THE GREAT COURSES®

Topic
Science

Subtopic
Physics & Chemistry

The Mind-Bending Physics of String Theory

Course Guidebook

Shamit Kachru

THE GREAT COURSES®

Printed in the United States of America

4840 Westfields Boulevard, Suite 400
Chantilly, VA 20151-2299
USA
1-800-832-2412
www.thegreatcourses.com

Shamit Kachru

Shamit Kachru is a Professor of Physics at Stanford University, where he has also served as the chair of the Physics Department. Additionally, he is the director of the Stanford Institute for Theoretical Physics. He earned a PhD in Physics from Princeton University, and he coedited the book *Strings, Branes and Gravity*. His research focuses on string theory and quantum field theory as well as their application to particle physics, cosmology, and condensed matter physics.

Table of Contents

1

What Led to String Theory?

To physicists, string theory is the theory of everything. It's the proposition that everything from light to matter—and from subatomic particles to the Great Pyramid of Giza—consists of infinitesimally small, vibrating strings. String theory is itself a theoretical rubber band that seeks to tie together Albert Einstein's notions of space and time in his general theory of relativity with the rules of quantum mechanics as outlined by his European peers Planck, Bohr, Heisenberg, and Schrödinger.

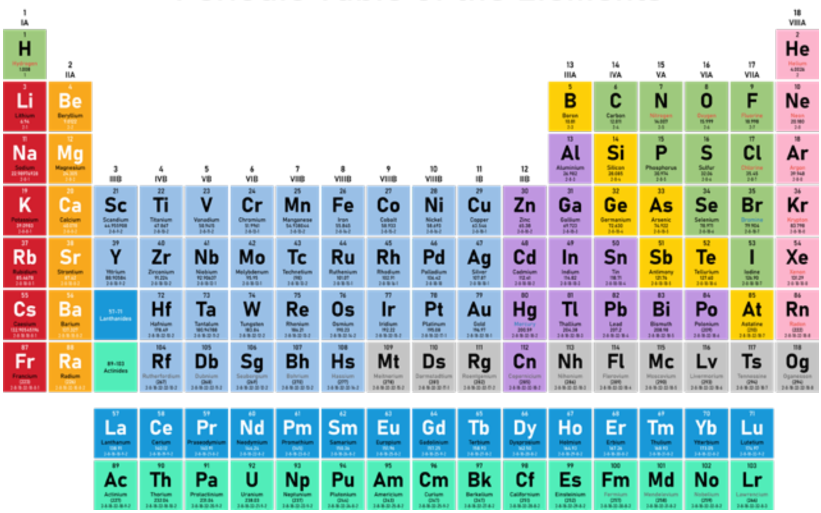
The Standard Model of Elementary Particles

Physics has sometimes been called the queen of the sciences. It concerns itself with the basic laws of nature—the foundations upon which chemistry and biology are built. It's also been crucial in the development of modern mathematics. At the heart of the physicist's quest to understand nature is the attempt to reduce everything to a few simple laws from which the rest follow. Albert Einstein's quest was to develop a single, unified theory capable of explaining all of the physical interactions that are seen in nature. String theory is humankind's most ambitious attempt in this direction to date.

Like Russian nesting, or *matryoshka*, dolls, the components of matter get smaller and smaller. The ancient Greeks hypothesized that tiny, indivisible constituents—atoms—existed, even if they couldn't be seen by the naked eye. From these tiny atoms, larger structures were assembled. So the trick to describing everything, as the ancient Greeks theorized, is to find fundamental building blocks out of which everything else is built.

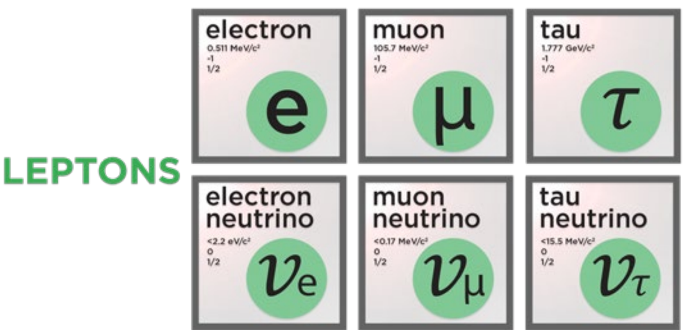
High school chemistry teaches that the basic building blocks of all matter are the different flavors of atoms in the periodic table of the elements.

Periodic Table of the Elements



These, in turn, are built from a short list of the elementary particles that were discovered in the early 20th century: protons with positive electrical charge and electrically neutral neutrons in a tight-knit nucleus, with negatively charged electrons orbiting the nucleus. Beginning in the 1930s, physicists built huge atom smashers and slowly learned that the proton and the neutron are not elementary. By the 1970s, it was discovered that they are composed of more fundamental particles, such as quarks, which carry a fractional electric charge.

Beyond the elementary electron are a pair of fatter cousins, the muon and the tau particle, which are identical to the electron in every way except that they carry a larger mass. The resulting table of elementary particles includes six quarks and three electron (or electron-like) leptons. In the standard model, three complementary neutrinos round out the leptons.



GAUGE BOSONS

But why do quarks stick together in protons and neutrons? Why do electrons orbit the protons in the nucleus? And don't forget the most familiar force of all: gravity. Why do things fall? Well, there is another set of particles: the force carriers. The photon transmits electric forces and light, and the gluons hold quarks together in the proton and hold protons and neutrons together in the nucleus. Finally, the massive vector bosons W and Z round out the picture. Their main role is to facilitate nuclear decay. As a real-world example, they are responsible for the clicks heard from Geiger counters used to measure radioactivity.

There are a few things to note about the standard model of elementary particles. First, it has a lot of ingredients—a list that starts to rival the periodic table itself in complexity! Can all this stuff really be thought of as "fundamental"? Second, the standard model of elementary particles is missing the graviton—the essence of what's thought to be the most basic force: gravity. Why is the graviton not included in the model of elementary particles?

The desire to establish a basic list of building blocks in a theory of everything led many physicists—starting with Einstein—into something of a rabbit hole of grand unification theories. Einstein was stymied in his attempt to construct such a theory in a few ways.

First, at the time he lived, scientists simply didn't know about some of the basic interactions between the existing particles or about the replicated family structure of similar particles in the modern standard model. In other words, Einstein didn't have a good enough picture of what he was trying to explain.

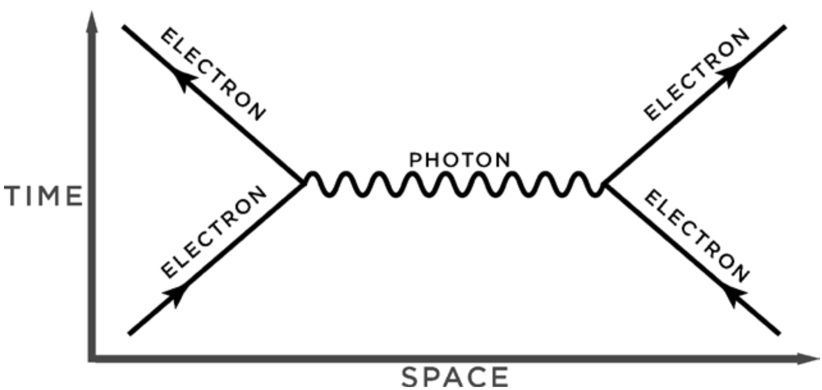
As for why the standard model is missing the graviton, a highly influential way of understanding the behavior of matter and energy that was discovered in Einstein's lifetime was a framework known as quantum mechanics. Gravity is thought of as only one of four fundamental forces of nature. Quantum mechanics describes the other three. And, as it turns out, getting gravity to play well with quantum mechanics is a tricky business.

Why is gravity special compared to the other forces? The role of gravity in the understanding of elementary particles is special in a few ways. First, in day-to-day life, even at a particle collider used to study particle physics, gravity is so weak that it's negligible. In the quantum realm of particle colliders, the electromagnetic force between two electrons is larger than their gravitational attraction by a factor of about 10^{40}. For this reason, gravitational interactions are never probed in any interesting way in modern particle colliders.

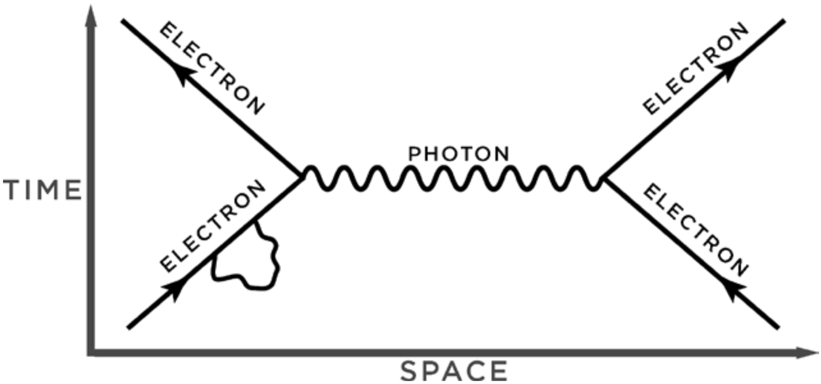
Feynman Diagrams

There is also another special aspect to the role of gravity in the understanding of high-energy particle physics. The quantum interactions between subatomic particles are very complicated to understand, so physicists expand them in a series of approximations, each one depicted by a suggestive drawing of the interactions it represents. These are called Feynman diagrams, after the theoretical physicist Richard Feynman, who first demonstrated them.

For example, in a diagram illustrating the relationship of four electrons with one photon, the solid and wavy lines in the diagram represent the matter and force-carrying particles, respectively. Space runs from left to right while time runs upward. In this example, two electrons can scatter off one another by exchanging a photon. Keep in mind that the diagram's lines and vertices are shorthand for precise mathematical expressions.



There are more involved diagrams, too. A more complete list would include one-loop diagrams. In this diagram, multiple electrons are exchanging a photon, and a virtual loop of electrons contributes to the exchange.



After decades of work, physicists today know how to evaluate these and even more complicated loop diagrams. And they contribute tiny—but precisely measurable—contributions to the understanding of electron-electron interaction. On the other hand, what happens with gravity? Why is the gravitational force between two electrons negligible?

For the case of quantum electrodynamics, the electron-electron scattering diagrams have various places where an electron emits or absorbs a photon. Each emission or absorption event, or vertex, should be sensitive to the electromagnetic charge of the electron. This results in a factor of the electron charge, e, being inserted when the diagram is translated into mathematics. Feynman diagrams can also be drawn where the electrons exchange gravitons instead of photons, and the result is profoundly different.

Newton's law states that the strength of an object's gravitational field depends on a constant—Newton's constant—times the mass of the object. Unlike the electron charge, e, Newton's constant is dimensionful: It has the dimension of inverse mass squared.

In fact, a famous mass scale built just out of fundamental constants of physics was discovered by Max Planck, one of the inventors of quantum mechanics. It is called the Planck scale, or Planck mass, in his honor. In detail, it is given by

$$m_{\text{Planck}} = \sqrt{\frac{hc}{G}},$$

where c is the speed of light, h is the Planck constant of quantum mechanics, and G is Newton's constant.

Newton's constant is given by the inverse square of the Planck mass. It therefore turns out that—to match Newton's law—in Feynman diagrams modeling graviton exchange between two electrons, each vertex should be accompanied by a factor of the electron mass divided by the Planck mass.

Finally, Einstein also discovered that energy equals mass times the speed of light squared, or $E = mc^2$. So, in a theory obeying Einstein's relation, the gravitational coupling of a particle is perhaps better thought of as being proportional to the particle's energy divided by the Planck mass. So at each vertex in a Feynman diagram involving graviton exchange, a factor of the particle energy E divided by the Planck mass should be inserted:

$$\frac{E}{m_{\text{Planck}}}$$

The Planck mass is about 10^{-8} kilograms, roughly the weight of a grain of sand. Typical particle masses and energies—even at today's biggest and most impressive high-energy physics colliders, such as the Large Hadron Collider at CERN—are many orders of magnitude below the Planck mass. This leads to two important facts:

1. The gravitational Feynman diagrams contribute a quantitatively tiny correction (of no importance) to particle scattering in any collider humans have built or can imagine building.

2 But if a collider of sufficiently high energy could be built, the quantum strength of the gravitational interactions would grow without bound. This is because the gravitational Feynman diagrams contain explicit positive factors of the particle energy E in their coupling constants.

Stated otherwise, gravity is negligible in modern quantum scattering experiments but incalculably strong in thought experiments about very-high-energy physics. Technically, the theory of gravity is non-renormalizable, which means that there are divergences in Feynman diagram calculations that are hard for theorists to tame.

This dual curse led high-energy physicists to neglect the study of gravity for most of the 20th century. Then, in the mid-1980s, an idea that could potentially unveil the structure underlying the standard model and tame the strength of quantum gravity came to the fore.

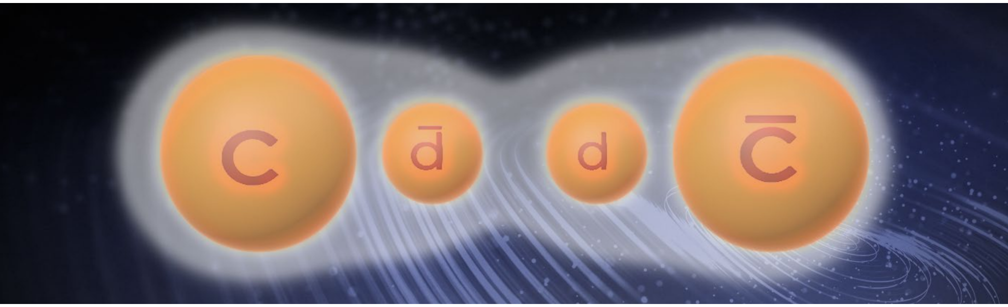
> A typical adult human has more than 10^{28} electrons and protons in their body.

The Beginnings of String Theory

It is now known that protons and neutrons are bound states of more-elementary quarks. Physicists had also discovered bound states of quarks other than the proton and neutron. Bound configurations of two quarks are known as mesons. But it became a problem to explain the complicated zoo of mesons that had been discovered. After a complex and rich intellectual history, the following picture evolved.

Say you have the particular meson made out of a charm quark (one of the six from the table) and its antiparticle. They are secretly bound into a meson by exchange of gluons, which are subatomic particles that are believed to bind quarks together. If you tried to unbind the quarks with an imaginary tiny forceps and drag them apart, a dense sheet of gluons exchanged between them expands with the distance, forming a binding flux tube.

When the flux tube carries enough energy of its own, it can create another pair of quarks, which can bind with the original charmed particles. The result is a process where quarks remain confined inside of bound states. Importantly, it suggests another, rather intuitive picture of the meson. Viewed this way, the meson is an open string stretched between two endpoints (the two quarks in the meson). And the string is made of the gluon flux tube.



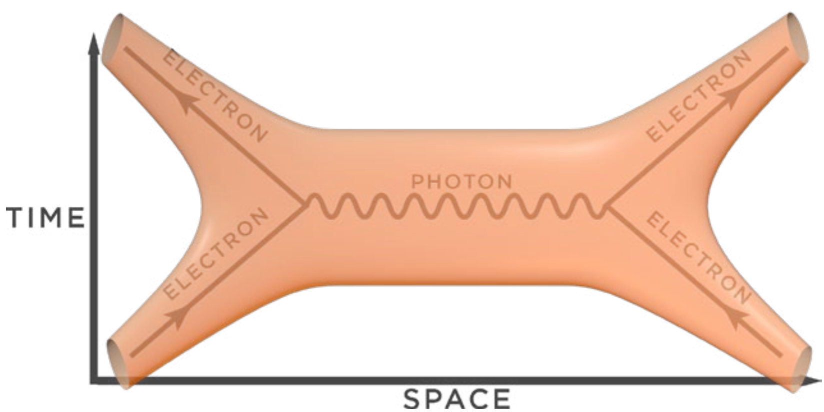
A flux tube can be thought of as a cylindrical space containing something like a magnetic field. Tension is generated by the energy in the tube, which can stretch and break open. It's also natural to imagine, then, closed strings that circle around and rejoin themselves instead of ending on some object, such as a quark. So, out of the study of meson dynamics came the notion of string theory, with open and closed strings as the possible level of structure underlying strong interactions.

Allowing for the possibility that the gluon flux tubes themselves vibrate and wiggle themselves leads to a spectrum of excited mesons. An experimenter could use this string model to try to match the phenomenology of mesons actually discovered in nature.

During the mid-1970s, another theory—the theory called quantum chromodynamics, which is now known to describe the strong interaction between quarks connected by gluons—came along to supplant the string model. But then an amazing fact was discovered by the French theoretical physicist Joël Scherk and the American physicist John Schwarz at the California Institute of Technology.

They found that if they included the closed-string loops in addition to the open-string mesons, the lowest energy state in the resulting system would describe a particle with all of the properties normally ascribed to a graviton. Out of a partially successful attempt to describe the rather messy phenomenology of mesons, instead a theory developed of possible fundamental constituents whose first prediction is that there should be gravitons (and quantum gravity) in the world it describes.

The first miracle of string theory isn't just that it gives rise to a graviton. The real miracle is that the string theory analogue of Feynman diagrams—with a string replacing the point particle traversing loops in the Feynman diagram—is manifestly finite. There are no infinities.



The singular vertices of the Feynman diagrams are smoothed out. The scattered string states are now joined together in a completely smooth diagram. In the full quantum theory, this absence of singular vertices translates to a finiteness property for the string scattering calculations, smoothing out the divergences of naive quantum Einstein gravity.

In intuitive terms, the size of the string is finite, with a string length of approximately 100 times the Planck length in common versions of string theory—and an associated energy scale. The would-be high-energy divergences of quantum gravity get cut off as the energy approaches the string scale. The result is a new class of theories of gravity that enjoy nice properties. These reduce to agreement with Einstein's theory when the energies are low compared to the Planck scale or when the distances are large compared to the Planck length. But they manifest new phenomena at high energy or short distance in such a way as to cure the more obvious pathologies of Einstein gravity.

Two Provisos of String Theory

It's important to mention two provisos. The first is that while string theory provides possible sensible theories of quantum gravity, there is no direct experimental evidence that string theory is relevant to the world. The nature of indirect evidence could be debated. Still, string theory does present an interesting theoretical structure that extends the current framework for understanding physics.

Second, string theory predicts that there are extra dimensions of space. If you think of the plane as representing the familiar three spatial dimensions of length, width, and height/depth, then above each point of this familiar space is a hidden dimension. It can be shown as a circle. If the circle is small enough, you wouldn't be able to notice it in everyday life—but it would still be there.

String theory enriches this picture in a few ways. For one thing, it requires many hidden dimensions, not just one. String theory typically predicts the existence of 10 space-time dimensions.

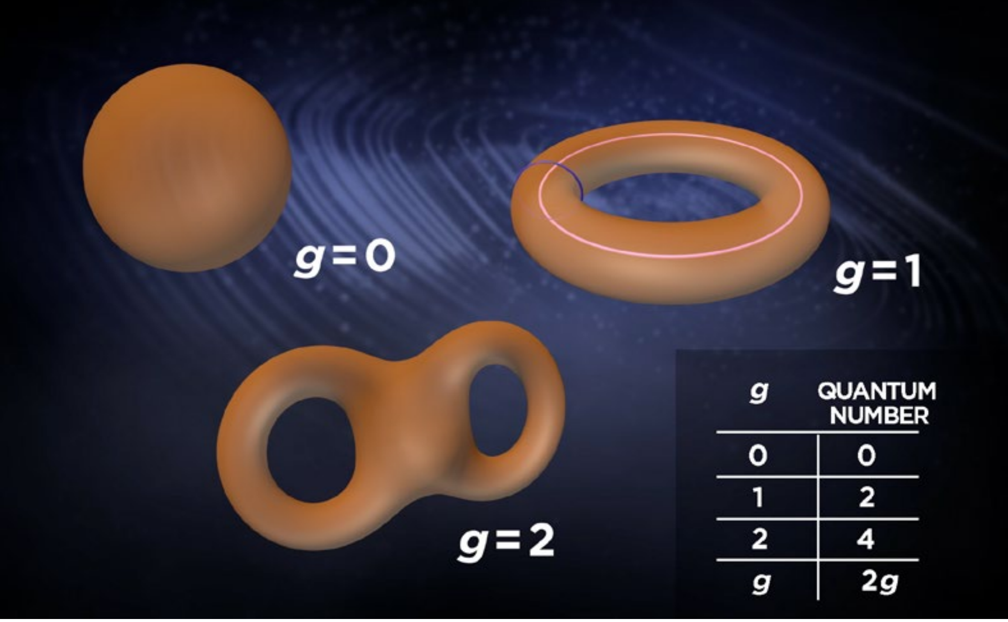
How does the theoretical existence of six extra dimensions (in addition to the three spatial dimensions plus time) affect anything? These other dimensions must be small enough to avoid daily notice. But even so, their existence has important implications for subjects ranging from particle physics to cosmology to black-hole physics in string theory.

Consider this in the context of particle physics. Suppose you want to make a table of the elementary particles that string theory predicts. How would the existence of extra dimensions matter?

Imagine a loop of string living in a world with an extra spatial circle. The string could make use of the circle in several ways. Most prosaically, it could ignore the extra dimensions and just wiggle around or oscillate in the big dimensions. Also, it could move with some velocity—either in space or around the tiny circle. And there is a qualitatively new thing it can do with the circle that it can't do in three Euclidean dimensions: It can wind.

Just as a rubber band can encircle a door handle any number of times, the string has an integer number characterizing how it winds the extra circle. This is a simple example in a field of mathematics known as topology, which examines geometric properties and spatial relations under such conditions as stretching, twisting, crumpling, and bending. Roughly, the strings detect the topology of a space in a way that point particles that have no size, shape, or structure cannot.

Topology can also give rise to new elementary particles by giving qualitatively different string states that are not wiggling around with string-scale energy. For instance, in a world with two compact dimensions, you could imagine a sphere, a torus, a double torus, and so forth. In these examples, g denotes the number of handles, or genus, of the two-dimensional space. There are no interesting ways for a string to wind a sphere (the $g = 0$ case).



On the other hand, on the $g = 1$ surface, known as a torus, or hollow donut, there are two integer numbers characterizing the number of times the string winds either around the donut hole or around the circular cross section. At $g = 2$, the number of relevant quantum numbers characterizing winding strings is 4, and simple induction tells you that there will be $2g$ such numbers for a genus g surface.

Each of these winding numbers corresponds to a conserved quantum number in string theory. It is not an exaggeration to say that each is like an electric charge the string can carry. Only winding states of string carry the charge. And just as the electron is stable because it is the lightest charged particle in nature, in string theory the lightest winding mode on a given kind of loop in the extra dimensions will be stable.

Reading

Greene, B. *The Elegant Universe*. New York: W. W. Norton & Company, 2003.

2

The Hidden Dimensions of the Universe

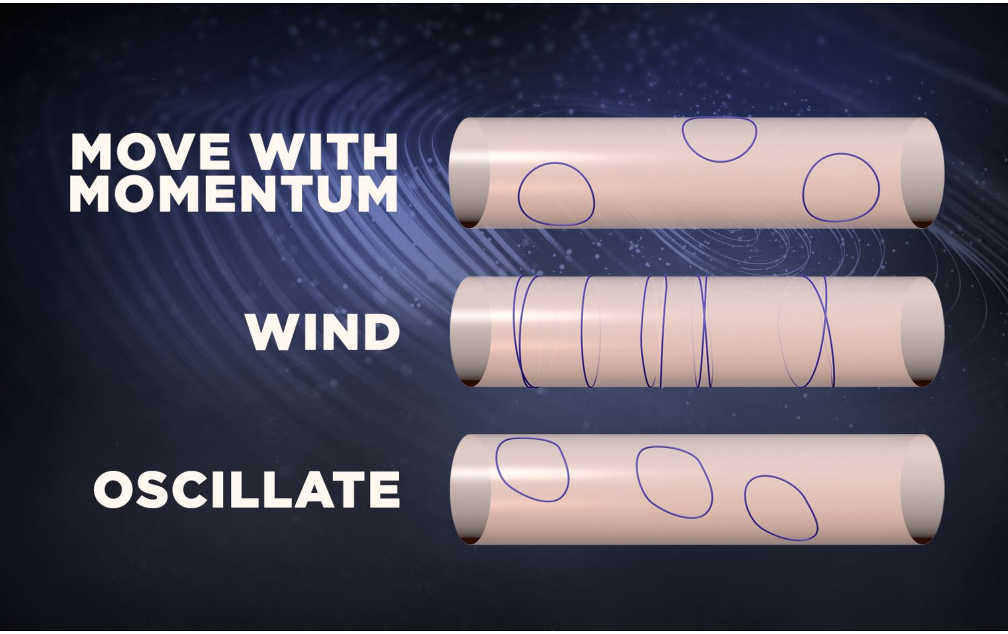
This lecture further explores the notion that string theory predicts the existence of extra spatial dimensions. The lecture describes a surprising fact about strings on a circle and introduces some more general geometries that are of interest as possible shapes of the hidden dimensions. It also returns to the concept of open strings and asks where they are in the theory.

An Extra Dimension Compactified on a Circle

In the simplest case, string theory's extra dimensions are curled up on a circle of radius R. Consider the spectrum of elementary particles in such a theory. Each type of elementary particle is just a reflection of an underlying string state, where you cannot see with the resolving power needed to resolve the tiny string inside the lump of energy.

There are three ways to excite a basic loop of string:

1. It can move with some momentum. This is like a ball you've thrown.
2. It can oscillate. This is what happens when you pluck a violin string.
3. It can wind. This is like what you can do with a rubber band on your finger, wrapping it many times, more and more tightly.



How does each of these processes contribute to the energy of the resulting string (or the mass of the particle it represents)?

1. In addition to motion in the usual spatial dimensions, consider momentum around the extra spatial circle of radius R. How does the fact that this space is a circle affect things? In quantum mechanics, particles—or strings—are represented by wave functions, which let you compute the probability that a particle is in a given place at a given time. When you consider a wave function of an object on a circle instead of in flat space, you need to make sure that the function is well defined on the circle. That means when you shift the location all the way around the circle, the wave function must come back to the same value. This basic mathematics translates in quantum mechanics to a statement that the momentum going around the circle has to be quantized in units of 1 divided by the circle's radius, R. The term *quantized* means it can only be an integer multiple of $1/R$. Therefore, the states carrying momentum around the circle have a momentum and carry an energy that goes like $1/R$.

2. On a string of tension set by the string scale, the oscillator modes cost an energy given by an integer multiple of the string mass (or the inverse string length). That's independent of R. You should imagine the strings as being very tense, so it costs a lot of energy to make them oscillate. The oscillator modes would correspond to very massive elementary particles.

3. If you wind a string of fixed tension around a circle of radius R, the energy cost will be the tension times the distance the string traverses. Since the distance around a circle of radius R (the circumference) is $2\pi R$, this gives an energy that is an integer multiple of $2\pi R$ in the appropriate string units.

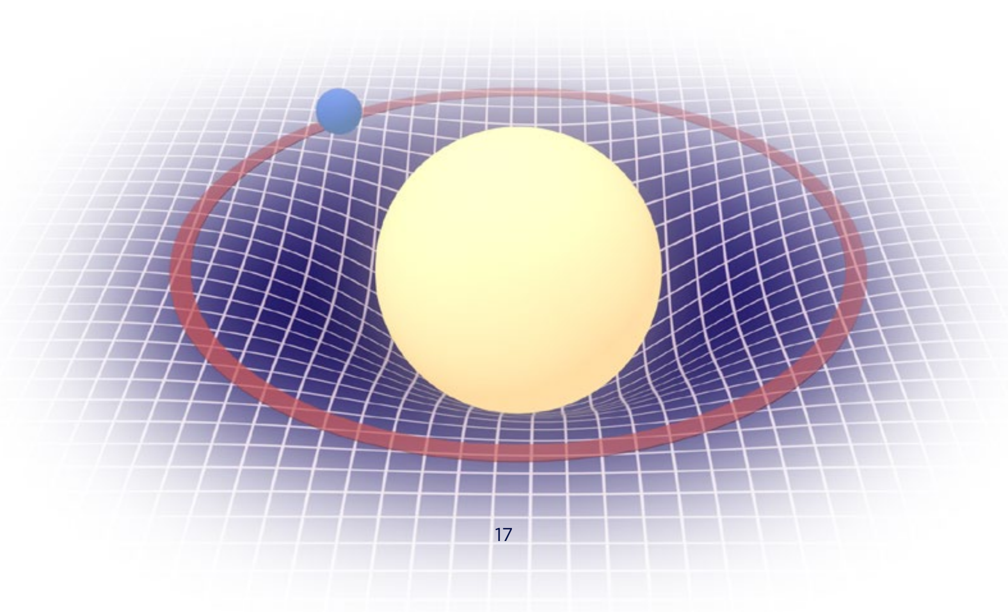
Notice that the momentum and winding modes have energies that depend suggestively in a precisely inverse way on the radius R. It suggests that there may be a duality in string theory where R and $1/R$ get exchanged and somehow strings on a small circle are identical to strings on a very large circle. Note that since point particle theories do not have the winding excitations of type 3 in the previous list, you can't wind a point particle on a circle any more than you can wind a bead on your finger, so particle theories can't enjoy such a relationship.

The equivalence of two naively distinct physical theories is called a duality, and this particular duality—between strings on a circle of radius R and strings on a circle of radius $1/R$—is called T-duality.

Possible Vacuum Solutions with Six Compact Extra Dimensions

So far, you've considered extra dimensions in the special case where there is one extra dimension compactified on a circle. While this is an interesting and instructive case, it is only a toy model, or simplified model, for the more general case. What is a more general case?

Assume for now that you're looking for vacuum (meaning there is nothing in it) solutions of the theory—solutions in the absence of any strings or particles or energy sources or anything else. Under this assumption, it is possible to argue that the equations of motion of string theory require that the compact dimensions satisfy the Einstein equations of general relativity, at least if the dimensions are big enough. More precisely, if the extra dimensions have a size R, then up to corrections in a $1/R$ expansion, they satisfy the vacuum Einstein equations.



In Einstein's theory of gravity, the instantaneous Newtonian force between two separated point particles is replaced with a more nuanced notion of gravity. In Einstein's theory, a particle of a given energy creates a curvature, or bending of space, proportional to its energy. Other particles then respond to its presence by moving in the straightest possible paths, or geodesics, on the curved space.

So here, the earth's closed orbit around the sun can be viewed as a result of the straightest-possible motion in a curved geometry, caused by distortion from the sun's enormous mass-energy. Rephrasing this more precisely, in a problem in Newton's mechanics where you study the earth-sun system, the basic quantity you keep track of—the dynamical data—is the position of the earth relative to the sun. In Einstein's theory of gravity, instead, the basic dynamical data is a metric on space-time. This metric is a function that gives the distance between any two points of the space.

A space with a given topology (like Euclidean space) can admit many different metrics. Topology is a coarser notion than geometry; the same topology of space can have many different distance functions defined on it. Einstein's theory picks out metrics that have special properties, satisfying Einstein's equations. The curvature is one property of such a metric. The basic fact that mass or energy sources cause curvature in Einstein gravity also says the converse: In the absence of sources, there should be no curvature.

So, the possible vacuum solutions of string theory with six compact extra dimensions should be in correspondence with six-dimensional spaces, often called manifolds, that admit a metric (or distance function) with vanishing curvature.

Curvature comes with different precise measures. A favorite in this context is the Ricci curvature, named after 19th-century Italian mathematician Gregorio Ricci-Curbastro. Because spaces with vanishing curvature are called flat, Ricci-flat six-dimensional manifolds are what you're after.

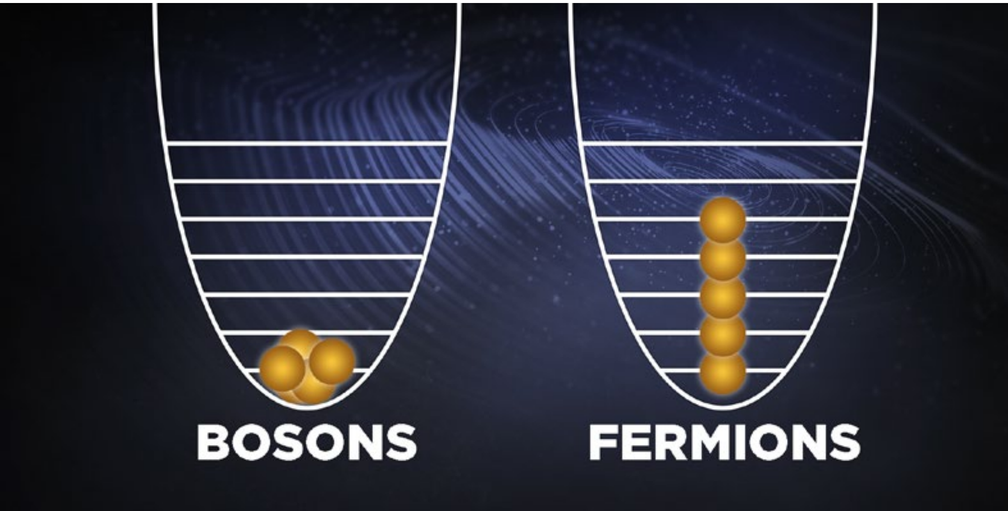
It is presently beyond reach to produce a list of all the possible Ricci-flat six-dimensional manifolds. At this stage, another simplification will be introduced to the already-very-simplified picture. The theories will be required to be very symmetric—in fact, supersymmetric.

Supersymmetry

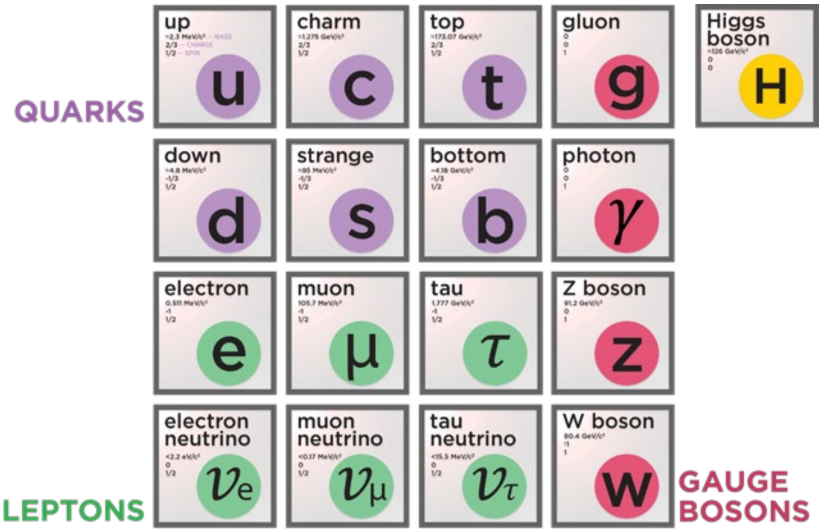
Supersymmetry is a way of introducing a spherical-cow model of particle physics. For physicists, a spherical-cow model is a model that abstracts away certain complications to get at the essence of a phenomenon. For instance, in a model launching a cow from a catapult, perhaps the most important features of the cow can be abstracted into its basic size (a radius) and mass, yielding a possible spherical model of a cow.

For the physics you're after, you don't need a fully realistic model of the world. This is good, because you're not going to get one. A simplification akin to the spherical cow is to assume that there is a symmetry in the theory known as supersymmetry.

What is supersymmetry? First, in quantum theories of elementary particles, there are two types of particles: bosons and fermions. Bosons are like familiar classical billiard balls: You can have as many identical bosons as you want, and they are easy to clump. Fermions, on the other hand, are purely quantum beasts. They obey the Pauli exclusion principle: Each fermion must occupy a unique quantum state. This rule—together with the existence of electron spin, a purely quantum phenomenon—leads to the structure of orbitals of the chemical elements.



Importantly, nature has some fundamental particles that are of each type: The quarks and leptons are fermions, while the force-carrying particles and the Higgs particle are bosons. In terms of the spin quantum number, which measures a purely quantum angular momentum of elementary particles, bosons have integer spin, while fermions have half-odd-integer spin.



In a supersymmetric world, things are stranger. Each boson has a corresponding fermion with exactly the same properties (apart from spin), so they differ in their boson/fermion nature. This is not true of the real world. There is no bosonic particle like the electron. But you can imagine a world where this were true. Such a world is supersymmetric. Supersymmetry is a symmetry that extends space-time symmetries like rotation and translations to include a Bose-Fermi reflection symmetry.

Because this is not true of the real world, the picture in any realistic model would have to have a slightly broken supersymmetric "fun house mirror" reflecting bosons into fermions. The particles and their reflection through the supersymmetry transformation aren't exactly the same in the picture, or in nature. They would have to exactly match in any supersymmetric theory of nature. For this reason, it is said that if supersymmetry exists, it is a broken symmetry.

Theories with unbroken supersymmetry provide beautiful spherical-cow models of fundamental physics. This adds one more condition to the extradimensional string theory setting. For now, supersymmetric solutions of string theory will be considered.

More on the Vacuum Solutions of String Theory

With the addition of supersymmetry to the list of assumptions, there is a lot more to say about vacuum solutions of string theory. In fact, assuming supersymmetry, the solutions are given by a class of spaces called Calabi-Yau manifolds. Eugenio Calabi conjectured in the 1950s that spaces satisfying a mild topological condition should admit a Ricci-flat metric and solve Einstein's equations. Shing-Tung Yau proved this while he was at Stanford more than 20 years later.

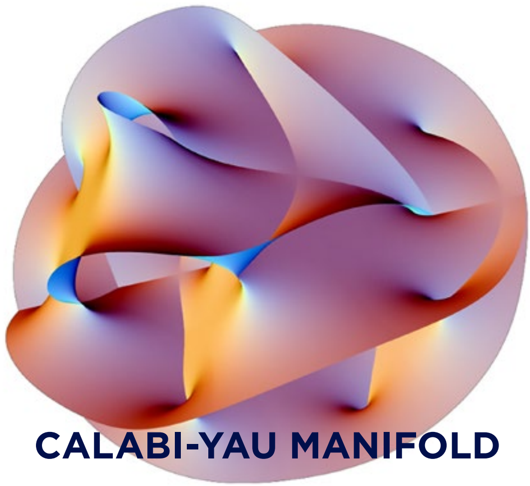
Because the condition Calabi placed on the spaces that bear his name is a mild topological one, it is easy to find ways to solve it. In fact, simple polynomial equations in many variables are implicitly known to give rise to solutions. On the other hand, the proof Yau gave of Calabi's conjecture is implicit, which means that while he knows that the Ricci-flat metric solving Einstein's equations exists on these spaces, he cannot write down the solution.

Finding such Calabi-Yau metrics remains a major challenge in geometry and string theory, though there have been significant steps using both analytical ideas and numerical methods.

So, string theory predicts the existence of extra dimensions. Under the spherical-cow assumption of supersymmetry, the vacuum solutions for such dimensions are in correspondence with a class of spaces mathematicians call six-dimensional Calabi-Yau spaces. And a mild topological condition suffices to determine whether a given space is in fact Calabi-Yau.

This simple specification of the data underlying a vacuum solution of string theory means that it is easy to write down examples. In fact, there are at least 500 million topologically distinct examples of such spaces.

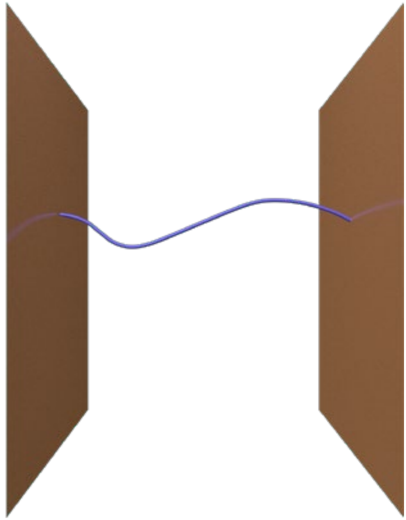
So, hundreds of millions of solutions of string theory can be written down (implicitly). Each such solution corresponds to a different shape of the extra dimensions. Because of the different ways that strings can move in the internal dimensions on each shape, the detailed four-dimensional physics that arises from string theory is very different in distinct Calabi-Yau compactifications.



This means that string theory gives rise to a range of supersymmetric, spherical-cow toy models of low-energy particle physics. The bumps and wiggles in the extra dimensions will turn out to be of significant physical importance in different constructions. But there is one more qualitative ingredient you need to understand before moving in this direction.

In a theory in the normal three dimensions of space, you can consider many types of objects. There are point particles (with no size or structure), but you can also imagine strings—like a network of cosmic strings stretching across a model universe. Finally, you can imagine membranes: objects with a two-dimensional surface that can lie flat like a sheet of paper or fluctuate into curved two-dimensional shapes.

In 10 dimensions, there are many more possibilities. In addition to points, lines, and planes, you can imagine *p*-dimensional objects, called *p*-branes, which have *p* spatial dimensions and exist for all time. Joseph Polchinski proved in the mid-1990s that string theory contains, among its excitations, such *p*-branes for all values of $p < 10$. (For larger values of *p*, it is impossible to fit them into the 10-dimensional spacetime.) But Polchinski's *p*-branes, called D*p*-branes for technical reasons, have a special additional property: Strings can break open and end on them. D*p*-branes are the potential endpoints of strings. For instance, here you can see an open string stretching between two separated 2-branes.



What Lives on a D*p*-Brane?

Now you have all of the ingredients you need to start doing some basic physics in string theory. You have extra dimensions, with a potentially rich geometry. You have closed strings, whose excitations include the graviton, giving rise to Einstein gravity. And you have D*p*-branes, whose physics includes open strings that stretch between a brane and itself or other branes.

There is one last question to answer before you can move on to the physics of strings, branes, and extra dimensions: What lives on a D*p*-brane? This question can be explained more clearly as follows. The case of a D2-brane, the $p = 2$ case of a D*p*-brane, can be used for illustration purposes since membranes are familiar and exist even in three-dimensional space.

Consider, then, a simple example of a membrane: an infinite straight sheet of paper. In what ways can the sheet of paper move? The basic motion it can perform (if, say, it lies along the *xy* plane in space) is to move transversely; at a given point on the *xy* plane, the sheet can fluctuate up or down in the *z* direction.

You can think of this possible excitation as parametrized by a scalar field that associates a number—the change in the z value of the location of the sheet of paper—to each point on the xy plane. You could say that there is a scalar field that lives on the membrane, parametrizing its transverse fluctuations.

A familiar example of a scalar field in the real world is the temperature—to any point of space you can associate a number, the local temperature. A more fundamental example is the Higgs field, which takes a numerical value at each point in space. Similarly, the fluctuations in the z location of a membrane living on the xy plane give a number for each point on the membrane, and hence a scalar field.

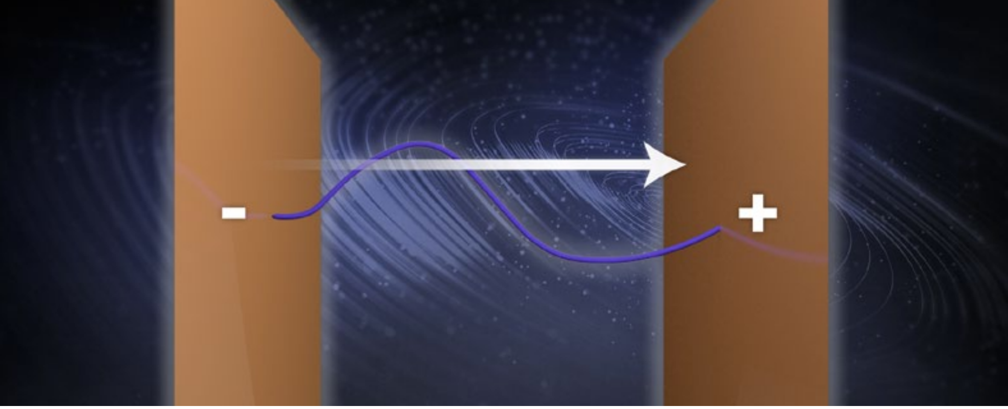
Now consider a Dp-brane in string theory. Since it is a p-spatial-dimensional object living in 9 spatial dimensions, it will have $9p$ transverse spatial dimensions. As in the thought experiment with a sheet of paper, it will then have $9p$ scalar fields living on its world volume; that is, there is a p-plus-one-dimensional theory with scalar fields living on the D-brane (the generic term for a Dp-brane of any p).

Scalars are conventional classical objects, such as the location in the z direction of a sheet of paper. But in a supersymmetric theory, each scalar comes with a fermion superpartner. So, on the Dp-branes, in a supersymmetric version of string theory, there will also be $9p$ fermion superpartners for the $9p$ scalar degrees of freedom just described.

Finally, though it doesn't follow from the qualitative logic just used, Dp-branes also have force-carrying particles living on their worldvolumes. In the simplest case, these particles are analogues of photons.

There is an alternative, more formal way to derive the spectrum of excitations of a Dp-brane. Instead of reasoning indirectly as before, the theory of open strings that can end on the D-brane can be quantized, much as the open string stretching between the two quarks in a meson can be quantized.

The simplest string theories have oriented strings (with an arrow of orientation along the line segment giving the string's spatial slice). One end of the string carries a +1 charge under the photon of the brane it ends on. The other end carries a –1 charge. The result is that strings between a brane and



itself are electrically neutral (under its photon), because the charges of the two ends cancel. But for strings stretching between distinct branes, you end up with a particle charged under each of the two photons of the brane theories. The rules for extending this logic to stacks of more than two branes involve more sophisticated concepts.

The ingredients that have been described can now be mixed and matched. In a general string model with D*p*-branes present and six dimensions compactified on a Calabi-Yau manifold, you might encounter the following kind of situation: Just as a rubber band can wrap your finger and a string can wind around a circle, a D*p*-brane could wrap around *p*-minus-three dimensions of a Calabi-Yau space, leaving its three spatial dimensions coincident with the usual (non-compact) three dimensions.

The degrees of freedom that live on this D*p*-brane would then live everywhere in the observable dimensions. For all intents and purposes, they would appear as elementary particles in the world. The force-carrying particles such as photons on the D*p*-brane could become literal photons in the three-plus-one-dimensional physics that you see.

Reading

Yau, Shing-Tung, and Steve Nadis. *The Shape of Inner Space*. New York: Basic Books, 2010.

3

Strings, Branes, and the Standard Model

This lecture builds on the picture developed in the first few lectures to figure out how something like the standard model of elementary particles might emerge in the framework of string theory. To do this, you will first need to develop a slightly better understanding of how the standard model of particle physics works. Then, you will be able to sketch how such a thing might arise.

More on Extra Dimensions of Space

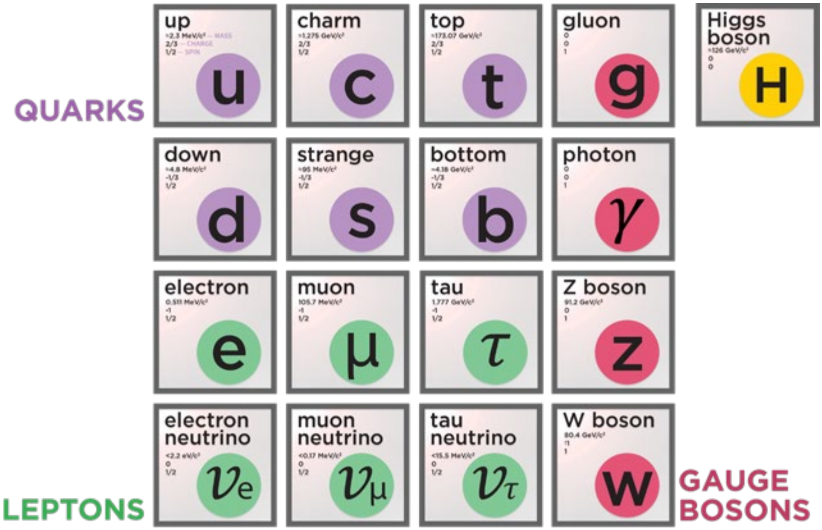
You learned some surprising things in the first few lectures. One is that modern theories posit that there are extra dimensions of space. The oldest such idea (dating to the 1920s) was that maybe space-time is five-dimensional, with four dimensions of space and one of time. The extra spatial dimension (beyond front/back, up/down, left/right) could live on a circle so small it can't be seen. Explorations of such a theory led to interesting ideas for unifying electromagnetism and gravity, but in the end, the framework was not useful for making contact with experiments. In fact, it is highly constrained by existing data.

In string theory, a modified version of this idea is revived—on steroids! The picture you were left with after lecture 2 was that a string world is 10-dimensional, with nine dimensions of space and one of time. Six of these dimensions are curled up on a Calabi-Yau manifold. The theory also has objects called D*p*-branes: *p*-dimensional extended objects on which strings can break open to end as open strings.

The particles or fields living on a D-brane can be obtained by quantizing the open strings that can end there. They give rise to scalar fields that correspond to possible fluctuations of the brane position in space-time. If you recall that the simplest theories you focus on are supersymmetric, then there are also fermion superpartners for these scalar fields. And, importantly, force-carrying particles like the photon are also found.

Analogues of electric charges live at the ends of the open string, with a positive charge at one end and a negative charge at the other. The string is oriented, carrying a little arrow. So a string with two ends on one brane gives something that is electrically neutral (not charged under that brane's photon). In contrast, a string stretching between two distinct branes gives a particle charged under the photons living on both branes (with opposite signs and charges).

This basic framework will be used to build toy models of particle physics. If this is the goal, you should first know what the current model of particle physics is.



The Standard Model in More Detail

As the table of particles shows, the particles come in two different types:

1. There are force carriers for the fundamental forces (other than gravity): the strong and weak nuclear interactions as well as electromagnetism.

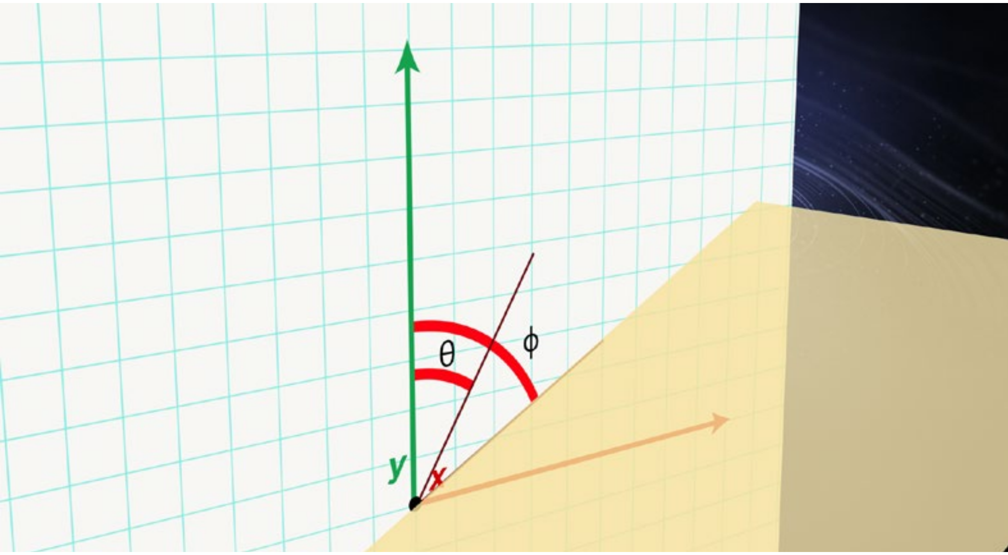
2. There are charged particles that participate variously in the interactions depending on their charges. For instance, just as the neutron is electrically neutral, the leptons are neutral to the strong force, and force carriers of one force do not carry charge under the others. Quarks do. The charged particles are divided into quarks and leptons, depending on whether they feel the strong nuclear force.

If you look in detail at how each particle experiences the various forces, there is a weird triplicate structure. For instance, the up (*u*), charm (*c*), and top (*t*) quarks feel all forces similarly—as do the electron, muon, and tau lepton. It is said that there are three generations of elementary particles, and each generation is like a copy of the others, as far as the forces are concerned.

This is a good first take on what the standard model of elementary particles is. To make further progress, you have to dig more into the details of force-carrying particles, charges, and so forth. To do this, the notion of a group has to be introduced.

A group is a mathematical object. It is a set with various elements and a law for composing those elements (i.e., taking two of the elements to produce a third). The groups that arise in physics have elements that are symmetries of real or abstract spaces.

A simple example of a group is the rotation group. Consider the Euclidean plane. You are free to rotate the plane through an angle theta (θ). Each point on the plane, except the origin, will move, but the plane as a whole comes back to itself after rotation. It is said that rotation through any angle theta is a symmetry of the plane. The rotations of the plane form a group. Given a rotation through an angle theta and another through an angle phi (ϕ), you can imagine composing them. The result is a rotation through an angle $\theta + \phi$. So, given two elements of the group, a third can be produced. This is the promised composition law for group elements. And this particular group is called $U(1)$.



Now, the surprise: The abstract mathematics of groups and symmetries is intimately tied to the way the standard model of particle physics is discussed. In fact, each of the fundamental forces of particle physics is in correspondence with a group of symmetries, with the force-carrying particles being in one-to-one correspondence with group elements that (taken together) generate the whole group through their compositions.

Each of the charged particle species is labeled by just a few numbers, then. One is the particle mass. Another is the spin: the intrinsic quantum angular momentum carried by each particle. Bosons have integral spin, while fermions have half-odd-integral spin.

But also, for each of the forces that exists, each particle carries internal quantum numbers. In a quantum-mechanical world, particles are represented by quantum wave functions. These internal quantum numbers indicate how that particle's wave function lives in the abstract internal space that is rotated by the group of symmetries corresponding to the force-carrying particles.

Formally, it can be said that to each force is associated a group G, and to each particle is associated a representation of G—a label indicating how the G rotations act on the internal degrees of freedom of the particle is also associated.

This offers a compact recipe for how to specify a model of particle physics. You specify a collection of groups G_1, G_2, and so forth, corresponding to the force-generating particles. And you specify a collection of representations of these groups, each to be attached to a given species of elementary matter particle.

To make this concrete: In the standard model, the electron is a spin-½ particle that is charged under electromagnetism. The group associated to electromagnetism is $U(1)$, and the representation of the electron can be said to be –1 in the following sense: Under an abstract rotation of the plane that $U(1)$ acts on by an angle θ, the electron wave function rotates by an angle $-\theta$. A positively charged particle, like the proton, would instead rotate by an angle $+\theta$.

The other forces correspond to other groups. Just as you can consider rotations of the plane, giving the group $U(1)$, you can consider rotations of other abstract, higher-dimensional spaces. It turns out that the groups corresponding to the weak and strong interactions can roughly be thought of as subsets of the possible rotations that could act on two or three complex dimensions.

These are harder to picture than the $U(1)$ rotations of the plane. In any case, those relevant rotation groups are called $SU(2)$ and $SU(3)$ for the weak and strong interactions, respectively. The result is that physicists sometimes say the standard model is based on the group $SU(3) \times SU(2) \times U(1)$.

A Caricature of the Standard Model

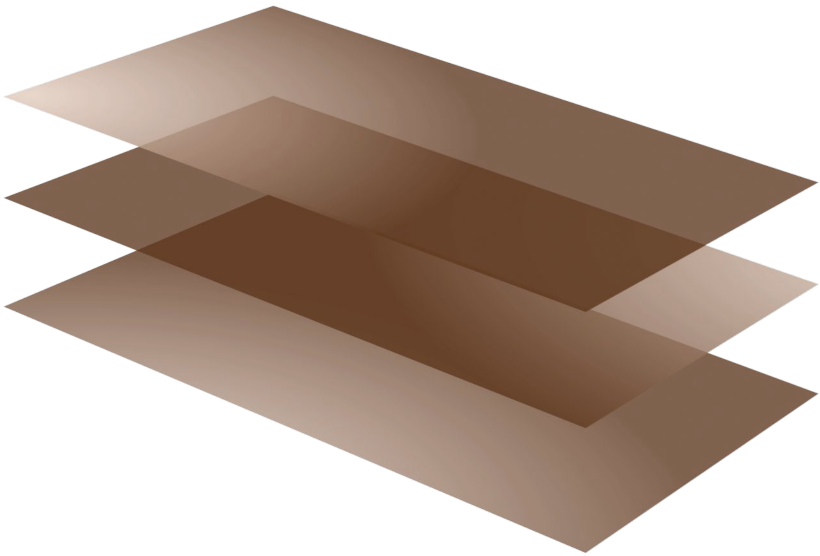
You've learned how groups are associated with forces in the standard model and how particles are really just representations of those groups. Now you are prepared to discover how a caricature of the standard model might emerge in string theory.

The first ingredient is that the 10-dimensional string theory must be compactified to four dimensions. This is accomplished through Calabi-Yau compactification, described in the previous lecture.

The second ingredient is that D-branes that fill the three-dimensional space on internal subspaces of the Calabi-Yau manifold must be wrapped. For instance, you could imagine D3-branes sitting at points on the compactification manifold and filling space-time. In this way, the particles and forces living on all the D-brane worldvolumes become the particles and forces evident in the world.

You can try to geometrize the groups and representations of the standard model in terms of D-branes (maybe wrapping parts of the Calabi-Yau) and open strings stretching between them. To do this properly, you need to learn several facts about D-branes, starting with ones you already learned: A single Dp-brane has a $U(1)$ force living on its worldvolume. Open strings with one end on the Dp-brane give rise to particles charged under this $U(1)$.

But what if you take multiple coincident branes instead? For instance, consider three D-branes that are parallel to one another. How do you think about the physics on these things as they approach one another and become coincident? Well, an open string ending on one D-brane gets one complex degree of freedom that rotates under the $U(1)$ force on that brane. When you have an open string ending on three coincident D-branes, it stands to reason that its endpoint has three complex internal degrees of freedom.



These can naturally be acted on by a rotation in three-dimensional complex space, and the result is that there is a natural $SU(3)$ group arising on the three parallel branes. (In fact, the group is the slightly different $U(3)$, which contains an extra factor of $U(1)$; this subtlety will matter.) So, the three parallel branes naturally have the force-carrying particles of the group $SU(3)$ living on them. You can get toy models of "quark flavors" charged under this toy strong interaction from endpoints of open strings on this brane stack.

Similar reasoning shows that with a stack of just two branes, you could naturally find an emergent $SU(2)$ with appropriate representations for the standard model particles. (Again, it is really $U(2)$, and the additional $U(1)$ factor will matter.)

More generally, a stack of N branes would give rise to a $U(N)$ gauge symmetry. The fact that you can try to identify the simplest cases that happen when $N = 2$ or 3 with groups that occur in the standard model is just a happy coincidence.

But in trying to examine a possible construction of the standard model from branes, you encounter a puzzle: Open strings only have two ends. If you tried to make the standard model on stacks (respectively) of three D3-branes, two D3-branes, and one D3-brane, giving the $SU(3) \times SU(2) \times U(1)$ symmetries you want, because open strings only have two ends, you could never write down particles charged under all three groups. No open string could have one end on each of the three brane stacks. But the standard model does have particles—such as the up quark—that interact with the force-carrying particles of all the forces.

This is where the subtlety of the extra $U(1)$ factors that are present on two– and three–D-brane stacks enter in a useful way. Including these extra $U(1)$ factors, there are in fact three different $U(1)$ rotations in the problem. By taking an appropriate combination of these three rotation groups to be identified with the $U(1)$ of the standard model, you can indeed arrange that the different sorts of open strings give the desired quark and lepton representations.

There are also extra particles not seen in nature yet that arise in the brane construction. These can be given a large mass and decoupled from the low-energy physics (but they would be seen at future colliders).

There is one last qualitative feature of the standard model that should be reproduced in this kind of Tinkertoy™ picture with D-branes: the possibility of finding multiple copies of particles with basically the same quantum numbers. This is like the triplication of the electron in the muon and tau or the three generations of quarks. How are you to get multiple copies of light particles charged under the force-carrying particles of different brane stacks?

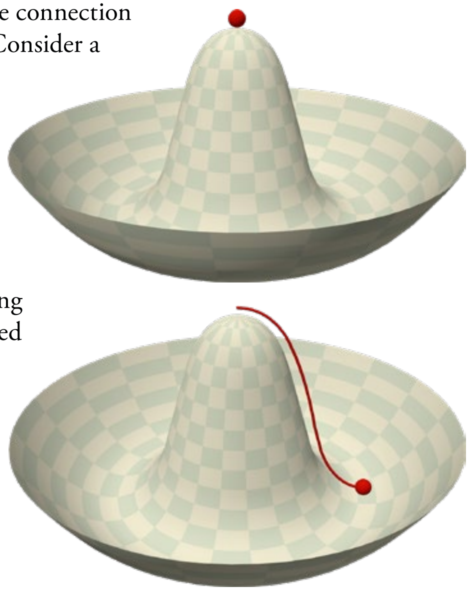
This is where the geometry of the extra dimensions can enter to help. While all the D-branes involved in this construction need to fill the three-plus-one space-time dimensions, they are free to wander on different subspaces of the Calabi-Yau space.

Light particles from open strings that stretch between branes can only arise from places where the different branes are close together. After all, the energy of an open string is proportional to the length of the string. So only short strings can give light particles.

How Branes Geometrize the Higgs Mechanism

So, the standard model has a few qualitative ingredients: force-carrying particles associated with various groups, matter particles with appropriate charges, and a multiplicity of generations of particles with the same charges. And there are simple geometric pictures of how these features can potentially arise in string theory. Branes also geometrize one of the most recently confirmed aspects of the standard model: the Higgs mechanism.

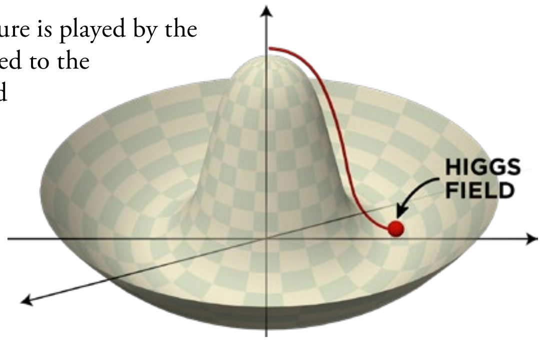
To understand this, let's revisit the connection between forces and symmetries. Consider a simple picture with a rotational, or $U(1)$, symmetry. You can consider this a picture of a potential energy landscape with the ball moving on the potential. You can rotate the picture by an arbitrary angle around the z-axis without changing the picture when the ball is perched at the top of the hill. But once it rolls down to the valley—regardless of where it lands—this rotational symmetry is broken. Its position on the rim would change under rotation.



This is an example of spontaneous symmetry breaking. Although the potential function itself is symmetric, the choice of any lowest-energy state breaks the symmetry. In the standard model, such symmetry breaking is a fundamental ingredient. The symmetry being broken is not a real space-time symmetry like rotation invariance but an internal symmetry.

The $SU(2) \times U(1)$ symmetry combining the electromagnetic and weak interactions into an electroweak symmetry is broken at sufficiently low energies in nature to just the $U(1)$ of electromagnetism. So a rotation symmetry, not in real space but in the internal space viewed by the quantum numbers of the elementary particles, is broken in a way analogous to the picture mentioned previously.

The role of the ball in the picture is played by the Higgs field—the field associated to the Higgs particle. The Higgs field rotates under rotations of the electroweak symmetry group. When it obtains a nonzero value (analogous to the ball rolling off the origin of the plane), it breaks the electroweak rotation symmetry.



Now consider a string theory setup with two nearly coincident D-branes. The symmetry group on each brane is $U(1)$, so the full symmetry in the problem is two copies of $U(1)$, which is written as $U(1) \times U(1)$. On the other hand, two coincident branes have a worldvolume symmetry group $U(2) = SU(2) \times U(1)$.

The difference between two coincident branes and two nearby but slightly displaced branes is a displacement of one of the branes along the axis separating them. This transverse motion of the D-brane is implemented by turning on one of the scalar fields on the brane. After all, the scalars parametrize the transverse position of the brane. But this motion is precisely a brane realization of the Higgs mechanism. As the nonzero value of the scalar field is turned on, the symmetry breaks. In this case, however, it is breaking from $U(2)$ to $U(1) \times U(1)$ instead of from $U(1)$ to nothing. Still, the basic idea is the same.

There is one important difference to the previous case, which exists because of supersymmetry: There is no potential function that tells the D-branes to separate or to be coincident. Instead, the Higgs potential for this "separation mode" of the branes vanishes.

There is a space of possible vacuum states parametrized by expectation values of this Higgs field. Such a condition is common in supersymmetric theories. The space of vacuum states itself is an interesting object called the moduli space of vacua.

> In string theory, it is the geometry of what is going on in the extra dimensions that helps scientists explain facts about theoretical physics—though admittedly in what are toy models.

Another Intriguing Possibility

The existence of models where the standard model particles are confined to branes opens up another intriguing possibility. In the real world, humans can resolve distances beneath a millimeter with their visual acuity, and the best microscopes see far beneath this scale. One way or the other, this direct "sight" involves using electromagnetic probes—photons impinging on people's eyes or the like.

Now, imagine that the photon is confined to a brane, and there is a compact extra dimension of space transverse to the brane. People can no longer "see" there using the photon; it simply is unable to move there, being confined to the brane. In this kind of picture, what constrains the physics of the bulk space off the brane? For instance, could it be meters across?

No. For one thing, gravity is not a force that is known how to localize to a brane. As in so many other situations in physics, gravity is the odd force out. Even in string theory, gravitational fields exist everywhere that there is space-time. This is almost tautologous, since in current thinking, gravity is fluctuations of the space-time geometry.

Constraints from the equivalence principle and precision gravitational tests can therefore constrain the size of any extra dimensions in a "brane world," where the photon lives on a brane. Amazingly, the strongest constraints today suffice to get humans beneath—but not extremely far beneath—a millimeter scale for the size of the extra dimensions of space. So, while many string theorists would expect that any extra dimensions to be found will be at much smaller scales, the conceptual space opened up by brane-world models allows many more audacious possibilities.

Reading

Lincoln, Don. "The Standard Model: The Most Successful Theory Ever." *Fermilab Today*, Nov. 18, 2011. https://www.fnal.gov/pub/today/archive/archive_2011/today11-11-18_NutshellStandardModelReadMore.html.

Zee, Anthony. *Fearful Symmetry*. Princeton, NJ: Princeton University Press, 2007.

4

Oscillating Strings and Supersymmetry

This lecture focuses on one of the most intrinsically stringy things about string theory: the highly excited, wiggly, oscillating string states. Strings with no oscillations result in the charged particles of the standard model, or the graviton. But what about the highly excited string states that appear beneath them? What can be said about them? And are they of any use in physics?

A Taxonomy of String States

In a typical setting in modeling nature, one usually imagines the string length and mass scale are a length of 10^{-30} centimeters and mass of about 10^{16} GeV. For reference, the mass of the proton is about 1 GeV. Since a single oscillation present on an excited string costs an order M in energy, only the states with no oscillation are relevant in modeling light particles like the electron or graviton. Nevertheless, this lecture will focus on excited states.

The first question in any endeavor is taxonomy. The most basic question in taxonomy is simply counting how much stuff there is. A familiar example is found in biodiversity, where animals are divided into genera, species, and so forth. The similar physics question to pursue is this: How many different string states are there of a given mass?

Probably, the basic answer to this question doesn't depend too sensitively on the shape or size of the extra dimensions of space. This is because a given string state is presumably small. The oscillations are occurring on a very tiny string, so it won't care about the global nature of the space around it—it can't detect the global nature.

So, let's count string modes in a d-dimensional space. Consider a string oscillating in one transverse dimension. Other than the straight ground state of string, there is a mode of lowest nonzero frequency as well as modes of higher-integer frequencies. The frequencies are quantized in units of the inverse length of the string. So, let's say you have modes with frequencies given by $1/L$, $2/L$, $3/L$, … , with L being the length of the string. Keep in mind that frequency counts number of oscillations per unit time. To get the right units, you should multiply these by the speed of light.

Classically, the oscillations each come with their own amplitude. You can pluck the lowest harmonic softly or hard. In classical mechanics, the amplitude squared would give the energy that the harmonic determines the energy it carries. Quantum mechanics changes this. As shown first by Max Planck and Albert Einstein, the excitations at a given frequency are quantized. Instead of a continuous amplitude, you get an integer number of quanta at that frequency.

Roughly, the amplitude squared counts the number of such quanta. A familiar example of this arises in electromagnetism. The classical theory talks about the amplitude of an electric or magnetic field. In the quantum theory, the energy in the electric field is really carried by a discrete number of quanta of light: photons.

Similarly, in the quantum string theory, the different oscillator modes can be excited with different numbers of quanta. So, in specifying an excited state of a quantum string oscillating in one transverse dimension, you get to choose the frequencies at which you are exciting modes and choose the number of quanta present at each such frequency.

The data, then, is a set of integers: n_1, n_2, n_3, … , with n_k counting the number of quanta present in the kth harmonic of the quantum string.

Now, what is the question you want to answer? One straightforward question of relevance in physics is this: How many states are there of a given energy E? So, let's search for the answer to this question. How many string states are there at energy E for a string oscillating in one transverse dimension?

Both the frequencies and the energy levels can be measured in units of $1/L$, the inverse string length. So, let's divide the $1/L$ and ask the following question: How many states have energy N in units of $1/L$?

The answer is implicitly found as follows. Given the counting of quanta previously, a string state characterized by n_k has mass squared calculated as the summation of kn_k:

$$m^2 = \sum_k kn_k.$$

For the Nth energy level of the string, then, you get a state for each way of writing N as the summation of kn_k for nonnegative integers n_k.

$$m^2 = N = \sum_k kn_k.$$

This a famous problem in mathematics—the problem of counting partitions of an integer *N*. Each distinct partition writes *N* as a number of occurrences of 1, of 2, and so forth—parametrized previously by the choices of the n_k.

You can easily tabulate the number of partitions of small numbers. For instance:

- one can be written as 1,
- two can be written as 2 or 1 + 1,
- three can be written as 3 or 2 + 1 or 1 + 1 + 1,
- four can be written as 4 or 3 + 1 or 2 + 2 or 2 + 1 + 1 or 1 + 1 + 1 + 1.

So, there's one partition of 1, two of 2, three of 3, and five of 4.

If you denote by $P(N)$ = number of partitions of *N* into integers, then you can express what was just stated by writing $P(1) = 1$, $P(2) = 2$, $P(3) = 3$, and $P(4) = 5$. But as *N* gets larger, it quickly becomes unwieldy to determine $P(N)$ by direct enumeration of all the ways to partition *N*. For instance, $P(N = 100)$ is about 190 million. You wouldn't want to try to write down all 190 million possibilities for partitioning 100 on a piece of paper!

The correct strategy instead is to try and determine an asymptotic formula that approximates $P(N)$ well as *N* becomes very large. In other words, you want an estimate for $P(N)$ such that the difference between the two, divided by $P(N)$ itself, becomes really small as *N* becomes really large. Then, fractionally speaking, you have done well in counting partitions, finding almost 100% of the possibilities.

This was a central problem in number theory in the early 20th century. It was cracked by one of the most famous collaborations in the history of mathematics—that between Srinivasa Ramanujan and G. H. Hardy. This collaboration inspired a major motion picture starring Jeremy Irons and Dev Patel: *The Man Who Knew Infinity*.

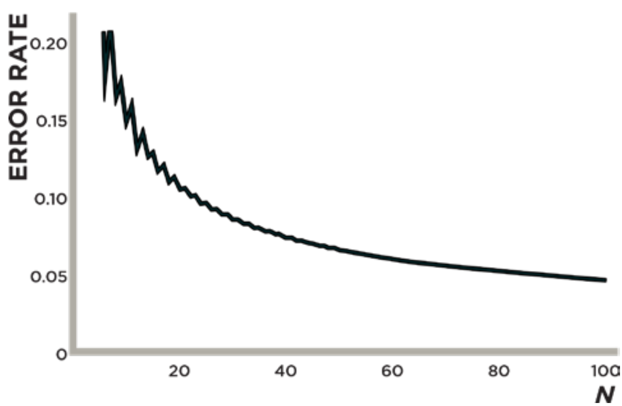
Ramanujan and Hardy

Srinivasa Ramanujan was an Indian mathematician, and G. H. Hardy was an eminent analyst at Cambridge University who invited Ramanujan to England and tutored him. Their singular achievement was to find a precise asymptotic formula for the number of partitions of an integer N. Their formula says that the number of partitions $P(N)$ goes as roughly $1/N$ times the exponential of π times the square root of $2N/3$:

> Cambridge University was the historical home of Isaac Newton and Charles Darwin.

$$p(N) \sim \frac{1}{4N\sqrt{3}} \exp\left(\pi\sqrt{\frac{2N}{3}}\right).$$

The formula starts off pretty accurate even for low values. The value of N is shown on the x-axis, while the relative error labels the y-axis. The leading Hardy-Ramanujan estimate is within 5% of the correct answer for N of 100. It is within 0.32% by the time N reaches about 20,000. In fact, they also provided a series of correction terms. Including these, their formula converges to the more precise answer quickly even for smaller values of N.



The important physics here is that the growth of states is exponential with the mass of the string state. This exponential growth in the number of states is called the Hagedorn growth of the number of string states. The detailed factors of π and such do not matter just yet. Later, when you learn about black holes, similar formulas will show up. There, even the numerical factors will be important.

There is one immediate, striking consequence of this formula. It implies that something bizarre happens when the theory is "heated up." To see this, let's recall some basic aspects of the way that thermodynamics and statistical mechanics work.

To begin a detour into statistical mechanics, consider a closed system that has some set of possible energy states. At each energy level, there will be different quantum states. A fundamental postulate of statistical mechanics says that in a closed system in thermal equilibrium at temperature T, a state with energy E has a probability of occupancy given by a formula of the rough form:

$$\text{prob(occupancy)} \sim \exp(-{}^{E}\!/_{T}).$$

There is an overall factor in front to make sure all probabilities add up to 1; the system must be in some quantum state (and, in fact, in precisely one).

In a string theory, the number of quantum states at energy E grows like the number at E goes like the exponential of some constant times E, with c some calculable constant set by the string energy scale:

$$\text{number}(E) \sim \exp(cE).$$

Putting these two ideas together, since each of the string states at energy E has the same probability of being occupied and there are some number, number(E), of them, the probability that the system has in energy E grows like the exponential of some constant times E minus energy over temperature:

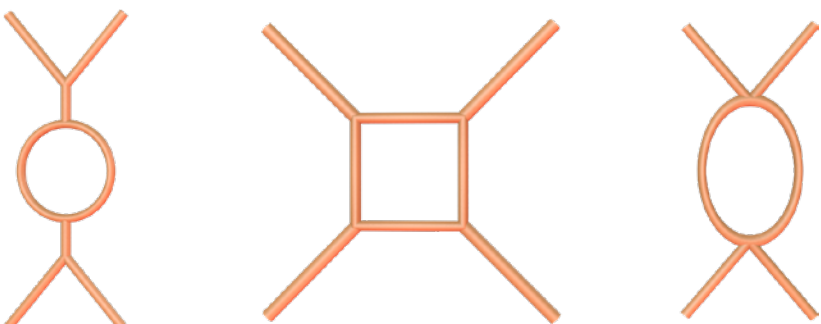
$$\text{prob(energy} = E) \sim \exp(cE - {}^{E}\!/_{T}).$$

This formula has a very odd feature: At some finite value of the temperature where $T = 1/c$ in the formula, the probability distribution becomes non-normalizable. In other words, no matter what number you put in front, you can't make the probabilities all add up to 1. The probability at higher temperatures is ever-growing at extremely high energies. This transition from conventional to bizarre thermodynamic behavior that occurs in string theory at some temperature set by the string scale is known as the Hagedorn transition.

This doesn't happen in the textbook cases you might learn about in a physics class. There, you are free to heat systems to high temperatures. They might melt or combust or whatever else—but surely the formalism of statistical mechanics won't break down. The interpretation of the Hagedorn transition depends on context; there are some cases where researchers have a very good idea what it means.

Now that you've learned about counting string states, let's start adding some nuances to the picture. So far, you encountered free, or noninteracting, strings. You were able to excite them to high energy without worrying about the possibility that the string state might decay to lower-energy string states. But in interacting physical theories, on the other hand, higher-energy objects decay to lower-energy ones as allowed by the quantum numbers in the problem. For instance, the muon and tau particles will decay quickly to electrons in the electroweak theory. That's why scientists have to work hard to produce them in particle colliders.

In string theory, the strength of string interactions controls the splitting and joining of strings. It is set by the value of a string coupling constant, g. A factor of g is associated with each split of a string into two in the fundamental "pants diagram" of string theory.



In general, many of the string states that have been counted at high energy may be unstable states in the interacting theory with $g > 0$ and will decay. They will typically decay in a time set by the string scale—so very quickly, unless g, the string coupling, is very tiny. But there are crucial exceptions: stable states.

Stable States

To figure out which states might be stable in a given theory, it helps to remember the electron. The electron is a poster-child example of a particle that is (thought to be) exactly stable. Why? There is a simple argument. Scientists believe that electric charge is exactly conserved in nature—it can't be created or destroyed. And the electron is the lightest particle that carries electric charge. Therefore, an isolated electron can't. No possible decay product could both carry its electric charge and conserve energy in the process.

The rich surprise in string theory is that in precisely the spherical-cow toy models this course wants to address—string compactifications with supersymmetry—the electron has many, many analogues. The reason comes back to the supersymmetry. A supersymmetric theory can contain more than one supersymmetry. It may have many different symmetry reflections that turn different bosons into different fermions.

You can prove that in a supersymmetric theory with a sufficient number of supersymmetries, there is a special function Z known as the central charge. For a given theory, Z takes as input all possible charges of a given particle under the symmetries of the theory. If there are symmetries $G_1, G_2, \ldots, G_k$, then it will take as input the charges $q_1, q_2, \ldots, q_k$ that a hypothetical particle in the theory has under all those symmetries. And it provides a bound on the mass that any such particle could have, just based on mathematical considerations of supersymmetry. The bound reads something like this:

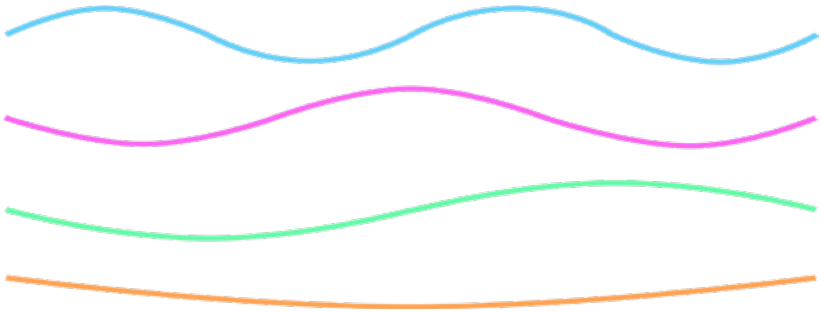
$$Z(q_1, q_2, \ldots, q_k) \leq M.$$

On the face of it, this may sound confusing and complicated. Why do you care about Z, and how are you to use this fact? But there is a simple route to use for this formula. Suppose you are studying the theory described previously. You identify a particle with charges $q_1, \dots, q_k$ and want to know whether it is a stable state or will decay. The theory may be one where perturbation theory in Feynman diagrams is very complicated. It may not be obvious how to answer the question.

However, in a theory with central charge, you are tempted to say the following: If the mass $M = Z$, the state is stable. Such states are called BPS states. If $M > Z$, the state is (very likely to be) unstable. Both statements have known caveats. In the first case, the state may have a marginal decay to products that have energy also adding up to exactly Z and conserving the charges. In the second case, the state may be stable for idiosyncratic reasons in the given theory but generically wouldn't be. But in general, these are very good rules to live by when studying supersymmetric theories. And they lead to a simple question that has proven immensely fruitful: Can you find and count the BPS states in interacting but supersymmetric string theories?

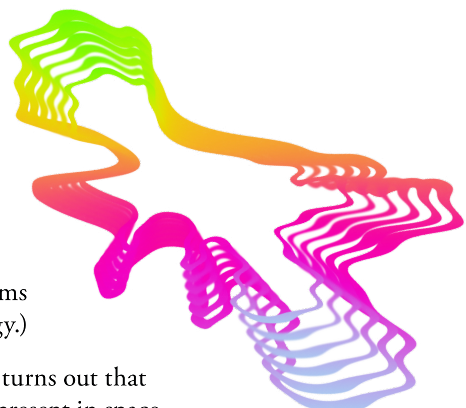
You often can, for another reason. Instead of carefully examining particle states one by one in a given theory, you can look for particle states in a supersymmetric theory that preserve some of the supersymmetry, instead of breaking all of it when you introduce them. It turns out that saturating the BPS bound (as a state with $M = Z$ does) is equivalent to preserving some fraction of the supersymmetry. And while it can be hard to a priori compute masses of a given state, it is often easy to identify particularly symmetric classes of states in a quantum theory—including those that preserve some amount of supersymmetry.

In a string theory, it turns out that there are often many, many BPS states—and, by extension, many, many stable charged-particle states. The reason is the following. Remember, you just went through the counting of string states at a given energy. You studied states consisting of superpositions of many modes with different frequencies and amplitudes (or numbers of quanta, in the quantum theory).



The picture shown above—and the count that you did—is the one relevant for an open string. In an open string, there is no such thing as a mode that moves only to the left or only to the right. A wave propagating to the right hits the end of the string and bounces back as a left-moving wave.

But you can instead consider closed strings, which also oscillate with modes of different frequency and amplitude. On such a string, you could consider both left- and right-moving waves. The previous considerations—counting partitions and so forth—would apply almost without modification. (The only difference is you could do partition sums for both a left- and right-moving energy.)



Returning to the idea of BPS states: It turns out that in string theory, the supersymmetries present in space-time can be associated with supersymmetries present in the worldvolume theory of the string itself. And since the string worldvolume has left- and right-moving modes, it stands to reason that some of the supersymmetries act on left movers (leaving right movers inert) and others act on right movers (leaving left movers inert).

That means that there is an easy way to make exactly stable BPS states in supersymmetric string theories: You simply take the supersymmetric ground state of the string and add either left- or right-moving excitations, but not both. The counting of oscillator excitations and energies for chiral (purely left- or right-moving) modes on a closed string closely parallels that for the modes on a single open string. All that happens in the open string is that you can consider left movers to be reflected into right movers at the left boundary of the string, and vice versa.

So, you are going to find a Hagedorn degeneracy of BPS states when you study simple supersymmetric vacua of string theory. This is a far cry from the single stable electron protected by its electric charge.

Reading

Collins, Graham. "Quantum Black Holes Are Tied to D-Branes and Strings." *Physics Today* 50, no. 3 (1997): 19–22.

Kanigel, Robert. *The Man Who Knew Infinity*. New York: Washington Square Press, 2013.

5

How String Theory Explains Black Holes

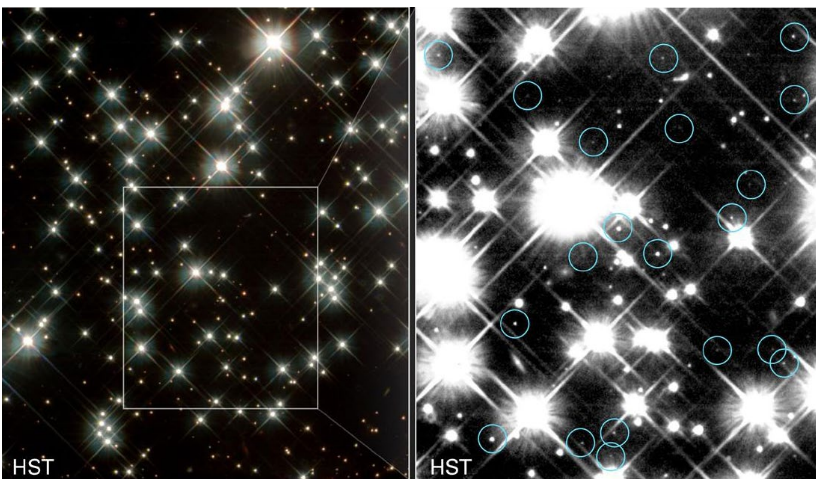
One of the main reasons for excitement about string theory is that it is a consistent quantum theory of gravity. It's important to have such a theory because there are situations in nature where quantum mechanics and gravity combine in very confusing ways. This lecture addresses one: the black hole.

Gravity and Black Holes

In lecture 2, you learned that Einstein's theory of gravity—general relativity—gives a precise, mathematical description of the way in which a massive body "curves up" the space around it. For instance, the sun creates a curved spatial geometry. The orbiting earth is just moving in the straightest-possible path in this geometry.

The sun is an object with a large but finite mass, *m*, about 10^{30} kilograms, trapped in a large region of space. You can think of it as a sphere with a radius of about 400,000 miles. It manages to curve up the space next to it pretty well. What would happen if you took a much more massive object of the same size, or maybe a much smaller object but with the same mass? Space would probably become even more curved up in its vicinity. Fortunately, nature has produced such objects.

The sun is held up (against its own attractive gravity) by nuclear reactions that take place deep in its core. As time passes, it will burn more and more of its nuclear fuel, until finally it runs out. What happens then depends a bit on details. For stars that weigh less than about 1.4 times the mass of the sun, it is believed that they collapse to form objects known as white dwarf stars.



WHITE DWARFS LOCATED IN M4

These are highly quantum balls of matter that are held up by the Pauli exclusion principle (which says that no two electrons—or other identical fermions—can be in the same quantum state). The dominant constituent of a white dwarf is a gas of residual electrons remaining from stellar collapse. The Pauli repulsion that prevents all these electrons from being in the same quantum state (just sitting at the center of the old star) is what holds the white dwarf up against gravitational collapse.

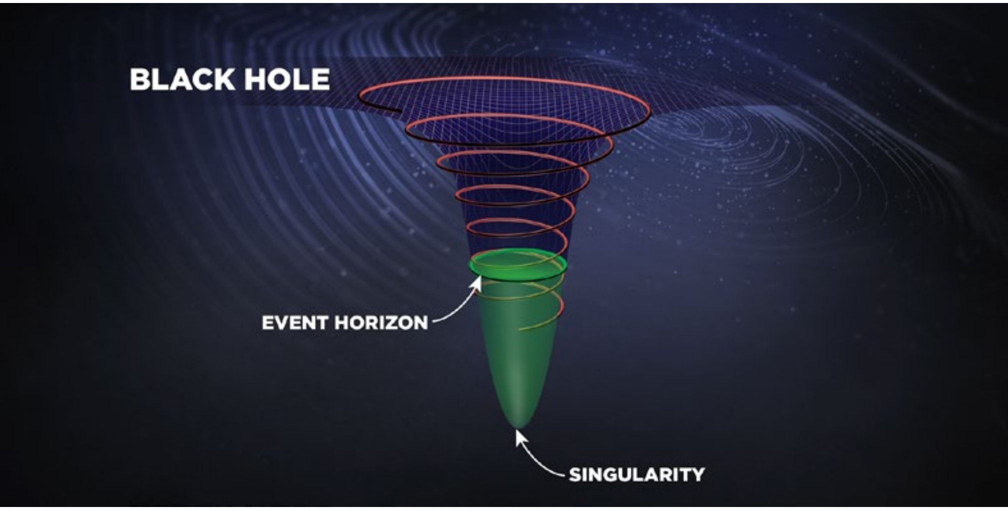
For stars that weigh more than 1.4 times the mass of the sun but less than about 3 solar masses, a similar fate befalls them. However, the collapse is more dramatic, and they manage to squeeze the remaining electrons and protons together in the collapse process. What results is a ball of neutrons supported by Pauli pressure—also known as a neutron star. Both white dwarf stars and neutron stars have been seen in nature. These predictions of Einstein's theory, together with basic quantum mechanics, have been verified. But what happens to stars weighing more than about 3 times a solar mass?

Their fate is to become black holes. These are objects so dense and massive that, starting at some distance from the object (known as the Schwarzschild radius), even light rays can't escape. The sphere at this distance from the center of the object, the Schwarzschild radius, is known as the event horizon. If you cross the event horizon of a black hole, you aren't coming back.

Explaining their curved geometry is difficult, but there is a simple Newtonian analogy. Earth has something called an escape velocity. If you throw up a ball or a rocket with velocity v, you know it has a kinetic energy (of $\frac{1}{2} mv^2$). It also has a gravitational potential energy that binds it to the earth (which explains the gravitational attraction). If the kinetic energy is large enough, it can overcome its gravitational binding to the earth and escape to outer space. This happens all the time; it's what NASA and SpaceX launches do.

To get this high kinetic energy, you need to crank up the velocity of the launch. Also, since the gravitational potential binding you to the earth is proportional to the earth's mass, M, the needed velocity gets larger as you try to escape a more and more massive object (planet, star, or whatever). You can imagine an object so massive (or with such a small radius) that the escape velocity from its surface would actually be faster than the speed of light. In fact, it was imagined by John Michell, a Cambridge professor, in the late 18th century.

There is now overwhelming evidence that black holes do exist in nature. In general relativity, their structure is as shown. The event horizon is as advertised previously. Past the horizon, at the core of the black hole, lies a region where Einstein's theory of gravity breaks down. The solution actually predicts diverging curvature of space. Any observer would be torn apart by the strong gravity long before reaching the final singularity. But once you pass the event horizon, your fate is sealed. You will be unavoidably drawn to the singularity.



These are very interesting objects—so interesting that, in fact, for years they were not taken seriously as predictions of general relativity. Although a black hole solution to Einstein's theory was written down in 1915 by Karl Schwarzschild, it was believed to be a mathematical curiosity.

Significant numbers of people first began to take the existence of these objects seriously in the 1960s. That is when astronomers began to see striking phenomena associated with other dense, compact objects—neutron stars in particular. By 1973, a black hole candidate, Cygnus X-1, was identified. It is now widely believed to be a black hole. In addition, it is now believed that almost all galaxies—including our own Milky Way galaxy—host supermassive black holes in their centers.

String Theory's Role

What does all this have to do with string theory? A role for string theory first emerges when theorists, fascinated by the enigmatic black holes now believed to exist, started to explore their formal properties in the 1970s. A powerful group of theorists at Cambridge and Princeton began to elucidate the properties of black hole solutions of Einstein's equations. They found a few basic facts.

First, a black hole is completely and uniquely characterized by three numbers: its mass, electric charge, and angular momentum. This is called the no-hair theorem for black holes. It is in striking contrast to other macroscopic objects in daily life.

The three numbers characterizing a black hole obey a relationship: A small change in the mass M has to be compensated by a small change in the area A of the event horizon, the charge Q, and/or the angular momentum J:

$$dM = \kappa dA + \mu dQ + \Omega dJ.$$

Also, the area of the event horizon can only ever increase. That's intuitive, as objects can enter the event horizon and increase the mass and event horizon radius of the black hole, but nothing can ever leave. So, dA, the change in A, is always positive:

$$dA \geq 0.$$

To most people, these may seem like a complicated set of rules with no particular familiarity or rhyme or reason. But to physicists, these rules smack of another set of very famous laws: the basic laws of thermodynamics that govern how heat flows. The most basic version of those laws, stated in terms of energy E, entropy S, and temperature T are these equations:

$$dE = TdS \quad \text{and} \quad dS \geq 0.$$

S, the entropy, is a measure of the disorder of the system. More precisely, it counts the number of different configurations the system could have while maintaining the same macroscopic properties.

This analogy between black holes and thermodynamics occurred to many people. It was written off as a peculiarity of the laws of black holes by all except Jacob Bekenstein, who felt that the analogy had some heft to it and should indicate something about black holes. His reasoning was as follows.

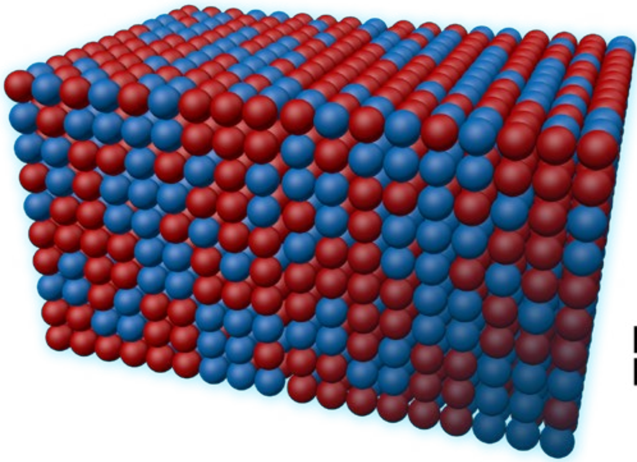
Consider two different states of the world: In state one, there's a hot gas and a black hole. In state two, the hot gas has been thrown into the black hole, leaving just a black hole. In the first state, the hot gas has a lot of entropy. Many different states of the constituent molecules that are whizzing around inside the box give a gas with the same temperature, pressure, and so forth—and so give the gas a large entropy. In the second state, the gas is just gone; there is only a black hole. So, has the second law of thermodynamics—which says that entropy always increases—been violated? The entropy of the universe, outside the black hole, seems to have decreased. Can the black hole somehow carry the entropy?

This is puzzling because the no-hair theorems suggest that black holes are basically unique objects. Bekenstein's insight was to realize that just as a box of gas with fixed macroscopic properties can have many underlying molecular states, leading to an entropy, a black hole with fixed mass, charge, and angular momentum may have many underlying microstates. In fact, if you identify area of an event horizon with an entropy, then you can save the laws of thermodynamics. The absorbed box of gas changes the event horizon area—by increasing the black hole mass—just enough to satisfy the second law.

Entropy Concretely

A box of hot gas has a large entropy. You can think of the entropy concretely as follows. The box has some very large number N that's much greater than 1 of molecules moving around inside it:

$$N \gg 1.$$



Each can be in any of several different states of motion (for instance, with slightly different velocity), while the macroscopic properties of the whole collection could remain the same. The different individual properties of the molecules characterize microscopic states, or microstates, of the gas, all of which correspond to the same macroscopic state, or macrostate.

The entropy S is a property of the macroscopic state, and it precisely measures the number of distinct microstates that could explain the same macroscopic gas. More precisely, you can think of the entropy as the log of the number of microstates:

$$S = \log(\text{microstates}).$$

For a box of gas, this makes sense. Gas has tiny constituents, so it's not surprising that it can be characterized by a large entropy.

But what about a black hole? The black hole solution of Einstein's field equations doesn't require any knowledge of underlying microscopic constituents. In fact, it arises as a solution of general relativity theory without making any assumptions about the existence of matter particles that could make up the black hole. On the other hand, the solution is, in a basic sense, incomplete. It has a singularity at the center and an event horizon that leads to mysterious physics.

A reasonable statement would be that any more complete theory of gravity should provide some explicit understanding of the microstates (the analogue of the individual molecules and their motion) whose collective solution "looks like" a black hole. And there should be precisely enough microstates for a black hole of horizon area A to give entropy equal to A:

$$S = A.$$

Can String Theory Explain the Black Hole Entropy?

The first four lectures introduced a candidate theory underlying Einstein's general relativity: string theory. A natural question is, then: Can string theory explain the black hole entropy? For realistic black holes in nature, this question remains beyond reach. Nobody knows if there is a solution of string theory that correctly describes the world. And, less ambitiously, nobody has sufficient theoretical control of string theory descriptions of the neighborhood of a realistic black hole to figure out whether the theory somehow captures its microstates.

Note that the second consideration is less ambitious than the first. Physicists might be able to locally model a realistic black hole, even if they can't make a complete model of the whole universe. And that's actually how physics has normally progressed. But there is no reason to despair. Let's return to the philosophical tool known as the spherical cow.

The spherical-cow solutions of string theory—supersymmetric solutions—were introduced previously. These produce imaginary worlds where bosons and fermions come in mirror pairs, enjoying all the same properties (except for their different bosonic-versus-fermionic nature). These corresponded, for instance, to situations where you compactify the 10 dimensions of string theory to a four-dimensional flat space-time, with a six-dimensional Calabi-

Yau above each point. Can you make spherical-cow black holes living within these spherical-cow toy models of string theory that you could use to explore the issue of black hole microstates?

Yes. The key is to realize that among the properties allowed by the black hole no-hair theorems is electric charge. In a supersymmetric world with suitable analogues of electromagnetism, you can find black hole solutions that are not realistic. They carry charge under various toy-world analogues of electromagnetism.

You can even find solutions that satisfy a property described in lecture 4. They saturate a bound of this form: mass M is greater or equal to Z as a function of q_1 up to q_k, where M is the mass of the object and the qs are its electric charges in this toy world that has k different types of photon:

$$M \geq Z(q_1, \ldots, q_k).$$

Z is the central charge of the supersymmetric theory, and states with $M = Z$ are the lightest states in the sector with a given charge. And as such, just like the electron in the real world, they are absolutely stable.

Now, these aren't realistic black holes. The real black holes in nature aren't supersymmetric. If they carried any electric charge, it would be quickly neutralized by attracting opposite charge from nearby in space. Electromagnetism is a tremendously powerful force compared to gravity. Nevertheless, these toy objects do genuinely deserve to be called black holes. They have an event horizon, they have a singularity, and they share many features of their real-world cousins.

Even better, the ability to construct these supersymmetric states can be used in the following clever way. Recall that string theory has two kinds of objects: strings and D-branes. Depending on the details of the string compactification you consider, this gives two different natural ways to construct charged states. In perhaps the simpler way, if there is a circle in the extra dimensions, you can simply wrap the string on the circle.

You can make a large number of such wound states by exciting either the left- or right-moving oscillator modes of the string, creating waves moving exclusively to either the left or the right. Such chiral oscillating strings preserve some supersymmetry. They can be used to make (what look like) charged particles in the transverse non-compact dimensions of space-time. These states saturate a bound of the form $M = Z$.

A second way to make particles in the extra dimensions is to take some of the p-dimensional branes of the theory and wrap them on p-dimensional subspaces of the compact Calabi-Yau space above some point in the non-compact space. The string described previously is an example of this with $p = 1$, but as long as suitable wrappable subspaces exist in the extra dimensions, you can also wrap Dp-branes with $p > 1$ on them. These states carry a charge of their own. It intuitively counts "wrapping number" of the given (topologically nontrivial) cycle in the Calabi-Yau space.

With either method, you obtain what looks—to an observer in the transverse flat space—like a charged particle. It carries some of the many possible electric charges of the supersymmetric theory.

But now consider the dials that can be used. As discussed in lecture 3, supersymmetric models come with moduli spaces of vacua; that is, they give rise to model universes parametrized by expectation values of scalar fields determining, for instance, the size and shape of the extra dimensions.

There is actually one preferred such scalar that can play a special role. It is the string theory coupling constant g, which controls the strength of interaction of strings. You can think of this as the rate at which strings split and join in Feynman diagrams. How will this help? You now need to be quantitative about two important notions.

First, you have learned that black holes form when there is too much mass in too small a region of space. How much mass in how small a region? It turns out that the event horizon radius (r_S) of a black hole of mass M is given by the following formula:

$$r_S = \frac{2G_N M}{c^2}.$$

If you compress matter of mass greater than M into a region smaller than this, a black hole will form.

Secondly, in string theory, the string coupling constant determines the strength of all fundamental interactions coming from the strings, including gravity. So, in string theory, Newton's constant—that is, G_N—can be derived and satisfies the statement that G_N is proportional to g^2:

$$G_N \sim g^2.$$

So, if you take one of your absolutely stable charged objects—made from wound oscillating strings or wrapped branes—and crank up the value of the string coupling constant, the mass of the object will remain the same, but the Newton constant will increase. By doing this, you can bring the object inside its own Schwarzschild radius. Because of the magic of supersymmetry (and the fact that the object has $M = Z$), it can't decay or do anything else unpleasant as you turn up the string constant. It has no choice but to persist and become a (charged) black hole.

String theorists have found that in many cases, charged black hole entropy can be equated to the log of the number of microstates of oscillating wrapped branes and strings:

$$\text{black hole entropy} = \log(\text{number of microstates oscillating wrapped branes and strings}).$$

This counts as a remarkable conceptual success of the theory.

Reading

Susskind, Leonard. *The Black Hole War*. New York: Little, Brown and Company, 2008.

6

How Strings Imply a Holographic Universe

In the previous lecture, you learned that the most basic black hole solutions of Einstein's theory describe states where matter has been compressed to an enormous density. As a result, it forms an object with such intense gravity that even light itself cannot escape. More precisely, for a black hole of a certain mass, there is a critical radius given by twice Newton's constant times that mass divided by the speed of light squared. An event horizon forms at that radius at the center of the black hole, and nothing that enters the event horizon can ever escape. Interestingly, black holes obey laws just like the conventional laws of thermodynamics. And that association indicates that to a black hole of a given Schwarzschild radius (the radius of the event horizon), an entropy given by the area of the event horizon should be associated. This lecture helps you digest this information.

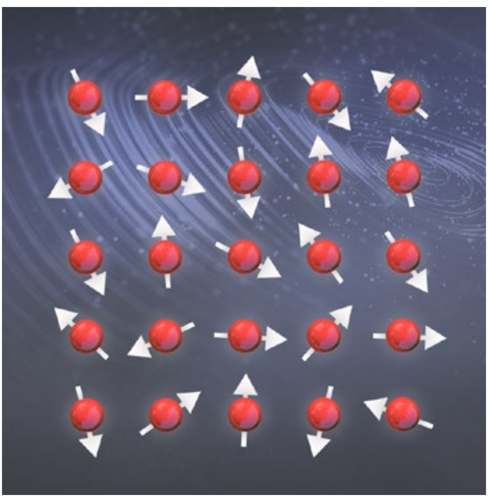
Entropy and Volume

How are the entropies of everyday objects expected to behave? Intuitively, say that you have a bunch of small masses (such as atoms or molecules) in a space of volume V. Then you'd expect them to have different configurations, where each one is exploring a different part of the volume.

It follows from thinking carefully about the fact that each small particle can be anywhere in the volume that the entropy S should be expected to scale like the volume. In other words, the entropy should be proportional to the volume:

$$S \sim V$$

That means if you double the size of the box while keeping the local physics similar, you should expect the entropy to grow by a factor of 2 × 2 × 2, or 8. That remains true when you think about quantum systems. For instance, quantum mechanics associates elementary particles with an intrinsic spin angular momentum, with the electron, for instance, having two possible spin states.



Imagine an array of these quantum spins in a lattice with the shape of a cube. Rows of these spins, with spin arrows pointing in different directions through each location, give a crude example of a magnetic system in solid-state physics. Locally, you can choose the direction of each spin. This gives an independent local choice for each unit of volume that contains a quantum spin. That results in an entropy that, again, scales with the volume.

Physical quantities characterizing a system that scale with the system size this way are called extensive. This is distinct from properties that remain fixed as you scale the system size, called intensive.

Entropy is a poster child for an extensive quantity in most of physics.

Returning to Black Holes

Black holes are formed out of conventional stuff that just happens to be in a state of high density. They have a natural size, given by the Schwarzschild radius. You'd then guess a priori that their entropy would scale like the volume; that is, the entropy of a black hole should go like the Schwarzschild radius to the third power:

$$S_{\mathrm{BH}} \sim R^3.$$

But this isn't the case. Instead, according to the laws of black hole mechanics discussed in lecture 5, the entropy of a black hole famously goes like the area, or the square of the Schwarzschild radius:

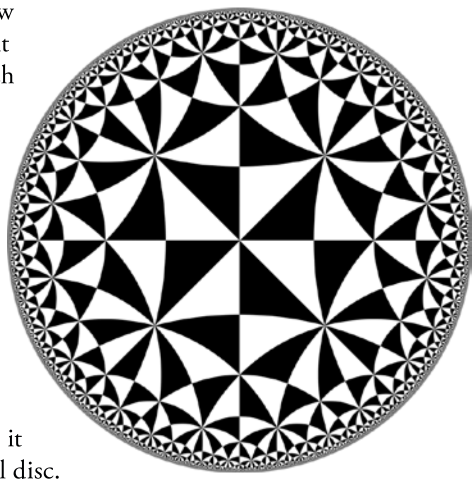
$$S_{\mathrm{BH}} \sim A \sim R^2.$$

Some of the black holes that are known of in nature are truly enormous. For instance, Sagittarius A*, the black hole in the center of our galaxy, has an estimated Schwarzschild radius of 12 million kilometers.

The relevant differences in entropy between the two formulas (or especially between the exponentials of the entropies, which measure available physical states of the system) would be enormous. But—as is often the case with good problems in theoretical physics—the problem here isn't really one of a numerical discrepancy. Often, you can understand the basic qualitative physics of a system while having a sufficiently hazy grasp of the details that precise computation remains elusive.

What is happening here is much more confusing. At its heart, the reason that entropy is extensive is that in a typical physical system, you are allowed to excite local degrees of freedom at each point in space. And the number of points in space scales like the volume of the region in question. So how would you make a physical system in volume V where this wasn't the case? That would mean that somehow even basic facts about how you are free to do different things at different points in space (which is deeply tied to the notion of locality in physics) would have to be wrong. That's a conceptual issue worthy of deeper understanding.

Let's approach a striking, modern view of what is happening at the black hole horizon through examining a suitable work of art. In a region of space near the center of the disc, each triangle is quite large, while as you move out toward the boundary of the disc a finite distance away, the number of triangles per unit area is growing very rapidly. In fact, if the artist were able to draw with perfect resolution, one could fit infinitely many tiny triangles in each fixed unit of area in the picture as one approached the boundary.



To change your perspective, you need to imagine two things:

1. Suppose you learned that this picture should be interpreted as follows: Each triangle drawn represents an equal unit of area. It is just hard to draw so it looks that way on the physical disc. This would mean that, in a real sense, all of the area of the disc would be concentrated very close to the boundary. So, while the disc is two-dimensional, a one-dimensional slice (with infinitesimal thickness) very close to the boundary would contain most of the area. Spaces like this really exist. This is a drawing of a spatial slice in something called the three-dimensional anti–de Sitter space. (The missing dimension is time.)

2 You can make similar mathematical objects (though drawing them is harder) also in other space-time dimensions. For now, let's stick to two spatial dimensions and a two-plus-one-dimensional space-time. But the concepts all generalize.

Unpacking Anti–De Sitter Space

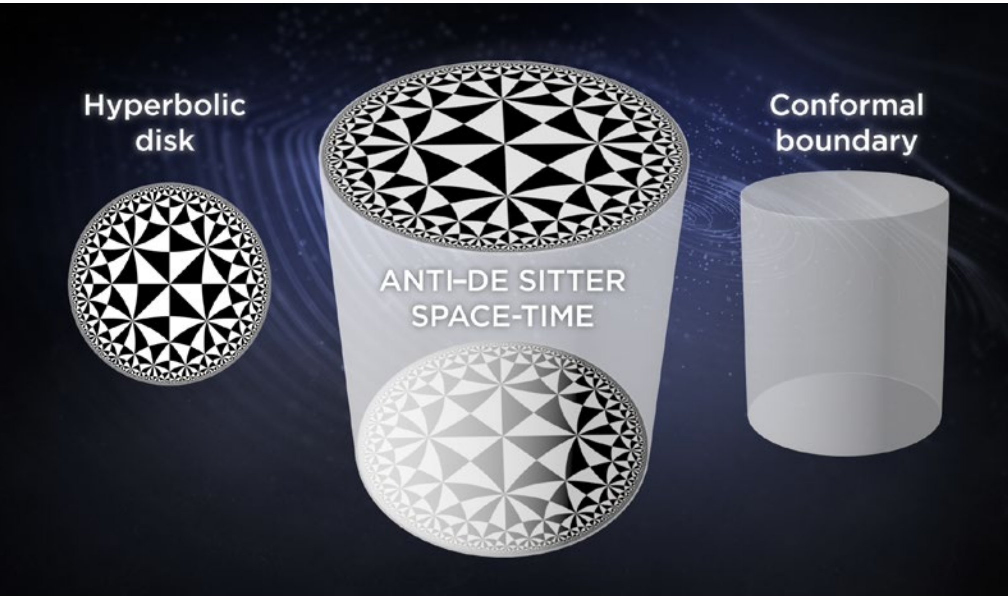
Willem de Sitter was a Dutch mathematician at the turn of the 20th century. Soon after Einstein's theory of gravity was discovered, de Sitter wrote down the simplest solution of the theory in the presence of a positive energy density in the vacuum (what today would be called a positive cosmological constant). The space de Sitter discovered is called de Sitter space.

> De Sitter space may be the single most important solution of Einstein's theory—because it may describe our universe today.

But de Sitter's solution describes a space where there is a positive energy density permeating space. What of its evil twin, where instead there is a negative energy density permeating space? This is aptly named anti–de Sitter space. That's actually the space-time that has been described previously: the simplest solution of Einstein's theory with a negative cosmological constant (though, for simplicity, the two-plus-one-dimensional version has been described).

A peculiar property of anti–de Sitter space has been uncovered: Almost all of its area lives at the boundary. This indicates that maybe a way of talking about physics inside of anti–de Sitter space can be found by formulating a description purely on the boundary.

In the three-dimensional space-time you've been studying, space-time would look like a soup can, with the horizontal slices of the can being space and time moving up along the can. The postulate of holography states that Einstein's gravity in anti–de Sitter space in D space-time dimensions can be exactly reformulated in terms of an ordinary quantum theory of matter and forces, without any gravity, in $D - 1$ dimensions. The $D - 1$ dimensions live at the boundary of the anti–de Sitter space, or the edge of the soup can.



Broadly, you've learned about what quantum theories of matter and forces look like in modern physics. One chooses a group G to specify the force-carrying particles and some representations R of G to specify the charges of the matter under the forces. The remarkable statement of holography is the following: Any theory of quantum gravity in the bulk of anti–de Sitter space can contain a huge list of dynamical objects in addition to the gravitational field itself, like a list of bulk matter particles and bulk force carriers.

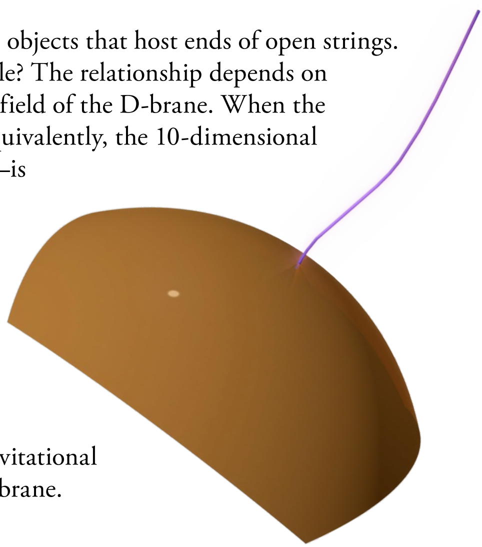
Then there is a dual description, capturing exactly the same physics, in terms of a normal quantum theory of matter and forces on the boundary (in one dimension less). This theory is specified by some G and choice of matter particles. The details of the G and matter particles depend on the details of the bulk theory of gravity. Both the magnitude of the negative cosmological term and the details of which kinds of forces and particles live in the bulk are important and are captured by the boundary theory.

How did anyone ever think of this duality? And what does this have to do with black holes and their "area law" entropy? Both questions can be answered at once: String theory (as interpreted by Juan Maldacena, an Argentinian theorist) both provided the duality and related it to black holes.

String theory, you'll recall, contains not just strings but also D-branes: the extended objects (with *p* spatial dimensions for a D*p*-brane) on which open strings can end. In a theory of gravity in 10 space-time dimensions, such as string theory, there is room for "black objects"—extended objects with horizons—of many different dimensionalities. You can have a black hole centered at a point whose horizon is then a sphere surrounding that point in 10 dimensions—so the horizon would be eight-dimensional.

But you can also imagine black objects where the point-like singularity of the black hole instead extends along a *p* spatial-dimensional sheet. Then, the horizon would actually be an eight-minus-*p*-dimensional sphere surrounding this extended singularity. These objects are called black branes, or black *p*-branes for a specific value of *p*.

D-branes have been described as objects that host ends of open strings. How is this related to a black hole? The relationship depends on the strength of the gravitational field of the D-brane. When the string coupling constant—or, equivalently, the 10-dimensional analogue of Newton's constant—is small, there is a garden-variety D-brane. However, as you dial up the strength of the Newton constant (or the string coupling constant), the energy density of the D-brane creates a stronger and stronger gravitational field. Eventually, it forms a curved gravitational object with a horizon: the black brane.



More precisely, if *N* D-branes are stacked on top of one another, they carry *N* times as much energy density. So the relevant quantity governing whether gravity is strong or weak is a product of the Newton constant and *N*, or $g_s N$.

Depending on whether $g_s N$ is very large or very small, you will obtain a gravitational object (a black brane) or a thin D-brane, best understood by quantizing the open strings that end on it.

String theorists such as Igor Klebanov noticed the following amazing fact: Imagine computing the scattering of strings off a black brane. It turns out that it behaves in a remarkably similar way to scattering computed in a certain quantum theory of matter and forces (the maximally supersymmetric Yang-Mills theory).

Maldacena was able to codify this in a precise way in 1998. He conjectured that if you looked at the gravitational solution very close to the horizon of the D-branes, you would obtain a theory that was dual to the maximally supersymmetric Yang-Mills theory. That just happens to be the quantum theory that you obtain by quantizing the open strings ending on the branes.

The geometry near the horizon of the D3-brane is in fact a copy of the five-dimensional anti–de Sitter space (times a sphere to fill up the rest of the 10 dimensions). Three of the dimensions of anti–de Sitter 5 (AdS_5) come from the spatial directions along the brane, and one comes from the time direction along the brane. The fifth is the radial direction describing moving away from the horizon.

Maldacena's conjecture is an example of holography. It says that quantum gravity in AdS_5, of the precise form given by string theory in the presence of D3-branes, is dual to an ordinary quantum field theory: the maximally supersymmetric Yang-Mills theory in three-plus-one dimensions. This observation has opened an amazing renaissance in the exploration of both quantum gravity and regular quantum theories without gravity. By using the duality, you can relate "easy" questions in one side of the duality to "hard" questions on the other side.

While you might think that quantum gravity is going to be the hard side for most purposes, this is not entirely correct. Many of the most mysterious phenomena in nature—from the confinement of quarks in the strong interactions to the behavior of nuclear matter in the dense core of neutron stars—involve strongly coupled quantum theories of matter and force. Using the duality, questions about these complicated quantum theories of matter and forces can sometimes be mapped to questions in a gravity theory in a large, weakly curved space-time. Then, classical Einstein gravity is a good approximation. So, classical gravity can be used to answer hard quantum questions on the other side.

The duality also clarifies the "area law" entropy of a black hole. For these (general) black p-branes, the volume inside the event horizon of the black object is a p-plus-one-dimensional volume. But the brane itself—or the boundary of the near-horizon AdS geometry—is only p-dimensional. By the duality, the entropy of the bulk p-plus-one-dimensional gravitational system must be captured by that of an ordinary quantum theory in p dimensions. This is the (p-dimensional analogue of the) area law.

But can't you easily show that this is wrong by taking a volume V of space in the gravitational theory and exciting little bits of energy independently in a bunch of small hypercubes within this volume? Then you'd be constructing states capturing a volume's worth of entropy.

The flaw with this argument rests in the fact that the theory is a theory of gravity. As you try to excite independent degrees of freedom within the volume, the energy density quickly becomes large enough that the system collapses to form a black hole. And the point where this happens never allows you to exhibit more than the area law worth of entropy. This argument was made precise by Jacob Bekenstein, who was a master of such thought experiments, in the 1970s.

Holography and the Big Bang

In this lecture, you've encountered an amazing new thing in physics: an exact equivalence between a physical theory in D space-time dimensions and one in $D - 1$ dimensions. The fact that one side is a quantum theory of gravity and the other isn't offers hope of a precise formulation of quantum gravity in terms of the better-understood rules of conventional quantum theories of matter and forces.

On the other hand, the place you've started to understand holography is anti–de Sitter space, which enjoys some very special features. Almost all its volume arises at its boundary. It has a natural boundary given by space at infinity. These facts don't generalize. The real world is much closer to flat Euclidean space (promoted to Minkowski space by including time) or, even better, to de Sitter's original expanding cosmology. In neither of these cases do physicists yet have nearly the kind of control over holography that they do in anti–de Sitter spaces. This arena of research has seen fascinating and promising developments. A future wave of advancement in quantum gravity will likely involve decisive progress on the problem of de Sitter holography.

A typical picture of de Sitter space describes an expanding universe. Running the clock backward, this means that in the far past, the universe was very small and highly curved. Einstein's theory has a natural scale built into it: the Planck length of about 10^{-32} centimeters. When the curvature radius of the universe approaches this tiny scale or the energy density exceeds the comparable measure of energy (which is 1 Planck mass per Planck volume), you can anticipate trouble. This is the regime of physics where the big bang occurs.

Reading

Klebanov, Igor, and Juan Maldacena. "Solving Quantum Field Theories via Curved Spacetimes." *Physics Today* 62, no. 1 (2009): 28–33.

7

The Origin of the Universe

String theory is a potential theory of quantum gravity. A new kind of theory to describe quantum gravity is needed for several reasons. One good reason is visible if you expand various calculations in a power series in Newton's constant. If you scatter gravitons off one another, as in the two-to-two graviton scattering process represented by a Feynman diagram, then, by dimensional analysis, the result has to be proportional to a positive power of the scattering energy. This is because each vertex in the diagram—which represents a gravitational interaction—is proportional to a factor of Newton's constant, which has dimensions of inverse mass squared, or inverse energy squared, in a relativistic theory where $E = mc^2$. Therefore, high-energy graviton scattering in naive quantum mechanics is divergent. A new kind of quantum theory—a sensible quantum theory of gravity—is needed to cure these divergences. There is strong evidence that string theory provides such a set of theories. However, there is another source of puzzles in classical gravity or general relativity.

Big Bang Cosmology

Recall that the black hole space-time has the following features: There is an event horizon that arises at a distance from the black hole center set by the scale of the mass that collapsed to form the black hole. There is a singularity at the center of the black hole. And as you approach the singularity, the curvature of space becomes very large. In practical terms, this means that any object sufficiently close to the singularity is torn apart by the diverging strength of the gravitational fields.

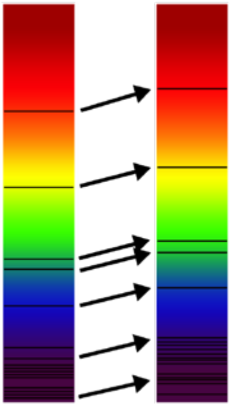
The singularity is disturbing as a point of principle. The laws of physics cannot be evolved past a point in time where an object hits the singularity. What happens is simply unknown. However, you could take the following attitude: The black hole singularity is cloaked by an event horizon that surrounds it. This makes its effects undetectable to an outside observer. So, why should they care what happens there? As long as all singularities are cloaked by horizons, you need not worry about their effects—unless you are unfortunate enough to be on the wrong side of the horizon.

However, in the current best models of physics, there is another place where singularities arise and where this same excuse doesn't obviously work. That is in the origin of the universe itself.

The study of the history and dynamics of the large-scale structure of the universe is called cosmology. Perhaps the foundational observation of modern cosmology—which emerged in the early 20th century—is that the universe is expanding. This was discovered through the physics of redshift.

You have likely heard the abrupt change in tone of a siren as an approaching ambulance or police car passes you and starts to recede. The sound goes from high to low pitched, corresponding to a change in the frequency of the sound waves. This is known as a Doppler shift.

A similar phenomenon holds with light waves. Different colors of light correspond to electromagnetic waves of different frequencies. Via the pneumonic ROYGBIV, you can remember that the lowest-frequency light is red, while the highest is violet. There is an electromagnetic analogue of the



Doppler shift: A receding light source shows up with a slightly redshifted color (depending on its velocity); one that is approaching instead has a corresponding blueshift.

In the late 1920s, Edwin Hubble, using the telescope at Mount Wilson outside of Los Angeles, found that light emitted from distant galaxies seems to have a frequency slightly shifted down—redshifted—with a magnitude of redshift depending on the distance of the galaxy. This admitted a straightforward interpretation: The galaxies are moving away from earth, with a velocity dependent on their distance.

This is the first foundational observation underlying what is known as big bang cosmology. Why "big bang"? Simply run the picture of the expanding universe backward. The simplest-possible inference is that the universe originated in a state of extremely high energy density at a finite time in the past. It "exploded" in a big bang from which expanding space itself emerged.

The Cosmic Microwave Background

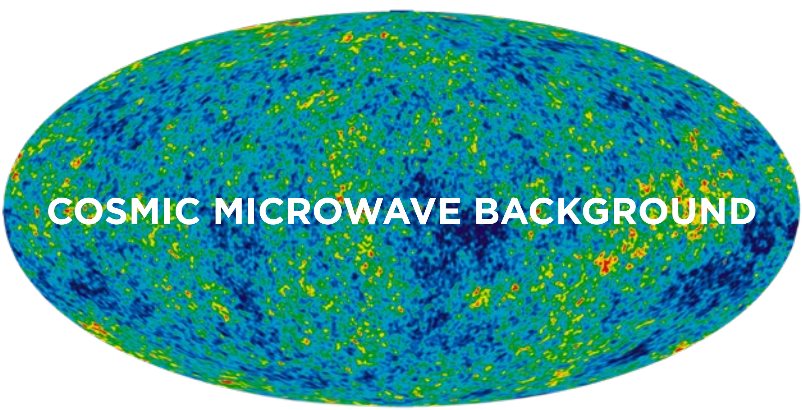
This basic picture—that the universe started in a hot, dense state and then expanded and cooled—enjoys several pillars of experimental support. The first is the redshift observations previously mentioned. A second is the observation of historical relic radiation from this hot, dense phase: the cosmic microwave background (CMB). The basic physics of the CMB is as follows.

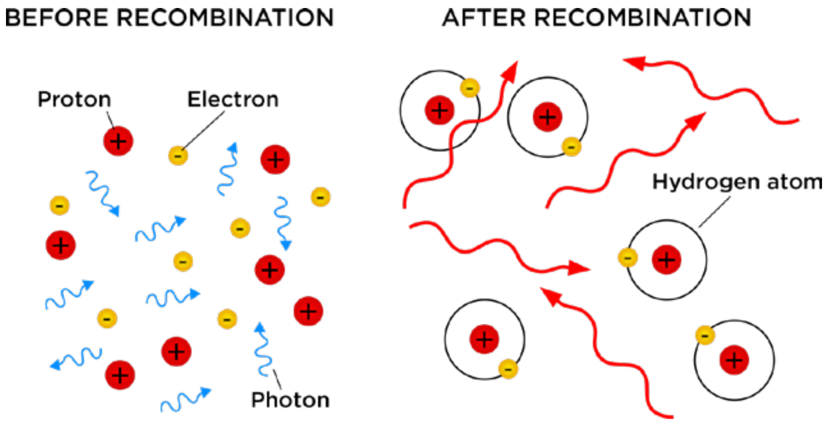
In the early hot, dense phase of the universe, the temperature was so high that bound states of electrons and protons—such as hydrogen and other more complicated atoms—could not exist. The energy binding the orbiting electron to the proton in hydrogen was simply too small to withstand the high temperatures. Any bound atoms were instantly broken apart by collisions with high-energy particles in the surrounding heat bath.

As the universe expanded, it cooled. At a certain point, the temperature fell enough that stable hydrogen atoms could form. At this point, the universe in bulk became electrically neutral. Because particles of light—photons—scatter off charged particles, before this time they bounced around like pinballs, scattered by one charge after another. But after this time, surrounded by neutral hydrogen gas, they moved in straight lines—more or less without disturbance. A lucky few such photons were captured by satellites, radio dishes, and other detectors.

Such relic photons are modern remnants of the big bang, traveling straight to earth from the time when protons and electrons combined into hydrogen for the first time. Amazingly, they were detected unintentionally at Bell Labs in 1964 by Arno Penzias and Robert Wilson.

The detailed properties of the CMB conform to a model of a hot big bang that was cooled by the universe's expansion. The CMB provides a second pillar girding the big bang model. A third support pillar is provided by the measured nuclear abundances. Just as you can understand the process by which electrons and photons combine to produce hydrogen in the expanding universe, you can also understand the process whereby protons and neutrons combine to form the nuclei of heavier atoms. They can only form nuclei after the temperature becomes sufficiently low that it doesn't overwhelm the nuclear binding energy holding the nucleons together.





It turns out that the light elements—such as hydrogen, helium, and lithium—were produced by this big bang nucleosynthesis process. The abundances with which these elements populate the universe is in good accordance with the big bang model. It turns out that the heavier elements were born later in the centers of stars or in even more extreme environments, such as those produced when neutron stars collide.

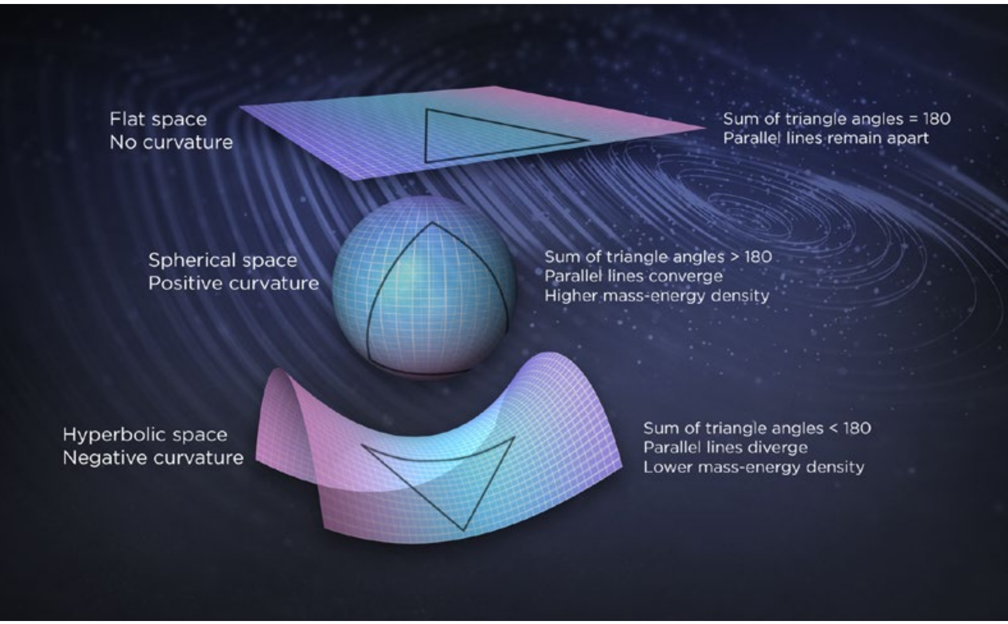
The Flatness Problem

So, on the assumption that the big bang model is basically correct, an initial singularity is still left to be explained. This singularity was not, in any obvious way, cloaked behind a horizon. The initial conditions out of which big bang cosmology emerges were, after all, a direct result of what came before—i.e., the nature of what banged, why it banged, and how it banged must be hidden in the physics of the singularity.

A direct frontal assault on the singularity problem has not proven fruitful. Happily, the universe has provided additional strong experimental evidence for the nature of some predecessor to the hot big bang. This evidence comes from two problems, known colloquially as the horizon and flatness problems. Let's focus on the flatness problem.

While curved-space geometry is an intricate subject, you can get some basic insight into it by considering the simplest cases: spaces with constant curvature (i.e., the same curviness everywhere in the space). In two spatial dimensions, the spaces of constant curvature are the sphere (for positive curvature), the Euclidean plane (for vanishing curvature), and the saddle (for negative curvature). A local way to measure this curvature is to draw a triangle (defined by a suitable set of intersections of straightest-possible lines).

It turns out that while the angles inside a triangle add up to the familiar 180° in flat space, they add up to more in positively curved space and less in negatively curved space. And the surfeit or deficit provides a measure of the curvature. While Euclidean flat-space geometry is taught in school, it is an experimental question whether space is flat or curved. And it turns out that the best present experiments find—to within astonishing accuracy—that space is flat.

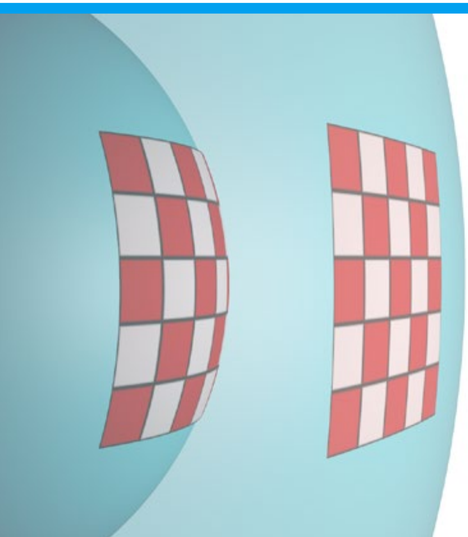


The trouble with flat space in expanding cosmology is that it isn't very stable. It is sort of like a pencil standing on its tip. The more precise version of this statement is as follows.

Suppose the early universe started with conventional matter in a space that was approximately (but not exactly) flat. Then, the expansion of the universe itself drove it further away from flatness. A slight positive curvature at early times became a huge positive curvature at late times—with the mirror fate for negative curvature. This happened, roughly, because the energy density in the matter diluted away with the expansion faster than the energy density associated with the curvature.

Cosmic Inflation

Present observations place the time of the big bang around 14 billion years ago. For the universe to remain flat in appearance so long after the big bang, physicists can estimate how exquisitely close to exactly flat it must have been at very early times. Expressed in terms of the tuned energy density required to support a flat geometry, tunings on the order of 1 part in 10^{60} are calculated. This problem and others suggest that the initial state of the universe must have been very finely tuned. The only out is if a robust mechanism that would produce such flat geometries more or less regardless of the initial state can be found.



The surface of the earth looks flat over small enough regions. An idealized model of the earth's surface is as flat as a balloon surface if the balloon had a radius of about 4,000 miles.

As it turns out, exactly such a mechanism has been proposed. It is growing into a pillar of its own in the modern view of early-universe cosmology. This is the idea of cosmic inflation. Suppose the initial state of the universe were highly curved. How do you produce a flat space—of the sort that's roughly observed—out of a curved space? An image of an expanding balloon provides a hint. As the balloon gets bigger, any portion of fixed area on the balloon surface looks closer and closer to flat.

If this trick is going to be used to explain why, to modern cosmologists, the entire visible universe looks approximately flat, a pretty significant expansion factor on the original (post–big bang) balloon is going to be needed. It is standard to measure such changes in size in *e*-foldings. The size of the universe must have expanded, post–big bang, by a factor of e^N, where N is much larger than 1, to adequately explain the flatness. It turns out that to solve the flatness problem—and other problems, such as the horizon problem—this expansion must have been much more rapid than the expansion that Hubble detected with redshifts and that is seen today. And, in fact, it must have been accelerating as it proceeded.

In trying to explain such a phase of accelerated expansion using only the typical ingredients of fundamental physics—things like quantum fields with various potential energy functions—one clever idea has clearly stood out. This is the idea underlying cosmic inflation, and it can be explained in a few steps.

First, suppose space had a constant positive energy density per unit volume, V. Then, it follows from Einstein's equations that the scale factor describing the size of the universe—usually denoted by a as a function of time, or $a(t)$—expands as $a(t)$ grows like e^{Ht}, or H as determined by the value of the energy:

$$a(t) \sim e^{Ht},\ M_P{}^2 \times H^2 = V.$$

With typical scales of fundamental physics appearing, V could be as large as the string scale, so H could be a very large energy scale, indeed.

What this means (remembering that, in dimensional analysis, time and energy have inverse units) is that the universe could potentially expand by many, many *e*-foldings (or factors of *e*) in even a small fraction of a second.

The important difference between the assumptions here and those that led to the flatness problem is that the energy density sourcing the expansion remains basically constant as the universe expands. It should not dilute, like conventional matter does.

Second, however, this also leads to a problem: There is no graceful way to end a phase of expansion with an exactly constant energy density of space. The space will inflate forever, or it will need to end by some dramatic process (such as quantum tunneling), which is unlikely to preserve the successful resolution of the problems that inflation was introduced to solve.

However, as a third point, there is no need for the potential V to be an exactly constant energy density of space. Consider instead a scenario that can produce close-to-constant energy density. A scalar field called the inflaton is added to the content of the physical theory. It has a potential energy function with a very flat region. Any part of the universe that starts on this part of the potential "thinks" it has an approximately constant energy density V (set by the magnitude of the inflaton potential at this plateau). The universe therefore exponentially expands. If the scale of the scalar potential is set by high-scale fundamental physics, this can offer a natural mechanism for raising large, flat regions of the universe.

The current cosmological horizon—the distance that light has traveled since the big bang—is about 10^{28} centimeters in size. The flat regions produced by inflation could even be many times larger than that. The beauty of this idea is that it uses familiar ingredients—a scalar field, a potential, and gravity—to solve the important and challenging flatness problem. But as an extra bonus, it also provides a simple potential graceful exit from inflation. This can smoothly match onto the conventional big bang cosmology.

The exit from inflation can occur because of the dynamics of the field phi. Suppose the inflaton isn't simply sitting still on its potential, but its value slowly evolves in time. It rolls toward the end of the flat plateau. When it reaches the end, it quickly "rolls down" to its potential true minimum. This causes the universe to naturally exit the phase of exponential expansion. And as the inflaton rolls to its minimum, if it enjoys couplings to the particles of the standard model, it can naturally decay to standard-model degrees of

> The interface of string theory with cosmology is a rich and exciting subject. String theory—or some other theory of quantum gravity—is necessary to derive a fully trustworthy theory of inflation.

freedom. This process—of the inflaton decaying to a hot gas of standard-model particles—is called reheating by inflationary theorists. Reheating could be the initial condition that underlies the hot big bang.

It is important to acknowledge that none of this resolves an even bigger question: What came at the start of the universe? Even inflationary cosmologies have a cosmic singularity in the distant past, before the time when inflation started. The problem of the initial singularity has been further hidden by the exponential expansion of space, but it is not solved. Still, inflation and the hot big bang provide an attractive potential picture of the early history of the universe.

Reading

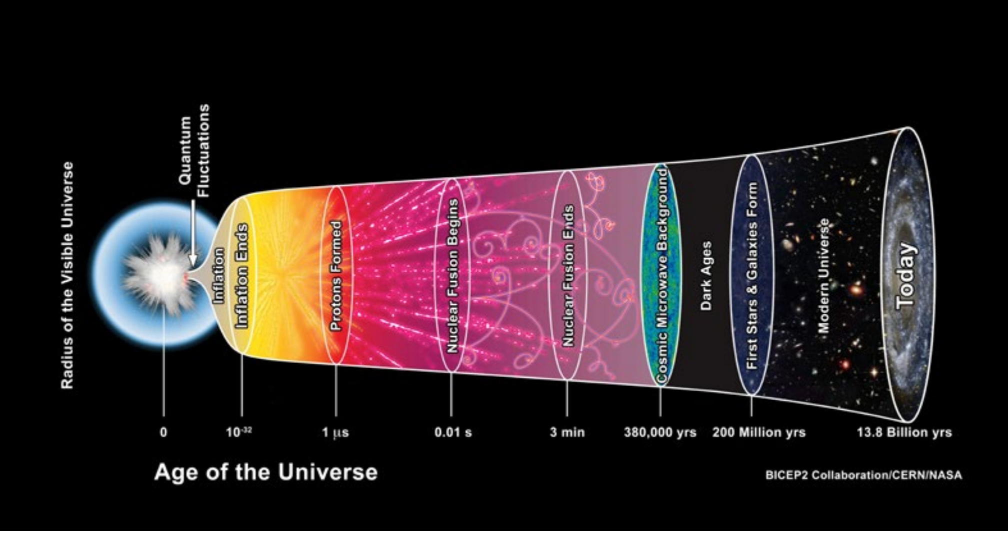
Guth, Alan. *The Inflationary Universe.* New York: Basic Books, 1998.

8

Strings and Inflation

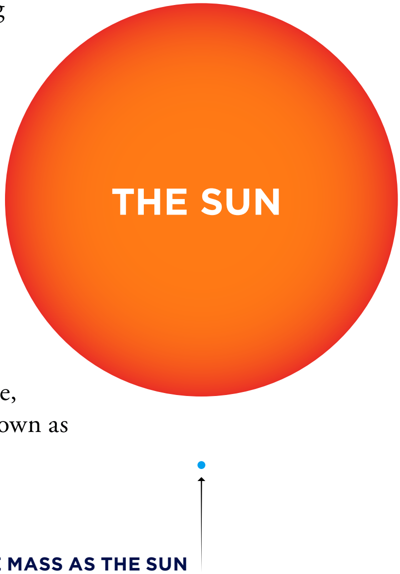
Theorems proved in the 1960s show that general relativity breaks down at very early times. During these times, it is thought that the universe expanded much more rapidly than it does today. In fact, it doubled in size every fraction of a second until it grew 25 orders of magnitude (or more) in size. This cosmic inflation is a leading theory of early-universe cosmology. It was supported by a new kind of matter—an inflaton field—which very closely mimics a constant energy per unit volume, filling all of space.

This lecture enriches this story—and its connection to string theory—in three ways. First, you will encounter another amazing cosmological discovery, known as the dark energy. Then, you will discover how both the dark energy and early universe inflation are sensitive to the full details of a quantum theory of gravity, like string theory. Last, you will learn about a startling and puzzling picture of the universe that seems to emerge from the best ideas that exist about how string theory might model universes like ours.



The Mystery of the Dimming Supernovae

Lecture 5 addressed the fate of "old" stars—stars that have burned all their nuclear fuel. Depending on their mass, one of a few different things happens to such a star. The known stable endpoints are white dwarf stars, neutron stars, and black holes. These objects are very dense. For instance, the sun has around 100 times the radius of the earth, but a white dwarf star with the same mass as the sun is about the size of the earth. It is much smaller and denser than any possible progenitor star. In the process of collapsing from their large, youthful size toward their ultimate denser fate, many stars give rise to catastrophic events known as supernovae.



Stars are held up against the crush of their own gravity by nuclear fusion. When they completely exhaust successive stages of their fuel, this process no longer works, and they begin to collapse. A supernova occurs when, as the object collapses, its hot and increasingly dense center can ignite previously inaccessible nuclear processes. These cause one final runaway stage of nuclear fusion. This releases so much energy that there is a titanic explosion. The results of such explosions can be seen in the universe today.

This should sound like an incredibly violent and difficult process to model. But one class of such events is so standardized in its mechanism and behavior that, in fact, people use the light from the resulting explosions as standard candles: astronomical objects whose features are so well known that their properties can be used to calibrate any peculiarities of the atmosphere (including, for instance, giving a precise measurement of its redshift). These standard-candle stellar explosions are known as Type Ia supernovae.

Careful studies of large classes of distant Type Ia supernovae in the late 1990s revealed that you could find their distance by comparing their brightness in the sky to their known absolute luminosity. After all, the amount of light reaching you from a distant object will fall as the square of your distance from the object. But also, using Hubble's idea that the redshift of the light measures the expansion rate of the universe, you could see how fast the universe was expanding at the time the supernova exploded.

Amazingly, combining these measurements revealed that the universe today is enjoying an accelerated expansion. This consequence follows from the fact that the supernovae at a given redshift are fainter than they'd be in a universe where gravity was slowing the expansion of the universe. So, something is causing an acceleration of its expansion. And it's called dark energy.

This is puzzling because you'd expect the gravity of all the normal matter in the universe—planets, stars, and gas—to pull together and slow the expansion, since it does have that effect in cosmology. But instead, the dominant source of energy density at the vast distance scales relevant for cosmology today, which is the dark energy, has to be exotic enough to have antigravitational effects.

In fact, Einstein had toyed with including just such an antigravitational source in his original general theory of relativity. It was an overall energy density for otherwise empty space, also called a cosmological constant. Such a constant—with a positive sign, the opposite sign from the one relevant for the anti–de Sitter spaces discussed previously—gives rise to something known as de Sitter space. This is an exponentially expanding cosmology.

Cosmology in the early 20th century was too crude a science to detect a cosmological term. Einstein later set this possible term in his equations to zero and called its introduction "my greatest blunder." In retrospect, the joke is on him—it appears that such a term is needed to correctly account for the behavior seen in modern precision cosmology. In other words, the observed dark energy seems to share the detailed properties of Einstein's cosmological constant. It is nonzero and positive. However, the presence (or even the absence) of a cosmological constant brings to the fore a central theoretical question: Why is the value of dark energy what it is?

In many cases, understanding the precise value of something is difficult, but you can get a good sense for its basic magnitude. In the simplest cases, the basic magnitude follows from dimensional analysis—i.e., just making units work out right.

There is a natural scale in Einstein's theory of gravity: the Planck mass. The Planck mass, for instance, is what determines the value of Newton's gravitational constant. So, it would be expected that the cosmological term—or the dark energy—would be given by an order of one number in the appropriate Planck units. But in fact, measured relative to its natural scale in Planck units, the dark energy that's seen is roughly 120 orders of magnitude too small. That is a pretty large error, even for a crude approximation. The basic problem here is that the vacuum energy can receive contributions from physics occurring at all possible energy scales.

> Origin myths for the universe are a feature of every human culture. It is hard not to extrapolate back in time and wonder how it all began.

Dark Energy and Inflation

A normal physical process is characterized by a typical size or energy scale of the objects involved. For instance, in the standard model, the strong force—described by quantum chromodynamics (QCD)—exhibits quark confinement at a typical energy scale of a few hundred MeV. This is the energy cost it would take to break apart a confined meson. At higher energies, around the TeV scale, the Higgs boson is believed to develop its nonzero expectation value.

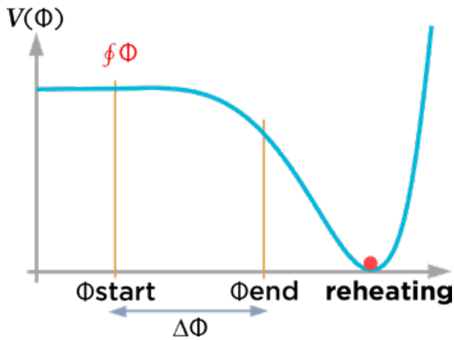
This breaks the electroweak symmetry down to the electromagnetic and weak interactions seen today. This is a process of spontaneous symmetry breaking, which was described in lecture 3. In QCD and the electroweak symmetry breaking, there are potential energies of the scale of (respectively) MeV and TeV. These could contribute to a nonzero vacuum energy. There is no reason they should cancel.

Such contributions could also occur from physics that appears at the scale where the known interactions hypothetically unify, known as the grand unification scale, or at the scale where extra spatial dimensions are rolled up ("compactified") in string theory. These are typically much shorter distance scales, associated with much higher energies, than the electroweak scale.

But the existing dark energy is not only 120 orders of magnitude smaller than the Planck scale; it is also 60 orders of magnitude smaller than the electroweak scale. In other words, all contributions to the energy density of flat space coming from processes at all energy scales above this must somehow conspire to cancel to many, many decimal places.

There is one known class of theories where this can happen naturally, because it is forced by a symmetry. This happens in theories with supersymmetry. However, supersymmetry cannot be exact beneath the electroweak scale. The required superpartners of the known particles are not seen. Supersymmetry cannot explain the small vacuum energy or cosmological constant in the real world.

A similar, but less dramatic, issue occurs in the models of inflationary cosmology discussed in the previous lecture. Recall that these are models where an exponential expansion of space is driven by an effective, almost-vacuum energy density. There's an inflaton field phi that has a potential with



a very flat plateau, and it slowly evolves along the flat part of its potential energy function. It fools the universe into believing that there is a large vacuum energy driving accelerated expansion. This explains why today's universe is so close to being flat. Just as a balloon blown up to many times its original size has a flatter and flatter surface, the same is true of the geometry of the universe.

Inflation is thought to occur at very high energy scales—that is, energies much higher than the value of today's dark energy, or even the electroweak scale. But even at such high energy scales, any new physics could lead to corrections to the shape of the inflaton potential. This is analogous to the way that any new physics can, a priori, correct the value of the measured dark energy.

This is puzzling. A notable and necessary feature of the inflaton potential is that it must be very flat as a function of the inflaton field phi. This is because only a very flat potential mimics a vacuum energy density of empty space (and hence drives the hypothesized model of early-universe inflation). Corrections to the potential where V of phi is changed to V of phi plus some ΔV of phi would generally spoil the desired flatness:

$$V(\phi) \rightarrow V(\phi) + \Delta V(\phi).$$

In quantitative terms, the measures of flatness involve products of the Planck mass and derivatives of the inflaton potential. This means that even corrections to the inflaton potential from quantum gravity—which are suppressed by the highest known scale in physics, the Planck mass—can shift the features of the inflaton potential in meaningful ways. Inflation requires a potential energy that is so flat it's sensitive to new physics—and corrections from quantum gravity or string theory—all the way up to the Planck scale. Adding a few Planck-mass particles with the wrong couplings to the inflaton field can spoil inflation.

In both the case of the dark energy and inflationary theory, this provides a tremendous opportunity. In many parts of theoretical physics, there is no good excuse to talk about quantum gravity or string theory effects. But here, discussing the full theory of high-energy physics—including the microscopic strings in string theory—is required to capture all the relevant physics.

So, how do you model something like dark energy or inflation in string theory? This general subject has been an area of tremendous activity in the last 15 to 20 years. An amazing—but tentative—picture of the large-scale structure of the universe has emerged.

The Large-Scale Structure of the Universe

Remember that a fundamental role in string theory is played by the need to compactify the extra dimensions the theory predicts. In some simple versions, above each point in the three-plus-one-dimensional space-time, there are six curled-up extra dimensions. In the best-studied models, the theory has exact supersymmetry, surviving all the way to low energies. The extra dimensions are not fixed in shape or size. As a consequence of the supersymmetry, in these toy models, the vacuum energy vanishes.

These are all properties of the vacuum solutions (solutions without extra sources of energy and density) that arise from string compactification on the Calabi-Yau spaces described previously. A priori, all of the points just discussed are problems for making a realistic world model out of string theory. They have consequences that are thought to be not true of the real world.

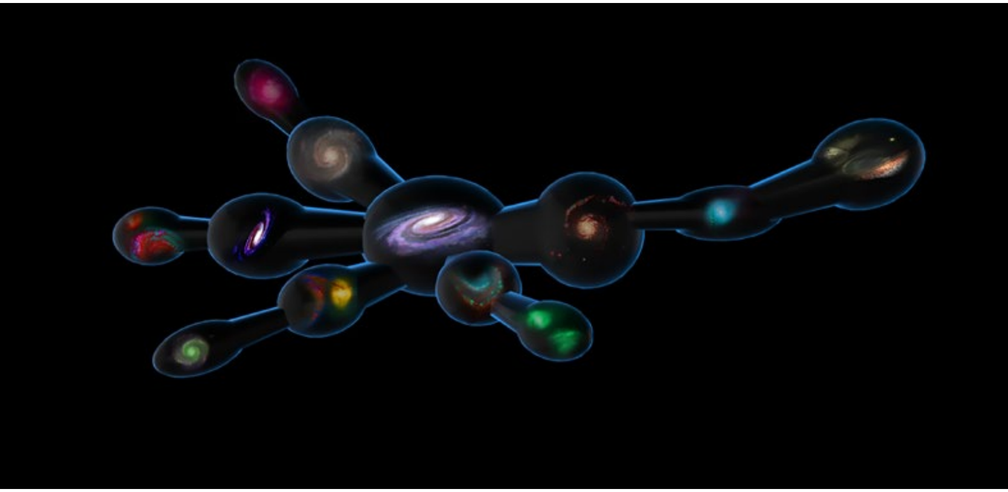
Another problem is that the Calabi-Yau solutions seem to be very nongeneric. The theory allows many background fields to be turned on that are absent in the Calabi-Yau solutions. In other words, there is no reason for the extra dimensions to characterize a vacuum solution; there could be energy sources and other stuff that lives in the extra dimensions. Why would nature pick a very special solution with no sources? Explorations in the early 2000s revealed that perhaps these problems are related.

The most obvious background sources you could turn on in string theory—to add background energy density to the Calabi-Yau solutions—are known as fluxes. In high school physics, you learn that in Maxwell's theory of electromagnetism, there are both electric (E) and magnetic (B) fields. In the presence of such fields, there is an energy density that goes like the square of the electric field plus the square of the magnetic field:

$$H \sim E^2 + B^2.$$

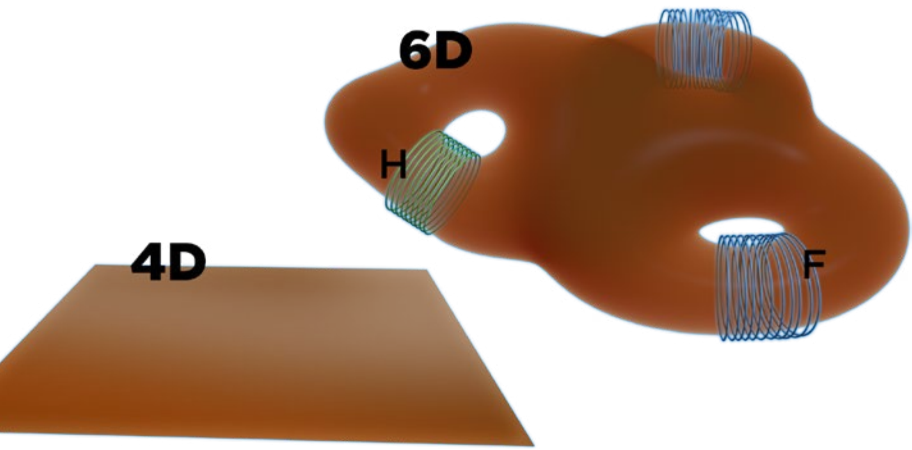
You can characterize a background magnetic field by its flux through various regions of the spatial geometry.

In the observed world, background fluxes of the electric or magnetic fields would not characterize anything that looks like our world at large scales. Among other things, these fields would break the translational and rotational invariance of our spatial dimensions. But in a world with extra dimensions, fluxes that thread the extra dimensions of space do not break the rotational or translational symmetries of "our" dimensions.



Higher-dimensional string theory contains analogues of the electromagnetic field. So, magnetic fluxes (analogous to the ones being discussed) could be threading two-dimensional cycles (providing cross-sectional areas) in the extra dimensions. But more generally, string theory also contains fluxes that can be generated by objects like the D*p*-branes described previously. The D*p*-branes come in flavors of various worldvolume spatial dimension *p*, and correspondingly they give rise to fluxes of fields that are appropriate to thread cycles of different dimension embedded in the six extra dimensions of space.

Roughly, the flux a D*p*-brane sources must be such that it would naturally pierce the higher-dimensional area of a sphere that surrounds the D*p*-brane in 10-dimensional space-time. When these fluxes of various values of *p* are turned on, the result is a picture of string compactification that has various fluxes of background fields (analogous to different magnetic fields of Maxwell theory) threading the extra dimension.



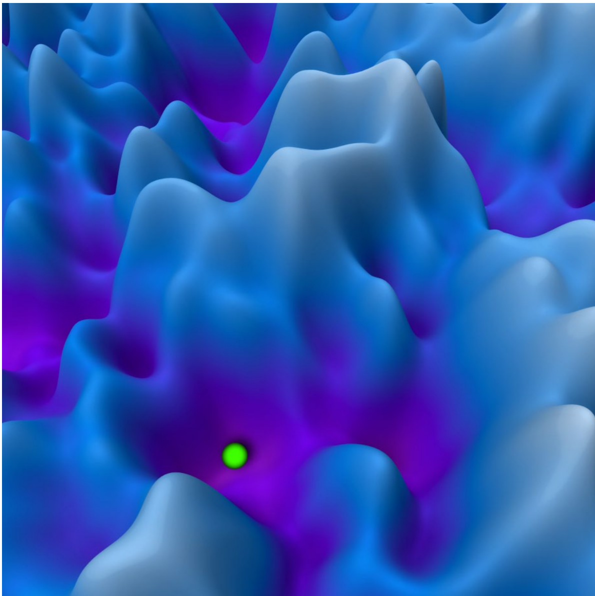
You might remember from high school that electric fields are sourced by charged particles. Charged particles are not needed in the extra dimensions to source the fluxes. A beautiful fact about topology is that in spaces with nontrivial-enough features, you can imagine field lines that wind around and produce flux with no charged source particle. On compact dimensions with nontrivial topology, you can have fluxes carrying energy density without any need for explicit charged source objects. The presence of these fluxes has a few interesting consequences:

- The fluxes typically break supersymmetry. That solves one of the problems.
- The fluxes typically cost an energy. This energy is minimized for some particular shape of the extra dimensions. It fixes their shape, solving another problem.
- The minimum of the energy determined by the fluxes is usually nonzero. There is a vacuum energy density from the perspective of a lower-dimensional observer.

These are all interesting, and somewhat generic, consequences of this more modern picture of string theory compactification. But there is one more feature that comes along for the ride.

There are many, many choices of possible background fluxes. In fact, for simple choices of the space on which one compactifies, the number of choices can give astronomical numbers like 10^{500}.

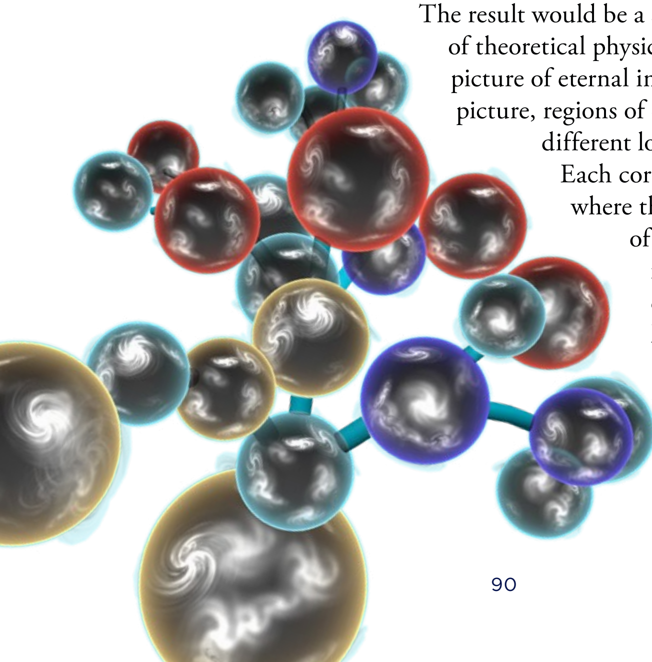
The resulting picture is one not of a single-string vacuum with broken supersymmetry and a vacuum energy, but of an "energy landscape" with many, many possible solutions. Perhaps early-universe inflation can even result as the shape of the extra dimensions "relax" to one of these minima.



To get to the modern picture of the string theory landscape, two more qualitative ideas are important.

- When the flux configuration and other effects are such that the vacuum energy in a given solution is positive, the resulting vacuum enjoys accelerated expansion. In other words, it is a de Sitter space with a positive cosmological term.
- In a quantum-mechanical theory, the distinct integer flux choices—even though they're discrete—are not disconnected from one another. If nothing else, quantum tunneling processes allow you to "jump" the value of the flux in a quantum tunneling event.

This picture of a string landscape with many vacua suggests a dramatic picture of the large-scale structure of the universe. Different regions in the universe may be trapped in different minima of the potential—characterized by different choices of the fluxes. Each such pocket universe may manifest what look like different low-energy "laws of physics," determined by the details of the vacuum solution governing that region. But on a large scale, quantum tunneling processes—which allow the amount of flux to jump on the internal dimensions—could connect regions with different amounts of flux.



The result would be a string theory realization of theoretical physicist Andrei Linde's picture of eternal inflation. In a simple picture, regions of different color manifest different low-energy physical laws. Each corresponds to a region where the extra dimensions of space have different fluxes, and maybe even different topology. Each has its own value of the vacuum energy. All regions with positive values are exponentially expanding with time.

It is hard to see how scientists would verify for certain the existence of other "bubbles" beyond our own in such a larger multiverse. It is also hard to know how researchers would find the particular configuration of the extra dimensions (and the fluxes) that is supposed to correspond to our own world. The qualitative picture seems robust, but making its details useful and predictive is an extremely challenging problem for the future. Could such a picture explain or accommodate the tiny value of the dark energy seen today? Only time will tell.

Reading

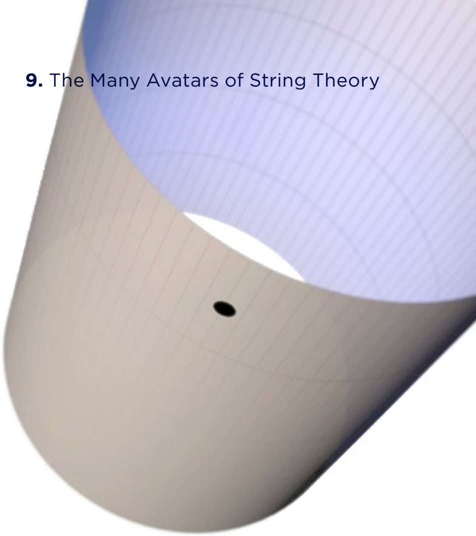
Susskind, Leonard. *The Cosmic Landscape.* New York: Little, Brown and Company, 2006.

9

The Many Avatars of String Theory

This lecture addresses the simplest features of string theory. You've learned from the start that string theories live, in some natural sense, in 10 dimensions. But where does that statement come from? And if there is a landscape of string theories in lower dimensions, then how many different string theories are possible in 10 dimensions? To get at these questions, you must consider the fundamental string itself.

The Fundamental String



The fundamental string itself a simple one-plus-one-dimensional object—a circular loop propagating through time. The simplest question you can ask is this: What would a very thin, one-dimensional ant crawling around on this cylinder think its world's fundamental particles and interactions were? If it were unaware of the larger bulk space-time in which the string is immersed, what would its one-plus-one-dimensional world look like?

To answer this question, you need to answer what the spectrum of particles and interactions is in the one-plus-one-dimensional quantum theory on the string worldsheet. Let's return also to the world of supersymmetric spherical cows and assume the string theory has supersymmetry reflecting bosons (on its worldsheet) into fermions. It turns out that with these assumptions, there are just a few possible choices.

First, for a string immersed in D space-time dimensions, there will be $D - 2$ transverse spatial fluctuations of the string. (After all, there are $D - 1$ spatial dimensions in total, and one of them is along the string itself.) That means that there are $D - 2$ scalar fields that live on the string worldsheet and parameterize its transverse fluctuations. Then, supersymmetry says there should also be $D - 2$ fermions—one fermionic partner for each scalar. That means that the theory being discussed will consist of $D - 2$ scalars and $D - 2$ fermions.

What values of D are allowed? Let's discuss this by doing some physics. The ultimate goal will be to take the theory of $D - 2$ scalars and fermions on the string worldsheet, let them fluctuate and quantize their fluctuations, and then read off a spectrum of elementary particles in the hypothetical D-dimensional world.

A subtlety arises in doing this. In quantum mechanics, the vacuum—the state of the quantum system when no particles are present—can itself be rather complicated. One way to think about this is that, due to the Heisenberg uncertainty principle, there is an energy-time uncertainty relation that says:

$$\Delta E \, \Delta t \geq \frac{1}{2} \, \hbar,$$

where h is Planck's constant and $\hbar$ is Planck's constant divided by 2π.

Therefore, at short times, a quantum system has quantum jitters—there's energetic motion that comes from virtual vacuum fluctuations. These vacuum fluctuations have one important effect: They shift the vacuum energy of the system by something known as the zero-point energy. Therefore, the energies of the lightest string states depend directly on D.

As you learned previously, quantizing the lightest string modes gives rise to a particle with the quantum numbers of a graviton (here, in D space-time dimensions). But you need to be careful about this statement. For $D = 10$, including the shift in vacuum energy from quantum jitters, the resulting gravitons are exactly massless. They come with the right multiplicity to be the graviton of 10-dimensional Einstein gravity. For other values of D, though, there's trouble. There are too many or too few potential graviton components to align with Einstein gravity in D dimensions unless $D = 10$. This is a sign of an inconsistency in the string propagation in flat space for other values of D. Therefore, you find that $D = 10$; that is, there must be 10 space-time dimensions in a supersymmetric string theory.

This logic is a bit too quick; it is both too refined and too coarse, in different ways. It is too coarse because it misses an important subtlety that is required in formulating the one-plus-one-dimensional quantum theory on a circle: Fermions have nonzero spin. So, to put them on the circle, you have to choose something called a spin structure.

This amounts to a choice of whether the fermions are periodic or antiperiodic as you go around the circle. There are two different consistent options for how you assign signs to the full set of fermions in the theory (or how you do the "sum over spin structures," in technical language). The two resulting theories are called the IIA and IIB theories.

Now, a brief digression: In two dimensions, you can imagine a massless particle on the string worldsheet that can only move "to the left." If the particle were a massive particle, it would be moving at a speed less than the speed of light. An observer moving even faster than that to the left would effectively see this particle relative to them as moving to the right. It would therefore have to be a particle with both "left- and right-moving" versions, or flavors.

But a massless particle moves at the speed of light. So, given a massless left-moving particle, you can never switch to a reference frame of an observer moving so fast that the particle appears to be moving to the right.

With this in mind, let's return to the logic that was used in finding the supersymmetric type II string theories. The logic was too refined. Supersymmetry was required in the theory on the worldsheet. But in a world where particles can be purely chiral (i.e., purely left or right moving), you can also have supersymmetries that are purely chiral. In other words, the mirror reflection of supersymmetry may take every right-moving boson to a right-moving fermion partner but could do nothing to the left-moving boson modes.

Allowing such supersymmetric but chiral theories, and once again making sure the sum over spin structures is consistent, yields a surprise: There are two more consistent string theories. Each has eight right-moving bosons and fermions, just as in the type II theories (eight for the eight transverse fluctuations of a string in 10 space-time dimensions).

But consistency of the sum over spin structures now requires that they have 16 left-moving fermions. Their worldsheet theories are therefore like those of the type II theory but with extra chiral fermions added on. These theories are called the heterotic string theories. You can think of the heterotic string as a crossbreed between the supersymmetric type II theories and another theory, the unstable so-called bosonic string. The right-moving particles look like those of a type II string, while the left movers look like those of the (unstable) bosonic string.

Heterosis is defined by Oxford Languages to be "the tendency of a crossbred individual to show qualities superior to those of both parents."

This may seem rather confusing and difficult to remember, but it becomes easier when you discover—as string theorists did in the 1980s—that the peculiar zoo of 10-dimensional string theories is in precise correspondence with another classification of 10-dimensional objects.

Classifying 10-Dimensional Objects

Physicists trying to classify possible supersymmetric theories of the fundamental interactions asked in the 1970s and 1980s, "What are all of the possible consistent structures containing supersymmetry and Einstein's relativity in flat space-time?" While it was difficult to answer this question in four space-time dimensions (and the answer remains unknown to this day), they had success in higher dimensions. The answer became simpler and simpler as they moved up in dimension. The most basic results of their classification were the following.

The highest dimension in which a supergravity theory can exist is 11. A unique supersymmetric theory exists there.

In 10 dimensions, there is a slightly richer spectrum of possibilities. In fact, there are four candidate theories. Two of them were called type IIA and type IIB supergravities, and they differ in the details of their fermion content.

The other two have fewer supersymmetries (or symmetries mapping bosons to fermions), but they compensated by enjoying a space-time gauge symmetry or a set of space-time force carriers in 10 dimensions. They have, respectively, symmetries given by the groups $SO(32)$ and $E8 \times E8$.

$SO(3)$ is a fancy name for the group of rotations of three-dimensional space, and $SO(2)$ or $U1$ is the rotations of the plane. $SO(32)$ can similarly be thought of as rotations in an abstract 32-dimensional space, which is not the actual physical space-time. $E8$ is an even more complicated "exceptional group."

So, there are four consistent string theories built out of supersymmetric (worldsheet) strings that live in 10 space-time dimensions. There are also four consistent supersymmetric theories of gravity (or supergravity theories, for short) in 10 space-time dimensions. It's not a coincidence that there are four of both.

Remember that in string theory, you can figure out the spectrum of elementary particles that emerge at low energies by looking at the least energetic fluctuations of the fundamental string. The string modes that carry the least energy—the string ground states—correspond to massless particles in space-time. And the four supersymmetric string theories turn out—certainly not by intentional design—to give rise to precisely the sets of particles that characterize the four 10-dimensional supergravity theories. The first two theories that were described give rise to the type IIA and type IIB supergravities, while the chiral theories mentioned afterward give rise to the supergravity theories with half as much supersymmetry but that enjoy space-time gauge symmetry.

The emergence of the most basic, fundamental supergravity theories as a limit of string theory in the 1980s was one of the early hints that string theory enjoys deep ties to other areas of theoretical physics.

It is important to emphasize, though, that this does not mean that string theory is the same as the supergravity theories mentioned. The supergravity theories only "know about" the lightest string modes—just as in Einstein's general relativity there's a massless graviton but no massive particles. And just like Einstein's gravity theory, the supergravity theories are not renormalizable, or their high-energy scattering gives nonsensical results.

All of the more excited modes of the string—with ever-wilder oscillations—describe particles whose mass is multiples of the string or Planck scale, and these particles are invisible to the low-energy supergravity theory. But they are crucial in allowing string theory to have the sensible behavior that it does at very high energies. Their presence (with the exact degeneracies and interactions they enjoy) is the string theory way of curing the non-renormalizability of supergravity.

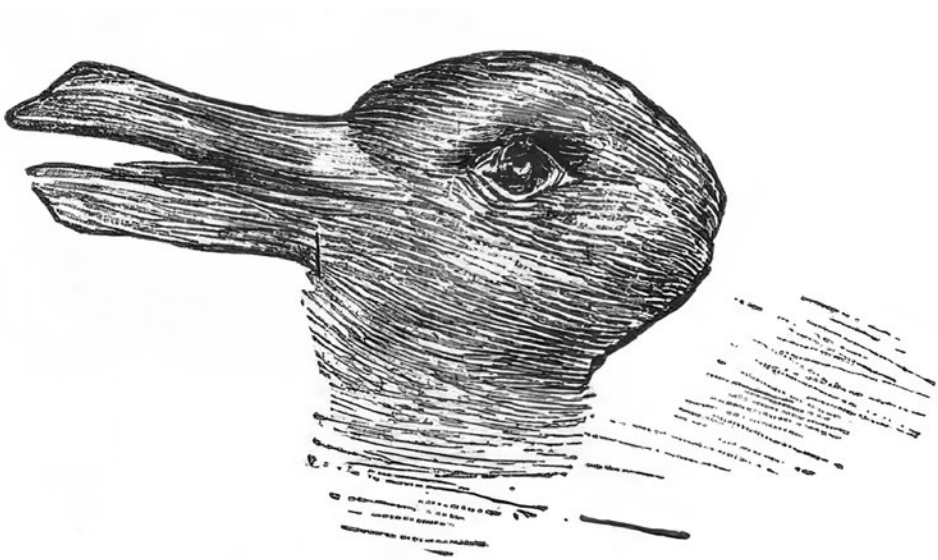
This leaves you with a few questions. First, you sort of want a single theory of unified interactions. It seems like, instead, you have four different structures in 10 dimensions. Is this four different possible sets of laws of physics and you have to choose one of them right at the start? Or is something more subtle going on?

Secondly, there is a supergravity theory in 11 space-time dimensions, but the corresponding string theory wasn't found. Does this mean that string theory—or whatever theoretical structure it is part of—doesn't even know about the 11-dimensional supergravity? Is there something wrong with the 11-dimensional supergravity theory?

The answer to both of these sets of questions ties to the notion of duality.

Duality

Two structures that look quite different may nevertheless be the same. The two different ways of looking at one object can result in rather different views, as in the duck-rabbit shown—where, depending on whether you view the two protuberances as the beak of a duck or the ears of a rabbit, you can choose to see either a duck or a rabbit. Could the objects that you're seeing be related in a similar way?



Let's first use this viewpoint to address the role of the 11-dimensional supergravity in string theory. As you learned in previous lectures, in addition to strings, string theories contain D-branes—objects on which open strings can end. Each type of D-brane comes with the spatial dimension *p* of its worldvolume (hence, D*p*-brane).

The type IIA string theory, in particular, has D-branes that enjoy the *p* value of 0, 2, 4, 6, and 8, while its cousin, type IIB theory, has the branes of *p* equal to 1, 3, 5, 7, and 9. If you consider the type IIA theory in 10-dimensional flat space, the only D-brane that can have finite extent—and therefore finite total mass—in its simplest flat configuration is the D0-brane, or D-particle. It has a mass given simply by the inverse string coupling: the mass of the D0 goes like 1 over the string coupling constant:

$$m(D0) \sim \frac{1}{g_s}.$$

It turns out that there are actually bound states of multiple D0-branes, just as there is a bound state of the electron and the proton in nature. The multiple D0-brane bound states have some remarkable properties. The first is there is precisely one bound state for each possible number *N* of D0-branes. The second is it is a bound state at threshold. This means that the binding energy—the energy that holds the objects together, as compared to the energy of *N* widely separated D0-branes—vanishes. This thing just barely exists as a bound state. It follows that for all *N*, there is a bound state whose mass is equivalent to that of *N* D0-branes, namely given by *N* over *g*-string in the appropriate units:

$$m\,(N\text{ D0 bound state}) \sim \frac{N}{g_s}.$$

These D0-brane bound states are also BPS states; the mass formula just stated is exact. That's a peculiar regularity for a set of bound states. What could it mean?

Now, a brief distraction: Consider 11-dimensional supergravity, but instead of studying it in 11-dimensional flat space, consider it instead on a space of R^9 flat dimensions times a circle times time:

$$R^9 \times S^1 \times \text{time}.$$

In 11 dimensions, there is a massless graviton (and its supermultiple—the full set of states that follow from the graviton by acting with the supersymmetries of the theory). If you add a circle of radius *R*, you can consider versions of the graviton that carry momentum around the circle.

Single-valuedness of the quantum wave function of the graviton restricts the momentum (and therefore the mass of the resulting object, viewed by a 10-dimensional observer) to be quantized in units of inverse radius of the circle, $1/R$. So, the momentum around the 10th spatial direction of the circle can be n/R, where *n* is 0, 1, 2, 3, or 4 and the mass of the resulting *n*th state is n/R:

$$\frac{1}{R} : p^{10} \; = \frac{n}{R},\; n = 0,\; 1,\; 2,\; 3,\; 4\; \ldots \rightarrow m(n) \sim \frac{n}{R}.$$

These graviton momentum modes on the circle are BPS; they preserve some supersymmetry, and the mass formula is exact.

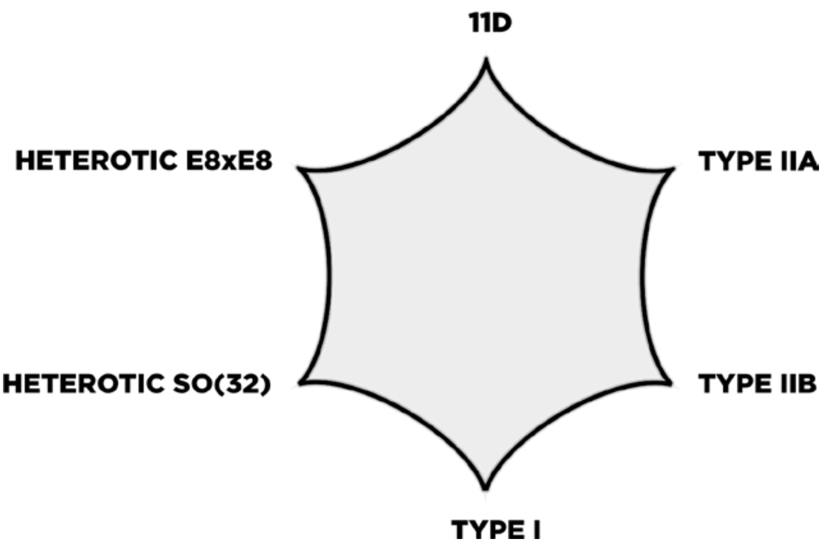
Now compare the formula for D0 bound-state masses to the formula for graviton momentum mode masses. If you identify the number of D0-branes with the number of units of momentum of the graviton and the radius of the 11-dimensional circle with the string coupling constant into *g*-string theory, for each D0-brane bound state of IIA string theory, there is a correspondingly unique graviton momentum mode of 11-dimensional supergravity with exactly the same mass (as seen by a 10-dimensional observer):

$$n = N,\; R \sim g_s.$$

This leads to a brave hypothesis: Maybe the IIA string theory at finite coupling is equivalent to an 11-dimensional theory on a circle whose radius is given by the coupling.

Another way to put this is that 11-dimensional supergravity (and whatever completes it as a theory that does not have non-renormalizable high-energy scattering) should be thought of as arising from the strong coupling limit of type IIA string theory.

This is the simplest example in a "duality web" that connects all of the 10-dimensional string theories—and 11-dimensional supergravity—to one another. The full picture for these simple theories looks something like the unifying web you see here. All the spokes of the blob correspond to the theories described in this lecture, together with another one called type I (which wasn't described). The blob in the middle is a space of values of scalar fields (parametrizing coupling constants or radii in the various duality frames of the different theories). For instance, the type I theory that wasn't described was omitted because it is simply the strongly coupled version of one of the heterotic theories that was described in 10 dimensions (the one that has $SO(32)$ symmetry).



These dualities that relate these different spokes of the web are cousins of the T-duality that you encountered previously in the course, which relates string theory on a circle of radius R to string theory on a circle of radius $1/R$. In fact, a more careful version of that statement is that T-duality relates type IIA string theory on a circle of radius R to type IIB theory on a circle of radius $1/R$. So, you have secretly known about some of the connections in the diagram for some time.

Reading

Witten, Edward. "Duality, Spacetime and Quantum Mechanics." *Physics Today* 50, no. 5 (1997): 28–33.

10

Duality: Which String Theory Is Fundamental?

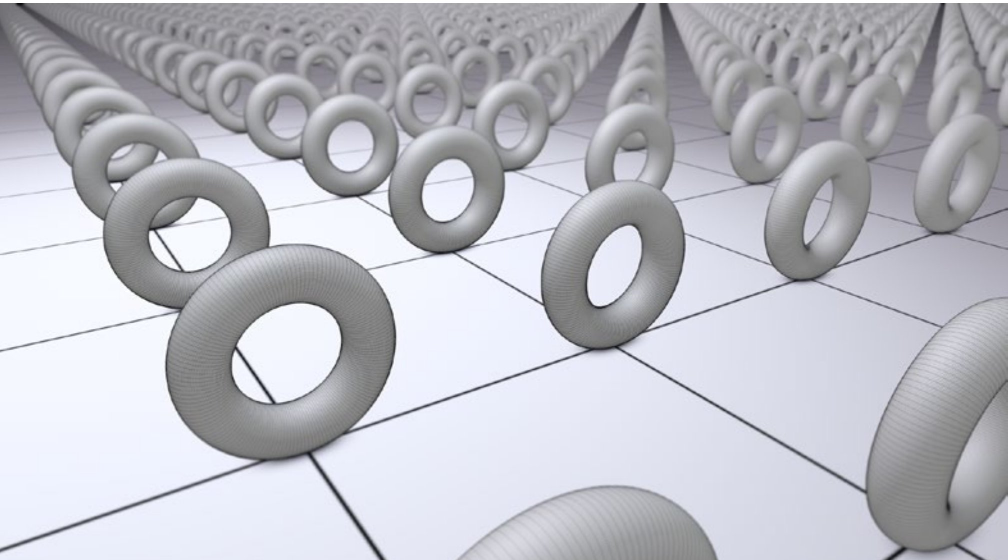
In the previous lecture, you discovered that there are a handful of naively different supersymmetric theories of quantum gravity in 10 and 11 space-time dimensions. This might seem to mean that you have a choice to make about which theory to explore, the true situation is more subtle and wonderful. The distinct theories fit together in a web in such a way that if you choose to explore one, you will inevitably be led to discovering all the others. They fit together in a "web" of theories, related by varying things like the string coupling constant or the sizes of compact dimensions. The example of this in this lecture will give you a flavor of how there can be deep interconnections between facts about pure mathematics on the one hand, string theory on the other, and conventional physics on a third. And these interconnections allow for the dualities that tie the string theory together in a web of theories.

Relating Theories

In the previous lecture, you learned how to relate the IIA and IIB theories—they are the same after you reduce on a circle by a T-duality. And you discovered how the 11-dimensional theory and the IIA theory are similarly related by circle compactification. But how are you to relate the other structures—the heterotic theories, with interesting gauge forces parametrized by groups like $SO(32)$ or $E8 \times E8$—to theories like the type II theories?

The heterotic theories carry half as much supersymmetry as their type II analogues. Theories with different amounts of supersymmetry can't possibly (even secretly, or in a subtle way) be the same. This is where the notion of compactification once again comes to the rescue. In making any kind of realistic model of string theory, you will have to compactify some (most likely six) of the space-time dimensions. So it is very natural to consider compactifications where some of the spatial dimensions are curled up on a compact shape, usually called a manifold by mathematicians.

While the simplest example of a compact dimension is just a circle—and it has higher-dimensional simple analogues, such as the two-dimensional torus, which is like a product of two circles—these won't help with the issue at hand. As it turns out, compactification of a supersymmetric theory on a circle or a product of circles like a torus can have many important physical



implications, but it does not change the number of supersymmetries—or symmetries relating bosons to fermions—that the theory exhibits. So, studying compactifications only on circles or on tori won't allow you to relate theories of different amounts of supersymmetry like the type II and heterotic theories.

So, then, is there an overall choice you have to make at the start of your study of string theory? Happily, the answer is no. The key is that you can compactify the extra dimensions in such a way as to break some of the supersymmetry, so the low-energy physics—beneath the energy scale where you can probe the compact dimensions—appears to have less supersymmetry.

The vacuum solutions—or solutions with no sources of energy density present—that allow this are those where string theory is compactified on Calabi-Yau manifolds. Such manifolds can be studied using the theory of complex variables, and since each complex variable has two real numbers hidden in it (its real and imaginary parts), they have an even number of dimensions. So, they can exist in 2, 4, 6, or higher dimensions. The case of most interest for real-world phenomena would be 6, where they could be used to compactify string theory from 10 to the 4 space-time dimensions that are observed. But as theoretical tools to allow the exploration of the physics of string theory, such cases can be useful in other dimensions as well.

Classifying the Calabi-Yau Manifolds

As was discussed previously, a useful beginning for any science is taxonomy. And classifying all of the Calabi-Yau manifolds will help you get a handle on all supersymmetric vacuum solutions of string theory. While that is a far cry from understanding the real world, it would constitute a first step—a solution of an analogue spherical-cow problem.

The first step isn't difficult. Two-dimensional compact spaces are classified by their number of handles, or genus. As vacuum solutions of string theory, Calabi-Yau manifolds have to admit a flat metric. You can show without much difficulty that the sphere doesn't (after all, it is round), and similarly that surfaces with more than one hole, or genus greater than 1, also don't. This leaves the case of genus 1, the two-dimensional torus. That is the simplest example of a Calabi-Yau manifold. However, a two-dimensional torus can be viewed as a product of two circles, so by the previous lecture, it's not going to break any supersymmetries when it is used to compactify a supersymmetric theory. You can't use it to find a relationship between heterotic and type II strings.

So, how about in dimension four, the next case? First, you can dispense with the simplest example, the four-dimensional torus. It is again Calabi-Yau, but not very interesting, being a product of circles. However, now there is an interesting example: the first example you've encountered of a nontrivial Calabi-Yau manifold. It is called the K3 surface—named after the mathematicians Ernst Kummer, Erich Kähler, and Kunihiko Kodaira, who did some of the most significant research uncovering its existence and properties. It's difficult to picture this space, but some simple things can be said about it.

First, it admits a flat metric (so it's a vacuum solution of the Einstein equations), but it is highly nontrivial. In particular, when it's used to compactify a theory to lower dimensions, it breaks some of the supersymmetry of the higher-dimensional theory.

Next, the flat metric on a K3 surface isn't unique. Just as a circle comes with a parameter (just the size of a radius) and a torus comes with two parameters (its shape and size), the K3 surface comes with parameters, which mathematicians call moduli, parametrizing its Calabi-Yau metric. In this case, the number of parameters is not 1 or 2—it's 20.

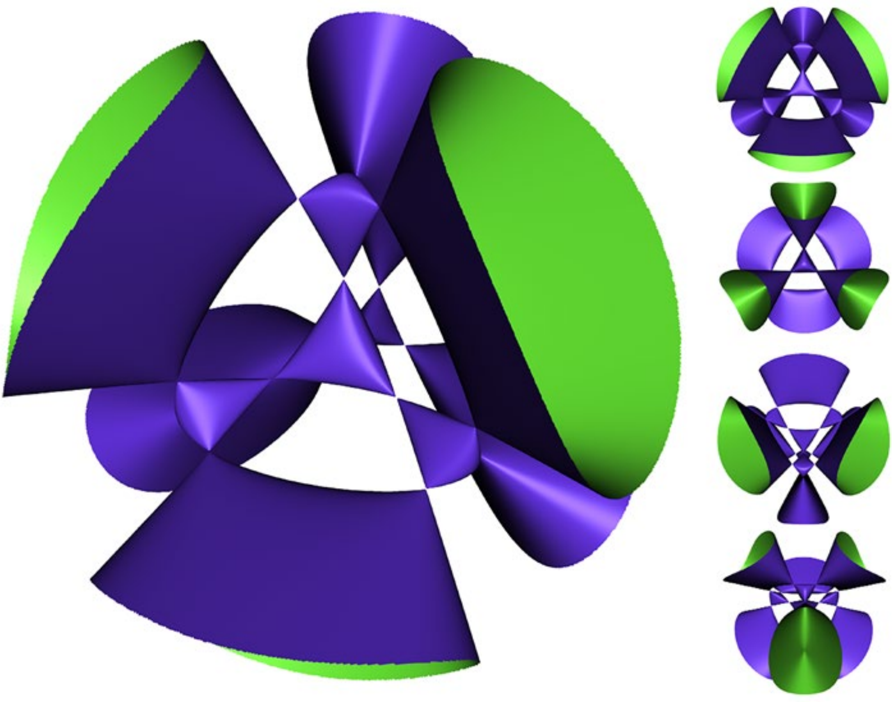
Third, while it is hard to write down an explicit example of a K3 surface in a way that makes the flat metric manifest obvious, you can easily write down spaces with the right topology, and then the theorem Calabi-Yau guarantees that the flat metric has to exist, but it isn't obvious.

As one simple construction, for instance, you can take four complex variables—z_1, z_2, z_3, z_4—and write a simple quartic equation that has terms like z_1^4 and z_2^4 and sets the sum of all these terms to zero:

$$z_1^4 + z_2^4 + z_3^4 + z_4^4 = 0.$$

In this construction, it becomes easy to understand what the 20 parameters giving K3 surfaces of slightly different shape are. Most of them simply appear by taking that quartic equation, which is a set of degree-four terms, and deforming it by other terms like to z_1^4 and z_2^4—you can add $z_1^2 \times z_2^2$. The different ways of deforming the equation give you the parameters in the Yau metric.

This is oversimplifying a bit. That is one equation on four complex variables, which would leave a six-dimensional manifold. There's also an overall scaling that identifies all points that differ by complex rescaling of the z variables. If you identify points on space as equivalent if they can be related by such a rescaling, you end up with a four-dimensional K3 surface. A picture of K3, obtained by cutting out slices of an equation like the previous one so that you can project it in two dimensions, looks like what you see here.



Now you are in business. If you compactify, say, the type IIA theory on a K3 surface, you will find a theory in the remaining six non-compact dimensions with the same number of supersymmetries as the heterotic theories. Of course, the heterotic theories live in 10 space-time dimensions. But you can get simple six-dimensional relatives by reducing them to six dimensions on a four-dimensional torus, preserving their supersymmetry.

This gives you two different natural families of supersymmetric theories in six dimensions: those you obtain from compactifying the heterotic strings on a torus or the type II strings on a Calabi-Yau manifold K3. A bit of counting shows that both come in 20-dimensional families. So, they have the same amount of supersymmetry and come with the same numbers of parameters, or moduli, that you can vary (which show up as scalar fields in the six-dimensional physics). Are they the same, giving you a duality relating the type IIA theory to the heterotic theories?

Unfortunately, there is a problem with this idea. In the classification of 10-dimensional supergravities, the heterotic theories appear as theories with big groups of symmetries, generalizing those of the standard model. Remember that such interactions are parametrized by a choice of group, and the relevant groups here are the groups $SO(32)$ and $E8 \times E8$. These are groups that are much larger than that of the standard model, but they happen to contain it as a smaller substructure. So, in particular, their presence means that the low-energy physics of heterotic theory exhibits many interesting forces generalizing the strong, weak, and electromagnetic forces in nature.

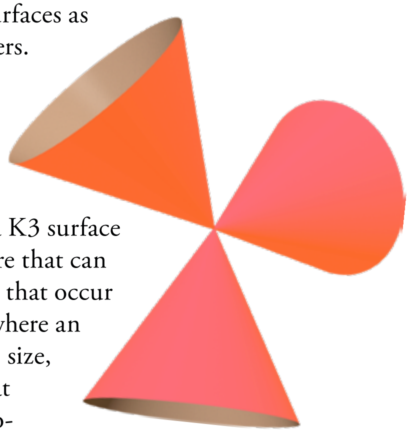
On the other hand, in the early days of string theory, it was proved that the type II theories cannot give rise to big symmetries like these. In fact, it was even shown that the standard model gauge group, or group of forces in four dimensions, can't arise as the spectrum of fluctuations of a type II string in a perturbative framework. This seemed to indicate—to physicists of that era—that of the bifurcation of choices between heterotic theories on the one hand and type II theories on the other, which seemingly lived in different theoretical worlds, the former should be chosen.

This also seems to provide an insuperable obstacle to the goal of unifying the heterotic and type II theories via duality. But before giving up on such an elegant idea, let's explore the space of K3 surfaces more thoroughly.

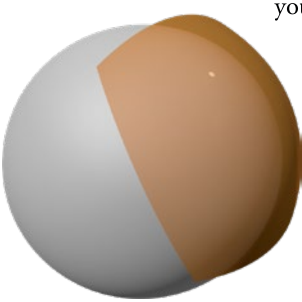
Exploring the Space of K3 Surfaces

One interesting thing that can happen as you deform a smooth geometry—say, taking a K3 surface given by the previous quartic equation and adding a quartic perturbation with a larger and larger coefficient—is the following.

Although by Yau's theorem the space admits a flat metric in the relevant sense for Einstein's equations (Ricci flatness), that still allows for local curvatures consistent with the Einstein equations. After all, space-time curvature is responsible for the perception of gravity in general relativity. And nothing guarantees that these local curvatures cannot grow large. (Examples where this happens include the big bang and black holes.) Exactly this phenomenon happens in K3 surfaces as you move around in their space of parameters. At various regions in the space of moduli, pinched regions on the K3 surface develop that look like conical singularities of the sort that are seen here.



It is hard to represent the full geometry of a K3 surface (which is four-dimensional) in such a picture that can be projected. The regions of high curvature that occur in K3 aren't quite places like in the figure where an encircling rubber band could shrink to zero size, as would occur in what is seen. Rather, what happens in these regions is that a whole two-dimensional curve within the K3 shrinks.



Strings are only one-dimensional, like rubber bands, and you can't wrap a rubber band or one-dimensional object around a shrinking two-dimensional curve. But type IIA string theory also has the D-branes of various dimensions. In particular, it has a D2-brane: a brane with two-dimensional spatial extent. Such a two-dimensional object has exactly the right property to wrap around a shrinking curve. (Think of wrapping a sheet of tinfoil around a basketball, for example.)

Since the D2-brane has a finite tension (determined by a ratio of the string tension and the string coupling), the resulting wrapped object gives rise—in the lower-dimensional or six-dimensional view—to an elementary particle whose mass is determined by the volume of the wrapped curve that the D2-brane is wrapping on the K3 surface. Normally, if the curve had a finite size, this would be a very massive particle (because the D2-brane tension is very large at finite string coupling). But if this curve shrinks to zero size—as it does at a true singularity of K3—then the resulting particle will become massless, and it should appear in the low-energy physics as it will to a low-energy six-dimensional observer.

Now, an interesting set of questions: What are the possible singularities of K3 surfaces, and what kinds of elementary particles do the resulting wrapped branes give rise to in the low-energy theory that a physicist uses to describe what is happening on each singular K3 surface?

Mathematicians have classified the allowed singularities of K3 surfaces. And amazingly, they have a classification that precisely lines up with the groups of symmetries, or gauge groups—including $SO(32)$ and $E8 \times E8$—that can appear in the heterotic string theory when it's compactified on a four-dimensional torus. To make the match precise, remember that in a gauge theory with charged matter particles (like the standard model), the gauge group can be "Higgsed" so that the gauge bosons become massive.

In the match between heterotic and type II theories, the singular K3 surfaces correspond to points in the heterotic theory where the corresponding symmetry is not Higgsed. And generic smooth K3 surfaces correspond to heterotic theories where all the interesting gauge symmetry (apart from a handful of photons) has been broken by the Higgs mechanism.

More on Relationships

What you've discovered so far in this lecture is that naively, there is a disconnect of string theories into two worlds: the world of heterotic strings and that of type II strings. The former has less supersymmetry but more gauge symmetry.

Just accounting for particles coming from perturbative fluctuating strings—the things made out of oscillating, winding, moving strings—you can't possibly find type II theories that match the properties of the heterotic theories.

But the type II theories also contain D-branes. Accounting for these—and, in particular, for the physics they can give rise to in singular geometries—you can, in fact, find a duality that relates type II and heterotic theories.

The two are different faces of the same underlying theoretical structure. If you consider a theory arising from one, you necessarily have the other as part of your framework. That shows that intricate and deep facts about the geometry of manifolds (aspects of what mathematicians call differential and algebraic geometry) can play an important part in understanding string theory. In fact, this relationship can also be flipped: In richer physical circumstances that enjoy less supersymmetry, you can use string theory to make striking predictions for differential and algebraic geometry.

You might wonder if you can demonstrate more directly the relationship between the type IIA string on K3 and the heterotic string on a torus. As was described, the two strings themselves have very different physics on the worldsheet: The type IIA string has 8 left- and right-moving bosons as well as their superpartners, whereas the heterotic string has a chiral structure. It has the same right-moving degrees of freedom as the type IIA string, but its left movers involve 8 bosons and a large number of (or, more precisely, 16 complex) fermions.

If the type IIA string theory is equivalent to the heterotic string theory, then since the heterotic string exists as an excitation on one side of the duality (namely, in the heterotic theory itself), there has to be some object that looks like a heterotic string also on the type IIA side. And in fact, it can be precisely identified.

The type IIA theory has p-branes for various values of p, including $p = 5$. If you consider wrapping its $p = 5$ (or five-plus-one-dimensional) brane over the entire K3 surface, then a five-dimensional object wrapping a four-dimensional space leaves a one-dimensional object or a string sticking out in the remaining six dimensions of space-time. And you can prove—by using detailed facts about the geometry of K3—that the degrees of freedom living on the remaining string precisely match those of the dual heterotic string.

In other words, the heterotic string moving around on a torus—which is the basis for a fundamental theory in its own right—can also be viewed as arising from studying the dynamics of a complicated excitation in the type IIA theory. It's not related to the fundamental string there. And in fact, the reverse is true: The type II string arises as a wrapped 5-brane in the heterotic string theory.

This raises an interesting question: If both the heterotic and type II strings arise as complicated excitations in the other, which one is fundamental? The lesson for physics—learned also in many other contexts—is that "fundamental" is in the eye of the beholder. In some regimes of the parameter space (for instance, when the K3 surface is very large and type IIA string coupling is weak), it may be easier to understand one description than the other (in that case, the type II description). But neither is more fundamental, and for other choices of parameters, the other description will be easier to understand.

Reading

Kachru, Shamit. "Learning to Count in String Theory." Simons Foundation Presidential Lecture, February 27, 2019. https://www.simonsfoundation.org/event/learning-to-count-in-string-theory/.

11

Finding Evidence for String Theory

This lecture discusses a highly speculative topic: How might string theory make contact with experiment? That is, how might evidence be discovered that string theory—instead of just being a consistent theoretical framework—actually describes the universe? String theory is a sophisticated theoretical framework. As it's been described in this course, it can capture physics in a variety of dimensions. It encompasses worlds with and without supersymmetry. And it seems to exhibit a bewildering landscape of different vacua. Given all this variety, how can it be decided which aspects of the theory are likely to yield clues in experiment? Instead of searching for smoking-gun evidence of the fundamental strings themselves, this lecture discusses more generic signatures: things that string theory suggests that could show up in the next generation of experiments.

Prediction I: There Are Extra Dimensions of Space-Time

The first generic prediction of string theory is that there are extra dimensions of space-time. This is, after all, maybe the first and most obvious, direct tension between string theory and the standard model—string theories naturally live in higher-dimensions, while the physics seen in experiment is resolutely in three-plus-one dimensions. However, could it be that there are extra dimensions to be found just around the corner in experiment and they might be discovered with modern-day experiments? To explore this possibility, let's examine two rather different cases.

In the first, the standard model itself is secretly a higher-dimensional theory. In other words, if you were able to view nature with a microscope and see a fifth dimension, each of the standard-model particles would actually be a particle that was free to move in all four spatial dimensions—that is, they could also move around the "tiny circle" hidden above each point of the visible space. In this scenario, the standard model itself is intrinsically a five- or six- or whatever-dimensional theory. Each of its particles—for instance, the electron—has a possible velocity in the fifth dimension as well as the (observable) velocity in the visible spatial dimensions.

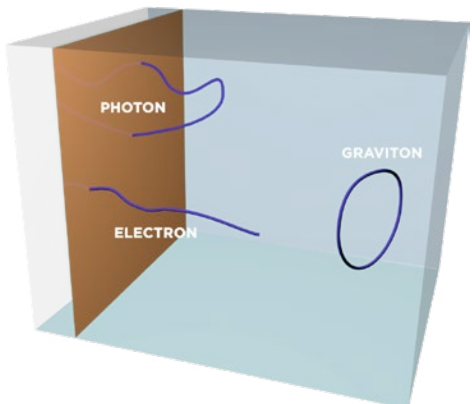
There is no evidence of such a fifth dimension. The fifth dimension would result in a tower of particles called a Kaluza-Klein tower that corresponds to each of the particles seen in nature. This is because, for instance, the electron could have a momentum of $1/R$, $2/R$, $3/R$, … , "around" the hypothetical fifth dimension of radius R. Such cousins of the electron are not seen; they would show up in the world as additionally charged particles. You might think that perhaps you can interpret the muon or the tau particle that way, but that interpretation doesn't work because the additional modes in the tower that the interpretation would require are not seen.

That doesn't mean that there can't be a fifth dimension like this. It just means that there must be a bound on its radius R. The fifth dimension has to be small enough—and the Kaluza-Klein tower of electrons heavy enough—that they aren't produced in modern experiments. In practice, a combination of the big collider experiments that can produce new charged particles put stringent bounds on such an extra dimension. At the very least, new cousins of the electron must have a mass above the TeV scale probed by experiments at Fermilab and CERN.

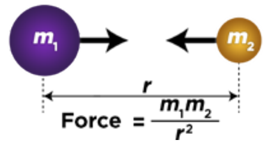
In terms of a size of the extra dimensions, this translates into a rough bound that any new dimension in which the standard model also lives has to be less than about 10^{-17} centimeters in size. That's pretty small. There is no especially good motivation to think that extra dimensions do exist around (or just below) that scale, but it certainly is a question of interest for present-day experiments, and in fact, modern collaborations put bounds on such dimensions.

There is another possibility for extra dimensions that is perhaps even more intriguing. String theory contains branes, which fill a subset of the spatial dimensions. These branes have particles—including both charged particles and force-carrying particles—that live purely on the brane worldvolume itself, instead of living in the "bulk" of the space-time. These localized particles come, for instance, from open strings that end on the brane.

So, consider a model of string theory where the standard model lives on D-branes in the extra dimensions. There could be transverse extra dimensions where none of the standard-model particles propagate. Since the string scale—which governs masses of oscillating open strings—can be as high as roughly the Planck mass, there would be no visible "fatter" cousins of the electron or other charged particles in this kind of scenario. So the methods that exist for finding extra dimensions in the first scenario don't apply here. There is no Kaluza-Klein tower of electrons with momentum in the fifth dimension.

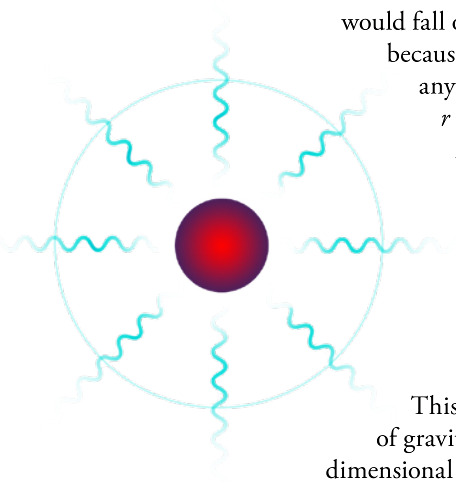


But not all is lost. One inevitable fact about string theory is that the graviton comes from the lightest closed-string mode. And unlike its open-string cousins, the closed string doesn't need to end on branes. So a closed string—including the graviton—can exist anywhere in space-time. If there are extra dimensions of size R, there need not be Kaluza-Klein electrons, but there will certainly be Kaluza-Klein cousins of the graviton. And their masses will go like the inverse radius $1/R$.



How would these show up in an experiment? Conceptually, the easiest way to think about this is in terms of the gravitational force itself. While Newton's law applies to well-separated masses in the familiar world, it could be modified in a higher-dimensional world.

In fact, a good way to think about the Newtonian force (or the electromagnetic force) is in terms of an argument due to Carl Friedrich Gauss. Suppose you draw lines of gravitational force emanating from the particle as shown below.



In a flat two-dimensional world, the density of such lines would fall off as $1/r$ a distance r away from the object—because all of the field lines have to go through any surrounding circle, and the circle of radius r has circumference that grows with them. A similar argument in a world with three spatial dimensions shows that the density of lines of force falls off as force feels like 1 over r squared:

$$F \sim \frac{1}{r^2}.$$

This is the origin of the inverse square laws of gravity and electromagnetism in the three-dimensional world.

But suppose that secretly space-time has five dimensions and there is a fourth spatial dimension. And say it has radius R. Then, there are two relevant regimes where you might probe the strength of gravity: at distances small or large compared to r. For particles separated by short distances less than r, the field lines don't even notice that the fifth dimension is compact with a finite radius. They will spread out as if they move in flat four-dimensional space. And as a result, the field lines and forces will fall as 1 over distance cubed for gravity at such short distances:

$$F \sim \frac{1}{r^3}.$$

Meanwhile, at longer distances—large compared to the size of the extra dimensions—the gravitational field lines will effectively ignore the compact dimension. They are too small to matter, relative to the large distance between the objects attracting each other via gravity. But there will be Kaluza-Klein gravitons—gravitons with momentum in the extra dimension. These can also be exchanged between massive objects, and they will contribute small corrections to the force that's normally contributed by the graviton. And in fact, the result will then be that the field line density in forces fall off as in an inverse square law with small exponential corrections that decay with distance:

$$F \sim \frac{1}{r^2}\left(1 + a \times \exp\left(-\frac{r}{R}\right) + \ \ldots\right).$$

It shouldn't be surprising that the gravitational force between separated objects is something that can be probed using precision experiments (though actually isolating the gravitational force is challenging because electromagnetic forces are so much stronger). You have to be exquisitely careful to electromagnetically shield your experiment because even a tiny induced charge can lead to electric forces that overwhelm any gravitational contribution.

Regardless, a combination of experiments by today's careful experimentalists constrain any extra dimensions even where only gravity propagates. In units of the micron, which is about 1/1,000 of a millimeter, the largest extra dimensions are thought to be bounded by about a few microns in size. While this is a significant constraint, it is impressive how much weaker the constraints become once you confine the standard-model particles to a brane.

Prediction II: There Are Strings

A second generic prediction of string theory—maybe a more obvious one—is that there are strings. Despite what you might think, this is not a 100% universal prediction of string theory. For instance, in the limit of the structure known as string theory that is described by 11-dimensional supergravity, there are no evident stable string states. Nevertheless, it is fair to say that strings are present in a great many of the realistic constructions coming out of string theory today. Could directly observing strings be a hope?

The main difficulty here is that in most proposed realistic versions of string theory, the string scale—or the size of strings—is somewhere around the Planck scale (which is 10^{-32} centimeters), or maybe larger by a factor of 10 or 100. That is far too small to hope to see in direct experiment in any foreseeable future.

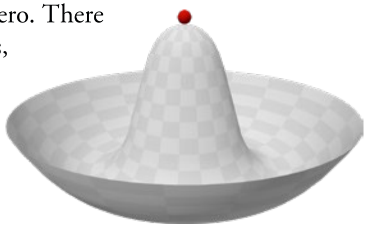
Bounds on charged cousins of the standard-model particle place a bound on how low you could push the string scale, for the same reasons described in discussing Kaluza-Klein modes of charged particles like the electron. So, even in a less motivated scenario where the string scale is as low as possible, it would be bounded by the TeV scale. Certainly, excited string modes have not been seen at the Large Hadron Collider, at least yet.

But there is another avenue for hope. Remember, in learning about early-universe cosmology, you discovered the fantastic physics of inflation. This is a period in the history of the early universe when the energy density was dominated by a potential energy for an inflaton field whose potential had a peculiar flat shape. The universe inflated as the inflaton rolled down its flat potential, and eventually (as it reached its minimum and relaxed to its ground state), the inflaton dumped all its energy into the standard-model particles. This created the hot big bang. It is interesting to ask whether models of inflation can be accommodated in string theory, and if so, what is the inflaton field?

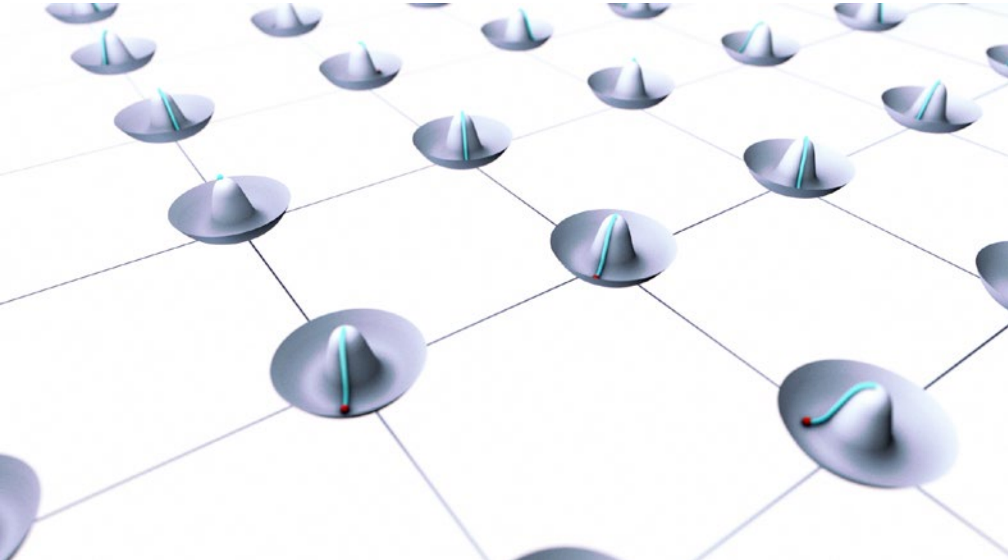
Some research in the early 21st century intensely explored classes of models where the inflaton is a scalar field parametrizing separation of D-branes in the extra dimensions required by string theory. Recall that a scalar field is just a field that associates a number to each point in space-time. (In this case, that number is the separation of the branes in the extra dimensions above the given point in space.)

In the right circumstances, the potential attracting the branes to one another can be suitably mild, resulting in the flat potential required for inflation. This potential arises by the combination of the exchange of gravitational and other force fields in the extra dimensions. At the end of inflation, the branes crash into one another. The resulting collision can result in the release of energy in many forms, including in the form of macroscopic strings that stretch across the (then-hot-and-dense) universe.

It is worth explaining why the strings can form and be macroscopic. This happens because at the end of inflation, the field that causes a smooth exit from the inflationary phase of cosmic history condenses in a potential that looks like a Mexican hat, or sombrero. There is a circle of ground states, or lowest-energy states, where this field can attain its final value. However, nothing says that the scalar field will roll to the same lowest-energy state—the same point on the rim of the hat—at different points in space. In fact, causality prevents that from happening instantly in all points in space.



As inflation ends in different parts of the universe, the scalar field will generically roll down to different minima in different parts of space. It is a subtle exercise—carried out by Tom Kibble in the 1970s—to see that this means that in a small part of the universe, the scalar will be forced to live "on top of the hill" and carry energy density instead of rolling down to the rim. Each region where the scalar encircles the rim once will eventually have a cosmic string produced by the resulting energy contained in the scalar field. These regions will form strings that will generically be the size of the cosmological horizon, or distance light can have traveled at the time of inflation, and these strings will expand with the universe as it evolves.

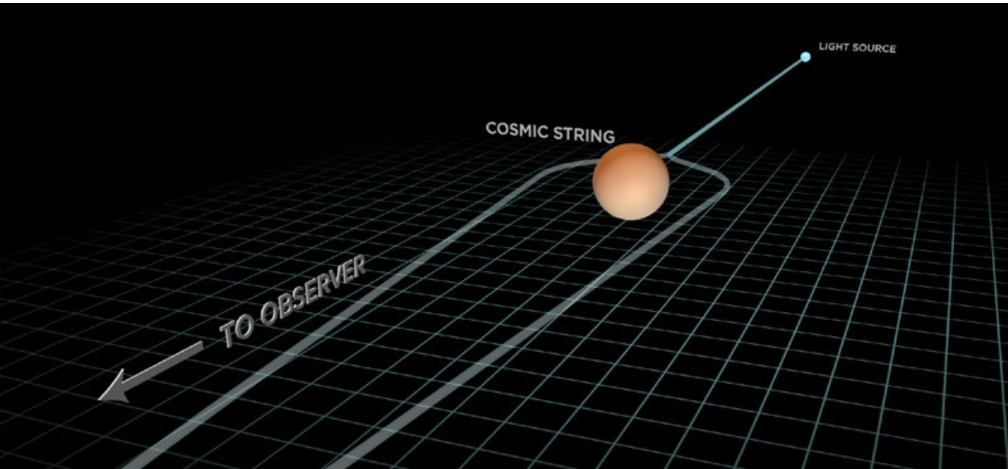


All that this mathematical argument means is that, generically, the end of brane inflation will produce cosmic superstrings. String theory actually contains different flavors of strings. In addition to the fundamental string after which string theory is named, there are also D1-branes, which are also stringlike objects. Brane inflation can result in the production of both fundamental and D-strings, and both are candidate cosmic superstrings that could be stretching across the sky.

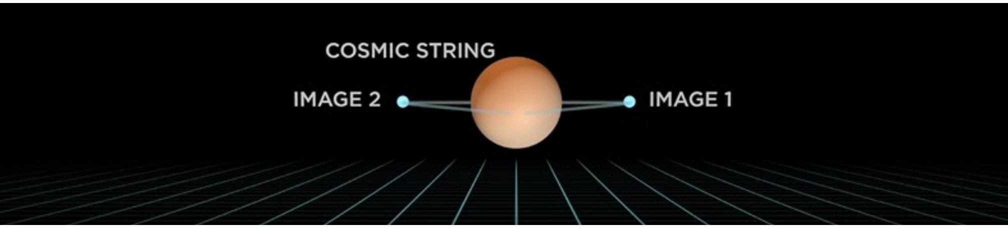
The idea that there are cosmic strings stretching across the sky may seem fanciful. But it is an idea with a long lineage that long predates the study of string theory models that exhibit this behavior. In fact, in the early 1990s, one of the exciting new ideas in cosmology was that such cosmic strings may have been the initial seeds of density fluctuations that eventually collapse into the galaxies and clusters of galaxies observed today.

That is no longer believed to be the case—perturbations from cosmic inflation have replaced cosmic strings as the leading model of structure formation, with much supporting evidence—but the properties and observability of string networks were studied intensely at that time. These studies were extended and updated by Edmund Copeland, Robert Myers, Joseph Polchinski, and others after 2005, when models of inflation based on branes became a subject of intense investigation. The upshot of their work is that cosmic strings may be visible in a few different ways.

The first discovery method involves imagining the image you'd see of a distant light source (a bright star, galaxy, or cluster) if it came to you from roughly behind a cosmic string source.



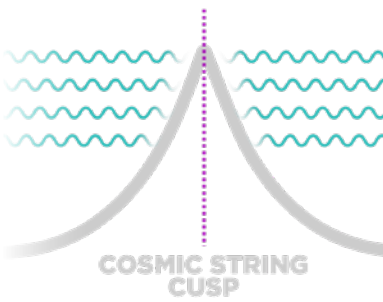
The light going around the string in either direction would be bent by its gravitational field, consistent with Einstein's general relativity. The result would be that you would see multiple images of the background source due to lensing of the cosmic strings. The search for such images in the cosmic background is a fascinating and ongoing activity. Every now and then, a rumor of such an object circulates, but unfortunately none have withstood scientific scrutiny yet.



A second discovery method involves the detection of gravitational waves emitted by the string network. Every so often, as a macroscopic string oscillates back and forth, it produces a sharp cusp on its worldvolume. This sharp feature results in the emission of significant gravitational wave signatures that could be detected by ongoing searches for gravitational waves. The size of this effect scales with the tension (μ) of the strings in the string network, with the dimensionless ratio of merit being Newton's constant times the tension of the string:

$$G_{\text{Newton}} \times \mu.$$

The tension of any network produced by cosmic inflation will depend on the energy scale at the end of inflation and therefore is tied to another parameter of significant microphysical interest. But to date, there are only bounds on the scale. No such strings have been detected through their gravitational wave signatures.



Quantum Engineering

So far, the methods you've encountered of making contact with string theory have followed the traditional path of discovery in physics. An observation of a new phenomenon in nature would lead to inference of new microphysics.

In the blossoming era of quantum engineering, there is another exciting possibility. A central discovery in string theory has been holography. A bulk theory of gravity in D-dimensional space-time can sometimes be equivalently described by a normal quantum field theory of matter in forces without gravity in $D - 1$ space-time dimensions. Can researchers turn this around and engineer a higher-dimensional theory of gravity by constructing the boundary quantum theory directly?

While scientists would see the fundamental quantum mechanical degrees of freedom in the lab, the experiments they would perform on these quantum bits of information (or qubits) would have their most ready interpretation in terms of a dual gravitational theory. This would be direct laboratory evidence for the validity of holography and the emergence of an anti–de Sitter space-time description for suitable strongly coupled quantum matter theories. Instead of discovering evidence of quantum gravity in nature, researchers would have engineered quantum gravity in the lab.

At present, this remains a pipe dream. The quantum theories that are dual to anything that is recognizable as a gravity theory are very peculiar beasts. They have many degrees of freedom; they have a spectrum of excitations, or excited states, that is very sparse at low energies, so there is very little to see; and they have very strong coupling. Making a system with a large number of degrees of freedom in the lab is trivial—after all, any tiny slab of metal has Avogadro's number of electrons in it. But arranging these degrees of freedom so that they interact strongly with a very sparse spectrum of low-energy excitations is a very difficult task. Absent these properties, you will not have produced a system whose best description is in terms of higher-dimensional gravity.

Nevertheless, with the impressive advances in quantum control evident in modern atomic physics laboratories—significantly motivated by the very different goal of building a quantum computer—you can anticipate tentative early explorations in this direction in the future.

Reading

Arkani-Hamed, Nima, Savas Dimopoulos, and Gia Dvali. "Large Extra Dimensions: A New Arena for Particle Physics." *Physics Today* 55, no. 2 (2002): 35–40.

Battersby, Stephen. "Cosmic String: The Search Continues." *New Scientist*, February 20, 2008. https://www.newscientist.com/article/mg19726441-700-cosmic-string-the-search-continues/.

12

Emergence: Can We Test String Theory?

This lecture discusses a subtle and controversial topic related to string theory. But obtaining a proper understanding of this topic also offers insight into many other aspects of physics. You will discover why it is true, and not at all surprising, that it has been so hard to verify or disprove string theory. After all, the fundamental postulate of the simplest limits of string theory—that the fundamental constituents of matter are made of tiny loops of string—sounds simple enough. Why can't scientists just look and check?

To make sure it doesn't sound like this lecture is telling a just-so story about string theory, the lecture addresses a few other cases in physics where what you see (as an observer of long-distance experiments) and what you get (as a fundamental underlying theory) are rather different. Of course, in all these cases, with the benefit of hindsight, the "right" answer for the underlying theory is already known. But the analogy should make it clear just how subtle and difficult it might be to see the strings underlying string theory.

Coulomb's Law

The underlying cause of possible confusion is that the laws of physics can appear to change in form as the distance scale at which they are probed is changed. So, consider, for instance, Coulomb's law, where two charges will attract or repel each other with a force proportional to the product of their charges divided by the distance squared:

$$F_e = \frac{kq_1q_2}{r^2}.$$

Why does anything change when this simple law is applied to real-world systems? There are at least two interrelated reasons.

The first is that macroscopic systems often contain many particles. Then, interparticle interactions with many pairs of particles have to be averaged. So, for instance, in a plasma (a gas of charged particles)—a bath of positive and negative charges interacting in a container—on average, each negatively charged particle will be screened at long distances by attracting a few of the positively charged particles a little closer to its location. That means that it can be hard to figure out what the charges and properties of the basic charge carriers are. Their apparent properties can be modified by their tendency to hide themselves behind clouds of slightly attracted partners of the opposite charge.

Secondly, quantum mechanics means that even pair-wise interactions between charged particles vary in a subtle way with distance. Heisenberg's uncertainty principle says, for instance, that you can't know both the position and momentum of a particle, or you can't measure energy to arbitrary precision over very short times:

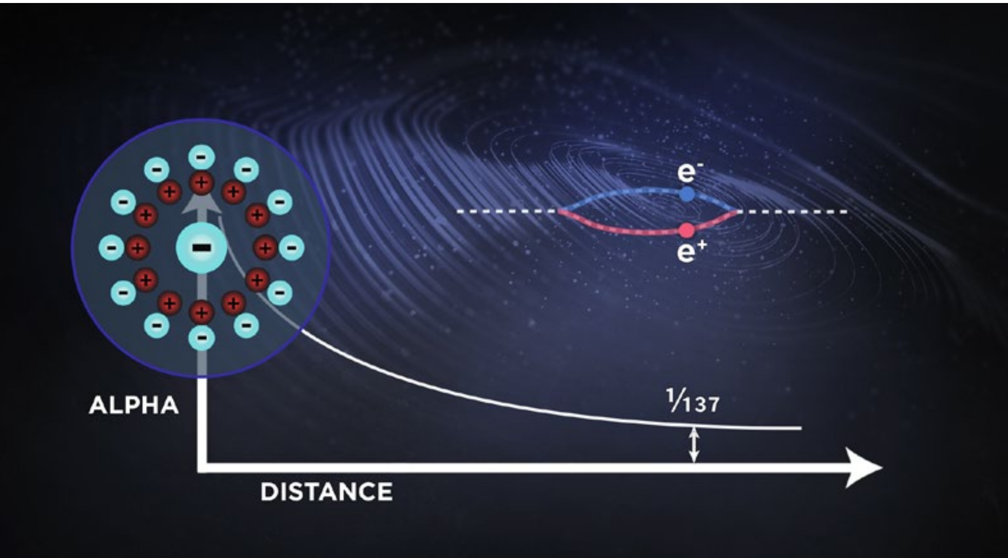
$$\Delta p \Delta x \geq \frac{1}{2}\hbar$$

$$\Delta E \Delta t \geq \frac{1}{2}\hbar.$$

In quantum mechanics, classical objects are fuzzed out by the Heisenberg uncertainty principle. Everything that is not forbidden is compulsory, and consistent with the rules of quantum mechanics, the vacuum can "borrow" virtual particle pairs for short amounts of time.

In other words, the energy-time uncertainty relationship allows you to borrow the energy to create, for instance, a virtual electron-positron pair where the two pop out a vacuum and then re-annihilate, and such pairs can contribute in appropriate Feynman diagrams that govern the electromagnetic interaction. The result is a screening of the interparticle interactions as you move to long distances. A real electron will attract the positron in the virtual pair a little bit and repel the electron a little bit, thereby slightly screening its charge. This is analogous to what happens in plasmas.

As a result, if you scatter electrons off each other, or off positrons at very short distances, the effective strength of the electromagnetic interaction will vary. This was seen directly in experiments both at the Stanford Linear Accelerator Center and CERN, where the effective coupling constant alpha in quantum electrodynamics varies, drawing to weaker and weaker values as you go long distances before it approaches the value 1⁄137. Putting both of these effects together—the presence of many particles and the existence of vacuum fluctuations—can result in profound alterations from what you might naively expect from reductionism in the style of the ancient Greeks.



The Physics of Metals

A slab of metal is comprised of an array of atoms arranged in a lattice. An isolated atom has a caricature as a solar system, where there is a central nucleus with protons and neutrons and electrons orbiting the central nucleus. In a metal, the nuclei arrange themselves in a regular lattice, but the naively bound electrons can roam free across the metal without being closely localized to a single nucleus. The mass of a free electron has been measured to high accuracy:

$$m_e = 9.10938291 \times 10^{-31} \text{ kilograms.}$$

You can also measure the force on charge carriers in a metal. One common technique to do this sticks the metal on a transverse magnetic field and looks at the cyclotron orbits of the electrons caused by their motion in the magnetic field. The electrons move in little circular orbits, and the frequency of these orbits can be measured. And that measurement determines the effective mass of the particle that is moving in the orbit.

The most naive picture would be that the properties of the electron in a metal—including its effective mass—would be the same as those of the electron in vacuum. But the actual results are much more interesting. It turns out that the effective mass of whatever's present in the metal that's supposed to be an electron can be renormalized by the medium by a factor of up to 30, depending on the medium, or the detailed properties of the metal. The emergent charge carrier, or quasiparticle, shares some of the properties of the electron.

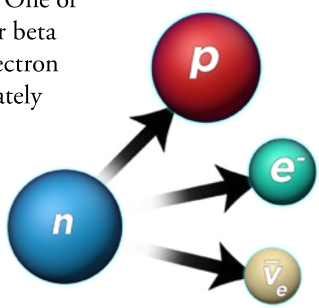
METAL	Ag	Au	Bi	Cu	K	Li	Na	Ni	Pt	Zn
m_e^*/m_e	0.99	1.10	0.047	1.01	1.12	1.28	1.2	28	13	0.85

A heuristic model of what's going on is as follows: Consider a horse running along a dusty trail. A detailed picture with a fine camera would reveal a horse and individual dust motes that it's kicking up. But far away, if the horse is running fast, you might instead see a vague object surrounded by a cloud of dirt. The moving object seen from far away will have its own effective radius (that of the dirt cloud that surrounds the horse) and its own mass (since the dust cloud adds some mass to that of the horse). That object might be called a quasi-horse, by analogy with quasiparticles—so it is with the dressed electron quasiparticles in a metal.

If emergent properties of something as prosaic as an electron in a metal can change so drastically, are even more drastic changes possible?

Nuclear Beta Decay

In the early 20th century, nuclear physicists began to experimentally characterize the processes that are allowed to interconvert among different nuclei characterizing different elements. One of the most mysterious was something called nuclear beta decay, where a nucleus could emit an energetic electron and transform to a different nucleus of approximately the same mass. As expertise about the world of elementary particles grew, this process was later recognized as one in which, for instance, a neutron in the nucleus decays to a proton, electron, and electron antineutrino.

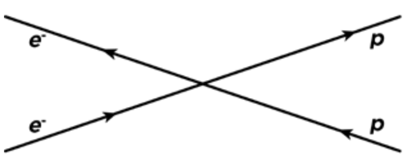


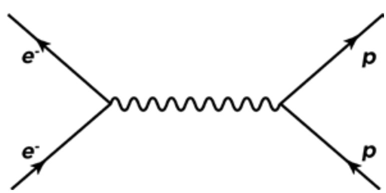
In a Feynman diagram, this kind of process could be represented by an interaction vertex where a neutron comes in and a proton, electron, and antineutrino all come out.

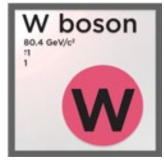
When you think about the coupling constant, called the Fermi constant, that would characterize such a process, governing the interconversion rate between nuclei, a slight problem emerges. In terms of dimensional analysis, you can show that because four fermion fields emerge from the vertex, the Fermi coupling constant has the dimensions of an inverse mass squared. From the prior discussion of the non-renormalizable nature of gravity—where the root problem was that Newton's coupling constant has dimensions of inverse mass squared—you can anticipate a problem. And indeed, there is one.

In the Fermi theory, while for a suitable choice of the Fermi coupling a good description of nuclear phenomenology is found, a prediction is that as you consider higher-energy processes, the rate for proton-neutron scattering will diverge as a power of the center of mass energy. Now that does not happen in nature. To understand what happens instead, you need to first consider electron-proton interactions. Because electrons scatter off protons in nature, you might expect that there is a four-fermion interaction with an incoming electron and proton and an outgoing electron and proton.

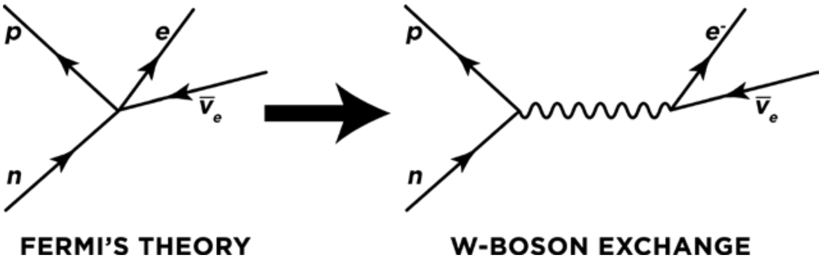
But quantum electrodynamics splits up this would-be analog of the four-fermion interaction by introducing an intermediate photon. Instead of a point-like interaction between two protons and two electrons, a photon gets exchanged between an electron and proton, mediating their scattering. The force-carrying particle, the photon, changes the behavior of the theory at higher energies. There is not a problem with electron-proton scattering growing without bound at high energies.







Could a similar thing happen for the weak interactions? Indeed, it was proposed in the 1950s that secretly the nuclear beta decay is mediated by exchange of another force-carrying particle. This would predict the existence of a new elementary particle: a force-carrying particle mediating the so-called weak interactions. Such a particle, the W boson—introduced completely by analogy to the photon—was later directly seen at experiments in CERN in the 1980s.



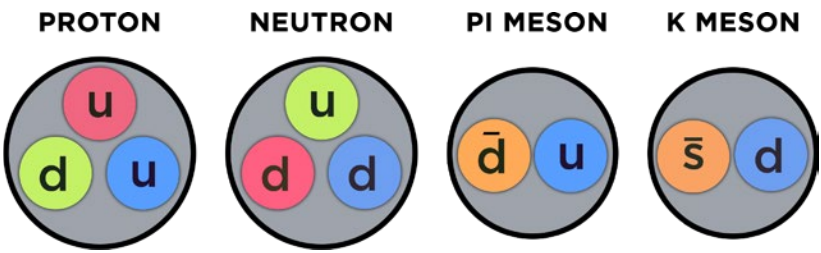
By measuring the rate of beta decay, you can measure the strength of the four-fermion coupling between protons, neutrons, electrons, and antineutrinos. Dimensional analysis relates this coupling to an inverse mass squared—and this mass scale can be identified with the mass of the missing force-carrying particle. Indeed, the W boson has the correct mass to be the missing particle in this story.

Why can the W boson and its mass be discovered this way, but the same thing doesn't happen with the massless photon? The reason is that the mass scale inferred from the Fermi interaction is very high compared to the energy of the particles involved in nuclear beta decay. This means that in the Feynman diagram describing beta decay, the W boson is a virtual particle that only exists for a very short time—a time much shorter than the other time scales involved in the beta decay process. If you blink, you can effectively ignore the W-boson exchange and see a four-fermion vertex. The only hint of its existence is in the mass scale that characterizes the four-fermion term. None of this is true in the case of electromagnetic interactions.

The Short-Distance Theory of Strong Interactions

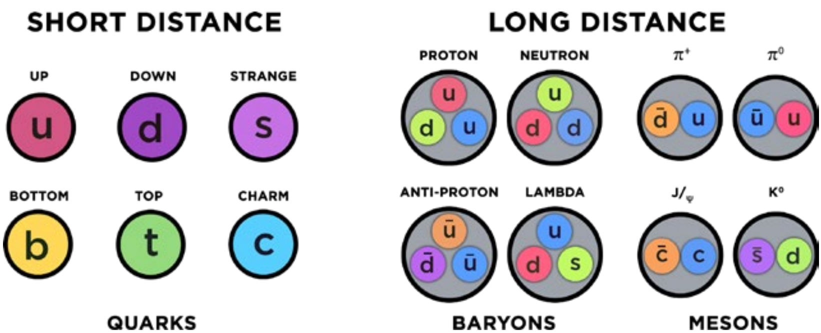
The basic constituents of the atom are the electron, proton, and neutron, arranged in a solar-system-like model with the electron orbiting a nucleus in order to make the atoms that are familiar from chemistry. But particle physicists use very-high-energy electrons in particle colliders as "microscopes," scattering them off nuclei to see the protons and neutrons in the nucleus. And the results of such experiments show that there is internal structure even inside these nucleons. The proton and neutron are each made up of three point-like constituents that are now called up and down quarks. The quarks attract each other with the strong nuclear interactions. In an opposite behavior to what is seen in electromagnetism, where charge was screened at long distances, the strong force exhibits anti-screening. It becomes stronger and stronger at long distances.

So, the short-distance theory of strong interactions is basically a theory of "almost free" quarks. But because of this anti-screening, the strong force strongly confines the quarks in proximity to one another as you try to separate them. So, at long distances, instead of seeing quarks, you find "strong neutral force" neutral objects: the protons, neutrons, and mesons that are seen in nature.



Now, suppose you were to try to isolate a free quark. Try as you might, you're going to fail. The flux tube of strong force carriers—or gluons—will tie the quarks together, almost like an open string holding its ends together. It is this flux tube of gluons that implements the anti-screening of the strong interactions, causing a linearly growing strength of the confining interaction as you try to separate quarks. This is different from atoms in electromagnetism, which you can break apart (or ionize). The quarks are confined. In this theory, in the short-distance description, you see the individual up, charm, top, and so forth, quarks. But at long distances—due to anti-screening—you get particles like the protons, neutrons, and mesons of quantum chromodynamics (QCD) phenomenology.

Going from the baryons and mesons (made of triplets and pairs of quarks, respectively) seen in experiment to the fundamental theory of the strong force—QCD—was quite an adventure. It spanned many decades and required several Nobel Prize–winning experimental and theoretical discoveries to uncover the story.



The Role of Emergence in Physics

This lecture has tried to offer some sense of the role of emergence in physics. It plays a key role in condensed-matter physics, where the microscopic electrons at short distances manifest in surprising quasiparticles (and even more surprising behaviors, such as superconductivity) at long distances. It plays a key role in particle physics, where the existence of W bosons at short distance (or high energy) gives rise to the striking phenomena of nuclear physics, such as nuclear beta decay.

And sometimes emergence means that the objects seen at long distances, where experiments are conducted, don't even have a simple relation to the basic constituent objects of the theory. This is the case, for instance, in the theory of the strong interactions. The fundamental quarks have at best a very indirect relationship to the basic objects seen at long distances in experiments.

These examples all illustrate some important, more general features of physics. Physics is organized by energy scale or distance. The degrees of freedom best for describing the system at short distances (such as electrons or quarks) may or may not be those best suited for describing its behavior at long distances (such as quasiparticles in a metal or the protons in an atom).

Sometimes the reorganization as you go from short- to long-distance scales is quantitative in nature. So, in a metal, the quasiparticle mass differs from the electron mass, but it is vaguely recognizable as a "fat electron." Happily, direct high-energy experiments can be done to see the electron, so it can be recognized when it's disguised in quasiparticle form.

Sometimes the reorganization is so complicated that modern mathematics had to be imported into theoretical physics to understand it. This was the case with the strong interactions, where the relationship between the quarks at short distance and the protons, neutrons, and their cousins at long distance required physicists to learn about the theory of Lie groups and their representations. Fortunately, after many decades, direct high-energy experiments could be done to produce the quarks, and hence they could be recognized in nucleon form.

In the case of a potential string theory of the real world, direct high-energy experiments cannot be done to "see" the strings. In most versions of string theory, observing the relevant energy scale to directly see excited strings would require a collider many orders of magnitude more powerful than any that can be imagined to be built, even in the most optimistic future scenario.

So, while string theory gives all evidence of being a unified theory of interactions much like those seen in nature, finding direct evidence for strings in the real world may remain a distant hope for the foreseeable future. Researchers will have to be really lucky. On the other hand, it would be inappropriate and even misleading to conclude these lectures by just saying string theory awaits confirmation of its fruitfulness from experiment and that researchers will need to be lucky to get that confirmation. In pure research, the fruitfulness of a research endeavor should be judged by the impact its ideas enjoy in neighboring areas. Does it produce insights that inform investigations in cognate fields?

With string theory, the past decades have seen this kind of interaction blossom with many other areas.

Highlights of String Theory Informing Other Fields

Most clearly, the inspiration from string- and brane-inspired models of extra dimensions has reverberated in studies of theoretical particle physics and early-universe cosmology.

Two of the best-studied ideas for understanding the so-called hierarchy problem of particle physics—the problem of why the Higgs physics of the standard model occurs so far beneath the Planck scale of gravity—involve physics in extra dimensions. One idea posits that there could be large extra dimensions of space and that their size gives a direct mathematical explanation of the hierarchy. Another hypothesizes that a slice of anti–de Sitter space-time—the hero in the discussion of holography—geometrizes the hierarchy and allows it to be understood directly.

In early-universe cosmology, models of inflation based on brane dynamics inspired some of the earliest work on understanding the detailed structure of so-called non-Gaussianities in the density fluctuations that seeded current-day galaxies. These non-Gaussianities, which could have resulted from the detailed nature of the inflation potential and could allow the reconstruction of aspects of inflationary history, are one holy grail of modern cosmological experiments.

But also—and maybe more surprisingly—string theory has led to tremendous developments in modern mathematics. As just one example, the study of string theory compactifications on Calabi-Yau manifolds has led to the discovery of startling new dualities, generalizing those addressed in prior lectures. These show, for instance, that strings can sometimes view topologically distinct manifolds as being equivalent (as their use in string compactification gives rise to completely equivalent, lower-dimensional physics).

This can turn into a powerful computational tool. Computing physical quantities in one side of the duality translates into mathematically powerful statements in the other side (where the computation is less tractable). This strategy applied to the duality known as mirror symmetry has yielded infinite series of predictions in enumerative geometry, a subject where mathematicians try to give counts of geometric objects having certain special properties.

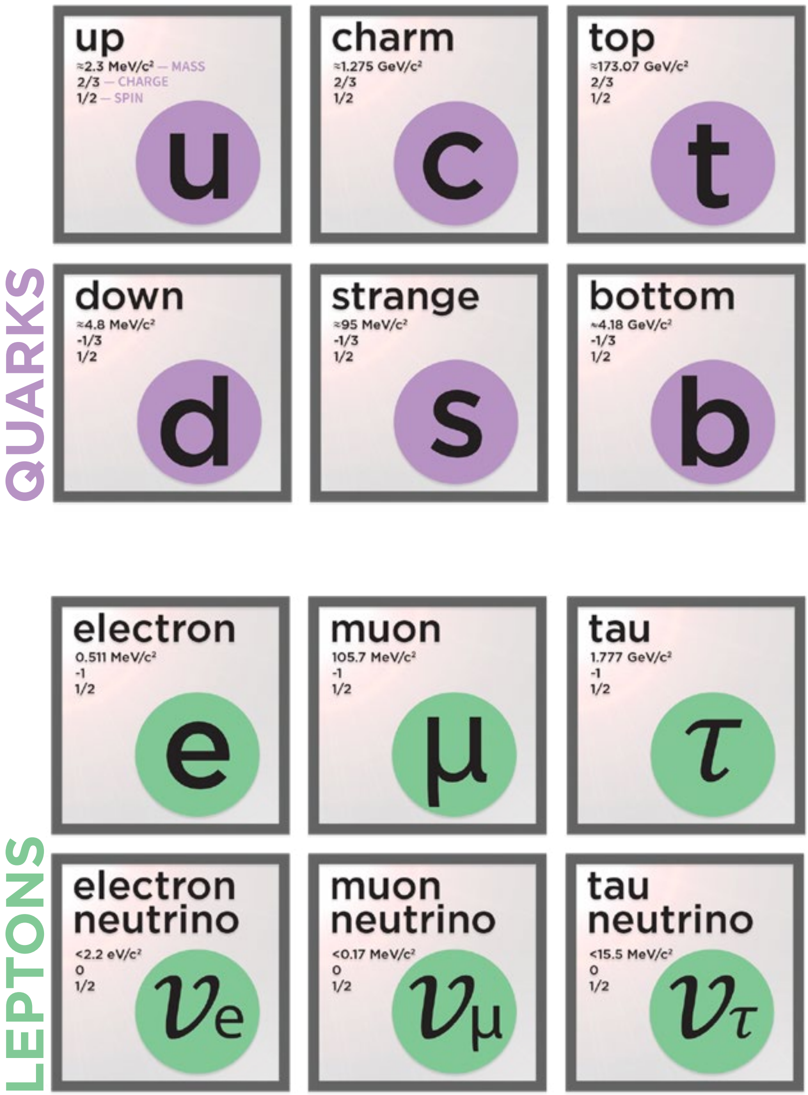
Finally, there has been a burgeoning desire to connect string descriptions of quantum gravity with modern ideas of quantum information theory. Recall that holography relates a quantum gravity theory in D dimensions to a conventional quantum theory of matter and forces in $D - 1$ spacetime dimensions. A basic slogan that many researchers have explored is that in holography, information about the state of the higher-dimensional quantum gravity theory is stored in the lower-dimensional dual in an analogue of a "quantum error-correcting code." The redundancy present in the error-correction mechanism is reflected in the precise map that allows reconstruction of the gravity state from the dual theory.

All of these connections promise continued fruitful exploration.

Reading

Laughlin, Robert. *A Different Universe.* New York: Basic Books, 2005.

The Standard Model



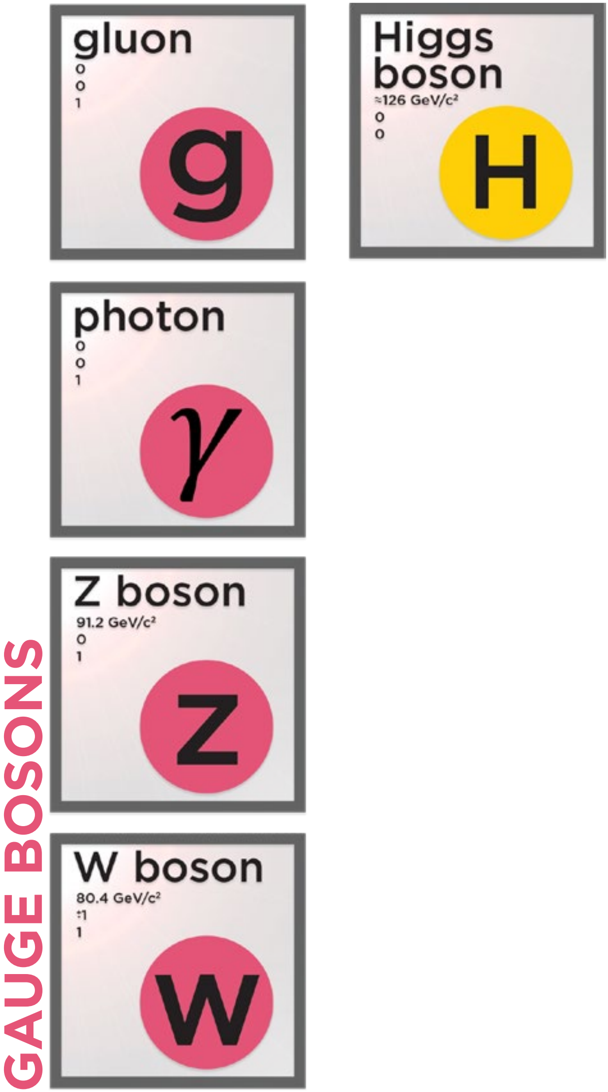

gluon
0
0
1
g
Higgs boson
≈126 GeV/c^2
0
0
H
photon
0
0
1
γ
Z boson
91.2 GeV/c^2
0
1
Z
W boson
80.4 GeV/c^2
±1
1
W
GAUGE BOSONS

Image Credits

i: Candace Boissy; 2: Getty Images; 22: Andrew J. Hanson/ Wikimedia Commons/CC BY-4.0; 50: NASA and H. Richer (University of British Columbia); 63: Roice Nelson/Public Domain; 73: NASA/WMAP Science Team; 81: TheAstronomyBum/Wikimedia Commons/Public Domain; 98: Hathi Trust Digital Library; 107: Claudio Rocchini/Wikimedia Commons/CC BY-SA 3.0