String Theory

Rainer F. Hauser
rainer.hauser@gmail.com

May 2, 2019

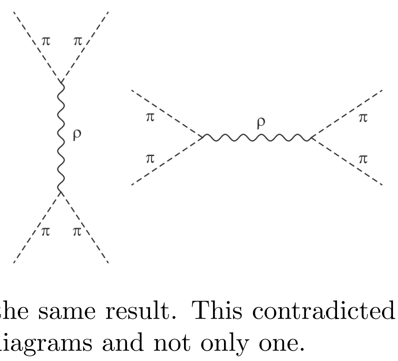
Abstract

String theory is an attempt to unify all forces of nature into a single mathematical framework. Particles are no longer point particles as in the standard model and its extensions, but are strings or D-branes in higher dimensions. Leonard Susskind from the Stanford University, who was himself involved in the development of string theory, held lectures in 2010 covering this topic. His lectures were available on YouTube at the time this transcript has been assembled and may as "String Theory and M-Theory" still be available today.

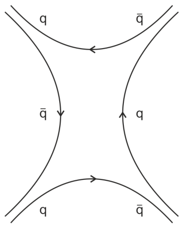
1 Introduction

1.1 The Historical Origins

String theory came initially from hadron theory with protons, neutrons and especially mesons, and it had nothing to do with gravity at that time when it was even not clear whether protons and neutrons consist of quarks and get connected by gluons. What was known around 1970 was that the number of particles states was large. When plotting the squared mass m^2 against angular momentum in so-called Regge plots, there were the nucleons with spin $\frac{1}{2}$, a next particle with spin $\frac{3}{2}$ and so on in steps of fermionic spins which were all on a straight line. This worked fine for all known hadrons with half spin. Also the mesons with integer spin appeared on one straight line. These straight lines are called Regge trajectories. The message that should have been taken from this fact and has been taken later, was that these particles must be composite particles.



Experiments with pion-pion scattering show that two pions can build a ρ-meson which will later decay into two pions again. Because of the principles of quantum field theory, this Feynman diagram can be turned by ninety degrees such that now two pions exchange a ρ-meson. However, when a ρ-meson can participate in these two processes, also the next excited state on the Regge trajectory can do that. There was something very peculiar going on, because when one adds up all the first Feynman diagrams on the left in the figure, one gets the experimentally found result, and when one added all the second Feynman diagrams turned ninety degrees on the right, one also gets the same result. This contradicted quantum field theory, because one should add both sets of Feynman diagrams and not only one.



When drawing the diagram in the figure on the left side, one can slice it symmetrically in two ways through its middle. Cutting horizontally gives the Feynman diagrams on the left side of the above figure and cutting it vertically gives the other Feynman diagram on the right side of the above figure. The next step was an ingredient just added for fun by asking the question what is holding the quarks together. The answer was that it may be something inside the four curves representing the quarks and antiquarks. There is an antiquark on the left side and a quark on the right side and something bridging between them, and this something – a string – is one-dimensional and connects the antiquark and the quark. This was a kind of starting point of string theory. It is not exactly where it came from, because different people thought about this in different ways. One of the consequences of the early string theory was, that if a hadron was a string with a quark and an antiquark attached to it, it can spin and have angular momentum.

The area in the above figure between the four curves of the quarks and antiquarks replaces the world line in the Feynman diagrams and is called a world sheet. This terminology is today standard and seems to have been coined by Professor Susskind himself. Horizontal cuts through the world sheet show the string standing still while inclining cuts correspond to strings in motion. A string standing still means that its center of mass is standing still, because strings wiggle a lot.

1.2 Current View

The present understanding today is that a gluon field is like a Maxwell field. The quarks are like poles of a bar magnet, and there are no monopoles. There are lines of flux spreading out from the two poles, and there may be a reason why the lines of flux build a narrow tube between the two poles. The current understanding of the connection between gluons and these strings goes as follows. Fields in quantum mechanics can be described either as particles or as fields. In the field description, the gluon field between a quark and antiquark would look like the field between two particles with opposite electric charge. In the case of quantum electrodynamics, the lines spread out and the field diminishes as one separates the charges. The non-linearities in quantum chromodynamics causes these lines to attract in a certain way, and the effect of it is that the flux lines form strings. As one pulls the quark and antiquark apart, the lines of flux do not spread out but get longer and longer, and these tubes can be seen as made out of gluons. Stretching a rubber band does not change the number of molecules and it will break at a certain time, but one can imagine a rubber band that whenever a gap appears between two molecules as one stretches it a new molecule is inserted in between. Such a rubber band can be stretched forever without breaking it. That is the nature of the gluon field between a quark and an antiquark. The energy from separating quark and antiquark goes into creating more and more gluons.

Because of these considerations, physicists started exploring the mathematics of strings and how they interact such that the physical laws of interacting hadrons result. Their thoughts took place in a scale of the size of a proton. This is large compared with the scale of quantum gravity which takes place in the order of a Planck length. It is possibly an accidental fact that the mathematics of string theory covers both cases because they are related to completely different length scales. The initial ideas of string theory were promising and are still promising, but things did not come out completely right for reasons which are in hindsight fairly clear. The mathematics was not quite right to study hadrons but was right for gravity. One kept getting particles with zero mass and spin two, but nobody wanted a graviton because researchers were looking for hadrons. Thus, maybe this is a theory of gravity and not of hadrons. Today, there is a string theory for hadrons which does work, but it is a little bit different.

1.3 Relativistic and Non-Relativistic Kinematics

The energy of a particle non-relativistically is $E = P^2/2m$, but one might add a constant term for the binding energy or the energy just because the particle is there. This binding energy does not depend on the motion, but it may vary from particle to particle. The total energy of a system is $E = \sum P_i^2/2m_i + B_i$. Relativistically, this constant energy would naturally be the energy when the particle is standing still, and the total energy of a system is $E = \sum \sqrt{P_i^2c^2 + m_i^2c^4}$. The energy of one particle can be written as

$$E = mc^2\sqrt{1 + \frac{p^2c^2}{m^2c^4}} \approx mc^2 + \frac{P^2c^2}{2m^2c^4}mc^2 = mc^2 + \frac{P^2}{2m}$$

by expanding the square root. Thus, the non-relativistic formula is a good approximation of the relativistic formula, but only when all particles move slowly compared to the speed of light.

There is an approach in which non-relativistic physics is an exact description of relativistic physics. It used to be called the infinite momentum frame and is now called the light-cone frame. Instead of looking at a system in its restframe, one takes a frame in fast motion such that it has a huge momentum along one axis. The energy is $E = \sqrt{P^2 + m^2} = \sqrt{P_x^2 + P_y^2 + P_z^2 + m^2}$ with $c = 1$. If P_z is huge, then P_x, P_y and m are fixed and much smaller than P_z such that one can write

$$E = P_z \sqrt{1 + \frac{P_x^2 + P_y^2 + m^2}{P_z^2}} \approx P_z \left(1 + \frac{P_x^2 + P_y^2 + m^2}{2P_z^2}\right) \quad E = \sum_i P_{z_i} + \sum_i \left(\frac{P_{x_i}^2 + P_{y_i}^2}{2P_{z_i}} + \frac{m_i^2}{2P_{z_i}}\right)$$

for one particle or a system, respectively. The first sum for the system does not change and can be dropped because only differences of energy are important. In quantum mechanics, E is the Hermitian operator $\mathbf{H}$ which is as the partial derivative of t times $i\hbar$ associated with time evolution. If the energy of a system is very small, changes take place very slowly. Thus, the faster a system is, the slower are changes because of time dilatation which rescales time. The first term in the second sum is $P^2/2P_{z_i}$. This looks very similar to the non-relativistic case with $P^2/2m$ except that P is here the momentum in the xy-plane, and the momentum along the z-axis has taken the place of the mass m. Mass is inertia, and the momentum along the z-axis is functioning as a kind of inertia with respect to forces in the perpendicular directions. Relativistically this is true, because a given source perpendicular to the direction of motion produces a smaller acceleration the larger the momentum. The second term in the second sum plays the role of the binding energy because it is independent of the state of motion in the xy-plane or the projection of the motion into the xy-plane, and it can be seen as internal energy.

This is very useful when studying particle dynamics and absolutely central to studying strings. This is the reason why string theory often describes a string non-relativistically. A non-relativistic string is a collection of point particles where the point particles get more and more continuous and where all of them are moving non-relativistically in the plane perpendicular to a relativistic motion of the whole string. As a postulate one can use strings by applying the two-dimensional analogy with non-relativistic physics to explore those strings as if they were conventional non-relativistic objects similar to stretchable rubber bands which can move, which can flap, and which can do everything an ideal rubber band can do. A string can be close or open, and an open string has two ends similar to a cut rubber band. There may or may not be something interesting attached to the ends, but here only the string itself is considered.

2 The Mathematics of Open Strings

2.1 The Energy

A string – here the string is assumed to be an open string – is a collection of point particles, and one takes limits. The mass of each point particle goes to zero, and the number of point particle goes to infinity. The energy is proportional to the kinetic energy and can be written as $E = m \sum (\dot{x}_i^2 + \dot{y}_i^2)/2$ where all points have equal mass. There must be also interactions, and the points are attracting each other because otherwise they would fly apart. Thus, in addition to the points one has to insert little springs between the points to connect them, and a string can be seen as little balls with springs in between. The energy together with these interactions is $E = m \sum \left((\dot{x}_i^2 + \dot{y}_i^2)/2 + k(x_i - x_{i+1})^2/2\right.$ and contains kinetic energy as a first term and Hooke's law as the second term.

To go to the limit, one makes the string denser and denser, the mass smaller and smaller, and the spring constant bigger and bigger. (The spring constant must become bigger to keep the flexibility of the string.) In the end, one gets an integral instead of the sum, and the integral

$$\int_0^\pi d\sigma \left[\frac{\dot{X}^2(\sigma)}{2} + \frac{1}{2}\left(\frac{\partial X}{\partial \sigma}\right)^2\right]$$

goes from $\sigma = 0$ to $\sigma = \pi$, where 0 is one end of the string and π the other end. The scale from 0 to π is obviously arbitrary and could have been chosen differently, but it is useful to pick π for the length of the string because a closed string will then have length 2π. The mass m has been included into the first term which corresponds to the kinetic energy and the spring constant into the second term which corresponds to the potential energy, and X contains the two spatial coordinates x and y in the xy-plane. This is the conventional energy of a vibrating string, and the Lagrangian $\mathcal{L}$ is the integrals with the kinetic energy minus the potential energy – the internal binding energy – as usual.

The string with its point masses and springs is a particle. All the vibrations and other motions of the quarks in a proton, for example, are vibrations and other motions of the strings. The center of mass of the string is interpreted as the position of the particle. The relative stretching and the relative vibrations are internal energy. This internal energy should not be related to the mass m, but to the squared mass m^2 of the entire assembly of the constituents of the string. If these constituents are adding up to something one would call a particle, this particle has a mass squared which is the sum of all the internal energies inside the particle. There is no motion of the string in the xy-plane and the string is therefore at rest in this plane except for vibrations and wiggling of the constituents. There are same factors of the speed of light which has been set to 1.

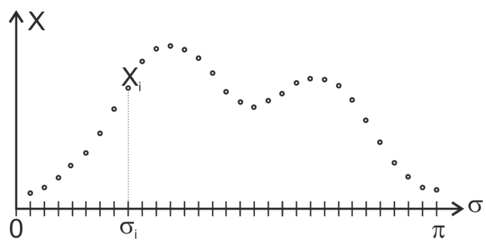
Another fact is that a string is not so different than a spring. If one looks at the spectrum of energies of a string then it is pretty much quantized the same way as a collection of springs. Springs have quantized energies which are – similar to the harmonic oscillator – multiples of something. So far, only classical physics and relativity theory have been used for the calculations, but the total mass squared of the string must be quantized as soon as quantum mechanics is introduced as well.

If one is stretching out the string such that it has length L, then the contribution of $\partial X/\partial\sigma$ is L/π and is proportional to the length L. The contribution of the whole term is therefore proportional to L^2. Because this is mass squared and length squared, by boosting the thing one got from L to L^2 and from m to m^2, and the restmass m is proportional to the length L. In the restframe it does not look like a Hooke's law at all. It looks like a string whose energy is proportional to its length. This fits with the picture of the lines of flux connecting quarks and antiquarks.

One can also imagine long lines of magnetic flux that is uniform along the tubes, and the energy density is therefore uniform along them which makes the energy proportional to the length. Superconductors have the property of repelling the magnetic field and pushing the magnetic field away. If one pushes a magnetic field, the lines of flux are squeezed into a narrow tubelike string which is called a fluxoid or a superconducting flux line, but a magnetic flux line and not an electric flux line. A gedankenexperiment shows how one could – at least theoretically – create such tubes in a superconductor with two magnetic monopoles, which may exist although they have not yet been detected, but could be simulated. Such a string would have uniform energy density, and the total energy would be proportional to the length of the string. In a superconductor, the superconducting condensate is made out of electric charge and it causes confinement of the magnetic charge. In quantum chromodynamics, quarks are confined, and there exists a kind of chromodynamic monopoles.

The flux lines between quarks and antiquarks which build tubes with uniform energy density can be seen as uniform lines of particles, and new particles get created to fill the gaps when stretching the flux lines. There are two kinds of strings. One follows the Hooke's law as these strings in the infinite momentum frame where the energy is proportional to L^2, and the other behaves like flux tubes with energy proportional to the length L such as the tubes between quarks and antiquarks.

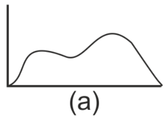
2.2 Some Mathematical Preliminaries



The first topic is repetition of a little bit of calculus. Functions in string theory are usually function $X(\sigma)$ where σ goes from 0 to π. A function – assumed to be smooth – can be approximated in discrete steps, and these steps can be made finer and finer. The discrete points of the function values are X_i where $i \in \{1, ..., N\}$. With $\Delta\sigma = \pi/N$ and $\Delta X = X_i - X_{i-1} \approx \frac{\partial X}{\partial \sigma}\Delta\sigma$, the limit

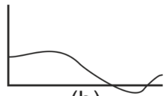
$$\Delta\sigma \sum_i X_i \to \int_0^\pi X(\sigma)\, d\sigma$$

leads to the integral.



If the function is a continuous function from 0 to π with certain boundary conditions, one can apply Fourier analysis and expand it by an infinite series of sine and cosine functions. However, there are additional boundary conditions needed which are called:

(a) Dirichlet boundary conditions where the function is zero at 0 and π
(b) Neumann boundary conditions where the first derivative is zero at 0 and π



(b)

The Dirichlet boundary conditions are used to describe, for example, a violin string whose ends are held fixed. The Neumann boundary conditions with $\partial X/\partial\sigma$ zero at 0 and π are used to describe, for example, organ pipes. Both kinds of functions allow Fourier decomposition, which is

$$X(\sigma) = \sum_{n=1}^{\infty} X_n \sin(n\sigma) \qquad\qquad X(\sigma) = \sum_{n=0}^{\infty} X_n \cos(n\sigma) \tag{2.1}$$

for functions with Dirichlet boundary conditions on the left side shown in figure (a) and Neumann boundary conditions on the right side in figure (b), respectively. The sum on the left side starts from 1 and the one on the right side starts from 0 because $\sin(0) = 0$ but $\cos(0) \neq 0$. The derivative of a function with Dirichlet boundary conditions has Neumann boundary conditions and vice versa. The integral is

$$\int_0^\pi \cos(n\sigma)\cos(m\sigma)d\sigma = \begin{cases} 0 & \text{if } n \neq m \\ \pi/2 & \text{if } n = m \neq 0 \\ \pi & \text{if } n = m = 0 \end{cases}$$

for the case of the Neumann boundary conditions.

The harmonic oscillator classically with the displacement X and units in which the mass m is one has the kinetic energy $\dot{X}^2/2$ and the potential energy $\frac{1}{2}kX^2$ where the spring constant and the frequence are related by $k = \omega^2$ because of the selection of units with $m = 1$. The energy E and the Lagrangian $\mathcal{L}$ are

$$E = \frac{\dot{X}^2}{2} + \frac{\omega^2}{2}X^2 \qquad\qquad \mathcal{L} = \frac{\dot{X}^2}{2} - \frac{\omega^2}{2}X^2$$

classically, and in quantum mechanics, the energy is quantized.

2.3 Difference Between Particles, Strings and Other Objects

Particles and strings are not the same. Strings consist of particles, but particles can also be strings. Particles have a location, but no particle is known to be a point. Even the electron is not a point particle. If one could look at it through a powerful microscope, one would see that the electron has some fuzz around it, and that fuzz would be virtual photons and so on. Certainly the protons are not points but are big, gigantic objects. A particle is not a point particle and can be made out of things. Even the photon is not a point particle because it is also surrounded be virtual photons, and a photon can split in a pair of electron and positron.

A box full of gas with many particles – or a cup of coffee – has a mass and location which is the position of the center of mass. The question is why is it not called a particle. The difference between what physicists call a particle and a highly composite object such as a cup of coffee has to do with the energy spectrum. Energy is equal to mass, and the difference has to do with the mass spectrum. The electron has a unique mass. One can add energy to the cup of coffee by shaking it, and energy is mass. With electrons however, there is nothing one can do to increase the mass of an electron, at least not in the laboratory at present. Today it is not possible to excite an electron into a state of higher mass. The proton has excited states, but they are pretty discretely different than the proton itself. It takes a couple of 100 GeV to spin up a proton or to cause it to oscillate. There is a kind of isolation of energy or mass. For a cup of coffee at zero Kelvin the next excited state is so close that one cannot distinguish it as a separate individual quantum state. There are zillions of excited states near by which are practically creating a continuum of energy levels. This is the difference between a particle and a mush.

The next question is whether a quantum string is a particle or not. This depends on the excitation spectrum of the energy levels above the ground state. If they are well separated for some reason, then it will behave like a particle. If they are extremely close together such that the experiment cannot distinguish them, then it would not behave like a particle. The energy that it takes to perturb a proton is a couple of 100 MeV or about 40% of its own mass, and that is pretty significant. For the kind of strings discussed there, this value is in the order of a Planck mass which is a huge mass from an experimental point of view. There is no hope to produce these excitations in the laboratory. That is why physicists call strings particles.

2.4 Model of a Relativistic Open String

The interesting point of the strings is that it is enough to know their projection onto the xy-plane when boosting the system in the z-direction. Thus, the light-cone frame with

$$E = P_z + \frac{P_x^2 + P_y^2 + m^2}{2P_z^2} \qquad\qquad (E - P_z)P_z = \frac{P_x^2 + P_y^2}{2} + \frac{m^2}{2} \tag{2.2}$$

is useful. The value $E - P_z$ is just the energy, because only energy differences are important, and the factor P_z in the right equation of (2.2) on the left side of the equal sign rescales the time due to time dilatation. Thus, the Hamiltonia $\mathbf{H}$ or the energy function $(E - P_z)P_z$ keeps track of the motion in the xy-plane and the internal motions. This is very non-relativistic, but the term responsible for the internal motion is not proportional to the mass m but to m^2.

The model of a relativistic string looks as in the figure on the right side that is wiggling around, moving, stretching and doing what strings can do in the two dimensions of the xy-plane which is perpendicular to the boosted motion of the the light-cone frame. The string is assumed to be composed of many mass point and springs. If N is the number of mass points, $N - 1$ is the number of springs between them.

The energy of the string composed of these mass points and springs is

$$E = \sum \left(\mu \frac{\dot{X}_i^2 + \dot{Y}_i^2}{2} + \kappa \frac{\Delta X_i^2 + \Delta Y_i^2}{2} \right) \tag{2.3}$$

with the kinetic energy containing $\mu = \frac{1}{N}$ on the left side and the potential energy containing the spring constant $\kappa = \frac{N}{\pi^2}$ on the right side. Going to the limit where μ becomes zero, N becomes infinity, and κ gets bigger and bigger, leads to the integral

$$E = \frac{1}{2\pi} \int_0^\pi \left(\left(\frac{\partial X}{\partial \tau}\right)^2 + \left(\frac{\partial Y}{\partial \tau}\right)^2 + \left(\frac{\partial X}{\partial \sigma}\right)^2 + \left(\frac{\partial Y}{\partial \sigma}\right)^2 \right) d\sigma \tag{2.4}$$

with the conveniently but arbitrarily chosen boundaries 0 and π such that $\Delta\sigma = \frac{\pi}{N}$ and with the proper time τ. The value μ is the non-relativistic analog mass such that its total non-relativistic analog mass is one. Therefore, this is a string with analog mass one. The spring constant of the composite string is smaller than the one of the individual stiff springs. The units have been selected such that further calculations become easier.

The energy E in (2.4) looks like a simple wave field and would satisfy a wave equation. If a wave moves on the string and comes to one of the ends, it will bounce off with either the Dirichlet or the Neumann boundary conditions. With Dirichlet boundary conditions, the wave flips to the other side, and with Neumann boundary conditions, it just gets reflected.

Newton's laws applied to the mass points determine the boundary conditions. A mass point is pulled by the springs from left and right except for the two mass points at the two ends of the string. The last mass point with number N is connected to the mass point with number $N - 1$ through a spring. The force acting on the mass point number N is according to Hooke's law proportional to the distance between the two mass points and this is ΔX in the X-direction. The spring constant is very large and is proportional to N. Thus, the force is in the order of $N\frac{\partial X}{\partial \sigma}\Delta\sigma$. Because $\Delta\sigma = \frac{\pi}{N}$, the factors N cancel and the force is proportional to $\frac{\partial X}{\partial \sigma}$, and this is the amount in which the string is stretched. If $\frac{\partial X}{\partial \sigma} = 0$ then the two mass points number N and $N - 1$ are on top of each other. If it is not zero, there is a gap between the two mass points. According to Newton's laws, the force is $F = \mu\ddot{X}$ where μ is the analog mass of one mass point which is $\frac{1}{N}$. This means that $\ddot{X}$ is proportional to $N\frac{\partial X}{\partial \sigma}$ which goes to infinity when N gets larger and larger. Thus, the Neumann boundary condition is the right boundary condition.

2.5 Quantum Mechanics of Open Strings

The limit of the string modeled as a collection of mass points and springs with the Lagrangian for the whole string and the Neumann boundary condition for the two ends

$$\mathcal{L} = \frac{1}{2\pi}\int_0^\pi \left(\left(\frac{\partial X}{\partial \tau}\right)^2 + \left(\frac{\partial Y}{\partial \tau}\right)^2 - \left(\frac{\partial X}{\partial \sigma}\right)^2 - \left(\frac{\partial Y}{\partial \sigma}\right)^2\right) d\sigma \qquad \frac{\partial X}{\partial \sigma} = \frac{\partial Y}{\partial \sigma} = 0 \tag{2.5}$$

determines the physics of an open string.

Because X and Y are functions of σ going from 0 to π, they can be expanded into series of cosines. This gives

$$X(\sigma,\tau) = \sum_{n=0}^{\infty} X_n(\tau)\cos(n\sigma) \qquad Y(\sigma,\tau) = \sum_{n=0}^{\infty} Y_n(\tau)\cos(n\sigma) \tag{2.6}$$

according to (2.1) where X_n and Y_n are time-dependent. The first term in the Lagrangian becomes

$$\frac{1}{2\pi}\int_0^\pi \left(\frac{\partial X}{\partial \tau}\right)^2 d\sigma = \frac{1}{2\pi}\left(\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \dot{X}_m \dot{X}_n \int_0^\pi \cos(m\sigma)\cos(n\sigma)\, d\sigma\right) = \frac{\dot{X}_0}{2} + \frac{1}{4}\sum_{n=1}^{\infty} \dot{X}_n^2$$

where X_0 is the average position making it the center of mass position. The term of the Lagrangian corresponding to the kinetic energy of the whole string is just the kinetic energy of the center of mass plus the internal wiggling represented by the sum. The second term in the Lagrangian becomes

$$\frac{1}{2\pi}\int_0^\pi \left(\frac{\partial X}{\partial \sigma}\right)^2 d\sigma = \frac{1}{2\pi}\left(\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} mn\, X_m X_n \int_0^\pi \sin(m\sigma)\sin(n\sigma)\, d\sigma\right) = \frac{1}{4}\sum_{n=1}^{\infty} n^2 X_n^2$$

such that the whole Lagrangian from (2.5) becomes

$$\mathcal{L} = \frac{\dot{X}_0}{2} + \frac{1}{4}\sum_{n=1}^{\infty}\Big(\big(\dot{X}_n^2 + \dot{Y}_n^2\big) - n^2\big(X_n^2 + Y_n^2\big)\Big) \tag{2.7}$$

which shows that this is the Lagrangian of two harmonic oscillators, one in X and one in Y, for each n from 1 to ∞. The different harmonic oscillators are not coupled with each other such that this is the sum of an infinite number of independent harmonic oscillators where the frequency of the n-th oscillator in X and Y is n. For $n = 0$, there is no restoring force because the center of mass can move arbitrarily, and for $n \geq 1$, there is kinetic energy and potential energy according to Hooke's law with a restoring force. The sum of kinetic and potential energy

$$\sum_{n=1}^{\infty}\Big(\big(\dot{X}_n^2 + \dot{Y}_n^2\big) + n^2\big(X_n^2 + Y_n^2\big)\Big)$$

in the energy which corresponds to the internal wiggling is what has to be associated with the m^2 in (2.2). The internal energy is the square of the mass in the light-cone frame.

By combining harmonic oscillators in X and Y, harmonic oscillators along different axes can be created. There are different energy level starting from the ground state $|0\rangle$ with no oscillator excited. This is not the vacuum, but a single string with no excitations. It has, as an ordinary harmonic oscillator of quantum mechanics, a gap to the next energy level, and it is therefore a particle.

The energy of a harmonic oscillator is $\hbar$ times the frequency ω times the number of quanta. Thus, the first excited energy level is one times $\hbar$ either in X-direction or in Y-direction. The second energy level above the ground state is two quanta for the first oscillator in X-direction or Y-direction or one in each direction, but one can also excite the second oscillator in X-direction or in Y-direction.

The quantum mechanical Hamiltonian is $\mathbf{H} = \mathbf{P}_n^2 + n^2\mathbf{X}_n^2/4 = (\mathbf{P}_n + in\mathbf{X}_n/2)(\mathbf{P}_n - in\mathbf{X}_n/2)$. The creation and annihilation operators have to be normalized because of $[a^-, a^+] = 1$ for creation and annihilation operators, $[\mathbf{X}_n, \mathbf{P}_n] = i$ for location and momentum, and $[\mathbf{P}_n + in\mathbf{X}_n/2, \mathbf{P}_n - in\mathbf{X}_n/2] = n$ for the two operators building $\mathbf{H}$ as a product all with $\hbar$ set to one. The resulting operators are

$$a_n^- = \frac{\sqrt{n}}{2}\mathbf{X}_n + \frac{i}{\sqrt{n}}\mathbf{P}_n \qquad\qquad a_n^+ = \frac{\sqrt{n}}{2}\mathbf{X}_n - \frac{i}{\sqrt{n}}\mathbf{P}_n$$

in X-direction and similarly for the operators $b_n^\pm$ in Y-direction. Adding a_n^+ and a_n^- as well as b_n^+ and b_n^-, respectively, gives

$$\mathbf{X}_n = \frac{a_n^+ + a_n^-}{\sqrt{n}} \qquad\qquad \mathbf{Y}_n = \frac{b_n^+ + b_n^-}{\sqrt{n}}$$

for the locations. One can write the Fourier expansions using (2.6) therefore in the form

$$\mathbf{X}(\sigma,\tau) = \sum_n \frac{a_n^+ + a_n^-}{\sqrt{n}} \cos(n\sigma) \qquad\qquad \mathbf{Y}(\sigma) = \sum_n \frac{b_n^+ + b_n^-}{\sqrt{n}} \cos(n\sigma) \tag{2.8}$$

as quantum mechanical operators.

2.6 Spin of Particles With and Without Mass

A particle with spin j has $2j + 1$ states, but massless particles are special. A massless particle with spin 0 has only one state as also have particles with mass, but particles with spin 1, spin 3 and so on have all two states instead of three, five and so on. The question is how this can be consistent. Any particle with spin not zero – massless or with mass – has a maximum and a minimum spin in the direction of motion. Right-handed spin in the direction of motion is considered positive and left-handed spin in the direction of motion is considered negative. If one direction of spin exists also the other direction must be there because of reflection symmetry. A particle with mass can be brought to rest with a spin in the former direction of motion. But if it has spin in this direction, it must also have spin in a direction perpendicular to this axis. Thus, a particle with mass and spin 1 must have all three spin states, but a massless particle cannot be brought to rest. Any Lorentz transformation will leave it with the speed of light.

Polarization of light is just either left-handed spin, right-handed spin or linear polarization which is a superposition of the two spin states. The difference between no spin and a superposition of both possible spin states is that there is no spin in the average but it is either spin state when measured. The polarization is always perpendicular to the direction of motion, and this means that it is transverse. The Maxwell equations state that the electric and the magnetic field – and therefore the polarization – are perpendicular to the direction of motion. If the photon travels in the z-direction, linear polarization can be either in the x-direction which can be described as state $|x\rangle$ or in the y-direction which can be described as state $|y\rangle$. Right-handed polarization is then $|r\rangle = |x\rangle + i\,|y\rangle$ and left-handed polarization is $|l\rangle = |x\rangle - i\,|y\rangle$. A circularly polarized photon is a superposition of the two linearly polarized states.

Gravitons are particles with spin 2, and they are also massless. They have therefore only two states. A graviton moving down an axis with maximal angular momentum which is two units of angular momentum is roughly speaking – and this is a good analogy – mathematically the same as two photons each with one unit of angular momentum moving along the same axes. One can imagine to take two photons, to stick them together and to let them move down the z-axis. They can both be circulating to the right and have two positive units of angular momentum, they can both be circulating to the left and also have two negative units of angular momentum, and they can be circulating in different directions. The first two possibilities correspond to the two states of the graviton, but the third possibility is not possible for gravitons. Thus, a particle with spin 2 can be imagined as two particles with spin 1 being in the same state. Gravitons are not two photons, but for their spin, this is a good analogy.

2.7 The Spectrum of Open Strings

If one adds up all energy of a string, one gets the square of the mass and not the mass. An open string at rest and therefore not moving in the perpendicular direction at all is just a collection of harmonic oscillators, and one can determine their states. The ground state $|0\rangle$ is the state for which $|a_n^- |0\rangle = 0$ and $|b_n^- |0\rangle = 0$ for all n, and it is the state where no oscillator is excited. The energy E_{GS} of the ground state is m_0^2. To get to the next energy state, either $a_1^+ |0\rangle$ or $b_1^+ |0\rangle$ or any linear combination of them can be used, and the energy is $m_0^2 + 1$. Using $(a_1^+ \pm ib_1^+) |0\rangle$ gives circular polarization because of the vector character of the particle. There are only the two states $a_1^+ |0\rangle$ and $b_1^+ |0\rangle$ but no third state. That means that there is no third polarization state. Only one conclusion is consistent with Lorentz invariance and this is that the two objects $a_1^+ |0\rangle$ and $b_1^+ |0\rangle$ must be massless. They must be like the massless photon. The problem is that $m_0^2 + 1$ must therefore be zero and the ground state $|0\rangle$ would get imaginary mass. An object with negative mass squared is called a tachyon and moves faster than the speed of light, but this is not the end of string theory because there are many string theories without tachyons.

For ordinary non-relativistic particles the relations $E(P) = P^2/2m$ and $\partial E/\partial P = P/m = v$ hold. This is actually a property of waves with its group velocity because E can be replaced by ω and P by k. Relativistically, these relations are $E = Pc$ for a massless particle, and, with c set to one, $E = \sqrt{P^2 + m^2}$ for particles with mass. For the massless particle, the relation $\partial E/\partial P = c$ is obviously the velocity, and for the particle with mass, $\partial E/\partial P = P/\sqrt{P^2 + m^2}$ is less than 1 which states that particles with mass move with a speed less than the speed of light. If the square of the mass is negative, the particle moves faster than the speed of light. That gave the tachyons their name.

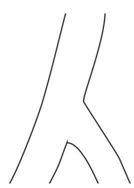
What a tachyon really means is not that signals can be faster than the speed of light but that the vacuum is unstable and that something can happen that destabilizes the system. With the speed of light set to one, a simple wave equation $\partial^2\Phi/\partial x^2 = \partial^2\Phi/\partial t^2$ which is equivalent to $\omega^2 = k^2$ or $\omega = k$ has waves as solution moving with the speed of light. Adding a mass term giving $-m^2\Phi + \partial^2\Phi/\partial x^2 = \partial^2\Phi/\partial t^2$ leads to the energy

$$\frac{1}{2}\left(\frac{\partial\Phi}{\partial t}\right)^2 + \frac{1}{2}\left(\frac{\partial\Phi}{\partial x}\right)^2 + \frac{m^2}{2}\Phi^2$$

of the wave. Increasing Φ would increase the last term with the mass, and therefore would increase the energy. The field would start to oscillate when one displaces it a little bit. It costs energy to increase Φ, and it would start to swing back. If m^2 is negative, this would turn the potential upside down and create an instability of the vacuum $\Phi = 0$. If the field is exactly balanced everywhere and it starts falling off in one point, this would spread out, and the domino effect would propagate but not with a speed faster than the speed of light. Thus, tachyons do not fly faster than the speed of light.

String theory has so far given just an instable vacuum. That is where superstring theory comes into the game. The simple string has found a massless particle with two polarization states. String theory has found a way to get rid of the tachyons, but no string theory has been able so far to get rid of these photonlike objects. At the time when string theory started, it was not supposed to be a theory for photons but for hadrons, and there are no massless mesons. Thus, these photonlike objects and these tachyons were seen as unwanted. In all these years of string theory, it was easy to get rid of the tachyons, but it was not possible to get rid of these photonlike objects.

2.8 Interactions Between Strings



So far, only free strings have been considered which wiggle and vibrate but did not interact with other strings. The only process apart from wiggling and doing its own thing – and everything is built out of it – is that two endpoints meet and join with a certain probability. This joining interaction can only take place when two endpoints of strings moving in spacetime touch as in the figure on the right side. This process corresponds to the Feynman diagrams of quantum field theory. Because anything that can happen in quantum mechanics can also unhappen through time reversal, there must be an amplitude for a string to break into two strings. This probability is called the string coupling constant and corresponds to the fine-structure constant which is the square of the electric charge and similar coupling constants. The probability that two endpoints fuse is the same as the probability that a string breaks.

Because a string does not know its endpoints, it can happen that the two endpoints of a string touch and fuse. This way a closed string can happen. Thus, it is inevitable for a theory with open strings that interact to also have closed strings. There are theories with closed strings that interact and fuse, but build again a closed string when fusing and never open strings. In other words, it is well possible to have theories with closed strings and no open strings, but it is not possible to have theories with open strings but no closed strings. It turns out that the lowest excitation of an open string is a photon and the lowest excitation of a closed string is a graviton. There are string theories without electromagnetism, but there is no string theory without gravity. There are many string theories, but the one thing they have all in common is the closed strings and the graviton.

Another interaction of two open strings is that they touch somewhere in the middle and rearrange in such a way that one part of one string and one part of the other string fuse and the remaining two parts fuse as well. Similarly an open and a closed string can fuse as well by touching somewhere. Therefore every open and closed string can absorb a closed string. This is – according to string theory – the reason why everything gravitates. Strings cannot only fuse and break but can also pass through each other. Because the probability per unit time is the same for all strings, the probability is smaller to fuse if two strings move faster than it is if they are moving slower.

3 The Mathematics of Closed Strings

3.1 Noether's Theorem

For every symmetry there is a conserved quantity, and in quantum mechanics, the conserved quantity becomes the generator of the symmetry. There is a Lagrangian $\mathcal{L}(q, \dot{q})$ which depends on some coordinates q and their time derivatives $\dot{q}$. The canonical momentum is $P_i = \partial\mathcal{L}/\partial\dot{p}_i$.

If there is a symmetry which involves a transformation on the qs, an infinitesimal symmetry which just shifts a little bit, then the variation is written as δq_i, and it might be something like $\delta q_i = f_i(q)\varepsilon$. If one makes a little change on one q_i, it may depend on all the other qs times a small ε. An example is rotation is space where all coordinates may change, and the change is proportional to the little angle. The preserved quantity – angular momentum in the example of the rotation – is

$$Q = \sum_i P_i f_i(q) \tag{3.1}$$

and is called the Noether charge or, in quantum mechanics, the generator of the transformation. This theorem is useful because there are all sorts of symmetries in string theory.

3.2 Model of a Relativistic Closed String

Some point of the closed string is labeled $\sigma = 0$. Half-way around the string is $\sigma = \pi$, and the length of the whole string is 2π. The point with $\sigma = 0$ and the direction in which σ increases have to be chosen once for all. A wave moving on the string can move to the left or to the right meaning in direction of decreasing or increasing σ. As a convention, a wave moving to the right goes into the direction of increasing σ. The coordinates $X(\sigma, \tau)$ and $Y(\sigma, \tau)$ describe the position of the point with σ at the proper time τ. For closed strings, there are no real boundary conditions, but there is something one could call a boundary condition and that is $X(2\pi, \tau) = X(0, \tau)$ and $Y(2\pi, \tau) = Y(0, \tau)$.

A wave moving on an open string gets reflected at the endpoints and changes therefore from left-moving to right-moving and vice versa. On a closed string, left-moving and right-moving waves stay what they are and continue to move into the same direction when they reached $\sigma = 2\pi$.

It is convenient to describe waves not by sines and cosines but by exponentials $e^{in\sigma} = \cos(n\sigma) + i\sin(n\sigma)$. One can also decompose them into Fourier series not with Dirichlet or Neumann boundary conditions but periodic

$$\begin{aligned} X(\sigma, \tau) &= \sum_{n\in\mathbb{Z}} X_n(\tau)e^{in\sigma} = \sum_{n>0} X_n(\tau)e^{in\sigma} + \sum_{n>0} X_{-n}(\tau)e^{in\sigma} + X_0(\tau) \\ Y(\sigma, \tau) &= \sum_{n\in\mathbb{Z}} Y_n(\tau)e^{in\sigma} = \sum_{n>0} Y_n(\tau)e^{in\sigma} + \sum_{n>0} Y_{-n}(\tau)e^{in\sigma} + Y_0(\tau) \end{aligned} \tag{3.2}$$

where $n > 0$ is a right-moving wave, $n < 0$ a left-moving wave. The coordinates for the motion of the center of mass are (X_0, Y_0), and the boundary conditions $X(2\pi, \tau) = X(0, \tau)$ and $Y(2\pi, \tau) = Y(0, \tau)$ are automatically fulfilled. Many calculations and equations for the closed strings are very similar to those for the open strings. The integrals go from 0 to 2π and so on, and the Lagrangian, for example, becomes

$$\mathcal{L} = \int_0^{2\pi} d\sigma \left(\left(\frac{\partial X}{\partial \tau} \right)^2 + \left(\frac{\partial Y}{\partial \tau} \right)^2 \right) - \left(\left(\frac{\partial X}{\partial \sigma} \right)^2 + \left(\frac{\partial Y}{\partial \sigma} \right)^2 \right)$$

such that it can be decomposed into a Fourier series similar to the open strings.

The energy is the same integral but with a plus sign, and the terms with X in the integral can alternatively be written as

$$\left(\frac{\partial X}{\partial \tau} \right)^2 + \left(\frac{\partial X}{\partial \sigma} \right)^2 = \frac{1}{2} \left(\frac{\partial X}{\partial \tau} + \frac{\partial X}{\partial \sigma} \right)^2 + \frac{1}{2} \left(\frac{\partial X}{\partial \tau} - \frac{\partial X}{\partial \sigma} \right)^2$$

which is useful because a wave moving in one direction is a function of $\sigma + \tau$ and a wave moving in the other direction a function of $\sigma - \tau$. Thus, the energy in this form is split up into the energy of the left-moving and the right-moving waves. Left-moving and right-moving waves are transparent to each other because they go right through each other.

Also here the X_n and Y_n are harmonic oscillators, but for each n there are now four oscillators because one going to the left and one going to the right for X and Y. They correspond to X_n, X_{-n}, Y_n, and Y_{-n}, and their creation operators are a_n^+, a_{-n}^+, b_n^+, and b_{-n}^+. As in the case of the open strings, the frequency of these four harmonic oscillators is n.

3.3 The Spectrum of Closed Strings

The ground state $|0\rangle$ has some energy m_0^2 as in the case of the open string. There are the four possibilities $a_1^+ |0\rangle$, $a_{-1}^+ |0\rangle$, $b_1^+ |0\rangle$, and $b_{-1}^+ |0\rangle$ in order to act on the ground state to get the first excitation. This is not yet the right picture as will be seen. There is no candidate for the angular momentum 0, and this must therefore correspond to massless particles as shown for the open strings. The theory so far has produced a doubling of the photon spectrum.

There are rules which forbid certain combinations, and the rule here is called level matching. It states that the right-moving energy and the left-moving energy must be the same. Because $a_1^+ |0\rangle$ has one unit of right-moving energy and $a_{-1}^+ |0\rangle$ has one unit of left-moving energy, both states do not satisfy the level matching rule, but going to the next level, $a_1^+ a_{-1}^+ |0\rangle$ satisfies it. Obviously, also $b_1^+ b_{-1}^+ |0\rangle$ satisfies it, but $a_1^+ b_{-1}^+ |0\rangle$ and $a_{-1}^+ b_1^+ |0\rangle$ do as well. These four states are the only states with two units of energy that fulfill the level matching rule.

The states $(a_1^+ + ib_1^+)(a_{-1}^+ + ib_{-1}^+) |0\rangle$ and $(a_1^+ - ib_1^+)(a_{-1}^+ - ib_{-1}^+) |0\rangle$ with angular momentum 2 and angular momentum -2, respectively, plus the two states $(a_1^+ + ib_1^+)(a_{-1}^+ - ib_{-1}^+) |0\rangle$ and $(a_1^+ - ib_1^+)(a_{-1}^+ + ib_{-1}^+) |0\rangle$ both with angular momentum 0 are linear combinations of the previous four states fulfilling the level matching rule and build another basis. This cannot be a particle with mass and spin 2 because there are no states with spin ± 1, but must be massless. The interpretation is that the first two states with angular momentum ± 2 represent gravitons, and the other two states correspond to the two massless particles dilaton and axion with spin 0 which have the common feature that they have never been discovered. There are possibilities to get rid of the dilaton and the axion without getting rid of the graviton.

To understand the level matching rule, one question to answer is whether the point $\sigma = 0$ on the closed string is really a physically special point. Another way of phrasing the question is whether the state a closed string is invariant under changing the origin of the σ-coordinate. Assuming the string consists of N point masses with coordinates $P_i = (X_i, Y_i)$, the wave function would be $\Psi(P_1, P_2, P_3 ..., P_N)$. If P_1 is not special, two wave functions $\Psi(P_1, P_2, P_3, ..., P_N)$ and $\Psi(P_2, P_3, ..., P_N, P_1)$ with different start point should be the same. In other words, the wave function should have a certain symmetry. This is certainly a reasonable property to ask for. Going from the discrete to the continuous string, the condition for the wave function becomes $\Psi(X(\sigma)) = \Psi(X(\sigma + \varepsilon))$ where all points on the string are relabeled by adding ε or, in other words, shifted by the same amount. Rewriting this equation to $0 = \Psi(X(\sigma + \varepsilon)) - \Psi(X(\sigma))$

leads to

$$\int \frac{\partial \Psi}{\partial X(\sigma)} \frac{\partial X}{\partial \sigma} \, d\sigma = 0$$

because it is a change for all values of σ and the ε is just a constant not influencing the fact that the integral is zero. Whenever a wave function in quantum mechanics is differentiated by a coordinate, this can be rewritten as the action of the corresponding momentum. The first factor in the integral becomes $P(\sigma)$ which is the velocity of the corresponding point such that

$$\int P(\sigma) \frac{\partial X}{\partial \sigma} \, d\sigma = \int \dot{X}(\sigma) \frac{\partial X}{\partial \sigma} \, d\sigma = \int \frac{\partial X}{\partial \tau} \frac{\partial X}{\partial \sigma} \, d\sigma = 0$$

becomes the condition for the fact that there is not a preferred point on the σ-axis. Because the energy has been found to be the sum of the left-moving and the right-moving energy

$$\int d\sigma \left(\frac{1}{2} \left(\frac{\partial X}{\partial \tau} + \frac{\partial X}{\partial \sigma} \right)^2 + \frac{1}{2} \left(\frac{\partial X}{\partial \tau} - \frac{\partial X}{\partial \sigma} \right)^2 \right)$$

above, the difference of the two energies must therefore be zero.

Thus, the fact that there is no distinguished point on a closed string gets rid of many states and leaves the photon as the lowest state above the ground state in the spectrum of the open strings and the dilaton, the axion and the graviton as the lowest state above the ground state in the spectrum of the closed strings.

3.4 Strings as the Most Fundamental Objects

The question whether the strings are the most fundamental objects in physics is, as physicists have learned, not a good question. As an example, one can look at the question whether the electron or the monopole is the more fundamental object. Assuming there are monopoles in quantum electrodynamics, it is easy to formulate quantum electrodynamics such that there are monopoles in it. The electric charge e times the monopole charge q has to fulfill the equation $eq = 2\pi$ in order for the Dirac string (the solenoid) which is connected to the monopole to be invisible. This is the condition that if you have a monopole connected to a long string – this is the only way to make a monopole mathematically – that charged particles which go around that string do not detect phase shifts. That means if the electric charge is very small one can do quantum electrodynamics with Feynman diagrams and so on. If the electric charge is very large, Feynman diagrams are not very useful because each Feynman diagram contains some vertices and each vertex has an e^2 in the probability such that the Feynman diagrams get bigger and bigger and do not converge. Thus, Feynman diagrams are a tool to study theories with small charges.

For $eq = 2\pi$, the magnetic charge q must be very big if e is very small. Thus, if there are magnetic charges, and the Maxwell equations are rather symmetric with respect to the electric and the magnetic field, the question is whether the electric or the magnetic charge is more fundamental. Because of the divergence problems of the Feynman diagrams for large charges, it makes sense to take the electric charge as more fundamental. The magnetic charge would emit photons with high probability and therefore be surrounded by many photons which would interact very strongly with pairs of electric charges. This would turn the magnetic charge into a very complex thing with all kinds of internal structure and would make it useless as a starting point for Feynman diagrams.

One could gradually make e bigger and q smaller. Whether one is more fundamental than the other depends on whether it is useful to think one way or the other. Whether one way or the other is more useful may depend on the parameter of the theory. There is no ultimate answer to the question whether the electric or the magnetic charge is more fundamental.

3.5 Discussion of Units

In a string theory there is only one parameter. A non-relativistic string stretched to length L behaves pretty much like a spring. It has a potential energy $E = kL^2/2$, which has been identified above with m^2. This means that the mass in the rest frame is proportional to the length L und the square root of the spring constant k. With the $T = \sqrt{k}$ called the linear tension in the string, and with units such that $\hbar = 1$ and $c = 1$, the unit of energy is one over the unit of length, and the unit of the tension is energy squared or one over length squared. Each time one excites a state by one unit, this adds essentially that tension to m^2. The thing which determines the units of the theory is this tension T. The energy jump between ground state and first excited state or first excited state and next excited state and so forth is controlled by this tension T. The string tension is energy per unit length, and energy per unit length is force because force times distance is work. Thus, tension is the force pulling back when one stretches the string. In other words, if one attaches the string on one side somewhere at the surface of the Earth, the tension is the weight that the string can support. Because energy E and force F satisfy the equations $E = T \cdot L$ and $F = E/L$, the force $F = T$ that the string can support is independent of the length. That is the character of these strings.

One can ask how much weight a meson, for example, could support, and the answer is in the order of magnitude of a truck. Thus, they are pretty strong. Of course, nobody has ever tested whether a meson can support a truck, but what one knows is how much energy it takes to excite a string, and from that one can figure out what the tension is. The strings corresponding to photons and gravitons could support the weight of a whole galaxy because their tensions are very large and the springs are extremely stiff such that it is very difficult to excite them. The energy to excite a string is somewhere near the Planck energy which is around 10^{-5} g times the speed of light squared.

In physics the three units mass, length and time are needed. The units kilogram, meter and second have more to do with the biology of the humans than with physics. Using units optimal for physics would make many formulas much simpler, but to make all of physics simpler and not just one field, the units must correspond to universal quantities. Three fundamental quantities that can be set to one are the speed of light c which is the same in every reference frame, the Planck constant $\hbar$ which appears in the uncertainty principle, and the gravitational constant G which makes the gravitational attraction independent of the kind of matter. They are all universal in the sense that they apply to everything. They can be set to one to make physics simpler.

The three values can be combined to get the units of mass, length and time. As an example, there are integers p, q, r such that $G^p \hbar^q c^r$ gives units of length squared. The unit of c is length over time, the unit of $\hbar$ is length squard times mass over time, and the unit of G is length cubed over mass times time squared. Solving the equations gives $q = p$, $r = -3p$ and $p = 1$ such that length squared has the same units as G times $\hbar$ over c^3. The value $1\,\mathrm{l_P} = \sqrt{\hbar G/c^3} \approx 1.6 \cdot 10^{-35}\,\mathrm{m}$ is called the Planck length. The Planck time is the time for a light ray to move a Planck length and is $1\,\mathrm{t_P} = \sqrt{\hbar G/c^5} \approx 5.4 \cdot 10^{-44}\,\mathrm{s}$. The Planck mass finally is $1\,\mathrm{m_P} = \sqrt{\hbar c/G} \approx 2.2 \cdot 10^{-8}\,\mathrm{kg}$. The observable universe has a radius of about $10^{60}\,\mathrm{l_P}$, and the age of the universe is also about $10^{60}\,\mathrm{t_P}$. The energy content of a tank of gasoline is about $1\,\mathrm{m_P}$.

String theory is in Planck units. The size of a vibrating string and the fluctuations due to the uncertainty are in the order of the Planck length, the period of the oscillations are in the order of the Planck time, and the mass needed to excite a gravition are expected to be in the range of the Planck mass. A Planck mass is in the order of 10^{19} proton masses, and that is a huge amount of energy. The difficulty is not to get a Planck mass of energy but to pump it into such a small volume needed to excite a graviton. Nobody believes that the length of a string is smaller than an Planck length, but it could be larger. Strings smaller than the Planck length would mean a mass larger than the Planck mass, and small things with a big mass are black holes. Predictions in this scale, however, are beyond direct verification. Experiments indicate that the ordinary quantum field theory probably holds to something like a distance scale of roughly a thousand times larger than the Planck scale.

3.6 Some Difficulties in String Theory

In order for the photons to come out right as massless when discussing open strings, the mass squared of the ground state had oddly to be minus one unit. The problem is the zero point energy of the harmonic oscillator which is $\frac{1}{2}\hbar\omega$. With $\hbar = 1$ and the frequency of the nth harmonic oscillator with frequency $\omega = n$, the ground state energy is $\frac{n}{2}$ in some units. Because there are oscillators for every integer n and one has to add up all this energy, it is not easy to imagine how this addition (the sum over all integers times some factor) can result in -1 with a mathematically convincing argument.

Going back to a system moving very fast down the z-axis with the overall momentum P which is a conserved quantity shows that the energy is $E = P + m^2/2P$. The momentum P is to be taken as the infinite limit, and $E-P$ in $2P(E-P) = m^2$ is a perfectly conserved quantity which never changes. It is not necessarily true that all this energy has to add up to minus one, because one can write $2PE = m^2 + 2P^2$, and the sum gets -1 plus the infinite term $2P^2$ which can be absorbed into something else. This is a trick of physics which takes place all the time. One gets some infinite answer and does not know what to do with it, but one realizes that the quantity that becomes infinite has already a constant piece in it such that one can add the extra infinity to the constant piece. In this case, the term $2P^2$ is already there, and the infinity of $\frac{1}{2}\sum n$ can just be added to it such that one gets -1 plus an infinity which partially was already there. As long as the infinite term is constant, it never effects anything.

The mathematics of the trick here is to add $1e^{-1\varepsilon} + 2e^{-2\varepsilon} + 3e^{-3\varepsilon} + ...$ instead of $1 + 2 + 3 + ...$ which is the same for $\varepsilon = 0$. If ε is a small number, the factor $e^{-n\varepsilon}$ overpowers the growing factor n in each term of the infinite sum such that the series converges. The goal is to find the infinite sum and let ε got to 0. To do so, the sum is turned into a converging geometric series

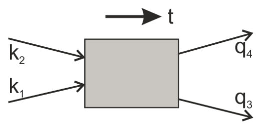
$$\begin{aligned}\sum_{n=1}^{\infty} n e^{-n\varepsilon} &= -\frac{\partial}{\partial\varepsilon}\sum_{n=1}^{\infty} e^{-n\varepsilon} = -\frac{\partial}{\partial\varepsilon}\sum_{n=0}^{\infty} e^{-\varepsilon}e^{-n\varepsilon} = -\frac{\partial}{\partial\varepsilon}\frac{e^{-\varepsilon}}{1-e^{-\varepsilon}} = -\frac{\partial}{\partial\varepsilon}\frac{1-\varepsilon+\frac{\varepsilon^2}{2}+...}{\varepsilon-\frac{\varepsilon^2}{2}-\frac{\varepsilon^3}{6}+...} \\ &\approx -\frac{\partial}{\partial\varepsilon}\frac{1}{\varepsilon}\frac{1-\varepsilon+\frac{\varepsilon^2}{2}}{1-\frac{\varepsilon}{2}+\frac{\varepsilon^2}{6}} \approx -\frac{\partial}{\partial\varepsilon}\frac{1}{\varepsilon}\left(1-\varepsilon+\frac{\varepsilon^2}{2}\right)\left(1+\frac{\varepsilon}{2}-\frac{\varepsilon^2}{6}+\frac{\varepsilon}{4}\right) \\ &\approx -\frac{\partial}{\partial\varepsilon}\frac{1}{\varepsilon}\left(1+\frac{\varepsilon}{2}+\frac{\varepsilon^2}{12}-\varepsilon-\frac{\varepsilon^2}{2}+\frac{\varepsilon^2}{2}\right) = -\frac{\partial}{\partial\varepsilon}\frac{1}{\varepsilon}\left(1-\frac{\varepsilon}{2}+\frac{\varepsilon^2}{12}\right) = \frac{1}{\varepsilon^2}-\frac{1}{12}\end{aligned}$$

using the Taylor expansion of the exponential function and $\frac{1}{1-\epsilon} = 1 + \epsilon + \epsilon^2 + \epsilon^3 + ...$ up to ε^2. Here, the term $\frac{1}{\varepsilon^2}$ gets absorbed into the infinite P^2 term which is an additive constant in the energy and can be ignored. The real answer is that in properly designed string theories the term $\frac{1}{\varepsilon^2}$ is not there. Thus, the answer is $-\frac{1}{12}$ – or actually $-\frac{1}{24}$ because of the factor $\frac{1}{2}$ in front of $\sum n$ –, but it is not -1. Because of the oscillators in X and Y, there is twice as much energy in the ground state. Thus, this quantity $-\frac{1}{12}$ is the energy of the ground state including the zero-point oscillations but throwing away a certain infinite term which has to be explained away later.

It is still not -1, and the solution is to say that there are not two dimensions but there are twenty-four dimensions plus the axis on which the system is boosted plus time. In total, there are therefore twenty-six dimensions. At the time around 1969, physicists were just exploring string theory, and this was certainly not a convincing argument at that time. For the closed strings, the ground state must have the energy -2. If the same theory has to have closed and open strings, the number of dimensions cannot be changed again and this is also not needed because closed strings have left-moving and right-moving oscillations doubling the energy.

4 Scattering of Strings

4.1 Scattering Experiments

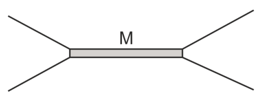


Particle physics always is about scattering because it is basically all one can do experimentally. Particles come in, something happens inside the box, and particles go out – not necessarily the same number of particles. The particles coming in carry momentum, spin and so on. Everything except momentum $(E, P_x, P_y, P_z) = k_\mu$ for the relativistic particles is ignored here, and the relation between energy, momentum and mass is $E^2 = P^2 + m^2$ or $P^2 - E^2 = -m^2$ with c set to one. This can also be written as $\vec{k}^2 - k_0^2 = k_\mu k^\mu = k^2 = -m^2$ in terms of the components of k. To distinguish incoming and outgoing four-vectors, the incoming are called k and the outgoing q. Energy and momentum conservation is $k_1 + k_2 = q_3 + q_4$ in the case of the figure on the right side.

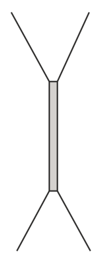
To simplify the picture, physicists prefer to write $q_3 = -k_3$ and similarly for the other outgoing particles. Energy and momentum conservation becomes now $k_1 + k_2 + k_3 + k_4 = 0$ and takes a completely symmetric form where the constraint $k^2 = -m^2$ is still satisfied. The probabilities and therefore the amplitude A in the scattering experiment is a function of the ks. There is some redundant information in $A(k_1, k_2, k_3, k_4)$ because of the $k^2 = -m^2$ and the $k_1 + k_2 + k_3 + k_4 = 0$. The sixteen variables represented by the four ks of the example in the figure can be reduced to two by going to a frame of reference where the center of mass is at rest, by rotating the frame such that the incoming particles move on the x-axis, by using the conservation laws for energy and momenta and by using the other constraints above. Only the energy E_{cm} of the center of mass and the angle θ of the outgoing particles relative to the x-axis remain as independent variables.

To create invariants from four-vectors, one can square them, but one can also add two and square the result to get $(k_1 + k_2)^2 = (\vec{k}_1 + \vec{k}_2)^2 - (k_{01} + k_{02})^2$. In the center of mass frame, $\vec{k}_1 = -\vec{k}_2$ with equal magnitude and opposite directions such that $(\vec{k}_1 + \vec{k}_2)^2 = 0$, and $k_{01} = k_{02}$ such that $(k_1 + k_2)^2 = -(2k_0)^2$ which is the square of the total center of mass energy and is called $-s$. Thus, $s = E_{cm}$, and this is the same as $(k_3 + k_4)^2$. All the particles have the same energy. They come in and go out with the same energy get scattered by an angle, and that is all that happens. For $(k_1 + k_3)^2$ which is the same as $(k_1 - q_3)^2$, $k_{01} + k_{03} = 0$ because $k_{01} = q_{03}$, and the incoming momentum $\vec{k}_1$ gets transferred to the outgoing momentum $\vec{q}_3$. The quantity $(\vec{k}_1 - \vec{q}_3)^2$ is called the momentum transfer which can be expressed in terms of the scattering angle as $(k_1 + k_2)^2 = 2(E^2 - m^2)(1 - \cos\theta)$. This angle is what is measured in an experiment. Thus, $s = E_{cm}$ and $t = (E^2 - m^2)(1 - \cos\theta)$ are the two important quantities. A third quantity is $-u = (k_1 + k_4)^2 = (E^2 - m^2)(1 + \cos\theta)$. The three quantities $-s = (k_1 + k_2)^2 = (k_3 + k_4)^2$, $-t = (k_1 + k_3)^2$ and $-u = (k_1 + k_4)^2$ are called Mandelstam variables, but are not independent, because there are only the two independent variables E_{cm} and θ. The interesting quantities are s and t.

4.2 The Veneziano Amplitude



If two particles collide, create a third particle with a different mass M, and then this particles decays again as in the figure on the right side, this is a Feynman diagram which has a value. This value is the product of two coupling constants g and is the propagator $1/(s - M^2)$ in between giving together $g^2/(s - M^2)$. That is the characteristic structure of a scattering amplitude, and it only depends on s, but not on t and u.



If, however, the two particles exchange the third particle as in the figure on the right side, where time also runs from left to right, the value of this Feynman diagram is $g^2/(t - M^2)$. There is also a Feynman diagram with leads to an equation with u, but it is not shown here. The equation with t contains the angle of scattering θ, but the one with s does not. In the process with the propagator $1/(s - M^2)$ every angle is equally probable. That seems odd on the first view, but the reason is very simple. The particles come in and form this compound state (the third particle), and when this compound state decays, they have forgotten from which direction they have come in. The Feynman diagram with t, however, depends on the angle θ and favors small angles.

This does not describe the scattering of mesons very well because there are many particles that can be produced when two mesons collide. The result of a collision of two mesons can have different mass and different angular momentum. The angular momentum changes the dependence of the angle θ. Thus, the formulas shown here are too simple to describe the scattering of mesons.

A more sophisticated formula than the simple addition of the s- and t-channel $g^2/(s - M^2) + g^2/(t - M^2)$ has been found by the young physicist Veneziano. It is

$$g^2 \frac{\Gamma(-s)\Gamma(-t)}{\Gamma(-s - t)}$$

and uses the gamma function $\Gamma(x)$ which is defined for integers as the factorial $\Gamma(n) = (n - 1)!$ but is generalized to real numbers. The formula is also symmetric with respect to s and t. It has all the important features for a scattering amplitude to have, it could be analyzed as if a whole bunch of particles were produced and then decayed, but it could also be analyzed for a whole bunch of particles being exchanged. This was new because before everybody would have added the contributions of the s-channel and the t-channel.

The question was what kind of physics gives rise to this formula. The answer turned out to be string theory. One starts with two open strings in spacetime which may with some probability coalesce into one string when the two ends touch. This string stays for a while together and may with some probability split into two strings again. If the first string has momentum k_1 and the second k_2, and if Ψ_0 is the wave function of the ground state, the wave functions to be multiplied

$$\Psi(X_1, ..., X_N) = e^{k_1 \frac{X_1 + ... + X_N}{N}} \Psi_0(X_1, ..., X_N) \quad \Psi(X_{N+1}, ..., X_{2N}) = e^{k_2 \frac{X_{N+1} + ... + X_{2N}}{N}} \Psi_0(X_{N+1}, ..., X_{2N})$$

describe the two strings without going to the continuous limit. With some probability the two endpoints merge when $X_N = X_{N+1}$. If they did, a new wave function describes the newly formed string. One uses the Hamiltonian to let it propagate in time by multiplying the initial state with $e^{i\mathbf{H}\tau}$. One lets it evolve and break up again to finally form two strings. To summarize, one starts with the two particles, constrains them to be on the same place, lets things evolve and projects them on the final state. This gives the amplitude for the two strings to coalesce for a time τ. This calculation is doable, and even not hard. The result is an integral summing according to Feynman over all possible paths. The integral is

$$\int_0^\infty e^{-\tau(-s-1)}(1 - e^{-\tau})^{-t-1} e^{-\tau} \, d\tau$$

using the Mandelstam variables. Defining $z = e^{-\tau}$ changes the integral to

$$\int_0^1 z^{-(s+1)}(1 - z)^{-(t+1)} \, dz = \beta(-s, -t)$$

which is completely symmetric with respect to s and t. This function is called the Euler beta-function, and it is equal to the Veneziano amplitude. The mathematics for closed strings is very similar.

This symmetry between the s- and the t-channel has the property that not only processes where particles coalesce are described but also somewhere buried in it also processes where particles exchange other particles. That was a surprise.

4.3 Spacetime View of String Theory

The trajectory of a particle can be written as $x(t)$ in classical mechanics, and as $X^\mu(\tau)$ in relativistic physics. In both cases, the motion of a particle is defined by an action − a principle of least action − and the action for a free particle is $S = \int dt \, \frac{1}{2} m \dot{x}$ in classical physics or $S = \int d\tau \left(\frac{\partial X^\mu}{\partial \tau} \frac{\partial X_\mu}{\partial \tau} \right)$ in the relativistic case. In quantum mechanics, one does something completely different with the action. One does not ask what the trajectory is given to the endpoints, but asks what the amplitude is to find the particle at a given point $x(t_2)$ if it has been at the beginning of the time interval at a given point $x(t_1)$. The amplitude is

$$A = \int e^{i \frac{1}{2} \int_{t_1}^{t_2} dt \left(\frac{dx}{dt} \right)^2}$$

where the outer integral goes over all paths and is therefore a monstrous object. For relativistic quantum mechanics, the inner integral is replaced by the corresponding action for relativistic motion. In relativistic physics, the amplitude is called the propagator for the particle to go from the startpoint to the endpoint. One integrates or sums over all Feynman diagrams.

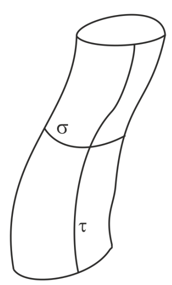
There is a cheating trick of physicists which is actually justified. The term $e^{i\cdots}$ in the amplitude has magnitude 1 for all paths, and they cancel each other in many cases. In the relativistic case, the variable τ is replaced by $\tau = \alpha s$ where α is a number chosen later. The integral becomes

$$\int_{s_1}^{s_2} \frac{\alpha}{\alpha^2} \, ds \left(\frac{\partial X}{\partial s} \right)^2$$

where one of the factors α cancel. With $\alpha = -i$ the amplitude magically looses the i, and the integral becomes

$$e^{-\frac{1}{2} \int \left(\frac{\partial X}{\partial s} \right)^2 ds}$$

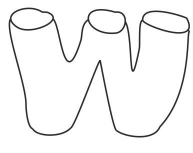
and can be treated as a real number although the endpoints of the integral are not real numbers. Later, one can remember that the endpoints are imaginary numbers. Integrating this expression over all possible paths is now much easier. For paths wiggling strongly or going far off and come back, the term $(\partial X/\partial s)^2$ gets big and therefore the action gets big. Any trajectory far from the simplest trajectory has a huge action with a negative sign. Because e to the minus a large number is a small number, these far off trajectories do not contribute much. Only the trajectories near the simplest one are important. This form of the integral over the paths is called Wiener integral and has been invented in a mathematically rigorous way before the Feynman path integration. In the end, the integral gives a function $F(s_1, s_2)$, where s_1 and s_2 are the imaginary startpoint and the endpoint. After evaluating this function, the value has to be analytically continued (or extrapolated) to imaginary values of s. This is known how it has to be done, and this is the only real way physicists do path integrals.



In string theory, the procedure is similar, but the particles are no longer points but strings such that world lines become world sheets for open strings and world tubes for closed strings. The position $X^\mu(\tau, \sigma)$ on the world tube as in the figure on the right side is given by the two parameters τ and σ where τ is a kind of time variable and could be proper time, but is here just a parameter similar to σ. Also in this calculation one creates an action with an integral over all possible tube surfaces

$$S = \int d\tau\, d\sigma \left(\frac{\partial X^\mu}{\partial \tau}\right)^2 - \left(\frac{\partial X^\mu}{\partial \sigma}\right)^2 \qquad \int e^{i \int d\tau\, d\sigma \left(\frac{\partial X^\mu}{\partial \tau}\right)^2 - \left(\frac{\partial X^\mu}{\partial \sigma}\right)^2}$$

which is the amplitude for the given startpoint and endpoint.



One can generalize the idea to two strings going to two strings or two strings going to three strings as an example in the figure on the right side shows. The amplitude is similarly the integral over all possible surfaces connecting the initial states with the final states. These surfaces can have holes in it. Because of the $e^{i\cdots}$ again each path contributes something with magnitude one, and the solution is again to analytically continue. The problem is not σ but τ. With replacing τ by $\pm i\tau$ whatever fits better, the amplitude becomes

$$\int e^{-\int \left(\frac{\partial X^\mu}{\partial \tau}\right)^2 + \left(\frac{\partial X^\mu}{\partial \sigma}\right)^2 d\tau\, d\sigma} \tag{4.1}$$

where the term inside the inner integral in the exponent is always positive. The wilder the surface is, the bigger is $(\partial X/\partial \tau)^2 + (\partial X/\partial \sigma)^2$. If it wiggles a lot, the term $(\partial X/\partial \sigma)^2$ is large, and if it vibrates very much, the term $(\partial X/\partial \tau)^2$ is large. If it stretches out far away, both will be large. (This action is called Polyakov action although Professor Susskind has invented it. There is also a second, more complicated action called Nambu–Goto action.)

Because the parameters σ and τ can be chosen in various ways, the transformation properties of the sum $(\partial X/\partial \tau)^2 + (\partial X/\partial \sigma)^2$ with its invariants are of interest here. Instead of looking at the action S on the left side, the equation of motion on the right side is used

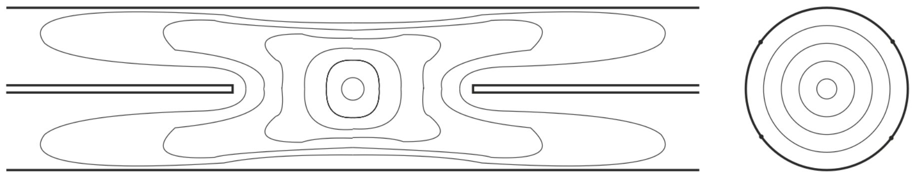
$$S = \int d\tau\, d\sigma \left(\frac{\partial X^\mu}{\partial \tau}\right)^2 - \left(\frac{\partial X^\mu}{\partial \sigma}\right)^2 \qquad \frac{\partial^2 X}{\partial \tau^2} + \frac{\partial^2 X}{\partial \sigma^2} = 0$$

which is called the Laplace equation. Whether one works with the action S or the equation of motion does not matter, because both give the same result. If $X(\tau_1)$, $X(\tau_2)$, $X(\tau_3)$ are neighboring points with distance ε, then the first derivative is approximately $X(\tau_3) - X(\tau_2)$ divided by ε and the second derivative $[X(\tau_3) - X(\tau_2)] - [X(\tau_2) - X(\tau_1)] = X(\tau_3) + X(\tau_1) - 2X(\tau_2)$ divided by ε^2. Similarly in two dimensions, the Laplace equation states that $X(P_1) + X(P_2) + X(P_3) + X(P_3) - 4X(P_0) = 0$ for the values at the four neighbor points P_1, P_2, P_3, P_4 around P_0. In other words, the value at a point is the average of its neighbors around it. (The same is true in higher dimensions.) Because the Laplace equation is invariant under rotation, one can take any infinitesimal square, and the value at the center is the average of the values of the four corners.

4.4 The Laplace Equation

The Laplace equation – and therefore the above action – is invariant under any transformation $\sigma'(\sigma, \tau)$, $\tau'(\sigma, \tau)$ that transforms every infinitesimal square into an infinitesimal square, and these transformations are called conformal mappings. Mappings which preserve angles between curves transform sufficiently small squares into squares. However, any shape with a closed boundary can be mapped by some conformal mapping to any other shape also with a closed boundary, and a large square can turn into anything else, because only infinitesimal squares are mapped to infinitesimal squares.

Going back to the two strings that join for some time and split again, this doubly-slit strip is topologically equivalent to a filled circle or disk as shown in the figure below with the four special points representing the endpoints of the strip at infinity, and this double-slit strip can be conformally mapped to a disk. It is still a Laplace equation. Thus, one can select parameters σ and τ in such a way that all values of X are on a unit disk, and the integration over all paths can be done on the unit disk with the four special points treated consistently.



In order to calculate the scattering amplitude, one has to do something with these four infinite points of the original doubly-slit strip. To set up and specify the initial and final states of the particles, their momenta, the particle types and so on, these information has to go somewhere, because there is more than the fact that there is a Laplace equation here. The points on the circle are a one-parameter family, and the time where the two strings are joined is also a one-parameter family. This time governs the place for the four points. The beta-functions comes from the four points which are called injection points. The limit where the two points left and the two points right are close together corresponds to the situation where the two strings are joined for a long time interval. The limit where the two upper points and the two lower points are close together corresponds to the situation where the two strings are joined only for a short period of time.

The two momenta k_1 and k_2 of the two strings on the left side give $s = (k_1 + k_2)^2 = E_{cm}^2$ and corresponds to one of the Mandelstam variables as well as to the square of the center of mass energy. This is the limit where the two points on the left side and the two points on the right side are close together. The one momentum k_1 on the left side and the one momentum k_3 on the right side of the doubly-slit strip give $t = (k_1 + k_3)^2$ and correspond to another Mandelstam variable. Thus, in this form it is clear that there is a symmetry between the s-channel and the t-channel.

As discussed above, this shows again that this process describes the situation where two particles collide and form another particle which later decays into the initial particles, but also describes the situation where two particles exchange a third particle. This is the duality of the s-channel and the t-channel.

4.5 Conformal Mappings and Analytic Functions

The Coulomb force in two dimensions is one over r and not one over r^2 as in three dimensions, and the potential energy between a pair of charged particles is logarithm, because the derivative of logarithm is one over r. In a world, where the lines of flux cannot escape into a third dimension, electrostatic problems would have two-dimensional versions of the electric field $\nabla\Phi = E$ and the charge density $\rho = \nabla E = \nabla^2\Phi$. This means that

$$\frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2} = 0$$

if there is no charge density. This equation is invariant under any change of coordinates which are angle-preserving or, in mathematical terminology, are conformal mappings. There is also a curl in two dimensions, but it has only one component which is

$$\frac{\partial E_x}{\partial y} - \frac{\partial E_y}{\partial x}$$

for $E = (E_x, E_y)$. The Laplace equation does not only appear in two-dimensional electrostatics, but also in fluids and other two-dimensional problems in physics.

Conformal mappings are mappings from the two-dimensional plane to the two-dimensional plane and can be written as mappings $\mathbb{Z} \to \mathbb{Z}$ with $z = x + iy$. They allows to derive from solutions of, for example, an electrostatic problem in a space with some boundary conditions an infinitesimal number of other solutions in other spaces with other boundary conditions by mapping the solution to the other space.

A conformal mapping can be seen as a coordinate transformation from $z = x + iy$ to $w = u + iv$ in the complex plane, and it is assumed to be a one-to-one mapping. Thus, any point in the plane can either be described by coordinates (x, y) or by coordinates (u, v), and $w(z)$ is a complex function. If one moves the point z to $z + \Delta z$, the point w moves to $w + \Delta w$. The derivative is the limit of $\Delta w/\Delta z$ for Δz going to zero. There is only one problem because one can approach the point z from any direction. It is not clear whether the result of this limit is not dependent on the direction from which one is coming even if the function is nice, continuous and with all kinds of good continuity properties. Thus, the question is what the conditions are that the derivative is well defined and independent of the direction.

The necessary condition for a well-defined derivative is that coming in from the x-axis or coming in from the y-axis gives the same result. Thus, the derivative must fulfill

$$\frac{dw}{dz} = \frac{du + idv}{dx + idy} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = \frac{1}{i}\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y} \qquad \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \tag{4.2}$$

and functions with this property are called analytic functions. Without proof, this is not only a necessary condition but also a sufficient condition. Because

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 v}{\partial y \partial x} \qquad \frac{\partial^2 u}{\partial y^2} = -\frac{\partial^2 v}{\partial x \partial y} \qquad \frac{\partial^2 v}{\partial y^2} = \frac{\partial^2 u}{\partial x \partial y} \qquad -\frac{\partial^2 v}{\partial x^2} = \frac{\partial^2 u}{\partial y \partial x}$$

the conditions for an analytic function can be written as

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \qquad \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0$$

in the form of the two-dimensional Laplace equations, but these conditions are not sufficient. Only if the two real functions u and v are linked together by the equations (4.2), called Cauchy-Riemann equations, the complex function $w(z)$ is analytic. Therefore, two real functions fulfilling the Laplace equations do not always give an analytic function.

A complex number can be represented by the real and the imaginary part but also in polar coordinates as $z = re^{i\vartheta}$. The ratio of two complex numbers z_1 and z_2 is $z_1/z_2 = (r_1/r_2)\, e^{i(\vartheta_1 - \vartheta_2)}$. Thus, the angles in a ratio get subtracted. Assuming that δz_1 and δz_2 are two small displacements of z in different directions with the corresponding displacements δw_1 and δw_2 of $w = w(z)$, and assuming that $w(z)$ is an analytic function, the derivatives $\delta w_1/\delta z_1$ and $\delta w_2/\delta z_2$ must be equal. This equation can be written as

$$\frac{\delta z_2}{\delta z_1} = \frac{\delta w_2}{\delta w_1}$$

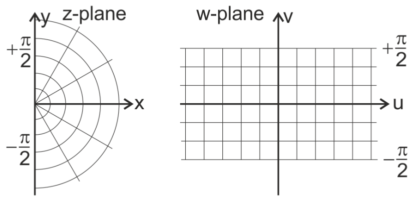
with the consequence that the angles between δz_1 and δz_2 on one side and between δw_1 and δw_2 on the other side are equal. In other words, the mapping is a conformal mapping. Because the Laplace equations do not change under conformal mappings, the analytic functions are very important in string theory.

Taking $w = z^2$ as an example, the equation $x^2 - y^2 + 2ixy = u + iv$ must hold. Thus, one gets $u = x^2 - y^2$ and $v = 2xy$. The Laplace equations are satisfied. Taking $w = z^*$ as another example, the equations $u = x$ and $v = -y$ cannot fulfill both Cauchy-Riemann equations, and the complex conjugate is not an analytic function. The third example $w = e^z$ gives $u = e^x \cdot \cos y$ and $v = e^x \cdot \sin y$. The two Cauchy-Riemann equations are satisfied. Because $w = \log(z)$ is the inverse function of $w = e^z$, and if a function preserves angles also the inverse function preserves angles, also the logarithm must be an analytic function.

4.6 Analytic Functions Mapping to Strings

The function $w = \log(z)$ is an important mapping in string theory. It is defined on the right half of the complex plane, and $w = \log(z) = \log(r) + i\vartheta$ with $z = re^{i\vartheta}$. In this form, ϑ goes from $-\frac{\pi}{2}$ to $+\frac{\pi}{2}$, and the lines with $\vartheta = \pm\frac{\pi}{2}$ in the z-plane are mapped to two horizontal lines.

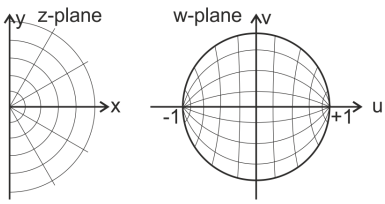
Straight lines leaving the origin in the z-plane become horizontal lines in the w-plane, and half-circles with the center in the origin of the z-plane become vertical lines in the w-plane as shown in the figure on the right side. In other words, the upper half of the y-axis in the z-plane corresponding to $+\frac{\pi}{2}$ marks the maximal v-value in the w-plane and similar for the lower half of the y-axis in the z-plane.



The w-plane has been used to model a string where σ goes vertically from $-\frac{\pi}{2}$ to $+\frac{\pi}{2}$ and τ goes horizontally from $-\infty$ to $+\infty$. If one shifts everything in the w-plane to the left, everything in the z-plane shrinks by a uniform factor, and if on shifts everything in the w-plane to the right, everything in the z-plane grows by a uniform factor. Thus, these so-called dilations which expand and contract in the z-plane transform into time-translations in the w-plane. The string world-sheet could be described as living on the half-plane with the string coming in at the origin.

If one maps not only the half-plane z but the whole plane, then this adds 2π horizontally to the w-plane, but the result is a cylinder because the horizontal line $v = \frac{3\pi}{2}$ is the same line as the horizontal line $v = -\frac{\pi}{2}$. Open strings can therefore be represented by a half-plane and closed strings as the whole plane.

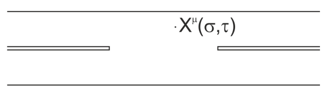
Another analytic function is $w = (z+1)/(z-1)$ where $w(0) = -1$, $w(\infty) = +1$, and $w(iy) = -(1+iy)/(1-iy)$ with $|w(iy)| = 1$ such that the right half-plane in the z-plane is mapped to the unit disk in the w-plane as shown in the figure on the right side. One can see that straight lines through the origin in the z-plane get mapped to parts of circles going through the points ± 1 in the w-plane, and the half-circles in the z-plane go to parts of circles as well but starting and ending on the unit circle in the w-plane. If not only the right half-plane of the z-plane is mapped but also the left half-plane, then the right half-plane goes to the inside of the disk, and the left half-plane goes to the outside of the disk. In general, straight lines and circles get transformed into straight lines and circles under linear fractional mappings such as $w = (z+1)/(z-1)$.



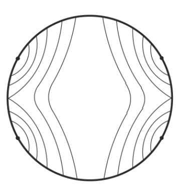
These mappings are useful when dealing with Laplace equations. Solving two-dimensional electrostatic problems, for example, can be done in the mapped geometry instead of the original geometry, and one can go back and forth between geometries.

4.7 String Scattering Concepts

The two strings joining for some time and the splitting again as in the figure on the right side consists of points where X^μ are their real location in spacetime, and σ and τ are their coordinates on the world-sheet. To get the scattering amplitude, one calculates (4.1). One question is how this integral over all possible path knows about the momenta k_1 and k_2. The answer is shown later. Another question is what the time interval is where the two strings are joined. The answer is that one integrates over all possible histories.



Here one can make use of analytic functions to find a more convenient conformal mapping. The two-slit string can be seen as having one single boundary with some infinite points on it. It can therefore be mapped to many other things with one boundary by a conformal mapping. It can especially be mapped to a disk as shown above using the combination of the logarithm and the linear fractional function. It is a bit more complicated because of the two slits, but can be done and looks like the figure on the right side. The lines in the figure correspond to vertical lines in the two-slit string with the four infinite points marked. The curves close to the four points correspond to vertical lines in the areas where the two strings have not yet joined or have already split. Where two of these curves touch corresponds to the point where the two string either join or split. The location of the four points on the boundary of the disk determine where the two strings join and split.



On this disk, it is much easier to calculate the integral (4.1) for the scattering amplitude. For each of the four points which represents an external particle, a factor is added which injects the momentum. This gives a product $\prod e^{ikX(z)}$ with a factor for each external particle where k is the momentum and z is the position of the corresponding point on the boundary of the disk. The path integral is

$$\int dz \int e^{-\text{Action}} e^{ik_1 X(z_1)} e^{ik_2 X(z_2)} e^{ik_3 X(z_3)} e^{ik_4 X(z_4)}$$

on the disk. To be precise, one should write $k_\mu X^\mu(z)$ instead of simply $kX(z)$. This is the whole set of rules for the theory of open strings. The amazing thing is that these integrals can be calculated. That is the good news. The bad news is that one almost always gets infinity, but only when one is in the wrong number of dimensions.

The answer in the right number of dimensions is quite simple. It is the solution of an electrostatics problem. One takes all the X^μ for every μ and thinks of each of them as an electrostatic potential. Each one is a separate electrostatics problem. There is a world of twenty-six different kinds of electric charge, twenty-six kinds of electric fields, and twenty-six kinds of electrostatic potentials if there are twenty-six dimensions. The electric charges are the components of the momenta, and each one is the source of its own electric field called X. There are not simply twenty-six charges of one kind but twenty-six different independent kinds of charges, each creating its own electrostatic field.

One takes a two-dimensional disk, puts the charges on the boundary and on each of these points on the boundary (corresponding to an external particle) one puts twenty-six different charges. The twenty-six components of k become the twenty-six charges. One can now calculate the electrostatic potential inside the disk, and from that one can calculate the field and therefore the energy. Calculating the electrostatic energy of such a configuration and integrating over the points corresponding to external particles which are moved around gives the scattering amplitude. There are twenty-six electrostatic energies which are added up. The coordinates satisfy the Laplace equation, and the Laplace equation is the basic equation for electrostatics. One does not have to call it electrostatics, but the mathematics is the same.

The same way one can calculate the scattering amplitude for more than just two incoming and two outgoing particles. Because conformal mappings allow to chose the position of three external particles on the boundary of the disk, there are only $n-3$ integrations needed over the position of the remaining external particles if there are n external particles. This is open string theory where the external particles are injected at the boundary of the disk.

5 Strings in Compact Dimensions

5.1 Strings and the Curvature of Space

Point particles can move in curved spaces, they move on a trajectory, and they satisfy an equation of motion which is a kind of Newton's equation for curved spaces. If there are no forces, the particle moves on a geodesic. With or without forces, finding the trajectory means solving a differential equation, and solving differential equations means breaking the trajectory in little pieces and go to the limit where these pieces get zero length. The question is how one knows whether the limit exists. Differential equations have been studied for a long time, and there is a lot of experience on how to solve them.

In quantum mechanics the question is a bit different. Given a starting point and an endpoint, one would like to know the probability to find the particle at the endpoint. There is also an action, but one calculates now the path integral. This creates also partial differential equations to be solved, and they are much harder, but there is also a lot of experience on how to solve them.

A point particle moving on a sphere with radius R has the kinetic energy $\frac{1}{2}mv^2 = P^2/2m$, but there is now a constraint that the particle has to stay on the sphere. The velocity vector and the momentum vector are now in the surface of the sphere, but otherwise it is just classical mechanics. Going to quantum mechanics, the momentum is quantized, because the angular momentum $L = PR$ is quantized. Sticking $P = L/R$ into $P^2/2m$ gives $L^2/2mR^2$ for the kinetic energy. With the moment of inertia $I = mR^2$, the kinetic energy becomes $L^2/2I$.

A string moving in a curved space, for example on a sphere with radius R, is also approximated by a finite number of point particles and going to the limit of infinitely many particles. The problem is that this limit does not exist in curved space. Because the string vibrates also in the ground state, it has a certain size. With equations (2.8), the average square size of an open string in the ground state at any point σ is $\langle 0|X^2|0\rangle = \langle 0|\sum_{nm}(a_n^+ + a_n^-)(a_m^+ + a_m^-)/\sqrt{nm}\cos(n\sigma)\cos(m\sigma)|0\rangle$. The term $a_n^+ a_m^+$ when acting to the right has two extra units of energy and that has no overlap with the ground state on the left at all. Also the terms $a_n^- a_m^-$ and $a_n^+ a_m^-$ gives nothing. The only term contributing is $a_n^- a_m^+$ if $n = m$. Thus, the average square size is $\langle 0|X^2|0\rangle = \langle 0|\sum_n (a_n^+ a_n^-)/n \cos^2(n\sigma)|0\rangle$. The $\cos^2$ term is always positive and its average is about 0.5, the average of $a_n^+ a_n^-$ is just one, and therefore $\langle X^2\rangle = \frac{1}{2}\sum_n \frac{1}{n}$. The $X(\sigma)$ in (2.8) is not completely correct, because the center of mass motion has been left out. If the points of the string are not at the center of mass, it means that the string is spread out. What has been calculated here is how much it is spread out. This is a problem, because this sum is infinite.

Interestingly this gives no trouble in flat space. One can terminate the infinite sum at n_{max} and later let the value of n_{max} go to infinity. As a function of n_{max}, the string occupies an area with radius r where $r^2 \sim \log(n_{max})$. In flat space this has an energy of $\frac{1}{2}mP^2$, and it does not matter that it is spread out. One does not have to remember that it is an extended object. This is not true in curved space. If the string is at the north pole, for example, and one starts with the state where there are no oscillating modes, the string would be a point and would move as an ordinary point particle with $L^2/2I$. It would move on great circles of the sphere and the motion would be controlled by the radius of the sphere because $I = mR^2$. When adding more modes of oscillation, the string starts to spread out, gets a size and grows. It fills up some region, and the center of mass moves on the same great circle. The moment of inertia about an axis is proportional to a sum or integral over the mass distribution and it is governed by the square of the distance to the axis. On the sphere, not all points have the same distance to the axis, and the moment of inertia gets smaller for points closer to the axis. In total, one can say that the string behaves as if it is a point mass that moves on a smaller sphere. If one adds even more modes of oscillation, the area covered by the string gets larger and larger, and the moment of inertia gets smaller and smaller. As n gets larger and larger, the string covers larger and larger parts of the sphere and the average moment of inertia gets closer and closer to zero. This is an effect of the curvatures. As one adds more and more modes of oscillation, the moment of inertia gets smaller and smaller, and this is unavoidable in all dimensions, and that is not a good theory.

The geometry with a metric tensor $g_{\mu\nu}(X)$ used for physics of strings should not change when increasing n_{max}. One can do the above calculation for an arbitrary geometry, and the change $\delta g_{\mu\nu}(X)$ in the geometry must depend on the curvature tensor $R^{\delta}_{\alpha\beta\gamma}$ in the form of the Ricci tensor $R_{\mu\nu} = R^{\alpha}_{\mu\nu\alpha}$, because in a flat space there is no change. Thus,

$$\delta g_{\mu\nu}(X) = -R_{\mu\nu} \tag{5.1}$$

where the minus sign comes from the fact that the sphere gets smaller when one adds more structure. This is the equation which shows how the geometry changes as one adds more and more fluctuations to the string. Mathematicians call this equation Ricci flow. It is a fuzzing out of the geometry because points have gotten fuzzed out by the fluctuations of the strings.

As one is interested in geometries which are stable and do not change when adding structure, and that means geometries with $R_{\mu\nu} = 0$. This condition – called Ricci flat – is obviously true for flat spaces, but there are other geometries which fulfill this condition. Ricci flat spaces allow strings to propagate sensibly on that geometry. In spacetime, this equation is the Einstein vacuum field equation. The Einstein field equation contains the Einstein tensor $G_{\mu\nu}$ for the gravity and is

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = T_{\mu\nu} \tag{5.2}$$

where $R = R^{\alpha}_{\alpha}$. The tensor $T_{\mu\nu}$ is the energy-momentum tensor containing everything except gravity. The Einstein field equation for $T_{\mu\nu} = 0$ is called the Einstein vacuum field equation which is similar to the Maxwell equations without charges and has gravitational waves as its solutions. There are simple gravitational waves and interacting gravitational waves which can do all kinds of complicated things. It is a whole infinite-parameter family of complicated solutions of these equations.

Setting $T_{\mu\nu} = 0$ is not yet the condition (5.1) for a Ricci flat space. The quantity R^{α}_{α} is the trace of the Ricci tensor. Writing the Einstein vacuum field equation in the form $R^{\nu}_{\mu} - \frac{1}{2}g^{\nu}_{\mu}R^{\alpha}_{\alpha} = 0$ where g^{ν}_{μ} is the Kronecker delta δ^{ν}_{μ}, the sum $R^{\alpha}_{\alpha} - \frac{1}{2}\delta^{\alpha}_{\alpha}R^{\alpha}_{\alpha} = 0$ becomes $R^{\alpha}_{\alpha} = 2R^{\alpha}_{\alpha}$ because $\delta^{\alpha}_{\alpha} = 4$. This means that $R^\alpha_\alpha = 0$, and the Einstein vacuum field equation is $R_{\mu\nu} = 0$. Thus, the conclusion is that the only geometries in which strings make sense are the solutions of Einsteins vacuum field equations. This is quite impressive. Out of the consistency condition for strings to have a well-defined geometry in the limit when the number of degrees of freedom becomes infinite, one finds Einsteins field equations.

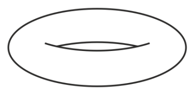
This is not the whole answer to the question of the consistency of string theory. The real answer is that the space must be Ricci flat and must have the right number of dimensions. The number of dimensions is ten for the superstring theory and twenty-six for ordinary string theory.

5.2 Compactification

Compactification is the process of getting rid of dimensions. One does not really get rid of them but one makes them small enough such that they become invisible except for very small things. These dimensions are rolled up into little manifolds. They have an effect on particles but cannot be detected with a coarse apparatus. In superstring theory there are ten dimensions of spacetime. There is one time and nine space dimensions, and there are therefore six additional space dimensions which one rolls up. (To imagine more than one time dimension is rather difficult and this idea is not pursued here.)

Particles in a one-dimensional world can only move in one direction, and they have a neighbor to the right and a neighbor to the left, but cannot pass each other. If one zooms in, there may be smaller things than these particles, and the one-dimensional world may turn out not to be a line but a cylinder such that these smaller things can move in two perpendicular directions. This is the most elementary kind of compactification.

Now assuming the world to be two-dimensional with an additional third compact dimension. This third direction can be a little bit of space between a two-dimensional floor and a parallel two-dimensional ceiling such that there is some thickness between floor and ceiling. Having edges like this makes troubles for strings because the fluctuate may get bigger than the distance between floor and ceiling. If, however, the floor and the ceiling are identified such that one can go from floor to ceiling and ends up at the floor again, then there is no edge anymore and the third dimension has the topology of a circle. This is an example of compactification of one out of three dimensions.



Compactification of two dimensions with avoiding edges identifies points on the floor with points on the ceiling and points on the back with points on the front. Similarly, any number of spatial dimensions can be compactified. This procedure is called toroidal compactification because a torus is the simplest way for compactifying dimensions. A torus in two dimensions can be cut one way to become a cylinder, and the cylinder can be cut again to become a rectangle. Thus, a rectangle is topologically the same as a torus if on identifies the left and right side to be the same as well as the top and bottom side to be the same. Depending on the angle respectively on the ratio of x- and y-direction, a particle moving on a straight line will either end at some time again on the same track if the ratio is rational or move for ever on different tracks if the ratio is irrational. There are geometries other than tori for compactifying two dimensions. A sphere, for example, can also be used, but spheres turned out to be not a good choice for compactification of two dimensions. In higher dimensions, tori also exist and can compactify more than two dimensions.

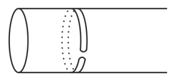
A two-dimensional torus can be described by three numbers called moduli. One can either take the width and the height or the area and the aspect ratio for the first two. Because two tori where width and height have been exchanged are geometrically the same, the aspect ratio can only vary from 1 to ∞ or 0 to 1, but not from 0 to ∞. The area and the aspect ratio are enough to specify a torus built from a rectangle, but a cylinder can be twisted before the round sides are glued together. This corresponds to a torus build from a oblique-angled parallelogram. The third modulus is the angle of the parallelogram.

All these tori have a flat geometry or – correctly stated – can be given a flat geometry. This is not true for a donut because it is an object in three dimensions and has positive or negative curvature depending on the area. A torus as a two-dimensional topology can be given a flat geometry and therefore also a Ricci flat geometry such that string theory is well-defined on a torus. It is by far the easiest way to hide dimensions, but it is not useful for the real world which is more complex.

There are other Ricci flat manifolds where one can compactify the six spatial dimensions of superstring theory. Especially, there are the Calabi–Yau manifolds which have very special properties where string theory is a good theory. They have enough complexity and lack of symmetry such that they look more like the real world. However, they are mathematically to difficult for these lectures.

For simplicity, the one-dimensional world with one curled second dimension is used here. A particle moving in this world can have a component in the ordinary dimension and a component in the compactified dimension. The momentum has therefore a component along the large dimension and a component in the curled dimension, but this second component is quantized because momenta on periodic spaces are always quantized. If $2\pi r$ is the circumference of this compactified dimension and P_c is the momentum in this direction, then $P_c r$ is like angular momentum which is quantized as $P_c r = n\hbar$ such that $P_c = n\hbar/r$ is also quantized. If the particle is massless then $E = c|P_c|$, and the energy is also quantized. If the particle is only moving in this curled direction, the particle seems to stand still but has energy. This energy is now the mass, and it is quantized. The smaller r is, the bigger is the spacing between the quantized mass levels because the mass is proportional to n/r. Theoretically, one could measure the spacing between mass levels and determine the radius r of the curled dimension.

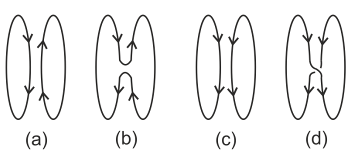
One can have a different kind of particle in the compact dimension which looks from the outside – the big dimension – just like a particle, but is wound around the compact dimension as a closed string. It can move up and down the big dimension. Its mass does not come from a motion around this compact dimension but from the energy due to the stretching. This energy is proportional to the length of the string and therefore to the radius r or the circumference of the curled dimension. Because the string can be wound several times around this curled dimension, the energy is proportional to nr where n is called winding number. If the string has an orientation, it can be wound positively or negatively around the curled direction such that the winding number can be positive or negative. This second kind of strings have mass not proportional to $1/r$ as the first kind of strings above but proportional to r. If both kinds are available, one has mass levels with big spacing for small values of r and one has small spacing for the same value of r. Thus, there are two different kinds of spectra for these different kinds of particles, one moving around the curled dimension called Kaluza–Klein particles and one wound around the curled dimension. The first kind has been proposed before the string theory was invented, and the second has been proposed by physicists working on string theory. Both complementary kinds of particles together would allow to determine properties of the compact dimension.



If there are compact directions in space, winding numbers cannot be avoided. A closed string can wind around the compact torus as in the figure on the right side. This takes energy for stretching the string, and one can collide Kaluza-Klein particles hard enough such that various things can happen including this kind of stretching. There is still no winding number, but the two ends may join now and build two closed strings with opposite winding numbers. They are independent and one can move far away such that one string remains with winding number one. This is similar to electric charge. The net electric charge might be zero in the whole universe, and the net winding number may also be zero in the whole universe, but that does not mean that there are no electrons.

5.3 T-Duality

In the following it is assumed that the compact directions form a torus and that strings are closed and oriented. If a string splits into two strings, both strings must have the same orientation. Orientation is similar to electric charge. One can compare two strings and determine whether they have the same or different orientation. One cannot say whether two strings with the same orientation have positive or negative orientation, but one can say whether they have the same or different orientation. Thus, if one replaces all strings with strings of opposite orientation, physics would be the same.



Strings can exist in two different ways on a compact dimension. Either they are positioned on the torus or they are wound around the torus. They can be wound in two direction, and one is said to have positive winding number and the other is said to have negative winding number. Two wound strings can join by going from (a) to (b) in the figure on the right side or split by going from (b) to (a) in the figure if they have opposite orientation, and they can join by going from (c) to (d) in the figure or split by going from (d) to (c) in the figure if they have the same orientation. The total winding number is preserved in both cases, because closed strings can only change their winding number in one of these ways. Wound strings have a certain tension which is energy per unit length. With the winding number w, the energy of a wound string is proportional to wr. Strings which are not wound have a quantized energy due to the component of momentum in the direction of the compact dimension which is proportional to n/r for an integer n.

If one tries to measure r, the radius of compactification, the difficulty is that one cannot distinguish between particles with momentum and particles with winding number given the energy level. Thus, there is an ambiguity, and it looks like a kind of symmetry where r is replaced by $1/r$ and momentum by winding number for closed strings. This symmetry of spectrum only exists for compact directions, of course, but it is a fundamental property of the theory of closed strings that momentum and winding number can be interchanged and that r cannot be made arbitrarily small because at a certain point momentum and winding number are interchanged. Remarkably, the entire theory with scattering amplitudes and the spectrum of particles is symmetric under this interchange. This symmetry is called T-duality where T stands for torus. Theories where momentum and winding number are interchanged behave exactly the same way.

A string moving consists of all its points, and they can have different velocities $dy/d\tau$ for the compact direction y. A string which is wound in the simplest way completely in the y-direction can be described by $\sigma = y/r$. In this case, $\partial\sigma/\partial y = 1/r$. Thus, the numbers n and w can be expressed as

$$n = r \int \frac{\partial y}{\partial \tau} \qquad\qquad w = \frac{1}{r} \int \frac{\partial y}{\partial \sigma} \tag{5.3}$$

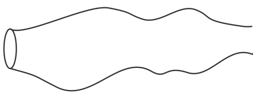
showing that T-duality means replacing n by w, r by $1/r$ and $\partial y/\partial\tau$ by $\partial y/\partial\sigma$. All observable quantities do not change if these three quantities are interchanged. This is an exact symmetry of string theory.

5.4 Closed Strings and Similarity to Electric Charge

The orientation is very similar to electric charge. Two strings with the same orientation will repel, and two strings with different orientation will attract. The attraction and repulsion comes from the electromagnetic field which is strongly linked to photons. In the case of orientation of strings, the attraction and repulsion comes from the gravitation, but gravitation in the compact dimensions and not the big dimensions. The electric and magnetic field are describable in terms of a vector potential A_μ in four dimensions.

In gravity, the metric tensor g_{MN} in a five-dimensional world with one compact dimension has the components $g_{\mu\nu}$ as the metric tensor in ordinary four-dimensional spacetime for the big dimensions and $g_{\mu 5} = g_{5\mu}$ and g_{55} in addition for the fifth dimension. The quantity $g_{\mu 5}$ is a four-vector and can be identified with A_μ, and g_{55} is a scalar called dilaton and is usually denoted by Φ. The metric tensor is a field which can vary from place to place, $g_{\mu\nu}$ is the usual gravitational field that can vary from place to place, $g_{\mu 5}$ becomes an analog to the electromagnetic field that can vary from place to place, and g_{55} is a scalar that can vary from place to place.

This scalar is the size of the fifth direction shown in the figure on the right side. The fifth direction is big if g_{55} is big, and is small if this value is small. One can imagine waves in space which are not electromagnetic or gravitational waves but waves of the varying size of the fifth dimension. The ordinary vacuum does not have such waves, and g_{55} is fixed, but in general it can vary from place to place.



The quantity $g_{\mu 5}$ is a gravitational field, and the sources of the gravitational field are energy and momentum. Especially, the values with mixed indices are created by momentum, and the source of $g_{\mu 5}$ is the flow of momentum in the fifth direction. Thus, the momentum quantum number n in (5.3) is analog to the electric charge, and it is connected to the graviton.

Also the winding number w has the property that opposite winding numbers attract and equal winding numbers repel similar to electric charge, but it is not the same electric charge as the momentum quantum numbers n. It is as if there are two kinds of electromagnetisms, two kinds of photons and two kinds of charges. The "photon" related to the momentum quantum number is a piece of the gravitational field and is just the graviton polarized along the $\mu 5$-direction. The "photon" related to the winding number has to do with the spectrum of the closed string. The operator a standing for the creation operator a^+

which creates a unit of excitation. Firstly, a has a direction i in space which can be either one of the three big dimensions x or the compact dimension y, and it is written as a^i. Secondly, a^i has a frequency n which tells how much energy is in the oscillation, and it is written a^i_n. Thirdly, the string has an orientation such that the wave can go in one or the other direction, and this is written here as $a^i_n(R)$ and $a^i_n(L)$ for left- and right-moving waves. The rule is that the amount of energy in the left- and the right-moving waves must be the same.

The first state is the one with no energy at all which is the tachyon and has minus two units of energy, and this is not wanted. The next level of excitation is $a^i_1(L)\,|0\rangle$ or $a^i_1(R)\,|0\rangle$ in any direction i, but they both do not fulfill the rule that the amount of energy in the left- and right-moving waves must balance. Thus, the first legitimate excitation is $a^i_1(L)a^j_1(R)\,|0\rangle$. One or both of these indices i and j can now be in this fifth compact direction. In the above discussion, these directions were only ordinary directions, and $a^i_1(L)a^j_1(R)\,|0\rangle$ corresponded to gravitons which have polarizations characterized by the two directions in space. If one of these direction is now compact, this expression corresponds to a photon but there can be two different types of objects corresponding to $a^i_1(L)a^5_1(R)\,|0\rangle$ and $a^i_1(R)a^5_1(L)\,|0\rangle$. The object corresponding to the sum of both is identified with the graviton, and the difference is identified with another field which has also the structure similar to a photon and which is the electromagneticlike field whose sources are the winding numbers. Thus, one linear combination is associated with the momentum quantum number n and the graviton which are the field quanta of the gravitational field $g_{\mu\nu}$, and the other linear combination is associated with the winding number w and a field called the Kalb–Ramond field $b_{\mu\nu}$ which is not the graviton.

To summarize, $a^i_1(L)a^5_1(R)\,|0\rangle + a^i_1(R)a^5_1(L)\,|0\rangle$ corresponds to the momentum quantum number and is identified with the gravitational field and the graviton, $a^i_1(L)a^5_1(R)\,|0\rangle - a^i_1(R)a^5_1(L)\,|0\rangle$ corresponds to the winding number and is identified with the Kalb-Ramond field, and $a^5_1(L)a^5_1(R)\,|0\rangle$ is the one that behaves like a scalar and corresponds to the dilaton as the field quanta of Φ. The T-duality is finally the interchange of n and w, of r and $1/r$, of $\partial x/\partial\sigma$ and $\partial x/\partial\tau$, and of $g_{\mu 5}$ and $b_{\mu 5}$. It is not clear at the moment whether the real world has such fields, and it is just a mathematical construction and an exploration of the mathematics of this theory.

5.5 D-Branes

The D in D-branes stands for Dirichlet, and the brane is related to membrane. There is also a value n in it for its dimension in space. A string is a one-brane, a membrane is a two-brane, a solid three-dimensional object which may be embedded in higher dimensions is a three-brane. They are not just made-up things. They were essential to the consistency of the theory. The result of knowing about them has been an enormous number of equivalences between different theories in the area of string theory which have been derived. This enormous number of equivalences found turned out to be an enormous numbers of equivalences between different kinds of geometric structures, and the mathematicians were surprised. D-branes have to do with open strings. Open strings do not have stable winding numbers. The question is what happens to the open strings when one applies the process of T-duality.

Open strings fulfill the Neumann boundary condition $\partial x/\partial\sigma = 0$ and $\partial y/\partial\sigma = 0$ at their two ends where x stands for all the ordinary dimensions and y for all the compact dimensions. This is because otherwise there would be a force acting on the last point particle of the string, and this last point particle gets lighter and lighter when going to the limit of an infinite number of point particles such that this force would accelerate this last point particle more and more and finally infinitely.

For closed strings, T-duality means replacing $\partial y/\partial\tau$ by $\partial y/\partial\sigma$ and this changes the Neumann boundary conditions to $\partial y/\partial\tau = 0$ at the ends of the string for the compact directions. It follows that the ends of the strings are not allowed to move, and therefore there must be objects around that can nail down the ends of the strings. One starts with a theory where open strings can move freely, and after T-duality their ends are stuck in the compact directions. The objects that hold the ends of the strings is called a D-brane. It is a movable object which acts as an anchor point for the ends of the strings. If it is movable and it is also bendable.

Assuming there are nine spatial dimensions and one of them is compact. The dimensionality of the D-brane would be eight because one has pinned down one direction and left the other eight directions free to move around. To understand it, one can take a table and attaches an object to it which is now free to move in the two directions on the table but is constrained in the third direction. Thus, if one constrains one of nine dimensions, this creates an eight-dimensional surface in which the string can move around, and each direction which is constrained removes one direction in which the string can move freely. If all directions are constrained, this gives a D0-brane which is a point and is also a new kind of particle.

If one shrinks the radius r of the compact direction more and more, the T-dual radius gets larger and larger until it is so big that it can be taken as a non-compact ordinary dimension. Thus, even in non-compact dimensions there must be anchors for strings.

A D0-brane is just a point where one or both ends of a string can attach. A D1-brane is a line which may not be a straight line because it has to be bendable. It is a one-dimensional object, but it is not the same as a string because D1-branes are as anchors which nail strings down much heavier. A D2-brane is called a membrane. A D3-brane in the three-dimensional space fills space, and strings attached to it can just freely move in space. It is often said that ordinary open string theory is string theory on a space-filling D3-brane.

D-branes are the source of many applications in string theory. A D2-brane is used here to illustrate this. A D2-brane with no strings attached is empty, and the one shown in the figure on the right side has two strings attached which are free to move around as long as their endpoints are still on the D2-brane. These strings are assumed to be oriented, and two strings with compatible orientation can join. They are constrained to the surface, but otherwise they can behave as strings do in ordinary open string theory. Creation and annihilation of particles can be described. The mathematics of the interactions of these particles looks very much like quantum field theory, and these particles behave like photons in a lower-dimensional theory. For low energies where the strings are not much excited it is simply quantum field theory where the particles are the open strings. One application of D-branes is to define quantum field theory.

There can also be two branes, for example, two parallel D2-branes. One can move them closer and closer together until they touch and form a compound brane. Another example is three parallel D2-branes which can have exaltations meaning ordinary strings. If one calls these three parallel D2-branes red, green, blue, an oriented string can start on red and end on red such that it may be called a red-red string, but it can also start on red and end on green or blue such that it may be called a red-green or red-blue string. In this case particles can be labeled by two colors. This looks like gluons in quantum chromodynamics, and the rules are exactly the same. A gluon has two ends representing a color and an anticolor. A quark is a string that has only one end fixed in a brane while the other end is at infinity or at least at a distant brane of a different kind. Thus, a quark is either a string coming in or a string going out of the brane. When a string going out and a string coming in meet each other, they can join and just disappear out of the brane, and this is an annihilation of a quark and an antiquark.

If there is only one D-brane, this is like quantum electrodynamics. It has only photons. Strings with both ends fixed in the brane are photons, and string with only one end fixed in the brane while the other end goes to infinity or so represent charge respectively electrons or positrons depending on their orientation. When an electron and a positron come together, they can annihilate. As mentioned above, there are D1-branes also called D-strings which are heavier than ordinary strings, and the question arises whether a D-string can also end on a brane like other strings. The answer is yes. They would represent magnetic monopoles. The universe is from the current point of view a D3-brane and therefore also the three colors of quantum chromodynamics are D3-branes. If a D-string which is also oriented comes from another dimension, people living in this universe see it as a magnetic monopole.

6 Conclusions and Outlook

6.1 Results of String Theory

String theory has been developed by many people over the last at least twenty-five years. Quantum chromodynamics today is studied in the context of string theory. Calabi–Yau manifolds allow more complicated asymmetries than the toroidal compactifications shown here. These asymmetries are needed to describe certain phenomena in physics. String theory as it stands today made a full circle from a theory of hadrons to a theory of gravitons, and now has become a theory of quantum chromodynamics in a ten-dimensional space with branes. However, it has not yet brought an understanding of the particle spectrum, and that is probably because it is just too complicated.

6.2 Reductionism

Reductionism is the philosophical assumption that things are made out of little things, and little things are made out of even more little things. Molecules are made out of atoms, atoms are made out of electrons and nuclei, nuclei are made out of protons and neutrons, protons are made out of quarks and so on. A further assumption is that things get simpler and simpler as one goes to smaller and smaller things. String theory and other developments in modern physics mark the end of reductionism.

Before one predicts the end of reductionism, one can see how far reductionism and especially the assumption that things get simpler when going to smaller and smaller things can lead. Particle physics should be simple, but it is not. One can ask how many different particle are there, and how many of them are unexplained. There are approximately seventy-five different particles in the standard model of which about twenty or a bit less are independent and have no tight relation between each other. There are about twenty distinct parameters in the sense of unexplained constants which have to go into the theory, but the theory is definitively incomplete because it does not have gravity, dark matter, particles that are necessary for inflation and it has a terrible fine-tuning problem. Theories like supersymmetry try to overcome these short-comings, and it is possible that supersymmetry is discovered one day, but if it does, it adds about a hundred plus n new parameters where n is an unknown number. Thus, either physicists have not yet reached the bottom, or this is the end of reductionism.

There is a more theoretical fact based on quantum field theory and string theory. When exploring models of quantum field theory with one dimension of time and one dimension of space, one can start with particles which are fermions and corresponding fermion fields Ψ. Putting two fermions close together gives a boson, and therefore a hydrogen atom, for example, consisting of a proton and an electron is itself a boson. Thus, there are also boson fields Φ, but one cannot build fermions with only bosons because any number of bosons gives a boson. If one plots the field Φ, a kink in the field where the field smoothly jumps represents a fermion. Kinks in a boson field are extended, thick, massive and much heavier than the bosons. The question is now whether the bosons or the fermions are the more basic objects. There are parameters in such a theory, and it turns out that for some ranges of parameters it is much more convenient to think of the fermions as the building blocks and for other ranges of parameters bosons are much more efficient for studying the theory. One parameter is the coupling constant g. Starting with low values for g makes fermions the better starting point because they are easy to deal with, easy to study and they are small objects because the kink is small. Increasing the coupling constant g makes the kink to spread out and gets less and less pointlike, but the bosons start to behave more and more like simple, elementary objects.

This situation in a world with a one-dimensional space is known for a long time. It is assumed that things like that exist in elementary particle physics with electric and magnetic monopoles. An electric monopole could be an electron with an electric field around it. The electric charge of an electron is very weak. Measured in photons it radiates when suddenly stopped it is about one photon per hundred electrons. This value is called the fine-structure constant. As a consequence, the electric field of an electron is also weak and does not do much to the surrounding space. It is also believed that magnetic monopoles exist, because they come up in so many theories, but they have never been discovered. A magnetic monopole is like the end of a bar-magnet where the bar-magnet is so thin that it cannot be seen. The magnetic field would come out of it like the electric field for the electron. The magnetic monopoles have a magnetic charge which is very large. The force between two magnetic monopoles would be ten-thousand times larger than the force between two electrons at the same distance. Therefore, a magnetic monopole creates a very strong field with heavy processes going on inside. They have not yet been discovered because they are to heavy.

The mathematics of quantum electrodynamics with electric and magnetic monopoles is completely unclear about the question which of them is the more elementary particle. It depends on the fine-structure constant α which is basically the square of the electric charge. If α is very small compared to one, the electric monopole is more basic because it is simple while the magnetic monopole has a complex structure.

If one could change α and increase it, the field around the electron would get stronger, the electric charge larger, and the electric field would become more complex. At the same time, the magnetic field would get weaker, the magnetic charge smaller, and the magnetic field would behave simpler. The electric monopole would get heavier, and the magnetic monopole would get lighter. For very large values of α, the two would just interchange.

Calculus assumes that functions are smooth and looking at smaller and smaller intervals make them simpler. Quantum fields are different because they fluctuate on every scale and are never smooth. This is one reason why quantum fields are so difficult. Calculus breaks down.

6.3 Building Blocks in String Theory

String theory has an answer to the question of what are the fundamental building blocks which is obviously strings. However, there are also D-branes which are mathematically needed by the theory and which are quite heavy.

Point particles are zero-dimensional object like D0-branes. Any number of ordinary strings can end on a D0-brane, and D0-branes typically have strings attached to them. If one would probe a D0-brane by scattering something of it, one would find out that it has a collection of strings attached to it. In fact, the harder one probes it the more strings one would find such that one would come to the conclusion that a D0-brane is made out of strings. However, it is heavy, much heavier than an ordinary string.

A D1-brane which is also called a D-string is like an ordinary string, but it is much heavier measured in units of length. Like a D0-brane, it is a place where ordinary strings can attach. Ordinary strings and D-strings are very much like electrons and magnetic monopoles. The question whether string theory has parameters or not is discussed later. Here it is assumed that string theory has a parameter g which is the coupling constant corresponding to the probability that a string wiggling around breaks into two strings. If g is very small, then ordinary strings are very light, thin and look very fundamental, and the D-string is heavy, complicated and full of ordinary strings. If one starts increasing g, the ordinary string tends to break off little pieces which do not disappear but hang around and form a kind of atmosphere around the string. It becomes more and more complicated and develops structure. At the same time the structure of the D-string gets simpler and thinned out because it becomes harder and harder for the D-string to produce ordinary strings. At some point when g is about one, they start to look exactly like each other. For even larger values of g, the D-string gets very simple and becomes a thin line-like object.

When g is very small, ordinary open strings form closed string, and as they form little loops, they create particles such as gravitons and photons. As g gets larger and larger, the ordinary strings get more and more massive, and eventually they turn into black holes. The remaining D-strings get lighter and lighter, and the gravitons are made out of the D-strings. The duality between ordinary strings and D-strings is called S-duality. One might say that the coupling constant is small such that it makes sense to take the ordinary strings as the fundamental building blocks, but the coupling constant can vary in space because it is a field and satisfies field equations. Thus, in this mathematically idealized string theory it is possible that in one place the ordinary strings are simple and fundamental, in other places the D-strings are simple and fundamental, and in between nothing is simple and fundamental.

In electrodynamics, an electron moving in time emits and absorbs photons according to the fine-structure constant. The photons between being emitted and absorbed may split into an electron and a positron for a while, and this electron emits and absorbs photons. This is an unending hierarchy at smaller and smaller distance scale of structure within the electron. As long as the fine-structure constant α as the coupling constant is weak, this is a relatively small effect and the electron in the laboratory looks pretty much pointlike and it is difficult to scatter electrons to see this structure, but at higher energies, this structure can be made visible.

6.4 Even-Dimensional D-Branes

There are different versions of string theory. In one of them all the D-branes are odd-dimensional. In another one all D-branes have to be even-dimensional. No string theory has both odd- and even-dimensional D-branes. Thus, in an even-dimensional string theory there are no D-strings, but there are pointlike D0-branes and there are ordinary strings. This theory does something very different as g starts to get large. The D0-branes could not become strings and vice versa, but something bizarre happens.

As g gets large, a new spatial dimension materializes, or better stated, a very small compact dimension starts to expand. To visualize it, it is assumed that there are two big dimensions of space plus one very small compact dimension which is not noticeable by physicists living in this world. When g gets larger and larger this compact dimension gets larger and larger until the physicists living in this world can no longer ignore it. When g starts growing the D0-branes get lighter and lighter similar to the D-strings in the string theory with odd-dimensional D-branes, and they turn into the ordinary gravitons. Thus, again the complex objects D0-branes turn into the simple objects gravitons. There are also D2-branes in this theory. They can exist as membranes in the two-dimensional world, but they can also be membranes with one direction in the compact dimension which look like strings in this space with two large dimensions. These strings turn into membranes when the third dimension becomes recognizable. They are very heavy because there is so much material there in this third dimension. Thus, the light strings turn into heavy membranes, and the D0-branes turn into very light gravitons.

By developing this string theory, physicists found theorems they could not prove, and mathematicians figured out that they have a proof for this theorems but did not know where to apply and use them. Because g can vary in space, it is possible that in one place the world looks two-dimensional with D0-branes and strings while in other places the world looks three-dimensional with gravitons and heavy membranes. In this example and in the previous examples it is not clear which objects are fundamental and are the building blocks for more complex objects, and which objects are the complex objects built by these fundamental objects. There may be more fundamental objects than strings and D-branes which build these things, but nobody knows so far.

6.5 The Duality of Electric and Magnetic Monopoles

If strings are attached to a D3-brane, this looks as if they live in a three-dimensional world, and the D3-branes behave in many respects as is they are ordinary three-dimensional spaces. The D3-brane can have wiggles on it and strings attached to it. Two strings can move around and join to form a single string, and then they can go back to be two separate strings. These strings scatter and do all kinds of things particles would do in ordinary quantum field theory in three dimensions. If an ordinary open string comes from the outside of the three-dimensional space of this D3-brane and attaches one end to it, people living in this world would see an electrically charged particle. Also a D-string can end with one side on a three-dimensional brane. It also looks like a particle, but a fatter and heavier particle which would be seen as a magnetic monopole. Thus, the S-duality which connects the ordinary strings with the D-strings is nothing but the duality which connects the electric charges with the magnetic charges of ordinary field theory.

In a spacetime with one time and several space dimensions of which two are compact and the others are big, a D2-brane with one direction in a big dimension and the other in a compact dimension looks like a string to physicists living in this world. If the compact dimension gets thinner and thinner, this object looks more and more like a string and gets lighter and lighter. A D2-brane with one direction in a big dimension and the other in the second compact dimension looks similar. If this compact dimension is much larger than the other one, the object to a physicist is a D-string. By changing the sizes of the two compact dimensions, ordinary strings become D-strings and vice versa. The two sizes of the compact dimensions can vary in space and in time. The fine-structure constant α is the ratio of these two sizes.

6.6 String Theory as a Network of Ideas

These are just examples of the huge network of ideas that emerged out of string theory. These ideas sometimes connect two different string theories, sometimes they connect a string theory with quantum field theory and so on, but most of it has very little to do with what physicists see in the laboratory. These precise features are features of an idealization which is the supersymmetric string theory which is also called superstring theory. With the symmetry between bosons and fermions and other unrealistic properties and simplifications, it does not describe the real world in a similar way as circular orbits do not describe the elliptical orbits of real planets.