received it does not specify a version number of the GNU Lesser General Public License, you may choose any version of the GNU Lesser General Public License ever published by the Free Software Foundation.

If the Library as you received it specifies that a proxy can decide whether future versions of the GNU Lesser General Public License shall apply, that proxy's public statement of acceptance of any version is permanent authorization for you to choose that version for the Library.

9.3 GNU Lesser General Public License

GNU LESSER GENERAL PUBLIC LICENSE

Version 3, 29 June 2007

Copyright © 2007 Free Software Foundation, Inc. <<http://fsf.org/>>

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

This version of the GNU Lesser General Public License incorporates the terms and conditions of version 3 of the GNU General Public License, supplemented by the additional permissions listed below. 0. Additional Definitions.

As used herein, "this License" refers to version 3 of the GNU Lesser General Public License, and the "GNU GPL" refers to version 3 of the GNU General Public License.

"The Library" refers to a covered work governed by this License, other than an Application or a Combined Work as defined below.

An "Application" is any work that makes use of an interface provided by the Library, but which is not otherwise based on the Library. Defining a subclass of a class defined by the Library is deemed a mode of using an interface provided by the Library.

A "Combined Work" is a work produced by combining or linking an Application with the Library. The particular version of the Library with which the Combined Work was made is also called the "Linked Version".

The "Minimal Corresponding Source" for a Combined Work means the Corresponding Source for the Combined Work, excluding any source code for portions of the Combined Work that, considered in isolation, are based on the Application, and not on the Linked Version.