Susskind Lectures on String Theory

March 15, 2024

version 2.3

These notes are from a series of lectures given by Leonard Susskind (one of the founders of string theory). They follow the logic he used when he first started thinking about strings. [1] I wrote these notes for graduate students interested in quantum gravity. These notes are not verbatim; I watched the lectures and reproduced them afterwards.

Part I of this document follows the lecture series by Susskind titled "String theory and M-theory." This is a basic introduction to bosonic string theory using the infinite momentum frame description. In this "light cone" frame one is able to quantize the non-relativistic degrees of freedom in the transverse plane using canonical quantization. The massless excitations are examined and found to contain photon-like and graviton-like particles, as well as tachyon states. We show that to obtain a zero mass for the photon-like state requires that the critical dimension of spacetime be D=26. The Veneziano amplitude is introduced in the context of meson scattering, and is, remarkably, shown to result from the scattering amplitude of two open bosonic strings. Part II follows the Susskind lectures titled "Topics in String Theory". Part II is incomplete.

If you find any errors or typos in this document please email them to me at majzoube@umsystem.edu. I hope you enjoy the notes.

[1] Susskind gave this series of lectures for retired engineers and other scientists in his continuing education course at Stanford. All of Susskind's continuing education lectures are easily found on YouTube. Even for advanced students these lectures are well worth watching.

Contents

List of Figures

Part I

Bosonic String Theory

1 Energy of Non-relativistic and Relativistic Particles

1.1 Non-relativistic particles

The energy of a "particle" is given the by the kinetic energy of the particle plus the rest mass of the particle plus whatever binding energy holds together the constituent parts. Classically, we write this as

$$E = \frac{p^2}{2m} + mc^2 + B. \tag{1}$$

Here B represents the internal binding energy of the particle. If the particle remains slow and is bound tightly, then this energy will never be observed and can be moved to the LHS of (1) with a redefinition of the energy $E' = E - mc^2 - B$ giving

$$E' = \frac{p^2}{2m}. \tag{2}$$

For example, the energy of a cup of coffee would be the momentum of the center of mass of the motion of the coffee plus the rest mass m of all the atoms in the coffee plus the binding energy B of the water in the coffee. The rest mass m of the coffee would include the internal motion of the atoms (heat). The binding energy of the coffee would be the latent heat needed to evaporate all of the coffee and this would go into B.

1.2 Composite particles

What separates "elementary" particles from composite particles? It is the relative magnitude of the energy required to increase the internal angular momentum of the internal constituents of the particle.[2] If this number is a large fraction of the rest mass

[2] In particle accelerator experiments a plot of mass squared m^2 vs angular momentum J of pions, for example, shows a linear relationship. All hadrons including for example the pions (two-quark particles) and deltas (three-quark particles with spin 3/2) show the same behavior with the same sloping trajectories.

of the particle it is considered to be an elementary particle. If the required energy is a very small fraction of the rest mass of the particle it is considered a composite particle. Examples here are given by the electron, an elementary particle, and the cup of coffee, which is clearly composite. An electron has spin $\pm\hbar/2$. Is it possible to "spin up" an electron to a higher angular momentum state similar to the way hadrons can be spun up, resulting in the Regge trajectories seen in particle accelerators? Classically the energy of a rotating object is given by the kinetic energy $E = I\omega^2/2$ and the angular momentum is given by $L = I\omega$. Quantum mechanics dictates that angular momentum is quantized in units of $\hbar$. For a large, classical object like a basket ball it is easy to increase the angular momentum by one unit of $\hbar$ by slightly increasing the spin rate ω. Clearly, this is a small fraction of the total energy of the basketball. On the other hand an electron's intrinsic angular momentum is $\hbar/2$, and increasing it requires at least one unit of $\hbar$. In contrast to a macroscopic object, spinning up the electron from $\hbar/2$ to $3\hbar/2$ triples its angular momentum. The change in energy would be given by $E_{3/2} - E_{1/2} = 8I\omega_0^2/2 = 8E_{1/2}$ or eight times the rest energy of the particle, if an electron could be considered as a tiny spinning object (it cannot!). Another way to look at this large change in energy is through the equation $E = L/2I$, the energy is inversely proportional to the rotational inertia. The size of the electron is known to be very small, on the order of 10^{-20} m, so that the energy required to increase the angular momentum by one unit must be very large (on the order of the Planck energy).

1.3 Relativistic particles and the infinite momentum frame

We have argued that composite particles can be "spun up" by increasing the angular momentum of their constituent parts. We would like to consider now the motion of a composite particle after it is given a large Lorentz boost. Starting from the Klein-Gordon relation and setting $c = 1$ we have

$$E^2 = p^2 + m^2 \tag{3}$$

we can approximate the energy in two ways. First we can assume that the particle momentum is low and so the rest mass dominates the energy. Then we write

$$E = \sqrt{p^2 + m^2} = m\left(1 + \frac{p^2}{m^2}\right)^{1/2} \approx m + \frac{p^2}{2m} \tag{4}$$

giving the usual relationship upon restoring c: $E = p^2/2m + mc^2$. In the second case we perform a Lorentz boost along the z-direction. Here the momentum along z dominates the energy and we expand as

$$\begin{aligned} E &= \sqrt{p_z^2 + p_x^2 + p_y^2 + m^2} \\ &= p_z \left(1 + \frac{p_x^2 + p_y^2 + m^2}{p_z^2}\right)^{1/2} \\ &\approx p_z + \frac{p_x^2 + p_y^2}{2p_z} + \frac{m^2}{2p_z}. \end{aligned} \tag{5}$$

We can assume that as $p_z \to \infty$ the only dynamics that are left are those in the perpendicular plane. In this case p_z on the RHS is simply a constant and can be moved to the LHS as an additive constant to the total energy. The Lorentz contraction along the boost direction flattens the dynamics to the perpendicular plane with the total momentum along the boost direction acting as an inertial mass

$$\frac{p_x^2 + p_y^2}{2p_z} \longleftrightarrow \frac{p_x^2 + p_y^2}{2M}$$

with p_z playing the role of M. A Lorentz boost of the system results in time dilation and apparently slow dynamics from the point of view of the stationary observer due to the large momentum. A rescaling of the time variable can effectively scale out the p_z or one can simply multiply (5) through by p_z. Redefining the total energy as $E' = (E - p_z)p_z$ we have

$$E' = \frac{p_x^2 + p_y^2}{2} + \frac{m^2}{2}. \tag{6}$$

Now the dynamics is given in terms of momentum in the plane perpendicular to the boost and the internal additive energy of the particle is given by $B = m^2/2$. At relativistic energies m^2 is something that is measured, giving rise to the Regge trajectories described earlier. The two important points to remember from this section are (1) the internal energy goes as m^2 in the infinite momentum frame, and (2) The motion can be described classically in the plane perpendicular to the boost direction.

2 Relativistic String in the Infinite Momentum Frame

2.1 Coupled spring system in the continuum limit

In this section we model the string as a collection of point masses connected by springs. Consider an *open* spring system of finite length consisting of N point masses connected by springs, with the equilibrium spacing a and spring constant κ. If the system is boosted along the 3-direction the total energy in the 1,2-plane is given by

$$E = \sum_{n=1}^{N} \frac{m\dot{x}_n^2}{2} + \sum_{n=1}^{N-1} \frac{\kappa(x_{n+1} - x_n)^2}{2}$$

where $x \equiv (x, y)$ is the position of mass n. In the limit as $N \to \infty$, $a \to 0$, and the spring constant $\kappa \to \infty$ such that the product is finite $a\kappa \to$ constant, we can write the energy as an integral over the length of the string. In particular

$$\sum_n \kappa a(a\Delta n)\frac{(x_{n+1} - x_n)^2}{a^2} \to \int d\sigma \left(\frac{\partial x}{\partial \sigma}\right)^2 (\kappa a)$$

where $d\sigma = \lim_{a\to 0} a\Delta n$, and $\Delta n = 1$, and in the first term we write

$$\sum_n \frac{m\dot{x}^2}{2a} a\Delta n \to \int d\sigma \frac{\dot{x}^2}{2}(m/a).$$

This clearly requires that $m \to 0$ such that $m/a \to$ constant. Redefining $x = \sqrt{\kappa a}x$ will give the total energy

$$E = \int_0^\pi d\sigma \frac{1}{2}\left[\left(\frac{\partial x}{\partial t}\right)^2 + \left(\frac{\partial x}{\partial \sigma}\right)^2\right]$$

where the length was defined such that σ goes from 0 to π. This is convenient in the Fourier expansion below, and is defined this way for the open string so that the integral for closed strings will go in a loop from 0 to 2π.

2.1.1 Canonical quantization of the string

Now we can quantize canonically using the Lagrangian $\mathcal{L} = \int_0^\pi d\sigma \frac{1}{2}\left[\left(\frac{\partial x}{\partial t}\right)^2 - \left(\frac{\partial x}{\partial \sigma}\right)^2\right]$. Susskind writes this with some constants in front that can be absorbed into the definition of the field and also uses τ, the proper time.

$$\mathcal{L} = \frac{1}{2\pi}\int_0^\pi d\sigma \left[\left(\frac{\partial x}{\partial \tau}\right)^2 - \left(\frac{\partial x}{\partial \sigma}\right)^2\right] \tag{7}$$

Note that this is the Lagrangian for a wave field and the Euler-Lagrange equation is the wave equation. In order to Fourier analyze the string one must make a decision on the boundary conditions. At the end of the string the last mass point has a force proportional to the stretch in the spring or $m\ddot{x} \propto \partial x/\partial\sigma$ but the mass points were tending to zero so the acceleration would be increasing without bound. Therefore the boundary condition chosen is $\partial x/\partial\sigma = 0$ or Neumann boundary conditions. Expanding in a Fourier series we have

$$x(\sigma,\tau) = \sum_{n=0}^{\infty} x_n(\tau)\cos(n\sigma) \tag{8}$$

$$\dot{x}(\sigma,\tau) = \sum_{n=0}^{\infty} \dot{x}_n(\tau)\cos(n\sigma) \tag{9}$$

$$x_{,\sigma}(\sigma,\tau) = -\sum_{n=0}^{\infty} n x_n(\tau)\sin(n\sigma) \tag{10}$$

Plugging these equations into the Lagrangian (7) and using the orthogonality of the sines and cosines[3] we will find an infinite collection of harmonic oscillators of frequency $\omega_n = n$. Explicitly

$$\begin{aligned}\mathcal{L} &= \frac{1}{2\pi}\int_0^\pi d\sigma \left[\sum_{n,n'} \dot{x}_n\cos(n\sigma)\dot{x}_{n'}\cos(n'\sigma) - \sum_{n,n'} n x_n \sin(n\sigma) n' x_{n'}\sin(n'\sigma)\right] \\ &= \frac{1}{2}\left[\dot{x}_0^2 + \frac{1}{2}\sum_{n=1}^{\infty}\left(\dot{x}_n^2 - n^2 x_n^2\right)\right]\end{aligned} \tag{11}$$

This is a set of harmonic oscillators of increasing energy. The oscillator labeled $n = 1$ has one unit of energy, $n = 2$ corresponds to two units. The zero point energy of a harmonic oscillator is $\hbar\omega/2$ so for oscillator n the ZPE is $n/2$. The total energy of the n-th oscillator in the m-th excited state is

$$E_m = n(m + \frac{1}{2})$$

[3] $\int_0^\pi dx\, \sin(nx)\sin(mx) = \delta_{nm}\frac{\pi}{2}$ and $\int_0^\pi dx\, \cos(nx)\cos(mx) = \frac{\pi}{2}\delta_{nm} + \frac{\pi}{2}\delta_{n0}\delta_{m0}$

To obtain the Hamiltonian from this Lagrangian one calculates the conjugate momentum

$$p_n = \partial L / \partial \dot{x}_n = \dot{x}_n / 2$$

and rewrites the Lagrangian as

$$\mathcal{L} = \frac{1}{2}\dot{x}_0^2 + \sum_{n=1}^{\infty}\left(p_n^2 - \frac{n^2 x_n^2}{4}\right). \tag{12}$$

The Hamiltonian is

$$\begin{aligned} H &= \sum_{n=0} p_n \dot{x}_n - \mathcal{L} \\ &= \sum_{n=1}\left(p_n^2 + \frac{n^2 x_n^2}{4}\right) \end{aligned} \tag{13}$$

where we have dropped the CM motion term. Let us write this in terms of the usual creation and annihilation operators. Focusing on one of the terms we follow the usual procedure for harmonic oscillator quantization

$$\left(\frac{n^2 x_n^2}{4} + \dot{x}^2\right) = \left(\frac{n x_n}{2} + i\dot{x}\right)\left(\frac{n x_n}{2} - i\dot{x}\right)$$

and identify each term with either the creation or annihilation operator. Using the canonical quantization condition $[x, p] = i$ and knowing we need $[a^-, a^+] = 1$ we calculate the commutator

$$\begin{aligned} \left[\left(\frac{nx}{2} + ip\right), \left(\frac{nx}{2} - ip\right)\right] &= \frac{ni}{2}\left([p, x] - [x, p]\right) \\ &= n \end{aligned} \tag{14}$$

so we define

$$a_n^{\pm} = \frac{\sqrt{n}}{2}\hat{x}_n \mp \frac{i}{\sqrt{n}}\hat{p}_n. \tag{15}$$

These are the creation and annihilation operators. By adding and subtracting them we can represent the position and momentum operators in terms of the creation and annihilation operators

$$\hat{x}_n = \frac{1}{\sqrt{n}}\left(a^+ + a^-\right) \tag{16}$$

$$\hat{p}_n = i\frac{\sqrt{n}}{2}\left(a^+ - a^-\right) \tag{17}$$

If we define the operators $a^{\pm}$ to correspond to excitations of the oscillators in x, we can define $b^{\pm}$ operators for the excitations of the oscillators in y. Then we can begin to discuss the excitations from the ground state of the string, denoted by $|0\rangle$. Note that $|0\rangle$ is not the vacuum state, but the ground state of an existing string. The first excited state is doubly degenerate and consists of the states $a_1^+|0\rangle$ and $b_1^+|0\rangle$. The second excited state is 5-fold degenerate and consists of the states $\{a_1^+a_1^+|0\rangle, a_1^+b_1^+|0\rangle, b_1^+b_1^+|0\rangle, a_2^+|0\rangle, b_2^+|0\rangle\}$. Remember here that a_n^+ excites the n-th mode of the oscillator in the x direction, so a_2^+ excites a mode with frequency $\omega = 2$, and $\left(a_1^+\right)^2$ excites the mode-1 oscillator twice, giving energy $1+1=2$.

String theory is not consistent in 3+1 dimensions and requires higher dimensions. These can be incorporated easily in our discussion by simply allowing the oscillator to move in D+1 dimensions and adding the appropriate number of terms to the Hamiltonian and defining their corresponding creation and annihilation operators. Here we are identifying the excitations of the string in the transverse plane that has only two dimensions. In real string theory there are more dimensions, but the remaining dimensions are compactified.

The excitations of the string will be described by the creation and annihilation operators acting on the ground state of the string $|0\rangle$. Keep in mind that the energy of the internal motion of the string is to be identified with the mass squared. If the string is in its ground state we would write $E_{GS} \to m_0^2$.

2.1.2 Helicity for massless particles

Massless particles that have spin J can only have two spin states instead of the usual $2J+1$ states for a massive particle. This is because massless particles move at the speed of light and cannot be brought to a rest frame in which a rotation operator can act on the spin state to orient it in a direction perpendicular to its motion. The two remaining spin states are $\pm J$ and are called helicity.

A photon is a spin-1 particle described by the vector potential A^{μ}. It has two spin states of its angular momentum, $|R\rangle$ and $|L\rangle$, for left and right circularly polarized. In classical electrodynamics we describe the electric field vector of Maxwell's equations as having two possible orientations along the direction of motion of the electromagnetic wave. These are the linear polarizations along the x- or y-direction, which we would label quantum mechanically as $|x\rangle$ and $|y\rangle$. To go from one description (linear polarization) to another (circular polarization) we use a linear combination with complex coefficients.

$$|x\rangle = |R\rangle + i\,|L\rangle \quad (18)$$
$$|y\rangle = |R\rangle - i\,|L\rangle \quad (19)$$

Keep in mind that a linear combination of $|x\rangle$ and $|y\rangle$ with real coefficients would simply be a linearly polarized light wave along some angle between the x- or y-direction.

2.1.3 Low energy excitations of the string

The energy of the quiescent string in the light cone frame is $E_{GS} \to m_0^2$. It will include zero-point oscillations, but those can be absorbed in m_0^2. The lowest energy excitations of the string will involve using the creation operators to give one unit of oscillation to the string,

$$a_1^+\,|0\rangle \quad (20)$$

for example. It would increase the energy of the string by one in some units. We would write

$$E = m_0^2 + 1 \quad (21)$$

The same would be true for exciting oscillations along the y-axis. We could combine the creation operators to create a linearly polarized oscillation along an arbitrary direction. For example, along a 45 degree angle would be $(a_1^+ + b_1^+)/\sqrt{2}$. Both a^+ and b^+ transform as components of a vector. It is the transformation properties of the operators that is important. Note that we could also create oscillations that correspond to the photon with circular polarization by using

$$(a_1^+ \pm i b_1^+)/\sqrt{2} \quad (22)$$

Importantly, we notice that there are only two states of this energy that are linear combinations of a_1^+ and b_1^+. In particular there is not a third state, and because these states transform as vectors under rotation they represent polarization vectors just as in the case of photons. One can have either linear polarization with linear combinations $(a_1^+ \pm b_1^+)/\sqrt{2}$ or circular polarizations with $(a_1^+ \pm i b_1^+)/\sqrt{2}$. So there are photon-like objects in the open string. But there is a problem, photons are massless. They must be massless for Lorentz invariance to hold. This is a problem because the energy of the string is to be associated with its mass squared, $E \sim m^2 = 0$ and implies that we have

$$m_0^2 = -1 \tag{23}$$

a negative ground-state mass squared. This is called a tachyon and was originally thought of as particles that had to move faster than the speed of light since $E = \sqrt{p^2 + m^2}$ implies the speed of the wave should be

$$\partial\omega/\partial k = \partial E/\partial p = p/\sqrt{p^2 + m^2} > 1$$

This is unreasonable of course, and the correct interpretation is that the negative mass squared means there is an unstable vacuum from the point of view of field theory. In a field theory Lagrangian, $L = (\partial\phi)^2 - m^2\phi^2$ with the potential $V(\phi) = +m^2\phi^2$, the negative mass term corresponds to a potential that has no ground state.

2.1.4 Open vs. closed strings

If string ends are allowed to interact, then two open strings can approach each other and join ends, making a longer string. Think of the quark model in hadronic physics where quark-antiquark pairs are joined by stings and may interact in this way by annihilating a quark-antiquark pair and making a longer string. If this interaction is possible, then the ends of a single string may interact and form a closed string. So open string interactions imply the existence of closed strings. This does not work the other way around; closed strings do not have allow opening and so closed string theory does not have to contain open strings. So string theory predicts the existence of closed strings. We will find that open strings behave like photons, and closed strings behave like gravitons.

If an open string can interact with a closed string it can open the closed string to form an open string; it can absorb a graviton. Closed strings can do the same. It appears that every kind of string-like object can absorb closed strings, and this is why gravitons interact with everything.

2.2 Closed string spectrum

Susskind starts the lecture (number 4) with Noether's theorem and states the fact that the charge of a conserved quantity is the generator of the transformation that gives the conserved quantity. This is standard basic QFT so I have not written it up here.

The closed string spectrum is found by a similar analysis of the Lagrangian (7) but instead of the Neumann boundary condition one uses periodic boundary conditions. If the string is closed then the waves can be considered to run either left or right around the string, i.e., in the positive or negative σ direction. Before we Fourier expand the coordinates let us rewrite the Lagrangian to reflect the left- and right-moving waves.

$$\mathcal{L} = \frac{1}{2\pi}\int_0^{2\pi} d\sigma \frac{1}{2}\left\{\left[\left(\frac{\partial x}{\partial \tau}\right) + \left(\frac{\partial x}{\partial \sigma}\right)\right]^2 + \left[\left(\frac{\partial x}{\partial \tau}\right) - \left(\frac{\partial x}{\partial \sigma}\right)\right]^2\right\} \tag{24}$$

In this equation the cross terms cancel and, except for the change in limits, we have (7). In this In this case we let the length of the string go from 0 to 2π for convenience and require that $x(\sigma + 2\pi) = x(\sigma)$. The solutions with this boundary condition include both sines and cosines, and we write

$$x(\sigma, \tau) = x_0 + \sum_{n>0}\left[x_n(\tau)e^{in\sigma} + x_{-n}(\tau)e^{-in\sigma}\right]$$

where right-moving waves are given by the exponential $e^{in\sigma}$ and left-moving waves are given by $e^{-in\sigma}$. Following the same canonical quantization procedure above we find a set of creation and annihilation operators

$$a_n^+ \qquad a_{-n}^+ \qquad b_n^+ \qquad b_{-n}^+$$

where we now have operators that create waves moving to the left or right, indicated by the subscripts. For the closed string we require that the right- and left-moving waves have the same energy. This requirement can be stated in the form of the integral

$$\Delta E = \int_0^{2\pi} d\sigma \frac{1}{2}\left\{\left[\left(\frac{\partial x}{\partial \tau}\right) + \left(\frac{\partial x}{\partial \sigma}\right)\right]^2 - \left[\left(\frac{\partial x}{\partial \tau}\right) - \left(\frac{\partial x}{\partial \sigma}\right)\right]^2\right\} = 0$$

which clearly implies that

$$\int_0^{2\pi} d\sigma \left(\frac{\partial x}{\partial \tau}\right)\left(\frac{\partial x}{\partial \sigma}\right) = 0. \tag{25}$$

This result also follows from reparameterization invariance of the position along the string. Mathematically the quantum mechanical wave function must be the same for a reparameterization of σ. Here we consider an infinitesimal change $\sigma \to \sigma + \epsilon$ and write

$$\psi\left[x(\sigma+\epsilon)\right]=\psi\left[x(\sigma)\right]. \tag{26}$$

We should write this as a functional since σ varies over a continuous range and any construction, of say the partition function, will be a functional integral over all possible $x(\sigma)$. In this case we expand (26), integrate over σ and find

$$\int_0^{2\pi} d\sigma\,\frac{\partial\psi}{\partial x}\frac{\partial x}{\partial\sigma}\epsilon=0.$$

This is an integral operator: $\left(\int_0^{2\pi} d\sigma\,\frac{\partial x(\sigma)}{\partial\sigma}\frac{\partial}{\partial x(\sigma)}\right)|\psi\rangle$. But $\frac{\partial}{\partial x(\sigma)}=-i\hat{p}(\sigma)$ and the momentum $p(\sigma)=\dot{x}(\sigma)$ is proportional to $\frac{\partial x}{\partial\tau}$ so we again arrive at (25).

The lowest energy excitations of the string must have the same energy in left-moving excitations as right-moving excitations. The lowest energy energy excitations are then given by

$$a_1^+a_{-1}^+\,|0\rangle$$

with a total of 2 units of energy, one left and one right. Similarly for $\{b_1^+b_{-1}^+, a_{-1}^+b_1^+, a_1^+b_{-1}^+\}$. Now write them in terms of circular polarization states:

$$(a_1^+ + ib_1^+)(a_{-1}^+ + ib_{-1}^+)\,|0\rangle \tag{27}$$

$$(a_1^+ - ib_1^+)(a_{-1}^+ - ib_{-1}^+)\,|0\rangle \tag{28}$$

$$(a_1^+ + ib_1^+)(a_{-1}^+ - ib_{-1}^+)\,|0\rangle \tag{29}$$

$$(a_1^+ - ib_1^+)(a_{-1}^+ + ib_{-1}^+)\,|0\rangle \tag{30}$$

Equations (27) and (28) correspond to 2 units of angular momentum, and equations (29) and (30) correspond to zero units of angular momentum. For a zero mass particle of spin-2 we expect the two states $m=\pm 2$. These are the gravitons. We do not expect zero angular momentum states. Equations (29) and (30) are particles called the dilaton and axion, respectively. Neither of these particles has ever been observed. These particles can be removed from the theory while keeping the gravitons.

In response to a question, Susskind discusses in some detail the question of what is or is not a fundamental particle. Consider the example of the electron vs. the magnetic monopole. From Dirac's argument that the charges satisfy $eg=2\pi$, the magnetic charge must be large if $e\ll 1$, and Feynman diagrams in QED converge because the probability for every vertex contains a factor of e^2. In contrast, if $g\ll 1$ then the Feynman diagrams would all diverge, and so one might expect that the electron is fundamental while the monopole is not. The argument can be turned around if one considers that Maxwell's equations have some interesting symmetry (duality) and that if the ratio of e/g were varied and $g \ll 1$ then one would consider the monopole to be fundamental and the electron composite. Are strings fundamental? Well, strings have a duality with D-branes, and they morph into each other as a parameter of the theory is varied; from some viewpoints the strings are fundamental and in others the D-branes are fundamental.

2.3 Critical dimension of the open bosonic string

The negative mass squared we found in (23) is problematic and needs an explanation. We have previously ignored the zero point energy of the string, and this is where it comes in. If we consider the photon-like excitation of the string that must be massless we are left with explaining how the zero point motion could give rise to $m_0^2 = -1$. The zero point energy for a single oscillator is $\hbar\omega/2$, and if our string can vibrate in $D-2$ dimensions we should have a ZPE of $(D-2)\hbar\omega/2$. The $D-2$ comes from D spacetime dimensions, minus one dimension for time, and minus one dimension for the coordinate which contained the boost to very high momentum. The number of dimensions in the transverse plane is $D-2$. Using $\hbar = 1$ and $\omega = n$ we have

$$\frac{(D-2)}{2}\sum_{n=1}^{\infty} n = -1. \tag{31}$$

This looks bad because the sum is formally divergent. However, infinite sums can be assigned numerical values. This is not trickery, there is physics involved. A similar situation arises in a mathematical analysis of the Casimir force between two conducting plates. In that case one needs to "regularize" the sum because arbitrarily high frequencies of electromagnetic waves will not be contained by the plates and the regularized sum is not divergent. It is, however confusing it may be, negative. There are more and less sophisticated ways of arriving at the sum and we will do it informally, using a mathematically inconsistent technique, but it will give the correct answer. The sum is

$$S = 1 + 2 + 3 + 4 + 5 + 6 + \cdots$$

To make an alternating and therefore more convergent, but not convergent series, we write

$$\begin{aligned} S &= 1+2+3+4+5+6+\cdots \\ T &= 0-4+0-8+0-12+\cdots = -4S \\ S-4S &= 1-2+3-4+5-6+\cdots \end{aligned}$$

The last series is the Taylor expansion of $1/(1+x)^2$ about $x=0$ and setting $x=1$. We then have

$$S-4S = \frac{1}{4}$$

or $S=-1/12$ giving

$$\sum_{n=1}^{\infty} n = -\frac{1}{12}. \tag{32}$$

Adding infinite series as we have done with the additional zeros in T is inconsistent and not mathematically rigorous.[4] A rigorous derivation of this result requires use of the Riemann zeta function. Plugging (32) into (31) we find

$$\frac{(D-2)}{2\cdot 12} = 1$$

or $D=26$. This means that if the string lives in 26 dimensions we can satisfy (23) in a consistent manner. This is called the critical dimension of the string theory. In supersymmetric string theories the number of critical dimensions is reduced. In some sense it means that there are not arbitrarily high zero point oscillations.

Susskind uses a better argument to obtain this result and explains that the divergent sum should be thought of as having a finite piece that can be combined with the infinite P piece of the $2P(E-P)$ that was obtained from the infinite momentum frame. Then he calculates the sum using a convergence term.

$$\sum_{n=1} n = \lim_{\epsilon\to 0}\sum_{n=1} n e^{-n\epsilon}$$

Let the limit be understood, then we can write

$$\sum_{n=1} n = -\frac{\partial}{\partial\epsilon}\sum_{n=1} e^{-n\epsilon}$$

[4]Although it is interesting to note that Ramanujan had this derivation in one of his notebooks!

where we note that the sum starts from $n = 1$. Performing the sum gives $e^{-\epsilon}/(1-e^{-\epsilon})$ and expanding the exponentials to third order gives

$$-\frac{\partial}{\partial \epsilon}\left[\frac{(1-\epsilon+\epsilon^2/2-\epsilon^3/6)}{(\epsilon-\epsilon^2/2+\epsilon^3/6-\epsilon^4/24)}\right] = -\frac{\partial}{\partial \epsilon}\left[\frac{1}{\epsilon}\frac{(1-\epsilon+\epsilon^2/2-\epsilon^3/6)}{(1-\epsilon/2+\epsilon^2/6-\epsilon^3/24)}\right].$$

Keeping terms to order ϵ^2, and using $1/(1-x) = 1 + x + x^2 + \cdots$, gives

$$-\frac{\partial}{\partial \epsilon}\left[\frac{1}{\epsilon}\left(1-\epsilon+\epsilon^2/2\right)\left(1+\epsilon/2-\epsilon^2/6+\epsilon^2/4\right)\right]$$

where the $\epsilon^2/4$ comes from $(\epsilon/2-\epsilon^2/6)^2$, the x^2 in the expansion. Then we have

$$-\frac{\partial}{\partial \epsilon}\left[\frac{1}{\epsilon}\left(1-\epsilon/2+\epsilon^2/12\right)\right] \longrightarrow \lim_{\epsilon\to 0}\frac{1}{\epsilon^2}-\frac{1}{12}$$

which contains the formally infinite piece (infinity!) plus the *finite* piece $-1/12$.

The fermionic string reduces the number of required dimensions to D=10, and removes the tachyon states.

3 Open String Scattering

3.1 Mandelstam variables

In particle scattering analysis it is useful to define the Mandelstam variables. Consider a 4-particle scattering event with four *identical* particles with incoming and outgoing four-momenta $\{k_1, k_2, q_3, q_4\}$. Momentum and energy conservation dictate that we write $k_1 + k_2 = q_3 + q_4$. If we relabel $q_3 = -k_3$ and $q_4 = -k_4$ we have

$$k_1 + k_2 + k_3 + k_4 = 0.$$

In this case, the figure would show all lines going in to the blob. The Klein-Gordon relation can be written $p^2 - E^2 = k_\mu k^\mu = -m^2$ for each of the particle and provide constraints. There are 16 independent variables in the four-momenta of each of the particles. The constraint equations reduce the number of independent variables to just two. In the center of mass frame there are ultimately only two variables that describe the amplitude of scattering, the total energy E_{CM}, and the angle through which the scattered particles depart, θ. There is only one relevant angle if all four particles are identical. To form invariants from the four-momenta there is only one

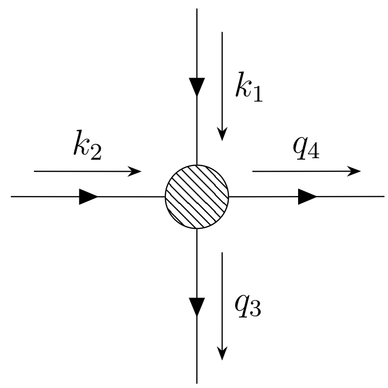

Figure 1: Four-particle interaction with incoming momenta k_1, k_2 and outgoing momenta q_3, q_4.

thing to do, form scalars by squaring them. Consider that k_1 and k_2 are the incoming particles. It is always possible to boost to the CM frame so that the particles are coming at each other directly. Then $\vec{k}_1 = -\vec{k}_2$ and

$$\begin{aligned}(k_1 + k_2)^2 &= (\vec{k}_1 + \vec{k}_2)^2 - (k_{01} + k_{02})^2 \\ &= 0 - (2E_{\mathrm{CM}})^2\end{aligned}$$

This is defined as $S = (2E_{\mathrm{CM}})^2$. Taking particles 3 and 4 would be the same as taking 1 and 2, gaining nothing new. Now form the invariant with particles 1 and 3. In this case $k_{01} + k_{03} = 0$ because we have defined 3 as an outgoing particle. The momentum transfer is then given by $\vec{k}_1 - \vec{q}_3 = \vec{k}_1 + \vec{k}_3$. Working out this quantity in terms of the angle of scattering gives

$$\left(\vec{k}_1 + \vec{k}_3\right)^2 = 2\left(E_{\mathrm{CM}}^2 - m^2\right)(1 - \cos\theta)\,.$$

The combination of particles 2 and 4 gives the same relation as 1 and 3. The only combination left is particles 1 and 4 giving

$$\left(\vec{k}_1 + \vec{k}_4\right)^2 = 2\left(E_{\mathrm{CM}}^2 - m^2\right)(1 + \cos\theta)\,.$$

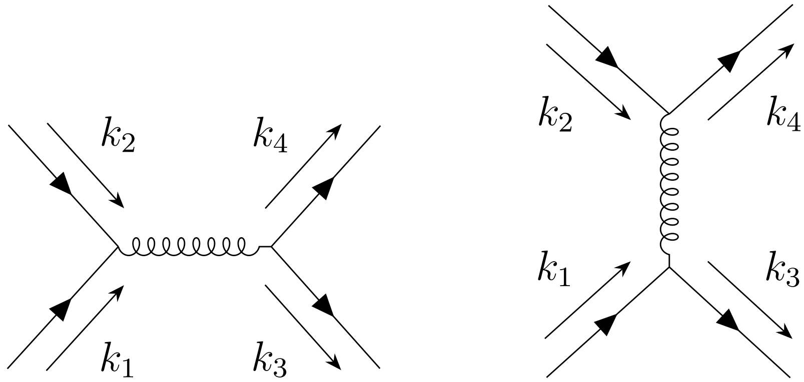

Figure 2: Feynman diagrams for S- (left) and t-channel (right) scattering with time running from left to right. Making particle 4 incoming and 2 outgoing corresponds to u-channel scattering. (The gluon line here represents many kinds of particles of different mass.)

Forming scalars from various combinations of the incoming and outgoing four-momenta, one defines the *Mandelstam variables*

$$\begin{aligned} -s &= (k_1 + k_2)^2 \\ -t &= (k_1 + k_3)^2 \\ -u &= (k_1 + k_4)^2 \end{aligned} \tag{33}$$

The final constraint is $s + t + u = -4m^2$, and the two independent quantities are usually taken to be s, and t. The resulting amplitudes for the processes in Fig. 2 are called s- and t- processes, and are given by

$$A_s = g^2 \left(\frac{1}{s^2 - M^2} \right) \tag{34}$$

$$A_t = g^2 \left(\frac{1}{t^2 - M^2} \right). \tag{35}$$

Note that in s-channel scattering the incoming particles form a new bound state that depends only on the CM energy. s-channel scattering is independent of outgoing angle, they are all equally probable. This makes sense because they from a compound state and then decay, and the compound state does not remember where it came from. A very different physical process is the t-channel scattering where particles 1 and 2 exchange a particle and go out as particles 3 and 4. The t-process depends on the angle of scattering, while s-processes do not. Note the intriguing symmetry between s- and t-channel scattering.

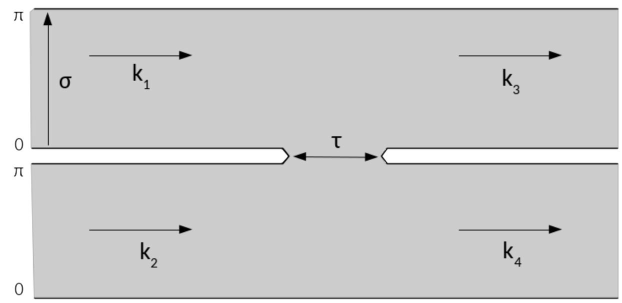

Figure 3: Bosonic open string scattering. Two open strings come together for a time τ and then separate. Time runs from left to right in this figure.

3.2 The Veneziano amplitude

Historically in the 1960's when meson scattering amplitudes were first being measured, the two amplitudes for s- and t-channel scattering were thought to add in a calculation of the amplitude. But the scattering amplitude from this calculation is twice what is seen in experiments. Physicists at the time were trying various functions for the amplitude that might result from an exchange of a large number of particles in the channels. The result was that one was discovered that matched the data perfectly, and has the same symmetry as the diagrams. It is called the Veneziano amplitude and is given by

$$A_{\text{Veneziano}} = g^2 \frac{\Gamma(-s)\Gamma(-t)}{\Gamma(-s-t)}. \tag{36}$$

How one arrives at the Veneziano amplitude from first principles is a triumph of string theory. The calculation goes as follows.

3.3 Bosonic open string scattering

We now consider two strings as shown in Fig. 3. Two strings moving freely through space come together and joint for a time τ. We need to calculate the scattering amplitude for this process. We take the top string to have incoming momentum k_1 and the bottom string incoming momentum to be k_2. We write the wave function of the top string in the ground state of all the harmonic oscillators that make up the string as

$$\psi_t(k_1; x_1, \cdots, x_N) = e^{ik_1 \frac{1}{N} \sum_{n=1}^{N} x_n} \psi_0(x_1, x_2, \cdots)$$

where x_l is the position of the l-th mass point on the string. Do not confuse this with the Fourier mode decomposition above where we quantized. The bottom string as

$$\psi_b(k_2; x_{N+1}, \cdots, x_{2N}) = e^{ik_2 \frac{1}{N}\sum_{n=N+1}^{2N} x_n} \psi_0(x_{N+1}, x_{N+2}, \cdots).$$

The state ψ_0 is the wave function of a collection of harmonic oscillators in their ground states and is just a product of exponentials.Note that the CM is given by $x_{\rm CM} = \frac{1}{N}\sum_{n=1}^{N} x_n$; we describe the momentum state of the string in terms of the motion of the center of mass. The crucial point in this analysis is the following. For the time period in which the strings are attached, we set $x_N = x_{N+1}$, providing a constraint, and describe the wave function of the combined strings as

$$\Psi(k_1, k_2; x_1, \cdots, x_N, x_N, x_{N+1}, \cdots, x_{2N}) = \\ e^{ik_1 \frac{1}{N}\left(x_{N+1} + \sum_{n=1}^{N-1} x_n\right)} e^{ik_2 \frac{1}{N}\sum_{n=N+1}^{2N} x_n} \psi_0(x_1, x_2, \cdots, x_N)\psi_0(x_N, x_{N+2}, \cdots). \quad (37)$$

At this point we must evolve the system in time using the Hamiltonian, $\Psi(\tau) = e^{-iH\tau}\Psi(0)$, which we have already expressed above. The amplitude for the event is given by projecting the time-evolved state $\Psi(\tau)$ onto the disconnected string states with outgoing momenta k_3 and k_4. According to the Feynman rules we must calculate the amplitude[5] by summing the projection for all possible times τ, giving

$$A = \int_0^\infty d\tau\, \psi(k_3, x_1, \cdots, x_N)\, \psi(k_4, x_{N+1}, \cdots, x_{2N})\, e^{-iH\tau}\Psi(0). \quad (38)$$

One can approach the calculation of $e^{iH\tau}\Psi(0)$ by considering a few simple cases. First, a one-dimensional quantum harmonic oscillator has ground state wave function $\phi_0(x) = \alpha e^{-m\omega x^2/2}$ with $\alpha = (m\omega/\pi)^{1/4}$. This is a stationary state so $H\phi_0 = \frac{\omega}{2}\phi_0$, and $e^{-iH\tau}\phi_0 = e^{-i\omega\tau/2}\phi_0$. In the case of the string we have a collection of mass points connected by springs and the Hamiltonian is given by

$$H = \frac{1}{2m}\sum_l^N \dot{x}_l^2 + \frac{1}{2}m\omega^2 \sum_l^N \left(x_{l+1} - x_l\right)^2.$$

[5]A path integral calculation of the scattering amplitude would involve an integral over all of the trajectories $X^\mu(\tau)$ for each of the mass points. Then the amplitude would be something like

$$A = \int [DX^\mu]\, e^{-\int d\tau\, L}$$

where exponential is the Wick-rotated action for the interacting strings.

Focus on the mass point located at x_l. This piece of the Hamiltonian will be

$$\begin{aligned} h_l &= \frac{1}{2}m\dot{x}_l^2 + \frac{1}{2}m\omega^2\left[(x_l - x_{l-1})^2 + (x_{l+1} - x_l)^2\right] \\ &= -\frac{1}{2m}\partial_l^2 + \frac{1}{2}m\omega^2\left(\Delta_{l-1,l}^2 + \Delta_{l,l+1}^2\right) \end{aligned}$$

where we have defined $\Delta_{l,l-1} = x_l - x_{l-1}$. The wave function for the string is

$$\Psi = e^{ik(\cdots+x_{l-1}+x_l+x_{l+1}+\cdots)/N} e^{-\frac{m\omega}{2}\left[\cdots+(x_{l+1}-x_l)^2+(x_l-x_{l-1})^2+\cdots\right]} \tag{39}$$

We need the action of the Hamiltonian on this piece of string moving with a momentum k, i.e.

$$e^{-ih_l\tau}\Psi(0)$$

so we need

$$h_l e^{ik(\cdots+x_l+\cdots)/N} e^{-\frac{m\omega}{2}\left(\cdots+(x_l-x_{l-1})^2+(x_{l+1}-x_l)^2+\cdots\right)}$$

to give us an eigenvalue so that the full exponential $e^{-ih_1\tau}\psi = e^{-i(\text{eval})\tau}\psi$. Calculating, we have

$$\begin{aligned} e^{(\text{all that})} &\equiv e^{ik(\cdots+x_l+\cdots)/N} e^{-\frac{m\omega}{2}\left(\cdots+\Delta_{l-1,l}^2+\Delta_{l,l+1}^2+\cdots\right)} \\ \partial_l e^{(\text{all that})} &= e^{(\text{all that})}\left[+m\omega\Delta_{l-1,l} - m\omega\Delta_{l,l+1} + ik/N\right] \end{aligned}$$

$$\begin{aligned} \partial_l^2 e^{(\text{all that})} &= e^{(\text{all that})}\times \\ &\quad \left\{\left[+m\omega\left(\Delta_{l-1,l} - \Delta_{l,l+1}\right) + ik/N\right]^2 - 2m\omega\right\} \end{aligned}$$

so that

$$\begin{aligned} h_l e^{(\text{all that})} &= \left(-\frac{1}{2m}\partial_l^2 + \frac{1}{2}m\omega^2\left[\Delta_{l-1,l}^2 + \Delta_{l,l+1}^2\right]\right)e^{(\text{all that})} = \\ &+m\omega^2\Delta_{l-1,l}\Delta_{l,l+1} + i\frac{k}{N}\omega\left(\Delta_{l-1,l} - \Delta_{l,l+1}\right) + \frac{k^2}{2N^2m} + \omega \end{aligned}$$

In the limit $N \to \infty$ the spacing between points goes to zero, and $\Delta_{l,l\pm1} \to 0$, leaving us finally with the result

$$h_l e^{(\text{all that})} \to \frac{k^2}{2N^2m} + \omega$$

so that we may write

$$e^{-ih_1\tau}\psi \to e^{-i\left(\frac{k^2}{2N^2m}+\omega\right)\tau}\psi.$$

We have $N \to \infty$ oscillators, giving

$$e^{-i\left(\frac{k^2}{2N^2 m}+\omega\right)N\tau}\psi = e^{-i\frac{k^2}{2Nm}\tau}e^{-iN\omega\tau}\psi.$$

In the limit, the mass points go to zero such that $Nm \to 1$ (a constant we define to be unity), finally giving

$$e^{-i\frac{k^2}{2}\tau}e^{-iN\omega\tau}\psi$$

The wave function $\Psi = \psi_t \psi_b$ will have evolution

$$\Psi(\tau) = e^{-iH\tau}\Psi(0) = e^{-i\tau\left[\frac{1}{2}(k_1+k_2)^2+N\omega\right]}\Psi(0).$$

Recall that $s = -(k_1+k_2)^2$. The amplitude we seek (38) will have another term with $(k_3+k_4)^2$ and also has $x_{N+1} = x_N$ in the initial state before time evolution. This particular term will give $(k_1+k_2-k_3)^2$. We still need to fill in the details here (Wick rotation?). If we Wick rotate, then the infinite term $N\omega$ gives $e^{-N\omega\tau} \to 0$ and we are left with terms like $e^{-\tau s}$. The calculation of the full amplitude (38) results in

$$A = \int_0^\infty d\tau\, e^{\tau(s+1)}\left(1-e^{\tau}\right)^{-t-1}e^{-\tau}$$

which upon setting $z = e^{-\tau}$ gives

$$\begin{aligned} A &= -\int_1^0 dz\, z^{-(s+1)}(1-z)^{-t-1} \\ &= \frac{\Gamma(-s)\Gamma(-t)}{\Gamma(-s-t)}. \end{aligned}$$

This is the Veneziano amplitude (36) for the scattering of two strings each in their ground state which we recall is a tachyon state. This function has a name in mathematics, the Euler beta function $\beta(a,b) \equiv \Gamma(a)\Gamma(b)/\Gamma(a+b)$. What we have just demonstrated is remarkable; the scattering of two strings produces the amplitude of s- or t-channel processes described by Veneziano for the scattering of mesons, and does so analytically. This strongly suggests that there is validity in the idea that strings are fundamental objects in nature.

Further confidence was gained by calculating the scattering amplitudes of the photon-like and graviton-like (excited state) strings with themselves and with other types of string states. In QFT the scattering amplitudes of photon-photon and graviton-graviton scattering are distinctive and string theory reproduces those amplitudes.

3.3.1 Euler Beta function

The beta function is defined as

$$\beta(n,m) = \int_0^1 dl\ l^{n-1}(1-l)^{m-1} \tag{40}$$

We can construct a solution to this integral using the gamma function $\Gamma(t) = \int_o^\infty dt\, e^{-t} t^{n-1}$. First we write the product of two gamma functions.

$$\Gamma(n)\Gamma(m) = \int_0^\infty dt\ e^{-t} t^{n-1} \int_0^\infty dq\ e^{-q} q^{m-1}$$

We want to manipulate this integral to contain the beta integral in (40). We need to cover the right upper quadrant that is represented by the integral over t and q. We will perform a coordinate transformation $(t,q) \to (z,l)$. To obtain the correct measure we use differential forms with the substitutions

$$\begin{aligned} t &\to zl \\ q &\to z(1-l) \end{aligned}$$

giving

$$\int_0^\infty dt \int_0^\infty dq \longrightarrow \int_0^\infty dz \int_0^1 f(z,l)\ dl$$

where we need to find the function $f(z,l)$. The Jacobian can be calculated in the usual way. Here we'll use differential forms and write

$$\begin{aligned} dt &= l\ dz + z\ dl \\ dq &= dz - z\ dl - l\ dz \end{aligned}$$

$$\begin{aligned} dt \wedge dq &= (l\ dz + z\ dl) \wedge (dz - z\ dl - l\ dz) \\ &= l\ dz \wedge dz - l\ dz \wedge z\ dl - l\ dz \wedge l\ dz \\ &\quad + z\ dl \wedge dz - z\ dl \wedge z\ dl - z\ dl \wedge l\ dz \\ &= -l\ dz \wedge z\ dl + z\ dl \wedge dz - z\ dl \wedge l\ dz \\ &= -lz\ dz \wedge dl + z\ dl \wedge dz + zl\ dz \wedge\ dl \end{aligned}$$

$$\Rightarrow dt\ dq = z\ dl\ dz$$

Our double integral is transformed to

$$\Gamma(n)\Gamma(m) = \int_0^\infty dz\, e^{-z}\ z^{n+m-2}\ z \int_0^1 dl\ l^{n-1}(1-l)^{m-1}$$

or

$$\Gamma(n)\Gamma(m) = \Gamma(n+m) \int_0^1 dl\ l^{n-1}(1-l)^{m-1}.$$

Finally we have the form of the beta function we desire

$$\int_0^1 dl\ l^{n-1}(1-l)^{m-1} = \frac{\Gamma(n)\Gamma(m)}{\Gamma(n+m)}.$$

The LHS is the beta function $\beta(n,m)$. Note that when doing this manipulation the substitutions must follow the rules $t+q=z$ and $t=z\,f(l)$, and $g=z\,g(l)$, where $f(l)$ and $g(l)$ are functions of l. This allows the gamma function to come out.

4 World sheet symmetries

The trajectory for a point particle is a world line $x^\mu(\lambda)$ where λ is a parameter along the world line of the particle. A string sweeps out a world sheet. Each point along the string is parameterized by $x^\mu(\tau,\sigma)$. The path integral is written as

$$\int [Dx]\, e^{iS}$$

with

$$S = \frac{1}{2}\int d\tau\, d\sigma \left[\left(\frac{\partial x^\mu}{\partial \tau}\right)^2 - \left(\frac{\partial x^\mu}{\partial \sigma}\right)^2\right] \tag{41}$$

In order to make the path integral converge we Wick rotate $\tau \to -i\tau$.[6] An integral over all surfaces is what defines the path integral for strings. For closed strings the world sheet of a single string is a tube and interactions among closed strings look like tubes that join into larger tubes, and separate again. This is the generalization of a single vertex in point particle QFT. If one allows the strings to have any arbitrary number of holes it becomes the generalization of the sum over all Feynman graphs containing any number of loops and so on. So the path integral is a sum over all possible surfaces that connect the starting positions and end positions of a set of strings.

The equation of motion for (41) yields the wave equation. After Wick rotation the signs of the derivative terms are the same and the equation of motion will instead be the Laplace equation

$$\frac{\partial x^{2\mu}}{\partial \tau^2} + \frac{\partial^2 x^\mu}{\partial \sigma^2} = 0$$

The symmetries of the Laplace equation include reparameterization of the coordinates $\sigma' = (\sigma,\tau)$ and $\tau' = (\sigma,\tau)$. In particular, the kinds of coordinate transformations that preserve the Laplace equation are conformally invariant transformations. Conformal transformations preserve angles, and infinitesimal squares are mapped

[6] The generic path integral in QFT is given by something like $\int [Dx]\, e^{i\int_{t_1}^{t_2} dt\, L(x,\dot{x})}$ where the weight for each path is $e^{iS(x)}$ with $S(x) = \int_{t_1}^{t_2} dt\, L(x,\dot{x})$ and therefore has the same magnitude (they are points on the unit circle in the complex plane). Generally, this integral is hard to define mathematically and is non-convergent. The standard prescription is analytically continuing the path integral by setting $t = i\tau$ so that the path integral becomes $\int [Dx]\, e^{-\int_{-it_1}^{-it_2} d\tau\, L(x,\dot{x})} = F(it_1, it_2)$ and then analytically continuing the result back.

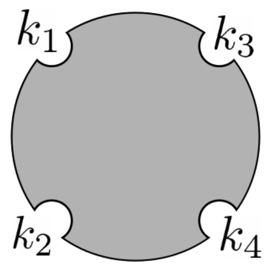

Figure 4: Conformal mapping of the two-string interaction in Fig. 3 to the unit disk. The interior of the disk is the interior of the shaded string region in Fig. 3 and the injection points are shown as exaggerated cutouts in the disk. These are mapped to infinitesimal points on the exterior of the disk.

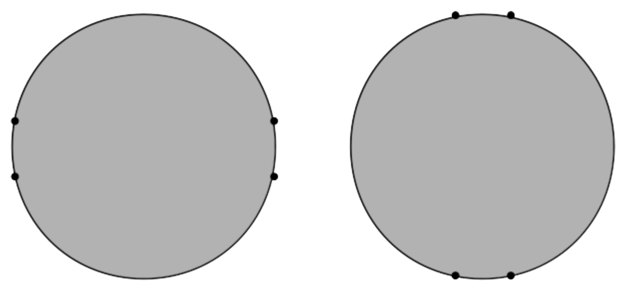

Figure 5: s- and t-channel scattering maps. The τ parameter determines the symmetric locations of the four vertex operators. For $\tau = 0$ the points would move closer and overlap on the left. For $\tau \to \infty$ the point on the right map would move closer and overlap.

to infinitesimal squares. In particular the shaded region in Fig. 3 is topologically equivalent to a unit disk with four special points (vertex operators, or insertions) corresponding to the string pieces that come in from minus infinity and go out to plus infinity. The special points are injection points for the momenta k_i. The s- and t-channel scattering

4.1 Conformal mapping

Conformal mapping is usually introduced in electrodynamics in physics to map 2D electrostatics problems into one another as an aid in obtaining solutions.[7] These

[7]Incidentally, for a charge confined to 2D, where the flux is also confined to the 2D surface containing the charge, the electrostatic force $F \sim 1/r$, and the potential is then $V \sim \ln r$. This is easy to see by contrasting with charges in 3D. The flux eminating from a 3D charge would be measured by surrounding the charge with a sphere whose area is $4\pi r^2$ and so the flux escaping a

techniques are standard methods used in elementary complex variables so we will summarize the basics quickly. Consider a mapping $w(z)$ that takes the complex z plane to the complex w plane. Where $z = x + iy$ and $w(z) = u(x, y) + iv(x, y)$. We assume that the mapping is one-to-one and onto (bijective). In defining the derivative of the mapping, $w'(z)$ we examine the quantity

$$w'(z) = \frac{dw}{dz} = \frac{du + idv}{dx + idy}.$$

For the derivative to be uniquely defined, it must be the same if approached along any direction in the z plane. It is easiest to come in along the x and y axes independently, so that $dy = 0$ and $dx = 0$, respectively. Setting these two derivatives equal yields the Cauchy-Riemann equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad\qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}. \tag{42}$$

Mappings that satisfy the Cauchy-Riemann equations are called *analytic* or *holomorphic*. Setting the mixed partials of these equations equal shows that both of the functions $u(x, y)$ and $v(x, y)$ satisfy the Laplace equation, $\nabla^2 u = 0$ and $\nabla^2 v = 0$.

We will now show that any analytic mapping is a conformal mapping, i.e. it will preserve angles. For this we consider two small arrows in the z plane, $\delta z = \rho e^{i\theta}$ and $\Delta z = \rho' e^{i\theta'}$ starting at the point z. The ratio of these displacements is $\delta z/\Delta z = (\rho/\rho')\, e^{i(\theta-\theta')}$, indicating that the angle between these displacements is $(\theta - \theta')$. For an analytic mapping $w(z)$, we stipulate that the derivative $w'(z)$ exists. Then $w(z + \delta z) = w(z) + \delta z\, w'(z)$ and $w(z + \Delta z) = w(z) + \Delta z\, w'(z)$. Therefore we have the following ratios equal

$$\frac{\delta z}{\Delta z} = \frac{\delta w}{\Delta w}$$

and the mapping is angle preserving. A few simple examples of analytic functions are $f(z) = z$, and any power of z^n, except $1/z$, which has a pole at the origin, but is analytic everywhere else. Typical functions such as e^z and $\sin(z)$, are analytic. An example of a non-analytic function is $f(z) = z^*$. It is easy to see that this function does not satisfy the Caucy-Riemann equations. However, it is important in string theory to consider functions that are *anti-analytic*. These are functions that are analytic in the variable z^*, often denoted with a bar as $\bar{z}$. The Cauchy-Riemann equations for anti-analytic functions differ by a minus sign from the usual definition

small patch of area da on the sphere must be $da/4\pi r^2$. In 2D, the flux would pass throug a circle of 'area' $2\pi r$.

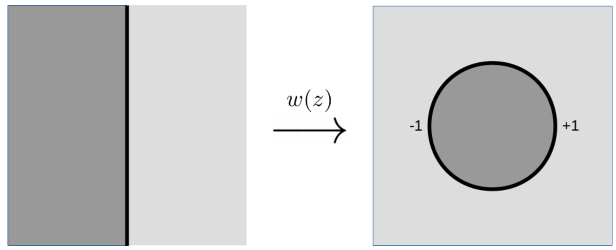

Figure 6: Linear fractional map $w(z) = \frac{z+1}{z-1}$.

(42). If $f(\bar{z}) = \bar{z} = x - iy$, then $u = x$ and $v = -y$, so that we must have $\frac{\partial u}{\partial x} = -\frac{\partial v}{\partial y}$.

There are two special mappings that we now consider. The first is the function $w(z) = \log(z)$, where we write $z = re^{i\theta}$ and consider the mapping of the right half-plane. From the properties of the logarithm we have $w(z) = \log r + i\theta$. The vertical line from the origin at $z = 0$ to $+i\infty$ takes $\log r$ from $-\infty \longrightarrow +\infty$ (this is the same for any radial line from the origin out to infinity) at angle $\theta = \pi/2$. In the w-plane this is a horizontal line at height $i\pi/2$. The vertical line from the origin to $-i\infty$ is mapped to a horizontal line at $-i\pi/2$. Thus the right half z-plane is mapped into a strip in the w-plane that extends horizontally from $-\infty$ to $+\infty$, going from $-\pi/2$ to $+\pi/2$ in the vertical direction. Half circles (fixed radius) in the right half-plane are mapped to vertical lines in the strip. This mapping is useful for open strings where the σ parameter ranges from 0 to π. If one maps the entire z-plane with the logarithm one obtains a strip in the w-plane that is 2π wide, and where the top of the strip is identified with the bottom of the strip. This describes a cylinder and is useful for closed strings.

The second mapping we consider is the linear fractional mapping $w(z) = \frac{z+1}{z-1}$. This maps the point $z = 0$ to $w = -1$, and maps $z \to \infty$ in any direction (the "point at infinity") to $w = +1$. The vertical y-axis is mapped to a half circle in the w-plane going from -1 to +1. This can be seen by writing $z = re^{i\theta}$ for $\theta = \pi/2$, then $w(z) = -\left(\frac{1+ir}{1-ir}\right)$. As $r \to \infty$ this goes to +1, and in between there is a small excursion into the positive complex part of the w-plane. The negative y-axis is mapped to a half circle in the negative imaginary direction. Along the positive x-axis the mapping hits a pole at $z = 1$. The mapping takes (0,1) to (∞,1) and takes (1,∞) to (-1,$-\infty$), never covering $u \in (-1, 1)$. Overall the mapping takes the left half-plane to the interior of the disc, and the right half plane to the exterior of the disc.

The defining path integral for the open string scattering amplitude can then be written as

$$A = \int \left(\prod_j dz_j \right) \int DX\, e^{-\int d\tau d\sigma \left[\frac{\partial X_\mu}{\partial \tau} \frac{\partial X^\mu}{\partial \tau} + \frac{\partial X_\mu}{\partial \sigma} \frac{\partial X^\mu}{\partial \sigma} \right]} \prod_j e^{ik_\mu^{(j)} X^\mu(\vec{z})} \tag{43}$$

where the integral over the z_j is the integral of the vertex operator positions on the boundary of the unit disc that inject strings of momentum $k^{(j)}$ into the world sheet. By conformal mapping three of these positions can be fixed, but the rest must be integrated over. The number of vertex opertors corresponds to the number of strings interacting. This is a genus 0 sheet, it captures all s- and t-processes and anything that can be deformed topologically without tearing the sheet. So this takes care of all tree-level processes.

The action in (43) is conformally invariant and the equation of motion for each field X^μ satisfies the Laplace equation. The path integrals are all Gaussian and can be done independently for each of the 26 fields. The problem is therefore reduced to solving the Laplace equation on the disc with the charges being the $k_\mu^{(j)}$. In the case where there are four strings interacting, we have $k_\mu^{(1)}, \cdots, k_\mu^{(4)}$ and 26 kinds of "charge," one for each dimension. What is being calculated is the analog of the electrostatic energy for each field X^μ. The final result is that the amplitude is given by

$$A = \int \left(\prod_j dz_j \right) e^{-\sum_j E_j(\vec{z})} \tag{44}$$

where $E_j(\vec{z})$ is the energy of the j-th type charge which depends on the locations of all the other injected charges of the same type. This is the simplest form of open string theory and (43) can be taken as a definition of this theory.

This result (44) is calculated for strings in 26 dimensions and the mathematics so far has been relatively simple. The real complexity in string theory comes in with the compactification of these extra dimensions to leave only the 3+1 dimensions that we see. The number of ways of compacting these dimensions is very subtle and mathematically difficult. The compactification allows one to put gauge group symmetries into string theory thus giving a connection the Standard Model. However, the number of ways of compactifying results in a large number of possible "standard models" of the order 10^{500}. The current problem with string theory is that there seems to be no way of picking one and that they may occur randomly.

4.2 Compactification

Let us contrast the difference between a point particle and a string moving on a surface embedded in a large flat space. A point particle that is constrained to remain in the surface with no other forces on the particle will move along a geodesic. Classically, we can divide up the trajectory of the particle into small pieces, and in the limit we have a continuous line. Quantum mechanically the picture is different.

On a sphere of radius R a point particle with no other forces would travel on a great circle. The energy of the center of mass of the particle would be $mv^2/2 = p^2/2m$. Because the angular momentum is quantized, and because $L = mvR = pR$, we have that

$$\begin{aligned} E &= L^2/2mR^2 \\ &= L^2/I, \end{aligned} \tag{45}$$

where the moment of inertia is $I = mR^2$. For a string moving on the geometry of a sphere things are much different. We have to consider the spatial extent of the string due to its zero point motion. We could calculate the root mean square by calculating $X_{\text{rms}} = \sqrt{\langle 0|X^2(\sigma)|0\rangle}$. Using the creation and annihilation operators we have

$$\langle 0| \sum_{nn'} \frac{1}{\sqrt{n}\sqrt{n'}} \left(a_n^+ + a_n^-\right) \left(a_{n'}^+ + a_{n'}^-\right) |0\rangle$$

and the only contributing term will be $a_n^- a_{n'}^+$ with $n = n'$. This will give $\sum_n \frac{1}{n}$ which is formally divergent. Again we regulate the sum using a cutoff $\sum_n^{n_{\max}}$ and look for the dependency of the energy in (45) on the regulator. The integral approximation to the sum is $\sum_n^{n_{\max}} \sim \ln(n_{\max})$. On a sphere, when the higher ZP modes are included in the string X_{rms} begins to cover the surface of the sphere and the CM of this "string tangle" moves closer to the center of the sphere and in the limit of including all the modes we find $I \to 0$, and (45) becomes divergent. We interpret this to mean that string motion in this geometry is not allowed; the sphere is a bad geometry for strings to move in. In order to find a suitable geometry we have to consider the metric, $g_{\mu\nu}(x)$, of the surface it moves in. Because the string itself has energy and can deform the metric of the space, we need the change in the metric of the space to be well-behaved with the regulator introduced above. This will require that the space be "Ricci flat." We can write this as

$$\delta g_{\mu\nu} = R_{\mu\nu\alpha}{}^{\alpha} = R_{\mu\nu}. \tag{46}$$

This is the variation of the metric with a change in the regulator. We want this to be zero.

$$\delta g_{\mu\nu} = R_{\mu\nu} = 0 \tag{47}$$

Now consider the Einstein equations for the vacuum,[8]

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 0.$$

The scalar curvature $R = R_\alpha^\alpha$ is the trace of the Ricci tensor. Raising one index in the free field equation we have

$$\begin{aligned} R_{\mu\nu} g^{\nu\alpha} - \frac{1}{2} g_{\mu\nu} g^{\nu\alpha} R &= 0 \\ R_\mu^\alpha - \frac{1}{2} g_\mu^\alpha R &= 0 \\ R_\mu^\alpha - \frac{1}{2} \delta_\mu^\alpha R &= 0. \end{aligned}$$

Now form the trace by setting $\alpha = \mu$ and summing, giving

$$\begin{aligned} R - \frac{4}{2} R &= 0 \\ R &= 2R. \end{aligned}$$

Clearly this can be satisfied only for $R = 0$, or the space is Ricci flat. This means that the consistency condition (46) for strings moving in a spacetime background requires that the background space be a solution of the free field Einstein equations. This is remarkable.

Compactification of a surface can be done many ways, and we begin with an example of a 2D spatial sheet of finite extent. If we take a rectangle and identify the left side with the right side we obtain a cylinder. If we take the same rectangle and identify the lhs with the rhs and the top with the bottom, we obtain a torus. Topologically the two are identical if we remember the identifications required on the rectangle. If we start with a 3D rectangular parallelopiped, infinite in one direction, and identify the front with the back and the top with the bottom (the compact

[8]The full Einstein field equations have the energy-momentum tensor $T_{\mu\nu}$ on the rhs that contains matter and everything other than the gravitational field. Solutions to the vacuum equations include, in particular, gravitational waves, e.g. think about the vacuum Maxwell equations that admit wave solutions. Other "solutions" include Schwarzchild, and FRW, but those have singularities. The de Sitter spaces are solutions to the vacuum equations provided there is an additional cosmological constant term.

directions), we have a torus at each point of the non-compact direction. This can be imagined as a 1D line with a torus attached at each point of the line. Similarly we could imagine another topology consisting of a line with a sphere attached at each point of the line. However, we have seen above that a sphere is not a good geometry for strings to move on. Toroids are the simplest and most direct way to compactify dimensions, and this process is called "toroidal compactification."

There are three parameters that describe a 2-torus. If we unwrap a 2-torus we first cut it and unwrap it to form a cylinder, and then cut the cylinder lengthwise to a rectangle of sides a and b. The three parameters are (1) the area ab, also called the Kahler modulus, (2) the ratio of the sides a/b, is the complex structure modulus, and (3) the identification "angle" or parameter that describes any twist applied to the cylinder before reconnecting it to form the 2-torus is called the Dehn twist. These are called the moduli, and make up the *moduli space* of the torus.[9]

Other Ricci flat manifolds called Calabi-Yao manifolds are good manifolds for strings to move on. They appear to provide more of a connection to the particle physics we know because they are less symmetric than tori, and the tori are too symmetric.

4.3 Energy spectrum and winding number

Consider a point particle moving on a 2D surface which is compact in one dimension, e.g. a long cylinder of radius r. It can have two components of momentum, one along the non-compact direction and one along the compact direction. Momentum along the compact direction is quantized. This can be seen using the usual periodic boundary condition for a plane wave, $e^{ikx} = e^{ik(x+2\pi r)}$, and so $2\pi kr = 2\pi n$ for some integer n. Therefore $k_m = n/r$. For a massless particle $E = |p|$, and so $E \sim n/r$ for $n \in \mathbb{Z}$. Because energy is mass, the mass spectrum of these particles will have spacing $1/r$. As $r \to \infty$, then the spacing between the states is very small. This particle can also be a string of small spread.

Now consider a string wrapped around the compact dimension. Its length has an energy density, so the number of times the string is wrapped around the compact dimension gives its energy, and this goes as nr, where n is called the *winding number*. Negative n correspond to the direction in which the string is wound. So we have two types of mass spectra, the non-wound string that goes as n/r, and the wound

[9] There are non-orientable surfaces such as the Klein bottle that can be made by identifying the cylinder edges before reconnecting, but they are not good surfaces for strings to move on. The surfaces must be Ricci flat.

string that goes as nr. Note that the spectrum of these two pictures exchange and go into each other as $r \to 0$ or $r \to \infty$. These pictures are dual to each other by exchanging the n of the quantized momentum with the winding number and exchanging $r \longleftrightarrow 1/r$. This is a symmetry of the theory. At the radius $r = 1$ the spectra are the same. As one reduces the compact dimension length scale below $r < 1$, the non-wound theory just looks like the string theory with winding number in a space with $r > 1$.[10] This is called T-duality where the T is for torus.

The winding number is a conserved quantity. To show this we remember that for closed oriented strings, there is a sense of direction along the string. Take two strings of winding number $w = 1$ and $w = -1$ wrapped around the compact direction in opposite directions. If the strings are allowed to interact and combine, we must follow the rule that the sense of direction of a string cannot change along the string. Then when the strings combine, the only way for them to do so is to form one string that is no longer wound around the compact direction. The total winding number 0 is conserved. Drawing a picture is easy and it will be obvious. If instead we wrap two strings both with $w = 1$, and let them combine to form one string it is easy to see that the winding number of the single string is 2.

The mathematics of the T-duality can be described as follows. For an unwound string with (quantized) momentum along the compact direction we write $E = p = n/r$, but $p \sim dy/d\tau$ where $y(\sigma)$ is the coordinate in the compact direction at the point σ along the string. So we can write

$$n = r \int d\sigma \frac{dy(\sigma)}{d\tau}$$

where the integral over $d\sigma$ accounts for different parts of the string moving at different velocity. Similarly, we can consider the winding number of a string. The coordinate along the string σ goes from $0 \to 2\pi$ for a closed string. The circumference of the compact space coordinate is $2\pi r$, therefore we can write $\sigma = y/r$ so that as y goes from $0 \to 2\pi r$, σ goes from $0 \to 2\pi$. Here the energy $E = w2\pi r$ is proportional to r and we write

$$w = \frac{1}{r} \int d\sigma \frac{dy(\sigma)}{d\sigma}.$$

T-duality takes $n \to w$, $r \to 1/r$, $\tau \to \sigma$.

[10]This implies that there is a smallest length scale.

5 The Kaluza-Klein Picture

Closed strings on a compact manifold exhibit repulsion and attraction; they behave like charged particles. If two strings attract or repel each other in this way it is natural to ask how. For simplicity we consider 5 dimensions, one time and four space, one of which is compact. This is the Kaluza-Klein (KK) picture that is able to combine electromagnetism with general relativity. The mathematics is also used in string theory. Here we are considering closed strings. We will see that closed strings with momentum number $+n$ in the compact direction attract those with momentum number $-n$, and likewise for strings with winding number $+w$ and $-w$. Repulsion occurs between objects whose momentum number or winding number is the same sign.

5.1 Kaluza-Klein metric

The generalization to an extra compact dimension of space was first proposed by Kaluza and later refined by Klein. The idea is simple. Instead of the usual GR metric $g_{\mu\nu}$ where $\mu, \nu \in \{0, 1, 2, 3\}$ we define a more general metric tensor

$$g_{MN} = \begin{pmatrix} g_{\mu\nu} & g_{\mu 5} \\ g_{\mu 5} & g_{55} \end{pmatrix}. \tag{48}$$

We want to find an analog electric or magnetic field to describe the attraction and repulsion of these strings. Recall from electromagnetism that the fundamental object is the vector potential $A_\mu = (\phi, A_1, A_2, A_3)$ where $\vec{B} = \nabla \times \vec{A}$, and $\vec{E} = -\nabla \cdot \phi - \partial_t \vec{A}$. The vector potential A_μ is a four-vector. The photon is a vector particle and has a vector index corresponding to the μ. The KK metric tensor contains two new fields, $g_{\mu 5}$, a four-vector that can be identified with A_μ, and g_{55}, a scalar Φ called the dilaton. The dilaton or g_{55} is the component of the metric around the 5-th direction and tells you how big the 5-th direction is. If g_{55} is big or small, then so is the 5-th direction. Both new fields are functions of the spacetime coordinates $g_{\mu 5} = g_{\mu 5}(x)$ and $g_{55} = g_{55}(x)$. If g_{55} can vary, then the compact dimension can vary and looks like a tube of varying radius. These represent waves in the compact direction or particles associated with the dilaton. The quiescent vacuum would have no such waves and the radius would be constant.

Recall the definition of the energy-momentum tensor $T_{\mu\nu}$, the time components T_{0j} correspond to the momentum flow in each spatial direction. In the KK metric the terms $g_{\mu 5}$ correspond to momentum flow in the 5-direction. The source of the analog vector potential is the component of momentum going around the 5-direction. The momentum quantum number $\pm n$ can be though of as electric charge corresponding to the vector field $g_{\mu 5}$. This electric field is associated with the graviton. The momentum number n, and the winding number w correspond to two *kinds* of electric charge, giving rise to n-photons, and w-photons. Momentum photons are a piece of the gravitational field $g_{\mu 5}$, gravitons which are polarized along the μ5-direction. The more general picture including photons emitted and absorbed by the winding number are as follows. Recall the spectrum of closed strings where we had to use level matching. The creation and annihilation operators can now create oscillations in the usual three directions of the non-compact space. These are $\left(a^i_{\pm n}\right)^{\pm}$ where $i \in \{x_1, x_2, x_3\}$ and the left and right operators are given by $\pm n$. The tachyon has -2 units of energy and we ignore it. To get a non-tachyonic state we must create two units of energy and form states like $a^{i+}_{-1} a^{j+}_{+1} |0\rangle$, corresponding to gravitons polarized in two directions in space. We can also form excitations like $a^{i+}_{-1} a^{5+}_{+1} |0\rangle$, which has one index and can be identified with the polarization of the photon since we have only one index in this state. We can also form $a^{i+}_{+1} a^{5+}_{-1} |0\rangle$. The correct thing to do is to form linear superpositions and form two kinds of photon fields

$$\left(a^{i+}_{+1} a^{5+}_{-1} \pm a^{i+}_{-1} a^{5+}_{+1}\right) |0\rangle$$

where $+ \leftrightarrow n$-photons, and $- \leftrightarrow w$-photons. The n-photons are identified with $g_{\mu 5}$, gravitons, and the Ramond field $b_{\mu 5}$ is associated with the w-photons or winding number. There is also a combination of creation operators that have no index structure and correspond to a scalar particle, $a^{5+}_{+1} a^{5+}_{-1} |0\rangle$, the field quanta of the dilaton field Φ.

T-duality can be stated

$$\begin{aligned} n &\leftrightarrow w \\ r &\leftrightarrow 1/r \\ \partial x/\partial \tau &\leftrightarrow \partial x/\partial \sigma \\ g_{\mu 5} &\leftrightarrow b_{\mu 5}. \end{aligned}$$

At the moment these are mathematical constructions and real particle candidates are not known for all of these particles. However, there is plenty of room in physics for these objects.

5.2 D-branes

T-duality for open strings leads to the idea of Dirichlet branes or D-branes, discovered by Polchinski. The D-branes can have different dimensions and a Dn-brane has n dimensions. Consider the space $\mathbb{R}^3 \times S^1$ where the compact space is S^1 with coordinate y and take it to be very small. Suppose a string lives in this space. The Neumann boundary condition on the string is $\partial y/\partial\sigma|_{0,\pi} = 0$, which means there is no stretch of the string at the very endpoints. If we insist that T-duality holds for the string, then in the dual space where S^1 is very large we have the dual condition that $\partial y/\partial\tau|_{\sigma=0,\pi} = 0$, or that the string end points cannot move; the string is pinned in the S^1 space. This object that has dimension 1 anchors the end of the string and can be thought of as a 1-dimensional object that is heavy, because strings attached to it cannot leave it. These objects have to be deformable to make sense from a general relativistic point of view, and so they are objects that have their own dynamics. A D0-brane is a point, a D1-brane is a string (though it is heavy and not like the strings that are attached to it). A D2-brane is a membrane, and a D3-brane is a 3-dimensional space where strings attach. If we think about open string theory in our 3-dimensional world we would say the strings are attached to a space-filling D3-brane where they move all about the space because the endpoints are attached to the whole space. One can say that open string theory in which T-duality is imposed predicts D-branes.

The strings attached to a D-brane have an orientation that gives rules for joining and breaking of strings for scattering amplitudes. Each string attached to a brane has an arrow along its length. If two strings attached to a D2-brane approach each other the only ends that can join are the two of opposite orientation. In other words two strings cannot join that have arrows coming out of the points. That would lead to a string that had two arrows along its length of opposite direction. This leads to strings conserving orientation number.

QCD has a beautiful connection to D-branes and string theory in the following sense. Consider three separate D3-branes labeled R, G, B that are embedded in a higher dimensional space. Then a string attached to the R3-brane has points labeled $r, \bar{r}$, where one has an arrow going into the brane and the other an arrow going out. Strings can attach at one brane and go to another type of brane so there could be $r\bar{b}$ strings, and so forth. Incidentally there are 8 gluons and not 9 because one linear combination can come together and disappear. For example the ends of an $r\bar{r}$ gluon can come together and lift off the sheet and thus is not stable. An isolated quark in this picture would be a string that starts one of the RGB sheets and goes off to some distant other brane of a different kind. These rules are the same as super Yang-Mills QCD rules.

Supposing we have only one D3-brane the structure of the theory is like electromagnetism. In this picture a string with both ends attached to the brane is a photon like object, and a string with one end in the brane and one end at infinity is a an electron. A positron would have the opposite orientation. The theory predicts in this picture that a D1-brane that ends on the D3-brane is a magnetic monopole. It is heavier than the usual string that ends on the D3-brane that represents the electron or positron.

Finally, if open strings are attached to D-branes, and if our physical world is a D3-brane with open strings attached. Then open strings may describe all of the Standard Model particles, and closed strings which do not attach to these branes may describe the gravitons. These gravitons can wander freely over all of the spaces and this may explain why gravitational forces are so much weaker than the others; gravity can "leak out" into the other dimensions.

Part II

Topics in String Theory

This part of the document follows the first several lectures of Susskind's "Topics in String Theory" lectures.

6 Fundamental Objects

In quantum field theory there is a well-known picture associated with bosonization that indicates there is difficulty in determining what objects are fundamental or not. Consider two spin-1/2 fermions with coupling g. A field ψ describes the fermion field. If $g \ll 1$ then the fermions move freely and are considered the fundamental particles. If the coupling is very strong, $g \gg 1$, then the fermions look like a composite particle that can have a ground state of spin 0 and look like a boson. In this picture the boson would be considered the fundamental particle. The boson field ϕ is an effective description of the fermions. The boson field creates pairs of fermions. The bosons cannot be paired together to look like fermions because any pair of two bosons is also a boson. However, there is a completely different way to think about the fermions in terms of the bosons in terms of *kinks in the boson field*. The field ϕ smoothly jumps over some small distance. A good example is a belt that is twisted 2π over some small distance. Kinks are heavier than the bosons that make up excitations in the field ϕ, and these kinks behave like fermions.

For some ranges of the parameters in the field theories ψ and ϕ it is easier to think of ψ is the fundamental field and its building blocks are fundamental, and for other parameters of the field theory the bosons are more useful to think of as the fundamental objects. Here the bosons would behave more simply for example.

6.1 The electron and the monopole

A more sophisticated example is given by quantum electrodynamics. Here we think of the electron as a fundamental particle. The fine structure constant $\alpha \sim e^2$, with a numerical value of $\approx 1/137$, is the coupling constant that describes the probability of an electron emitting a photon if it is accelerated. The electron only emits photons perhaps 1% of the time. It takes very large energies to probe the structure of the electron and infer the cloud of virtual photons and e^+e^- pairs that surround it. From Dirac's famous calculation of the charge of the magnetic monopole, $g = 2\pi/e$, we see that the magnetic monopole will interact with photons much more strongly. In doing so there will be a complicated cloud of virtual photons, and electron-positron pairs that surround it. The monopole is therefore much heavier than the electron. If we consider the electric charge e as a parameter, then as we vary the parameter from very small to very large, the electron and the monopole trade places in the complexity of the cloud that surrounds them. The monopole would become very light and the electron very heavy. It is therefore not entirely clear then what object is considered fundamental.

6.2 D-branes and S-duality

In string theory the presence of D-branes is part of the mathematical framework and these objects must exist as discussed in the section above. We will now discuss two versions of string theory. In one theory there are only even dimensional D-branes and in the other only odd dimensional D-branes. If we consider first the odd dimensional version, then there are no D0-branes, but there are D1-branes which we discussed as being much heavier than fundamental strings. In string theory the parameter that allows for duality is the string coupling constant, g. This constant controls the probability of strings breaking and connecting.

In a theory that contains D1 branes the same behavior is present, the D1 string has lots of fundamental strings (F-strings) attached to it, and so it is a heavy object. In this type of theory a fundamental string and a D1 string will morph into each other based on the value of the coupling constant g of the theory. If $g \ll 1$ the F-strings are very light. As g becomes large the F-string gets heavier and so the D1 string cannot "emit" as many F-strings and so it starts to look more simple. At the same time any bits of string that break off from the F-string will form closed loops and be attracted around the F-string. The closed loops form the particles like the graviton, etc. As $g \gg 1$ these become very massive and turn into black holes. The D1 strings then get lighter and they make up the gravitons. This is called S-duality for "strength duality", the strength of the coupling constant. In string theory g can vary in space and it satisfies field equations.

In a string theory that contains D0 branes they behave like heavy objects because they have lots of fundamental strings attached to them. The look like little loops that come out of and return to the D0 brane, like a fluffy ball. Imagine a 3D space where one of the dimensions is compactified that looks like a thick sheet. As g starts to increase, the compact dimension decompactifies. The D0 branes become lighter and turn into the gravitons in the decompactified space. In 10D string theory this results in 11 total dimensions. The F-strings in the original 2D space (with compactified 3rd dimensioin) that form a loop for example will, upon g getting larger, form ribbons that are stretched in the direction that is becoming decompactified. These are D2 branes. There are also D2 branes that float between the two noncompact spaces. This is the nature of the dualities in string theory.

D3-branes can act as our usual 3D space. One can imagine that two strings attached to the D3-brane can come together and join, move around, and then break again. This looks like the particle scattering that we are familiar with from quantum field theory.

There are other objects in string theory in addition to D-branes. There are fluxes, and other objects. The spaces of real interest in string theory are Calabi-Yao manifolds. They have handles or holes. A relatively simple Calabi-Yao manifold with genus 500, then around each handle we can wrap fluxes, D-branes, and so forth. This configuration space is what gives the roughly 10^{500} possible ground states for string theory that result in the many different possible particle physics we see. It could be that particle physics is so complicated because there are lots of moving parts in the string theory, even though the underlying rules in the string theory are relatively simple. If string theory is correct then particle physics may be so complicated that we will not be able to completely unravel. The figure 10^{500} is actually a conservative number and of this large number there may be very many that look similar and so we may never be able to tell the difference between individual choices, but we may be able to only know the region of the configuration space we may be in.

Do we live in a compactified space? We don't know. Old-fashioned cosmological theories were that we live on the 3D surface of S^3, so all dimensions were compact! Observationally we can only say how flat the universe looks.

7 The Schwarzchild metric

Let us consider first the Euclidean and Minkowski metrics. The Euclidean metric is $ds^2 = dx^2 + dy^2 + dz^2$ and is the usual generalization of the Pythagorean theorem. There is no time in the Euclidean metric. The Minkowski metric is $ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2$ where time and space have different "signature." In terms of proper time we would write $d\tau^2 = dt^2 - (dx^2 + dy^2 + dz^2)/c^2$.[11] It is seen here that the speed of light is just a conversion factor for the dimensions. Light rays follow null trajectories and in the Minkowski metric this means that $d\tau^2 = 0$, and in one space dimension we have $0 = dt^2 - dx^2/c^2$ or $dx/dt = \pm c$ for left or right moving light rays. *The universal statement in physics is that the proper time along a light ray is zero.* We have just seen in Minkowski space, which defines special relativity, that this implies the constancy of the speed of light.

From here on we set $c = 1$. Writing the Minkowski metric in spherical coordinates we have $d\tau^2 = dt^2 - (dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta\, d\phi^2)$ which we write as $d\tau^2 = dt^2 - dr^2 - r^2 d\Omega_2^2$ where $d\Omega_2^2$ is the metric for a unit 2-sphere $d\Omega_2^2 = d\theta^2 + \sin^2\theta\, d\phi^2$. The Schwarzchild metric is a spherically symmetric line element describing a black hole from the point of view of a stationary observer at spatial infinity. The time for the observer at infinity is called coordinate time. The metric is given by

$$d\tau^2 = \left(1 - \frac{2MG}{r}\right) dt^2 - \frac{1}{\left(1 - \frac{2MG}{r}\right)} dr^2 - r^2 d\Omega_2^2 \tag{49}$$

where $r_S = 2MG$ is the Schwarzchild radius and is the radial location of the horizon of the black hole. We can examine how a radial light ray behaves by looking at the

[11]There are two conventions when writing metrics. The invariant interval is what it is, invariant. However if one is interested in time-like events one uses the proper time convention $d\tau^2 = dt^2 - dx^2 - dy^2 - dz^2$, so that the proper time comes out positive. The proper-distance convention is for investigating space-like phenomena and one puts in a sign change, writing $ds^2 = -dt^2 + dx^2 + dy^2 + dz^2$, so that the proper distance is positive.

null trajectory and ignoring the angular part of the metric. We have then

$$\frac{dr}{dt} = \pm\left(1 - \frac{r_S}{r}\right)$$

and outside the horizon light rays, both ingoing and outgoing, have smaller and smaller velocities in coordinate time. The constancy of the speed of light is broken in curved space! From the perspective of the observer at spatial infinity, ingoing objects never cross the horizon but rather get ever closer; the horizon has a "sediment" of flattened objects just outside. Even a light ray shined into the black hole never crosses the horizon from the perspective of the observer at spatial infinity. From the perspective of the observer falling into a sufficiently large black hole the experience is unremarkable and they simply cross the horizon. At $r = r_S$ there is a type of singular behavior but the real coordinate singularity is at $r = 0$ where the tidal forces become infinite. We will come back to the horizon and what happens there later.

7.1 Black holes

7.1.1 Information and entropy

The idea of entropy goes back to the nineteenth century and the work of Boltzmann who put it on firm mathematical grounds. The more recent ideas in information theory began with the work of Claude Shannon in the 1940's. The information content of a system can be determined by asking a series of "yes" or "no" questions about the system and forming a string of 1's and 0's to represent the answers. For a macroscopic object this list would be very large and could contain for example a list of whether there is a molecule or not in each small subdivision of the object.

The essential concept behind entropy is that it is the amount of inaccessible information content in a physical system. Take the following example. Consider two large tanks that are half full of water. In the first tank one drips in water in Morse code that contains the Feynman lectures on physics, and in the other tank one drips in the MTW treatise on gravitation. After some time the tanks become quiescent. By all appearances the water in the tanks looks identical.[12] In fact, the information is still in the water, but it is contained in degrees of freedom so numerous and so difficult to measure that it is for all practical purposes inaccessible. The entropy, temperature, and energy of a system are related by $dE = TdS$. A modern definition

[12]The rule in physics is that information never disappears it simply moves from one form to another. For this rule one needs the concept of ergodicity in the phase space of the system where phase space trajectories never cross; the trajectory is required to be microscopically time-reversible.

of the temperature of a system is the following. *The temperature of a system is equal to the change in energy of a system when adding one bit of information.* If $\Delta S = 1$, then $\Delta E = T$.

A black hole that absorbs energy (information, if thought of as the bits of information that are arranged in the Planck squared array on the surface of the hole) will eventually radiate that energy out. So the information that fell into the black hole is evaporated back out through Hawking radiation. This fact has to be reconciled with the notion that nothing can escape from behind the horizon. This is a major puzzle of quantum theories of gravity.

The conservation of information is more fundamental than the conservation of energy or momentum. It is so fundamental that most people forget about it. Without conservation of information there would be no conservation of energy or momentum. Consider an object in empty space where there is no friction whatsoever. Suppose that the natural law was that objects given a push will come to rest because they followed equations of motion that said they did. It follows then that no matter what direction the object was pushed it will come to rest without any interaction with anything. If the object "forgets" where it came from there is no time reversal symmetry. From time reversal symmetry follows energy conservation, so there is no energy conservation.[13] Also, there is no momentum conservation if it comes to rest. Is translation invariance broken? No, but then the statement that momentum conservation follows from translation invariance is broken. For momentum conservation it is necessary to have both information conservation and translation invariance.

In short, information conservation is probably the most fundamental idea in physics. It is similar to, but distinct from, the idea of causality. These are both meta-principles without which there is essentially no place to begin thinking about physics. In classical physics the statement of information conservation is the same as Liouville's theorem that the phase space volume of a system does not change. In quantum mechanics the conservation of information is called unitarity: orthogonal states remain orthogonal with time, which means that distinctions are always preserved. Unitarity is deeply connected with energy conservation in quantum mechanics.

7.1.2 Black hole quantum mechanics

Black holes are objects. Are they objects that violate quantum mechanics? Do things fall into a black hole? If they go in then all the information they take into the black

[13]Clearly if an object simply comes to rest and there is no friction to slow it down, then there is no energy conservation.

hole is lost inside. If things don't fall in it appears to be stuck at the horizon. However black holes evaporate over time, so what happens to the information? Does it get radiated out, is it stuck in the remnant of the evaporating black hole? There are many such questions that were largely resolved in part by string theory. The Beckenstein argument for an evaporating black hole is the following. First we must establish that a black hole has a temperature and an entropy.

Using our fundamental definition above, that the temperature of a system is the change in energy of a system due to adding one bit of information, we calculate the change in energy of a black hole. If we can calculate this, then the change in the radius of the horizon is simply related to the change in mass. To add one bit of information to a black hole it is most simple to consider throwing in one photon. It contains one bit of information in its polarization state. But if the black hole is large, say a kilometer, we don't want to throw in a photon of visible light that has a small wavelength. Then one could ask whether the photon entered at an angle θ with the black hole, etc. Instead we use light of wavelength near the size of the black hole so that its location of entry is uncertain.[14]

Proceeding with the calculation we need the Schwarzchild radius $r_S = 2MG/c^2$. The energy of a photon of this radius is $E = h\nu = hc/\lambda$ or the change in energy of the black hole is $\Delta E = hc^3/2MG$. Thus the temperature of a black hole of mass M is $T_{\text{BH}} = \frac{hc^3}{2MG}$. A proper calculation gives

$$T_{\text{BH}} = \frac{\hbar c^3}{8\pi MG}.$$

Using $\Delta E = \Delta Mc^2$ we have the change in mass of the black hole upon introducing one bit of information is $\Delta M = hc/2MG$. The change in radius of the black hole is

$$\begin{aligned}\Delta r_S &= 2\,(\Delta M)\,G/c2 \\ &= 2\left(\frac{hc}{2MG}\right)G/c^2 \\ &= \frac{h}{Mc}\end{aligned}$$

[14]It is interesting to note that photons of wavelength larger than the size of the black hole will not enter, but will be reflected out. See Susskind's book on black holes for this calculation

and the change in area of the black hole is $\Delta A = 8\pi r_S \Delta r_S$ or

$$\Delta A = 8\pi \frac{2MG}{c^2}\frac{h}{Mc} = 16\pi \frac{Gh}{c^3}.$$

The length associated with this change in area is $L = \sqrt{Gh/c^3}$, the Planck length. This means that the entropy of a black hole of finite size is given by the area of the horizon divided by the Planck length squared. A proper calculation getting all of the factors of 2 and π correct gives

$$S_{\rm BH} = \frac{Ac^3}{4G\hbar}. \tag{50}$$

This is one of the most remarkable equations in physics. Unlike the entropy defined in classical statistical mechanics that varies with the system volume, the quantum mechanical entropy of a black hole is proportional to the area of the horizon, not the volume of the black hole. Note that the Beckenstein-Hawking entropy has $\hbar$ in the denominator. In classical physics $\hbar \to 0$, so classically a black hole can have infinite entropy. In fact, in classical physics, any amount of entropy can be stored in a vanishingly small energy because $\hbar \to 0$. It is the presence of $\hbar$ in the denominator that prevents the black hole from having an infinite entropy. Finally, it is important to point out how large is the black hole entropy. It is vastly larger than anything of comparable size that is not a black hole.

How long does it take for a black hole to radiate away its mass? Using the Stefan-Boltzmann law we can calculate the luminosity

$$-\frac{dE}{dt} = AT^4$$

and using $A \propto r_S^2 \sim m^2$, $T \propto h/\lambda \sim 1/m$, and $E = m$, gives $-m^2 dm/dt = 1$ or $t \sim m^3$. A solar mass black hole takes approximately 10^{43} seconds to evaporate, much longer than the age of the universe.

7.1.3 Hidden degrees of freedom

What is the mechanism by which a black hole has entropy? Putting together a set of principles including quantum mechanics and thermodynamics the result is that there are many hidden degrees of freedom that give the black hole its entropy. These degrees of freedom are not apparent in the classical description of black holes or gravity. The understanding of classical general relativity is in a sense like fluid dynamics. Fluid dynamics describes the smooth flow of a uniform fluid. It can also describe turbulent flow and so forth, but it does not take into account the microscopic structure of fluids. Only in special cases is this considered. Ordinary fluid flow is described by the Navier-Stokes equations, a set of equations that are the effective theory. One does put entropy in the flow, but one does not ask where it came from. But the simple fact that there is entropy in a fluid tells us that there is a hidden microscopic structure; it is a hint that there is something more. The temperature and entropy we have discussed so far with black holes is also a hint that classical Einstein gravity is missing hidden degrees of freedom to be found in a proper quantum theory of gravity.

7.2 Metric near the Schwarzchild radius

The space near the surface of the horizon of a black hole is unremarkable. We want to show now that the metric there is that of flat space by several coordinate transformations of the Schwarzchild metric (49).

First we draw some inspiration by working with some simple metrics. Consider 2D flat Euclidean space with metric $ds^2 = dy^2 + dx^2$. In polar coordinates $x = \rho \sin\theta$ and $y = \rho\cos\theta$, the metric becomes $ds^2 = d\rho^2 + \rho^2 d\theta^2$. In Minkowski space we have $ds^2 = -dT^2 + dx^2$ and here we need hyperbolic functions. Defining $x = \rho\cosh\omega$ and $T = \rho\sinh\omega$, we again have[15]

$$ds^2 = -\rho^2 d\omega^2 + d\rho^2. \tag{51}$$

The coordinate ω is the "time". This is called the Rindler metric for reasons that will become clear below.

Now let us write the Schwarzchild metric in terms of proper distance so that we can make our analogy with flat Minkowski space. We write

$$ds^2 = -\left(1 - \frac{2MG}{r}\right) dt^2 + \frac{1}{\left(1 - \frac{2MG}{r}\right)} dr^2 + r^2 d\Omega_2^2 \tag{52}$$

[15] In polar coordinates in flat space the right triangle with hypotenuse ρ drawn from the origin has x, y coordinates as given here. Similarly for the Minkowski space one can draw the same right triangle, but the "radius" ω draws out a hyperbola. Dropping the perpendicular to the x-axis, one again has the sides of the right triangle as $x = \rho\cosh\omega$ and $T = \rho\sinh\omega$.

and consider the case where $r \approx 2MG$. Writing $r = 2MG + \delta$ we have

$$\begin{aligned}\left(1 - \frac{2MG}{r}\right) &= \left(\frac{r - 2MG}{r}\right) \\ &\approx \delta/2MG\end{aligned}$$

and we could expand around this, but let us instead do a variable change instead keeping in mind that δ is a positive number; we are expanding outside the horizon. Dropping the angular part of (52) we have

$$\begin{aligned}ds^2 &= -\left(1 - \frac{2MG}{r}\right) dt^2 + \frac{dr^2}{\left(1 - \frac{2MG}{r}\right)}. \\ &= -\frac{1}{r}(r - 2MG)dt^2 + \frac{r dr^2}{r - 2MG} \\ &\approx -\frac{1}{2MG}(r - 2MG)dt^2 + 2MG\frac{dr^2}{r - 2MG}\end{aligned} \tag{53}$$

If we use $d\rho = \sqrt{2MG}\, dr/\sqrt{r - 2MG}$ for a variable change then the second term goes to $d\rho^2$ and we are halfway there. Integrating we find

$$\begin{aligned}\int d\rho &= \sqrt{2MG} \int^r \frac{dr'}{\sqrt{r' - 2MG}} \\ &= \sqrt{2MG} \cdot 2\sqrt{r - 2MG}\end{aligned}$$

or

$$\rho^2 = 8MG(r - 2MG).$$

Then (53) goes to

$$\begin{aligned}ds^2 &= -\frac{1}{16M^2G^2}\rho^2 dt^2 + d\rho^2 \\ &= -\rho^2 d\omega^2 + d\rho^2\end{aligned}$$

where we have defined $\omega = t/\sqrt{16M^2G^2}$. Note that this metric is the flat metric (51). Therefore spacetime is approximately flat near the event horizon in a small local frame. This means that nothing crazy happens at the horizon for an observer falling through.

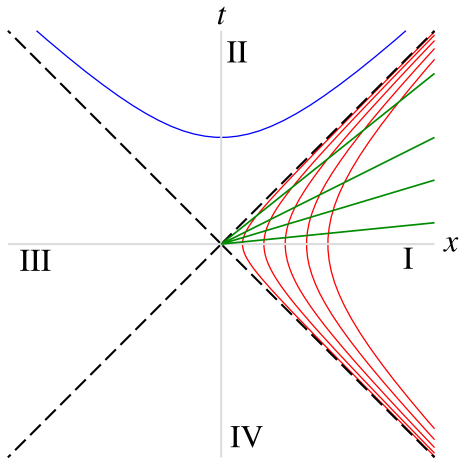

Figure 7: Rindler spacetime diagram. The horizon of a black hole is the 45-degree line defining the light cone from the origin.

Figure 8: Rindler metric, near and far from the horizon.

7.2.1 Rindler space

Rindler was the first to recognize that the metric in (51) could be interpreted in terms of a horizon in a simple diagram called a Rindler space diagram. In hyperbolic coordinates, the constant ρ lines (red) that asymptotically approach the light cone are the accelerated observers. They come in from minus infinity and go out to positive infinity at a fixed acceleration. They turn around when they reach the x-axis. The 45-degree line, which is just a light-cone in the usual description of this diagram, is called the acceleration horizon. This is because the accelerated observers shown in quadrant I can never receive a light signal from anything behind that light cone. In this sense it represents a black hole horizon.

7.3 Penrose diagrams

description of Penrose diagrams

7.4 Black hole formation

collapsing light cone

7.5 Counting black hole microstates

The following is a simplified argument for obtaining the entropy of a black hole from string theory. In units where $c = \hbar = 1$ we have, dimensionally, length = time and energy = 1/time, respectively. In particular this means that energy ~ 1/length, which we know from simple formulas like $E = hc/\lambda$ for the energy of a photon of wavelength λ. Newton's constant in natural units is then $[G] = [Fr^2/M^2] = El/E^2 = l^2$ or units of length squared. This is the Planck length, so we write $G \sim l_P^2$. In black hole physics or GR physics, the Beckenstein entropy formula is

$$S_{\text{BH}} \sim A/l_P^2 = A/G, \tag{54}$$

with A the area of the horizon. We would like to obtain this formula from stringy arguments. If the string coupling constant is g and this constant gives the amplitude for a string to "emit" a closed loop (graviton) for absorption by another string, then the probability of a graviton exchange will be g^4. The amplitude for the exchange process is g^2 because there are to events. If this is related to gravity we should be able to connect these ideas by writing $g^2 \sim G$ but g is dimensionless. The only thing it can be related to with dimension is the characteristic string length, l_S, which is the length of the wiggles on a ball of string, not the size of the ball of string. Then we can write

$$g^2 l_S^2 = l_P^2 \tag{55}$$

or just $gl_S = l_P$. This should make some sense in that the coupling constant should be less than one, and we have $g = l_P/l_S$ and $l_S > l_P$.

We now need the entropy of a string, or approximate entropy of a string. The easiest way to approximate it is to consider a simple cubic lattice of points that make up space. The string can be any configuration from one point on the lattice to another and can even go back on itself. Then in one space dimension the string can move one of two ways from each point. In two space dimensions, the string can move in any one of four directions. Clearly in d dimensions the string can go in any of $2d$ directions. If the string length is $L = l_S n$, n links long, then the number of configurations is $(2d)^n$. This gives an entropy for the string of $S_{\text{S}} = \ln{(2d)^n} = n \ln 2d$. This is essentially $S_{\text{S}} \sim n$ and from the string length equation we have

$$S_{\text{S}} \sim n = L/l_S. \tag{56}$$

Note the similarity with the Beckenstein formula, the characteristic parameter for a black hole is the area of the horizon, while for a string it is the string length. Simply divide out the characteristic lengths to obtain a dimensionless quantity and we arrive at the entropy.

The entropy of the black hole can be written $S_{\mathrm{BH}} = M^2G^2/G = M^2G$ using the Schwarzchild radius $R_S \sim MG$. Can we write the string entropy in this way? The mass of the string is $M_S = n/l_S$ where the mass of a segment of the string is $1/l_S$. Then we can write $M_S = L/l_S^2$, and the entropy of the string as $S_{\mathrm{S}} = L/l_S = l_S M_S$. Using (55) and $G = l_P^2$we can write $S_{\mathrm{S}} = m_s M_S G$.