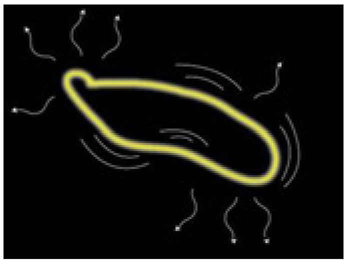
Physics for the 21st Century

Unit 4: String Theory and Extra Dimensions



© Matt DePies, University of Washington.

Unit Overview

This unit continues our movement from experimentally proven understanding of nature's four fundamental forces to the theoretical effort to develop a "theory of everything" that brings all four forces under the same conceptual umbrella. The most prominent aspect of that effort is the family of string theories that envision the basic units of matter as minuscule stretches of threadlike strings rather than point particles. Introduced in the mid-1970s, the string concept has stimulated a great deal of theoretical excitement even though it has no connection to experiment so far. The unit introduces string theory in the context of quantum gravity and outlines its inherent multidimensional nature; the most promising approach involves a total of ten dimensions. The unit then covers the relationship of string theory to particle physics and introduces the idea of "branes," related to strings. Next, the unit focuses on cosmological issues arising from our understanding of the Big Bang, outlines the way in which the concept of rapid inflation very early in the universe can solve some major issues, and details links between string theory and cosmic inflation. Finally, the unit summarizes the understanding that string theory brings to fundamental understanding of gravity.

Content for This Unit

Sections:

Section 1: Introduction

The first two units of this course have introduced us to the four basic forces in nature and the efforts to unify them in a single theory. This quest has already been successful in the case of the electromagnetic and weak interactions, and we have promising hints of a further unification between the strong interactions and the electroweak theory (though this is far from experimentally tested). However, bringing gravity into this unified picture has proven far more challenging, and fundamental new theoretical issues come to the forefront. To reach the ultimate goal of a "theory of everything" that combines all four forces in a single theoretical framework, we first need a workable theory of quantum gravity.

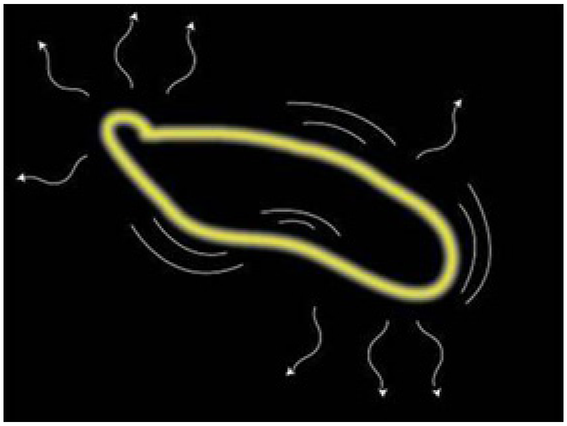


Figure 1: The fundamental units of matter may be minuscule bits of string.
Source: © Matt DePies, University of Washington.

As we saw in Units 2 and 3, theorists attempting to construct a quantum theory of gravity must somehow reconcile two fundamentals of physics that seem irreconcilable—Einstein's general theory of relativity and quantum mechanics. Since the 1920s, that effort has produced a growing number of approaches to understanding quantum gravity. The most prominent at present is string theory—or, to be accurate, an increasing accumulation of *string theories*. Deriving originally from studies of the strong nuclear force, the string concept asserts that the fundamental units of matter are not the traditional point-like particles but miniscule stretches of threadlike entities called "strings."

One of the most striking qualitative features of the string theories is that they predict the existence of extra spatial dimensions, with a total of 10 spacetime dimensions in the best-studied variants of the theory. This multitude of extra dimensions, beyond the familiar three of space plus one of time, suggests new approaches to account for some of the unsolved puzzles in the Standard Model of particle physics.

String theory also has the potential to provide insights into the ultimate puzzles of cosmology, such as the nature of the Big Bang and the origin of dark matter and dark energy that we will learn more about in Units 10 and 11.

However, candidates for an acceptable framework that combines gravity with the other three fundamental forces have one characteristic that separates them, and this unit, from all of the other topics covered in this course: Quantum gravity theories in general, and string theories in particular, have virtually no connection (as yet) to experimental evidence. There is, so far, no evidence that string theory is the correct modification of Einstein's theory, which would render it compatible with quantum mechanics in our world. String theories, or at least most models that follow from string theory, are only predictive at energy scales far from what can be probed with current particle physics and cosmological observations. This is not surprising; it follows from basic dimensional analysis, which we will describe in this unit, and which suggests that we will need a great deal of luck (or a very big accelerator) to directly test any approach to quantum gravity. Enthusiasm for string theory has been based, instead, on the theoretical richness and consistency of the structure it gives rise to, as well as the fruitful connections it has enjoyed with many other areas of physics. But one should keep in mind that its proponents will eventually need to make experimental predictions that can be tested to confirm or deny the validity of the approach as a literal description of quantum gravity in our universe.

In the following sections, we will see some current frontiers of physics where the quantum properties of gravity may be visible in near-term experiments studying the interactions of elementary particles at very high energy, or the physics of the very early universe. We hope, then, to gain one or more experimental windows (albeit very indirect ones) into the world of quantum gravity.

Section 2: *The Origins of Quantum Gravity*

In the early 20th century, physicists succeeded in explaining a wide range of phenomena on length scales ranging from the size of an atom (roughly 10^{-8} centimeters) to the size of the currently visible universe (roughly 10^{28} centimeters). They accomplished this by using two different frameworks for physical law: quantum mechanics and the general theory of relativity.

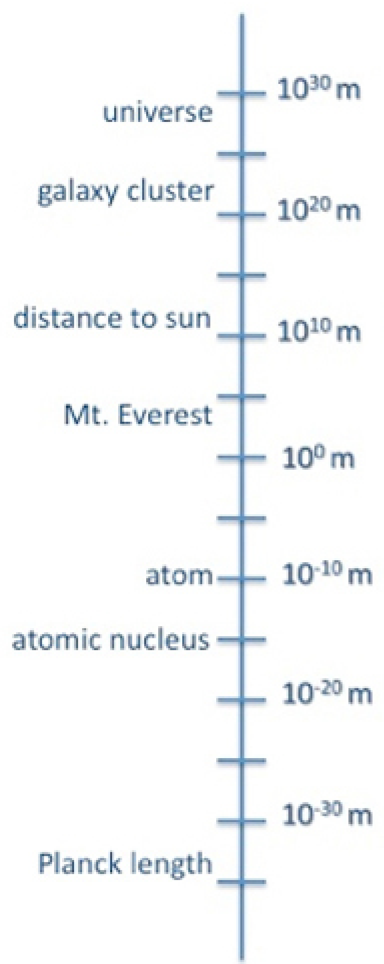


Figure 3: Quantum effects meet general relativity somewhere near the center of this length scale.
Source:

Built on Albert Einstein's use of Max Planck's postulate that light comes in discrete packets called "photons" to explain the photoelectric effect and Niels Bohr's application of similar quantum ideas to explain why atoms remain stable, quantum mechanics quickly gained a firm mathematical footing. ("Quickly" in this context, means over a period of 25 years). Its early successes dealt with systems in which a few elementary particles interacted with each other over short distance scales, of an order the size of an atom. The quantum rules were first developed to explain the mysterious behavior of matter at those distances. The end result of the quantum revolution was the realization that in the quantum world—as opposed to a classical world in which individual particles follow definite classical trajectories—positions, momenta, and other attributes of particles are controlled by a wave function that gives probabilities for different classical behaviors to occur. In daily life, the probabilities strongly reflect the classical behavior we intuitively expect; but at the tiny distances of atomic physics, the quantum rules can behave in surprising and counterintuitive ways. These are described in detail in Units 5 and 6.

In roughly the same time period, but for completely different reasons, an equally profound shift in our understanding of classical gravity occurred. One of the protagonists was again Einstein, who realized that Newton's theory of gravity was incompatible with his special theory of relativity. In Newton's theory, the attractive gravitational force between two bodies involves action at a distance. The two bodies attract each other instantaneously, without any time delay that depends on their distance from one another. The special theory of relativity, by contrast, would require a time lapse of at least the travel time of light between the two bodies. This and similar considerations led Einstein to unify Newtonian gravity with special relativity in his general relativity theory.

Einstein proposed his theory in 1915. Shortly afterward, in the late 1920s and early 1930s, theorists found that one of the simplest solutions of Einstein's theory, called the Friedmann-Lemaître-Robertson-Walker cosmology after the people who worked out the solution, can accommodate the basic cosmological data that characterize our visible universe. As we will see in Unit 11, this is an approximately flat or Euclidean geometry, with distant galaxies receding at a rate that gives us the expansion rate of the universe. This indicates that Einstein's theory seems to hold sway at distance scales of up to 10^{28} centimeters.

In the years following the discoveries of quantum mechanics and special relativity, theorists worked hard to put the laws of electrodynamics and other known forces (eventually including the strong and weak nuclear forces) into a fully quantum mechanical framework. The quantum field theory they developed describes, in quantum language, the interactions of fields that we learned about in Unit 2.

The theoretical problem of quantizing gravity

In a complete and coherent theory of physics, one would like to place gravity into a quantum framework. This is not motivated by practical necessity. After all, gravity is vastly weaker than the other forces when it acts between two elementary objects. It plays a significant role in our view of the macroscopic world only because all objects have positive mass, while most objects consist of both positive and negative electric charges, and so become electromagnetically neutral. Thus, the aggregate mass of a large body like the Earth becomes quite noticeable, while its electromagnetic field plays only a small role in everyday life.

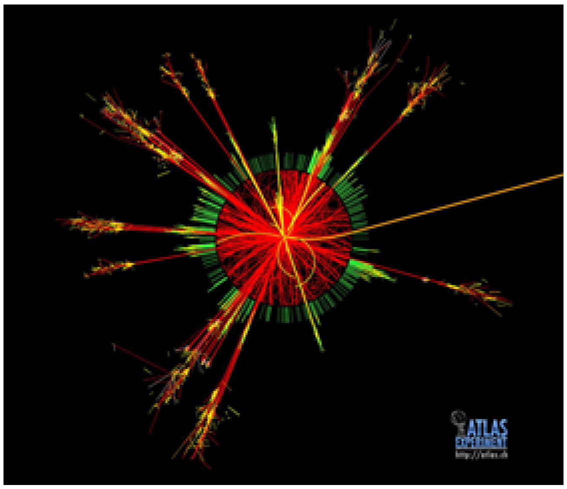


Figure 4: The rules of quantum gravity must predict the probability of different collision fragments forming at the LHC, such as the miniature black hole simulated here.
Source: © ATLAS Experiment, CERN.

But while the problem of quantizing gravity has no obvious practical application, it is inescapable at the theoretical level. When we smash two particles together at some particular energy in an accelerator like the Large Hadron Collider (LHC), we should at the very least expect our theory to give us quantum mechanical probabilities for the nature of the resulting collision fragments.

Gravity has a fundamental length scale—the unique quantity with dimensions of length that one can make out of Planck's constant, Newton's universal gravitational constant, G, and the speed of light, c. The Planck length is 1.61 x 10^{-35} meters, 10^{20} times smaller than the nucleus of an atom. A related constant is the Planck mass (which, of course, also determines an energy scale); is around 10^{-5} grams, which is equivalent to ~ 10^{19} giga-electron volts (GeV). These scales give an indication of when quantum gravity is important, and how big the quanta of quantum gravity might be. They also illustrate how in particle physics, energy, mass, and 1/length are often considered interchangeable, since we can convert between these units by simply multiplying by the right combination of fundamental constants. See the math

Since gravity has a built-in energy scale, M_{Planck}, we can ask what happens as we approach the Planckian energy for scattering. Simple approaches to quantum gravity predict that the probability of any given outcome when two energetic particles collide with each other grows with the energy, E, of the collision at a rate controlled by the dimensionless ratio $(E/M_{Planck})^2$. This presents a serious problem: At some energy close to the Planck scale, one finds that the sum of the probabilities for final states of the collision is greater than 100%. This contradiction means that brute-force approaches to quantizing gravity are failing at sufficiently high energy.

We should emphasize that this is not yet an experimentally measured problem. The highest energy accelerator in the world today, the LHC, is designed to achieve center-of-mass collision energies of roughly 10 tera-electron volts (TeV)—15 orders of magnitude below the energies at which we strongly suspect that quantum gravity presents a problem.

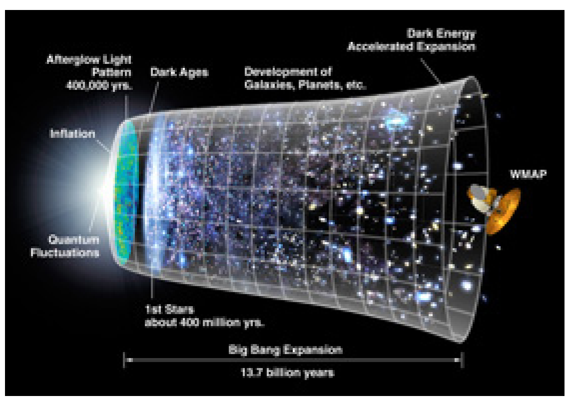


Figure 5: General relativity is consistent with all the cosmological data that characterizes our visible universe.
Source: © NASA, WMAP Science Team.

On the other hand, this does tell us that somewhere before we achieve collisions at this energy scale (or at distance scales comparable to 10^{-32} centimeters), the rules of gravitational physics will fundamentally change. And because gravity in Einstein's general relativity is a theory of spacetime geometry, this also implies that our notions of classical geometry will undergo some fundamental shift. Such a shift in our understanding of spacetime geometry could help us resolve puzzles associated with early universe cosmology, such as the initial cosmological singularity that precedes the Big Bang in all cosmological solutions of general relativity.

These exciting prospects of novel gravitational phenomena have generated a great deal of activity among theoretical physicists, who have searched long and hard for consistent modifications of Einstein's theory that avoid the catastrophic problems in high-energy scattering and that yield new geometrical principles at sufficiently short distances. As we will see, the best ideas about gravity at short distances also offer tantalizing hints about structures that may underlie the modern theory of elementary particles and Big Bang cosmology.

Section 3: String Theory

Almost by accident in the mid 1970s, theorists realized that they could obtain a quantum gravity theory by postulating that the fundamental building blocks of nature are not point particles, a traditional notion that goes back at least as far as the ancient Greeks, but instead are tiny strands of string. These strings are not simply a smaller version of, say, our shoelaces. Rather, they are geometrical objects that represent a fundamentally different way of thinking about matter. This family of theories grew out of the physics of the strong interactions. In these theories, two quarks interacting strongly are connected by a stream of carriers of the strong force, which forms a "flux tube." The potential energy between the two quarks, therefore, grows linearly with the distance between the quarks. See the math

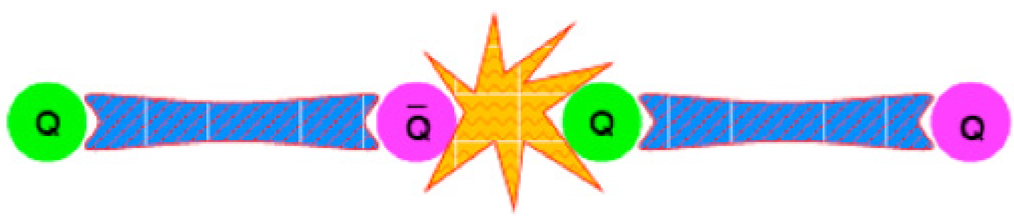


Figure 6: As quarks are pulled apart, eventually new quarks appear.
Source: © David Kaplan.

We choose to call the proportionality constant that turns the distance between the strongly interacting particles into a quantity with the units of energy "T_{string}" because it has the dimensions of mass per unit length as one would expect for a string's tension. In fact, one can think of the object formed by the flux tube of strong force carriers being exchanged between the two quarks as being an effective string, with tension T_{string}.

One of the mysteries of strong interactions is that the basic charged objects—the quarks—are never seen in isolation. The string picture explains this confinement: If one tries to pull the quarks farther and farther apart, the growing energy of the flux tube eventually favors the creation of another quark/anti-quark pair in the middle of the existing quark pair; the string breaks, and is replaced by two new flux tubes connecting the two new pairs of quarks. For this reason and others, string descriptions of the strong interactions became popular in the late 1960s. Eventually, as we saw in Unit 2, quantum chromodynamics (QCD) emerged as a more complete description of the strong force. However, along the way, physicists discovered some fascinating aspects of the theories obtained by treating the strings not as effective tubes of flux, but as fundamental quantum objects in their own right.

Perhaps the most striking observation was the fact that any theory in which the basic objects are strings will inevitably contain a particle with all the right properties to serve as a graviton, the basic force carrier of the gravitational force. While this is an unwanted nuisance in an attempt to describe strong interaction physics, it is a compelling hint that quantum string theories may be related to quantum gravity.

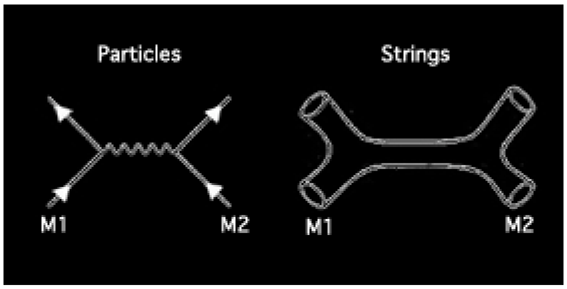


Figure 7: Gravitons arise naturally in string theory, leading to Feynman diagrams like the one on the right.
Source:

In 1984, Michael Green of Queen Mary, University of London and John Schwarz of the California Institute of Technology discovered the first fully consistent quantum string theories that were both free of catastrophic instabilities of the vacuum and capable in principle of incorporating the known fundamental forces. These theories automatically produce quantum gravity and force carriers for interactions that are qualitatively (and in some special cases even quantitatively) similar to the forces like electromagnetism and the strong and weak nuclear forces. However, this line of research had one unexpected consequence: These theories are most naturally formulated in a 10-dimensional spacetime.

We will come back to the challenges and opportunities offered by a theory of extra spacetime dimensions in later sections. For now, however, let us examine how and why a theory based on strings instead of point particles can help with the problems of quantum gravity. We will start by explaining how strings resolve the problems of Einstein's theory with high-energy scattering. In the next section, we discuss how strings modify our notions of geometry at short distances.

Strings at high energy

In the previous section, we learned that for particles colliding at very high energies, the sum of the probabilities for all the possible outcomes of the collision calculated using the techniques of Unit 2 is greater than 100 percent, which is clearly impossible. Remarkably, introducing an extended object whose fundamental length scale is not so different from the Planck length, $\ell_{string} \sim 10^{-32}$ centimeters, seems to solve this basic problem in quantum gravity. The essential point is that in high-energy scattering processes, the size of a string grows with its energy.

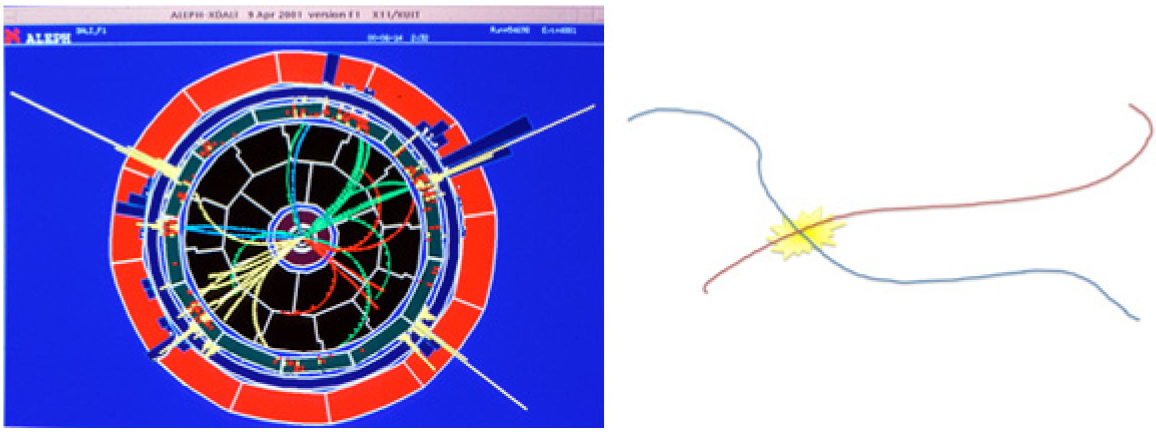


Figure 8: String collisions are softer than particle collisions.
Source: © Left: CERN, Right: Claire Cramer.

This growth-with-energy of an excited string state has an obvious consequence: When two highly energetic strings interact, they are both in the form of highly extended objects. Any typical collision involves some small segment of one of the strings exchanging a tiny fraction of its total energy with a small segment of the other string. This considerably softens the interaction compared with what would happen if two bullets carrying the same energy undergo a direct collision. In fact, it is enough to make the scattering probabilities consistent with the conservation of probability. In principle, therefore, string theories can give rise to quantum mechanically consistent scattering, even at very high energies.

Section 4: Strings and Extra Dimensions

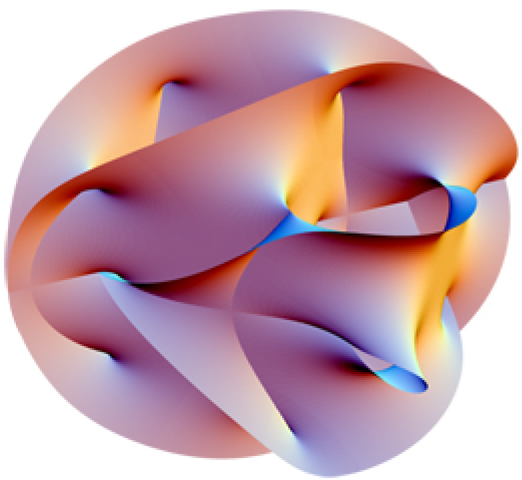


Figure 9: A depiction of one of the six-dimensional spaces that seem promising for string compactification.
Source: © Wikimedia Commons, CC Attribution-Share Alike 2.5 Generic license. Author: Lunch, 23 September 2007.

We have already mentioned that string theories that correspond to quantum gravity together with the three other known fundamental forces seem to require 10 spacetime dimensions. While this may come as a bit of a shock—after all, we certainly seem to live in four spacetime dimensions—it does not immediately contradict the ability of string theory to describe our universe. The reason is that what we call a "physical theory" is a set of equations that is dictated by the fundamental fields and their interactions. Most physical theories have a unique basic set of fields and interactions, but the equations may have many different solutions. For instance, Einstein's theory of general relativity has many nonphysical solutions in addition to the cosmological solutions that look like our own universe. We know that there are solutions of string theory in which the 10 dimensions take the form of four macroscopic spacetime dimensions and six dimensions curled up in such a way as to be almost invisible. The hope is that one of these is relevant to physics in our world.

To begin to understand the physical consequences of tiny, curled-up extra dimensions, let us consider the simplest relevant example. The simplest possibility is to consider strings propagating in nine-dimensional flat spacetime, with the 10th dimension curled up on a circle of size R. This is clearly not a realistic theory of quantum gravity, but it offers us a tantalizing glimpse into one of the great theoretical questions about gravity: How will a consistent theory of quantum gravity alter our notions of spacetime geometry at short distances? In string theory, the concept of curling up, or compactification, on a circle, is already startlingly different from what it would be in point particle theory.

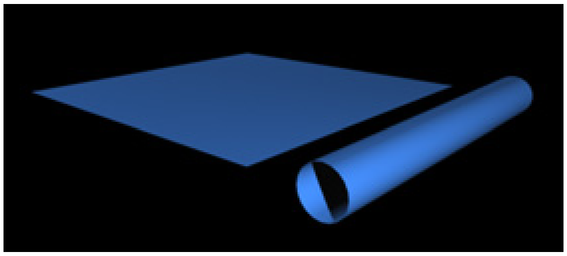


Figure 10: String theorists generally believe that extra dimensions are compactified, or curled up.
Source:

To compare string theory with normal particle theories, we will compute the simplest physical observable in each kind of theory, when it is compactified on a circle from ten to nine dimensions. This simplest observable is just the masses of elementary particles in the lower-dimensional space. It will turn out that a single type of particle (or string) in 10 dimensions gives rise to a whole infinite tower of particles in nine dimensions. But the infinite towers in the string and particle cases have an important difference that highlights the way that strings "see" a different geometry than point particles.

Particles in a curled-up dimension

Let us start by explaining how an infinite tower of nine-dimensional (9D) particles arises in the 10-dimensional (10D) particle theory. To a 9D observer, the velocity and momentum of a given particle in the hidden tenth dimension, which is too small to observe, are invisible. But the motion is real, and a particle moving in the tenth dimension has a nonzero energy. Since the particle is not moving around in the visible dimensions, one cannot attribute its energy to energy of motion, so the 9D observer attributes this energy to the particle's mass. Therefore, for a given particle species in the fundamental 10D theory, each type of motion it is allowed to perform along the extra circle gives rise to a new elementary particle from the 9D perspective.

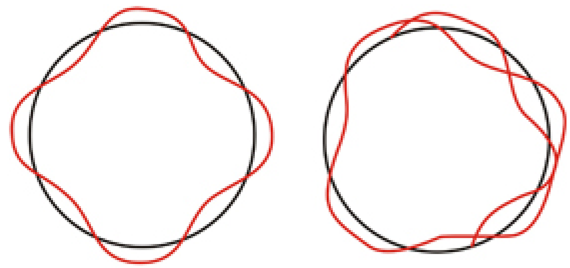


Figure 11: If a particle is constrained to move on a circle, its wave must resemble the left rather than the right drawing.
Source: © CK-12 Foundation.

To understand precisely what elementary particles the 9D observer sees, we need to understand how the 10D particle is allowed to move on the circle. It turns out that this is quite simple. In quantum mechanics, as we will see in Units 5 and 6, the mathematical description of a particle is a "probability wave" that gives the likelihood of the particle being found at any position in space. The particle's energy is related to the frequency of the wave: a higher frequency wave corresponds to a particle with higher energy. When the particle motion is confined to a circle, as it is for our particle moving in the compactified tenth dimension, the particle's probability wave needs to oscillate some definite number of times (0, 1, 2 ...) as one goes around the circle and comes back to the same point. Each possible number of oscillations on the circle corresponds to a distinct value of energy that the 10D particle can have, and each distinct value of energy will look like a new particle with a different mass to the 9D observer. The masses of these particles are related to the size of the circle, and the number of wave oscillations around the circle:

$$m_0 = 0, m_1 = 1/R, m_2 = 2/R...$$

So, as promised, the hidden velocity in the tenth dimension gives rise to a whole tower of particles in nine dimensions.

Strings in a curled-up dimension

Now, let us consider a string theory compactified on the same circle as above. For all intents and purposes, if the string itself is not oscillating, it is just like the 10D particle we discussed above. The 9D experimentalist will see the single string give rise to an infinite tower of 9D particles with distinct masses. But that’s not the end of the story. We can also wind the string around the circular tenth dimension. To visualize this, imagine winding a rubber band around the thin part of a doorknob, which is also a circle. If the string has a tension $T_{string} = 1/\alpha'$, (the conventional notation for the string tension), then winding the string once, twice, three times ... around a circle of size R, costs an energy:

$$m_1 = R/\alpha', m_2 = 2R/\alpha', m_3 = 3R/\alpha' \ldots$$

This is because the tension is defined as the mass per unit length of the string; and if we wind the string n times around the circle, it has a length which is n times the circumference of the circle. Just as a 9D experimentalist cannot see momentum in the 10th dimension, she also cannot see this string's winding number. Instead, she sees each of the winding states above as new elementary particles in the 9D world, with discrete masses that depend on the size of the compactified dimension and the string tension.

Geometry at short distances

One of the problems of quantum gravity raised in Section 2 is that we expect geometry at short distances to be different somehow. After working out what particles our 9D observer would expect to see, we are finally in a position to understand how geometry at short distances is different in a string theory.

The string tension, $1/\alpha'$, is related to the length of the string, ℓ_{string}, via $\alpha' = \ell_{string}^2$. Strings are expected to be tiny, with $\ell_{string} \sim 10^{-32}$ centimeter, so the string tension is very high. If the circle is of moderate to macroscopic size, the winding mode particles are incredibly massive since their mass is proportional to the size of the circle multiplied by the string tension. In this case, the 9D elementary particle masses in the string theory look much like that in the point particle theory on a circle of the same size, because such incredibly massive particles are difficult to see in experiments.

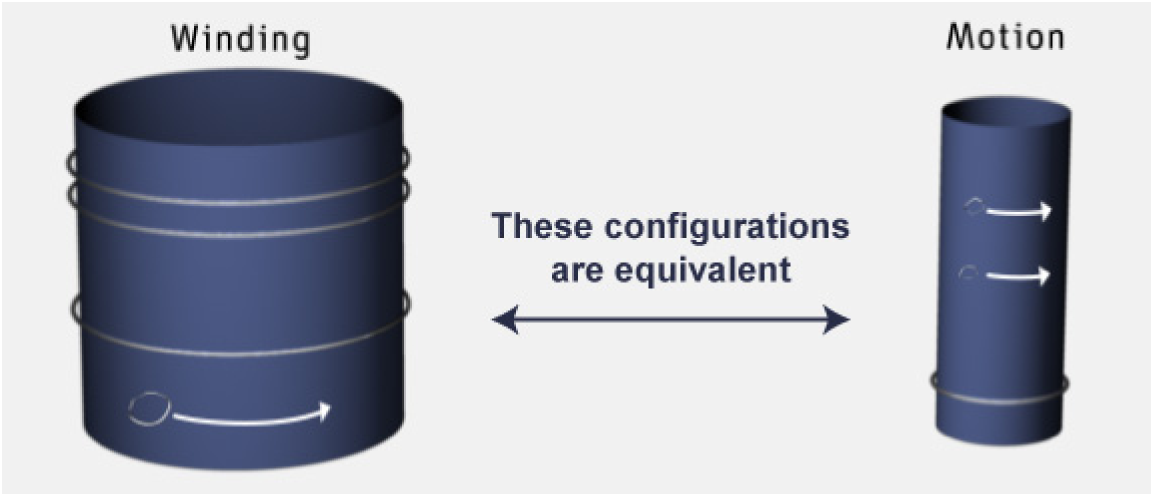


Figure 12: The consequences of strings winding around a larger extra dimension are the same as strings moving around a smaller extra dimension.
Source:

However, let us now imagine shrinking R until it approaches the scale of string theory or quantum gravity, and becomes less than ℓ_{string}. Then, the pictures one sees in point particle theory, and in string theory, are completely different. When R is smaller than ℓ_{string}, the modes m_1, m_2 ... are becoming lighter and lighter. And at very small radii, they are low-energy excitations that one would see in experiments as light 9D particles.

In the string theory with a small, compactified dimension, then, there are two ways that a string can give rise to a tower of 9D particles: motion around the circle, as in the particle theory, and winding around the circle, which is unique to the string theory. We learn something very interesting about geometry in string theory when we compare the masses of particles in these two towers.

For example, in the "motion" tower, $m_1 = 1/R$; and in the "winding" tower, $m_1 = R/\alpha'$. If we had a circle of size α'/R instead of size R, we'd get exactly the same particles, with the roles of the momentum-carrying strings and the wound strings interchanged. Up to this interchange, strings on a very large space are identical (in terms of these light particles, at least) to strings on a very small space. This large/small equivalence extends beyond the simple considerations we have described here. Indeed, the full string theory on a circle of radius R is completely equivalent to the full string theory on a circle of radius α'/R. This is a very simple illustration of what is sometimes called "quantum geometry" in string theory; string theories see geometric spaces of small size in a very different way than particle theories do. This is clearly an exciting realization, because many of the mysteries of quantum gravity involve spacetime at short distances and high energies.

Section 5: Extra Dimensions and Particle Physics

The Standard Model of particle physics described in Units 1 and 2 is very successful, but leaves a set of lingering questions. The list of forces, for instance, seems somewhat arbitrary: Why do we have gravity, electromagnetism, and the two nuclear forces instead of some other cocktail of forces? Could they all be different aspects of a single unified force that emerges at higher energy or shorter distance scales? And why do three copies of each of the types of matter particles exist—not just an electron but also a muon and a tau? Not just an up quark, but also a charm quark and a top quark? And how do we derive the charges and masses of this whole zoo of particles? We don't know the answers yet, but one promising and wide class of theories posits that some or all of these mysteries are tied to the geometry or topology of extra spatial dimensions.

Perhaps the first attempt to explain properties of the fundamental interactions through extra dimensions was that of Theodor Kaluza and Oskar Klein. In 1926, soon after Einstein proposed his theory of general relativity, they realized that a unified theory of gravity and electromagnetism could exist in a world with 4+1 spacetime dimensions. The fifth dimension could be curled up on a circle of radius R so small that nobody had observed it.

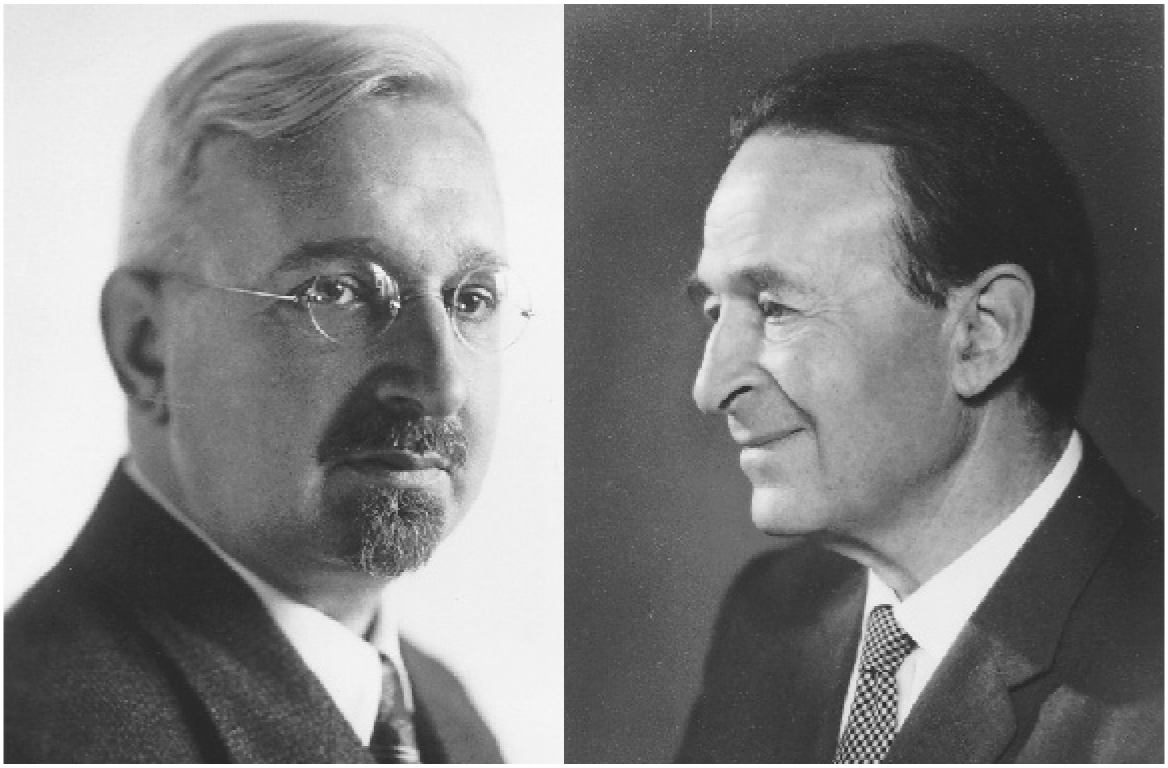


Figure 13: Theodor Kaluza (left) and Oskar Klein (right) made a remarkable theoretical description of gravity in a fifth dimension.
Source: © Left: University of Göttingen, Right: Stanley Deser.

In the 5D world, there are only gravitons, the force carriers of the gravitational field. But, as we saw in the previous section, a single kind of particle in higher dimensions can give rise to many in the lower dimension. It turns out that the 5D graviton would give rise, after reduction to 4D on a circle, to a particle with very similar properties to the photon, in addition to a candidate 4D graviton. There would also be a whole tower of other particles, as in the previous section, but they would be quite massive if the circle is small, and can be ignored as particles that would not yet have been discovered by experiment.

This is a wonderful idea. However, as a unified theory, it is a failure. In addition to the photon, it predicts additional particles that have no counterpart in the known fundamental interactions. It also fails to account for the strong and weak nuclear forces, discovered well after Kaluza and Klein published their papers. Nevertheless, modern generalizations of this basic paradigm, with a few twists, can both account for the full set of fundamental interactions and give enormous masses to the unwanted additional particles, explaining their absence in low-energy experiments.

Particle generations and topology

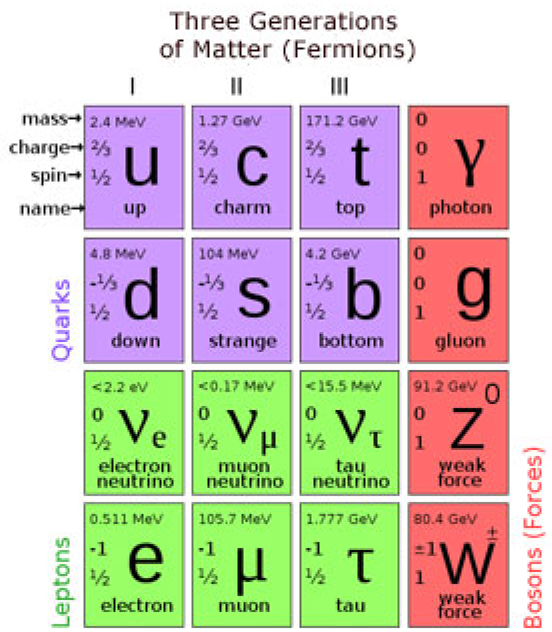


Figure 14: The Standard Model of particle physics.
Source: © Wikimedia Commons, Creative Commons 3.0 Unported License. Author: MissMJ, 27 June 2006.

One of the most obvious hints of substructure in the Standard Model is the presence of three generations of particles with the same quantum numbers under all the basic interactions. This is what gives the Standard Model the periodic table-like structure we saw in Unit 1. This kind of structure sometimes has a satisfying and elegant derivation in models based on extra dimensions coming from the geometry or topology of space itself. For instance, in string theories, the basic elementary particles arise as the lowest energy states, or ground states, of the fundamental string. The different possible string ground states, when six of the 10 dimensions are compactified, can be classified by their topology.

Because it is difficult for us to imagine six dimensions, we'll think about a simpler example: two extra dimensions compactified on a two-dimensional surface. Mathematicians classified the possible topologies of such compact, smooth two-dimensional surfaces in the 19th century. The only possibilities are so-called "Riemann surfaces of genus g," labeled by a single integer that counts the number of "holes" in the surface. Thus, a beach ball has a surface of genus 0; a donut's surface has genus 1, as does a coffee mug's; and one can obtain genus g surfaces by smoothly gluing together the surfaces of g donuts.

Figure 15: These objects are Riemann surfaces with genus 0, 1, and 2.
Source: © Left: Wikimedia Commons, Public Domain, Author: Norvy, 27 July 2006; Center: Wikimedia Commons, Public Domain, Author: Tijuana Brass, 14 December 2007; Right: Wikimedia Commons, Public Domain. Author, NickGorton, 22 August 2005.

To understand how topology is related to the classification of particles, let's consider a toy model as we did in the previous section. Let's think about a 6D string theory, in which two of the dimensions are compactified. To understand what particles a 4D observer will see, we can think about how to wind strings around the compactified extra dimensions. The answer depends on the topology of the two-dimensional surface. For instance, if it is a torus, we can wrap a string around the circular cross-section of the donut. We could also wind the string through the donut hole. In fact, arbitrary combinations of wrapping the hole N1 times and the cross-section N2 times live in distinct topological classes. Thus, in string theory on the torus, one obtains two basic stable "winding modes" that derive from wrapping the string in those two ways. These will give us two distinct classes of particles.

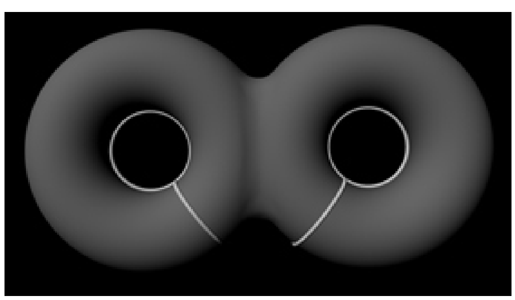


Figure 16: Strings can wind around a double torus in many distinct ways.
Source:

Similarly, a Riemann surface of genus g would permit 2g different basic stable string states. In this way, one could explain the replication of states of one type—by, say, having all strings that wrap a circular cross-section in any of the g different handles share the same physical properties. Then, the replication of generations could be tied in a fundamental way to the topology of spacetime; there would, for example, be three such states in a genus 3 surface, mirroring the reality of the Standard Model.

Semi-realistic models of particle physics actually exist that derive the number of generations from specfic string compactifications on six-dimensional manifolds in a way that is very similar to our toy discussion in spirit. The mathematical details of real constructions are often considerably more involved. However, the basic theme—that one may explain some of the parameters of particle theory through topology—is certainly shared.

Section 6: Extra Dimensions and the Hierarchy Problem

Figure 17: The weakness of gravity is difficult to maintain in a quantum mechanical theory, much as it is difficult to balance a pencil on its tip.
Source:

At least on macroscopic scales, we are already familiar with the fact that gravity, is 10^{40} times weaker than electromagnetism. We can trace the weakness of gravity to the large value of the Planck mass, or the smallness of Newton's universal gravitational constant relative to the characteristic strength of weak interactions, which set the energy scale of modern-day particle physics. However, this is a description of the situation, rather than an explanation of why gravity is so weak.

This disparity of the scales of particle physics and gravity is known as the hierarchy problem. One of the main challenges in theoretical physics is to explain why the hierarchy problem is there, and how it is quantum mechanically stable. Experiments at the LHC should provide some important clues in this regard. On the theory side, extra dimensions may prove useful.

A speculative example

We'll start by describing a speculative way in which we could obtain the vast ratio of scales encompassed by the hierarchy problem in the context of extra dimensional theories. We describe this here not so much because it is thought of as a likely way in which the world works, but more because it is an extreme illustration of what is possible in theories with extra spatial dimensions. Let us imagine, as in string theory, that there are several extra dimensions. How large should these dimensions be?

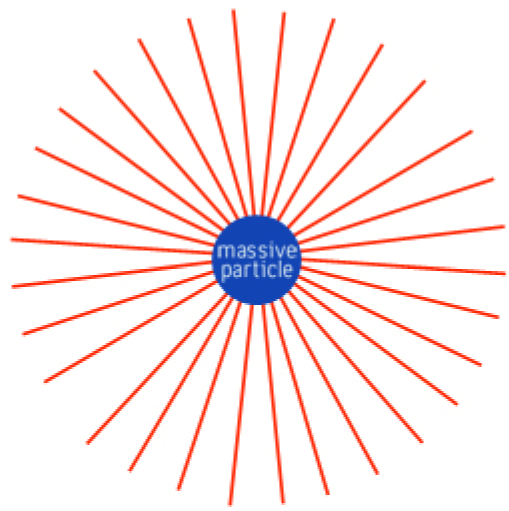


Figure 18: The gravitational field lines spread out radially from a massive particle.
Source:

First, let us think a bit about a simple explanation for Newton's law of gravitational attraction. A point mass in three spatial dimensions gives rise to a spherically symmetrical gravitational field: Lines of gravitational force emanate from the mass and spread out radially in all directions. At a given distance r from the mass, the area that these lines cross is the surface of a sphere of radius r, which grows like r^2. Therefore, the density of field lines of the gravitational field, and hence the strength of the gravitational attraction, falls like $1/r^2$. This is the inverse square law from Unit 3.

Now, imagine there were k extra dimensions, each of size L. At a distance from the point mass that is small compared to L, the field lines of gravitation would still spread out as if they are in 3+k dimensional flat space. At a distance r, the field lines would cross the surface of a hypersphere of radius r, which grows like r^{2+k}. Therefore the density of field lines and the strength of the field fall like $1/r^{2+k}$ —more quickly than in three-dimensional space. However, at a distance large compared to L, the compact dimensions don't matter—one can't get a large distance by moving in a very small dimension—and the field lines again fall off in density like $1/r^2$. The extra-fast fall-off of the density of field lines between distance of order, the Planck length, and L has an important implication. The strength of gravity is diluted by this extra space that the field lines must thread.

An only slightly more sophisticated version of the argument above shows that with k extra dimensions of size L, one has a 3+1 dimensional Newton's constant that scales like L^{-k}. This means that gravity could be as strong as other forces with which we are familiar in the underlying higher-dimensional theory of the world, if the extra dimensions that we haven't seen yet are large (in Planck units, of course; not in units of meters). Then, the relative weakness of gravity in the everyday world would be explained simply by the fact that gravity's strength is diluted by the large volume of the extra dimensions, where it is also forced to spread.

String theory and brane power

The astute reader may have noticed a problem with the above explanation for the weakness of the gravitational force. Suppose *all* the known forces really live in a 4+k dimensional spacetime rather than the four observed dimensions. Then the field lines of other interactions, like electromagnetism, will be diluted just like gravity, and the observed disparity between the strength of gravity and electromagnetism in 4D will simply translate into such a disparity in 4+k dimensions. Thus, we need to explain why gravity is different.

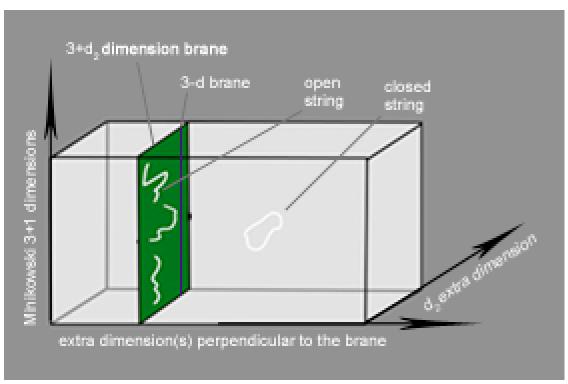


Figure 19: Strings can break open and end on a brane.
Source:

In string theories, a very elegant mechanism can confine all the interactions except gravity, which is universal and is tied directly to the geometry of spacetime, to just our four dimensions. This is because string theories have not only strings, but also branes. Derived from the term "membranes," these act like dimensions on steroids. A p-brane is a p-dimensional surface that exists for all times. Thus, a string is a kind of 1-brane; for a 2-brane, you can imagine a sheet of paper extending in the x and y directions of space, and so on. In string theory, p-branes exist for various values of p as solutions of the 10D equations of motion.

So far, we have pictured strings as closed loops. However, strings can break open and end on a p-brane. The open strings that end in this manner give rise to a set of particles which live just on that p-brane. These particles are called "open string modes," and correspond to the lowest energy excitations of the open string. In common models, this set of open string modes includes analogs of the photon. So, it is easy to get toy models of the electromagnetic force, and even the weak and strong forces, confined to a 3-brane or a higher dimensional p-brane in 10D spacetime.

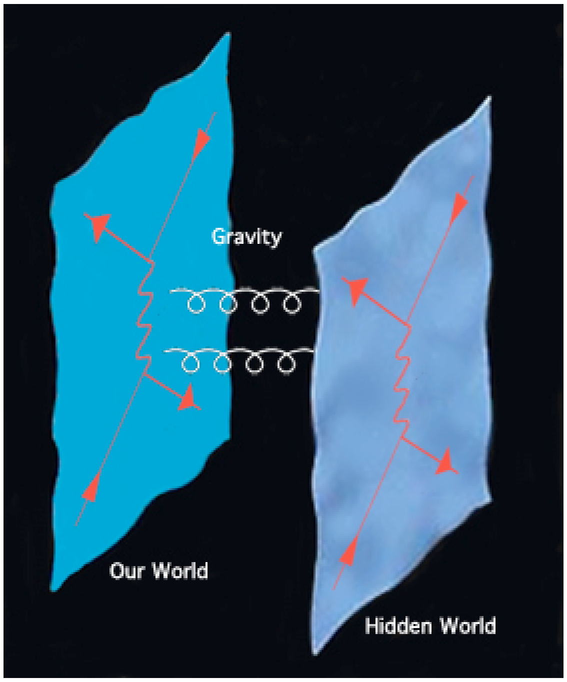


Figure 20: Gravity leaking into extra dimensions could explain the hierarchy problem.
Source:

In a scenario that contains a large number of extra dimensions but confines the fundamental forces other than gravity on a 3-brane, only the strength of gravity is diluted by the other dimensions. In this case, the weakness of gravity could literally be due to the large unobserved volume in extra spacetime dimensions. Then the problem we envisioned at the end of the previous section would not occur: Gravitational field lines would dilute in the extra dimensions (thereby weakening our perception of gravity), while electromagnetic field lines would not.

While most semi-realistic models of particle physics derived from string theory work in an opposite limit, with the size of the extra dimensions close to the Planck scale and the natural string length scale around 10^{-32} centermeters, it is worth keeping these more extreme possibilities in mind. In any case, they serve as an illustration of how one can derive hierarchies in the strengths of interactions from the geometry of extra dimensions. Indeed, examples with milder consequences abound as explanations of some of the other mysterious ratios in Standard Model couplings.

Section 7: The Cosmic Serpent

Throughout this unit, we have moved back and forth between two distinct aspects of physics: the very small (particle physics at the shortest distance scales) and the very large (cosmology at the largest observed distances in the universe). One of the peculiar facts about any modification of our theories of particle physics and gravity is that, although we have motivated them by thinking of short-distance physics or high-energy localized scattering, any change in short-distance physics also tends to produce profound changes in our picture of cosmology at the largest distance scales. We call this relationship between the very small and the very large the "cosmic serpent."

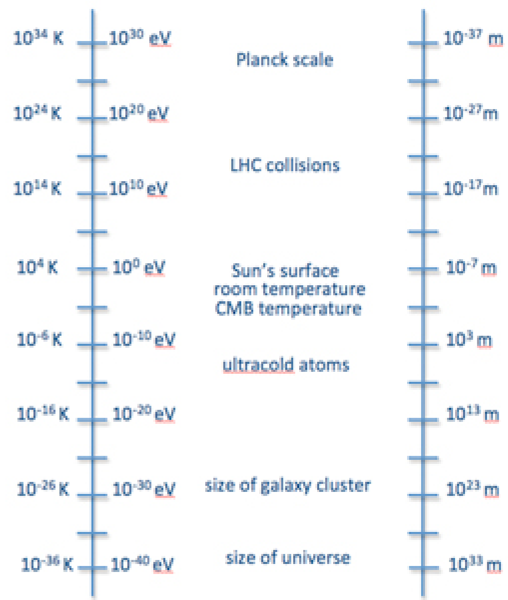


Figure 21: Changes in short distance physics—at the Planck scale—can produce profound changes in cosmology, at the largest imaginable distances.
Source:

This connection has several different aspects. The most straightforward stems from the Big Bang about 13.7 billion years ago, which created a hot, dense gas of elementary particles, brought into equilibrium with each other by the fundamental interactions, at a temperature that was very likely in excess of the TeV scale (and in most theories, at far higher temperatures). In other words, in the earliest phases of cosmology, nature provided us with the most powerful accelerator yet known that attained energies and densities unheard of in terrestrial experiments. Thus, ascertaining the nature of, and decoding the detailed physics of, the Big Bang, is an exciting task for both particle physicists and cosmologists.

The cosmic microwave background

One very direct test of the Big Bang picture that yields a great deal of information about the early universe is the detection of the relic gas of radiation called the cosmic microwave background, or CMB. As the universe cooled after the Big Bang, protons and neutrons bound together to form atomic nuclei in a process called Big Bang nucleosynthesis, then electrons attached to the nuclei to form atoms in a process called recombination. At this time, roughly 390,000 years after the Big Bang, the universe by and large became transparent to photons. Since the charged protons and electrons were suddenly bound in electrically neutral atoms, photons no longer had charged particles to scatter them from their path of motion. Therefore, any photons around at that time freely streamed along their paths from then until today, when we see them as the "surface of last scattering" in our cosmological experiments.

Bell Labs scientists Arno Penzias and Robert Wilson first detected the cosmic microwave background in 1964. Subsequent studies have shown that the detailed thermal properties of the gas of photons are largely consistent with those of a blackbody at a temperature of 2.7 degrees Kelvin, as we will see in Unit 5. The temperature of the CMB has been measured to be quite uniform across the entire sky—wherever you look, the CMB temperature will not vary more than 0.0004 K.

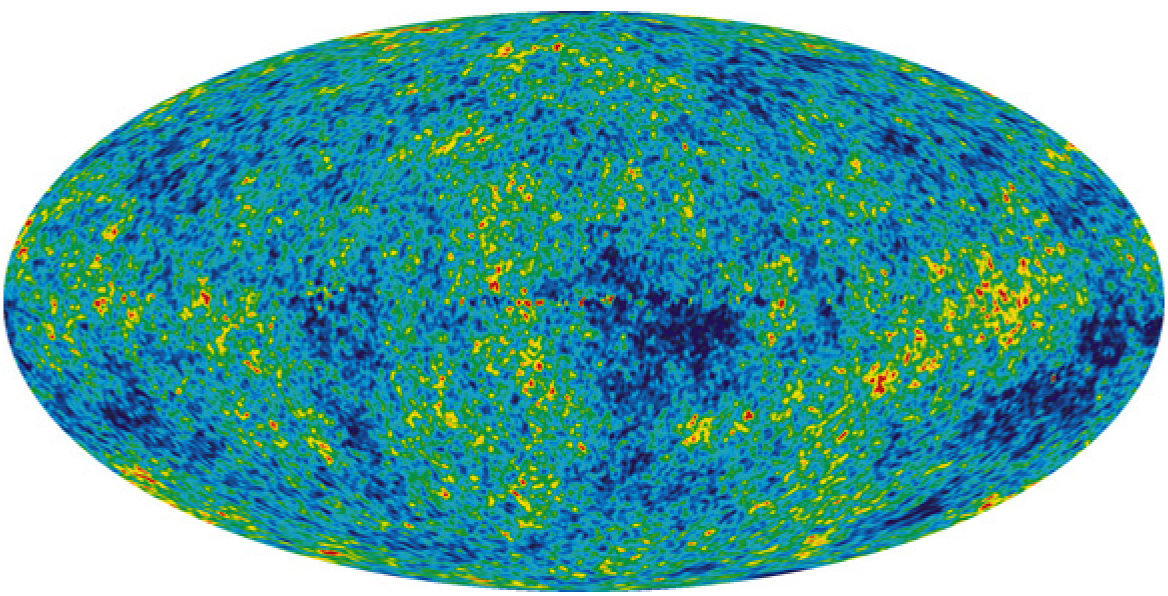


Figure 22: Map of the temperature variations in the cosmic microwave background measured by the WMAP satellite.
Source: © NASA, WMAP Science Team.

So, in the cosmic connection between particle physics and cosmology, assumptions about the temperature and interactions of the components of nuclei or atoms translate directly into epochs in cosmic history like nucleosynthesis or recombination, which experimentalists can then test indirectly or probe directly. This promising approach to testing fundamental theory via cosmological observations continues today, with dark matter, dark energy, and the nature of cosmic inflation as its main targets. We will learn more about dark matter in Unit 10 and dark energy in Unit 11. Here, we will attempt to understand inflation.

Cosmic inflation

Let us return to a puzzle that may have occurred to you in the previous section, when we discussed the gas of photons that started to stream through the universe 390,000 years after the Big Bang. Look up in the sky where you are sitting. Now, imagine your counterpart on the opposite side of the Earth doing the same. The microwave photons impinging on your eye and hers have only just reached Earth after their long journey from the surface of last scattering. And yet, the energy distribution (and hence the temperature) of the photons that you see precisely matches what she discovers.

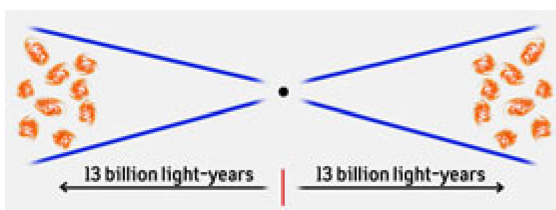


Figure 23: Both sides of the universe look the same, although light could not have traveled from one side to the other.
Source:

How is this possible? Normally, for a gas to have similar properties (such as a common temperature) over a given distance, it must have had time for the constituent atoms to have scattered off of each other and to have transferred energy throughout its full volume. However, the microwave photons reaching you and your doppelgänger on the other side of the Earth have come from regions that do not seem to be in causal contact. No photon could have traveled from one to the other according to the naive cosmology that we are imagining. How then could those regions have been in thermal equilibrium? We call this cosmological mystery the "horizon problem."

To grasp the scope of the problem, imagine that you travel billions of light-years into the past, find a distribution of different ovens with different manufacturers, power sources, and other features in the sky; and yet discover that all the ovens are at precisely the same temperature making Baked Alaska. Some causal process must have set up all the ovens and synchronized their temperatures and the ingredients they are cooking. In the case of ovens, we would of course implicate a chef. Cosmologists, who have no obvious room for a cosmic chef, have a more natural explanation: The causal structure of the universe differs from what we assume in our naive cosmology. We must believe that, although we see a universe expanding in a certain way today and can extrapolate that behavior into the past, *something drastically different happened in the far enough past*.

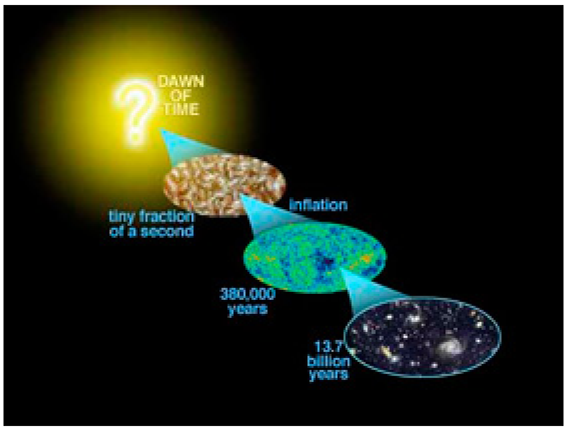


Figure 24: During the period of inflation, the universe grew by a factor of at least 10^{25}.
Source: © NASA, WMAP Science Team.

Cosmic inflation is our leading candidate for that something. Theories of cosmic inflation assert that the universe underwent a tremendously explosive expansion well before the last scattering of photons occurred. The expansion blew up a region of space a few orders of magnitude larger than the Planck scale into the size of a small grapefruit in just a few million Planck times (where a Planck time is 10^{-44} seconds). During that brief period, the universe expanded by a factor of at least 10^{25}. The inflation would thus spread particles originally in thermal contact in the tiny region a few orders of magnitude larger than the Planck length into a region large enough to be our surface of last scattering. In contrast, extrapolation of the post-Big Bang expansion of the universe into the past would never produce a region small enough for causal contact to be established at the surface of last scattering without violating some other cherished cosmological belief.

Inflation and slow-roll inflation

How does this inflation occur? In general relativity, inflation requires a source of energy density that does not move away as the universe expands. As we will see in Unit 11, simply adding a constant term (a cosmological constant) to Einstein's equations will do the trick. But, the measured value of the present-day expansion rate means the cosmological constant could only have been a tiny, tiny fraction of the energy budget of the universe at the time of the Big Bang. Thus, it had nothing to do with this explosive expansion.

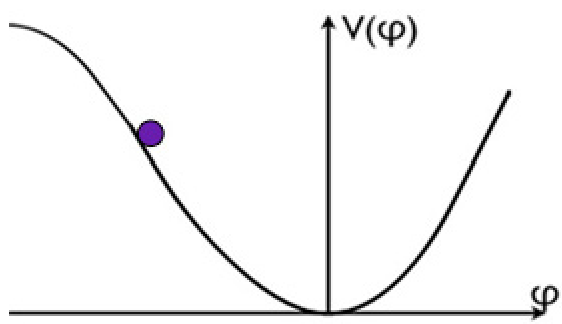


Figure 25: A potential like the one shown above for the inflaton field could have caused inflation.
Source:

However, there could have been another source of constant energy density: not exactly a cosmological constant, but something that mimics one well for a brief period of a few million Planck times. This is possible if there is a new elementary particle, the inflaton, and an associated scalar field φ. The field φ evolves in time toward its lowest energy state. The energy of φ at any spacetime point is given by a function called its "potential." If φ happens to be created in a region where the potential varies extremely slowly, then inflation will proceed. This is quite intuitive; the scalar field living on a very flat region in its potential just adds an approximate constant to the energy density of the universe, mimicking a cosmological constant but with a much larger value of the energy density than today's dark energy. We know that a cosmological constant causes accelerated (in fact, exponential) expansion of the universe.

As inflation happens, φ will slowly roll in its potential as the universe exponentially expands. Eventually, φ reaches a region of the potential where this peculiar flatness no longer holds configuration. As it reaches the ground state, its oscillations result in the production of the Standard Model quarks and leptons through weak interactions that couple them to the inflaton. This end of inflation, when the energy stored in the inflation field is dumped into quarks and leptons, is what we know as the Big Bang.

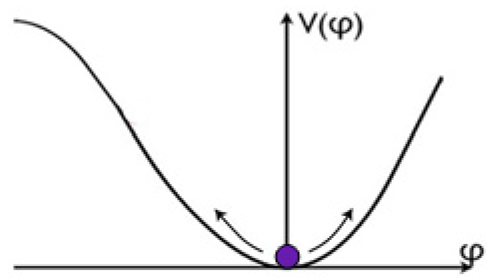


Figure 26: All the Standard Model particles could have been produced by the inflaton oscillating around its ground state like a ball rolling around in a valley.
Source:

We can imagine how this works by thinking about a ball rolling slowly on a very flat, broad hilltop. The period of inflation occurs while the ball is meandering along the top. It ends when the ball reaches the edge of the hilltop and starts down the steep portion of the hill. When the ball reaches the valley at the bottom of the hill, it oscillates there for a while, dissipating its remaining energy. However, the classical dynamics of the ball and the voyage of the inflaton differ in at least three important ways. The inflaton's energy density is a constant; it suffuses all of space, as if the universe were filled with balls on hills (and the number of the balls would grow as the universe expands). Because of this, the inflaton sources an exponentially fast expansion of the universe as a whole. Finally, the inflaton lives in a quantum world, and quantum fluctuations during inflation have very important consequences that we will explore in the next section.

Section 8: Inflation and the Origin of Structure

Figure 27: While the universe appears approximately uniform, we see varied and beautiful structure on smaller scales.
Source: © NASA, ESA, and F. Paresce (INAPF-AIASF, Bologna, Italy).

On the largest cosmological scales, the universe appears to be approximately homogeneous and isotropic. That is, it looks approximately the same in all directions. On smaller scales, however, we see planets, stars, galaxies, clusters of galaxies, superclusters, and so forth. Where did all of this structure come from, if the universe was once a smooth distribution of hot gas with a fixed temperature?

The temperature of the fireball that emerged from the Big Bang must have fluctuated very slightly at different points in space (although far from enough to solve the horizon problem). These tiny fluctuations in the temperature and density of the hot gas from the Big Bang eventually turned into regions of a slight overdensity of mass and energy. Since gravity is attractive, the overdense regions collapsed after an unimaginably long time to form the galaxies, stars, and planets we know today. The dynamics of the baryons, dark matter, and photons all played important and distinct roles in this beautiful, involved process of forming structure. Yet, the important point is that, over eons, gravity amplified initially tiny density fluctuations to produce the clumpy astrophysics of the modern era. From where did these tiny density fluctuations originate? In inflationary theory, the hot gas of the Big Bang arises from the oscillations and decay of the inflaton field itself. Therefore, one must find a source of slight fluctuations or differences in the inflaton's trajectory to its minimum, at different points in space. In our analogy with the ball on the hill, remember that inflation works like a different ball rolling down an identically shaped hill at each point in space. Now, we are saying that the ball must have chosen very slightly different trajectories at different points in space—that is, rolled down the hill in very slightly different ways.

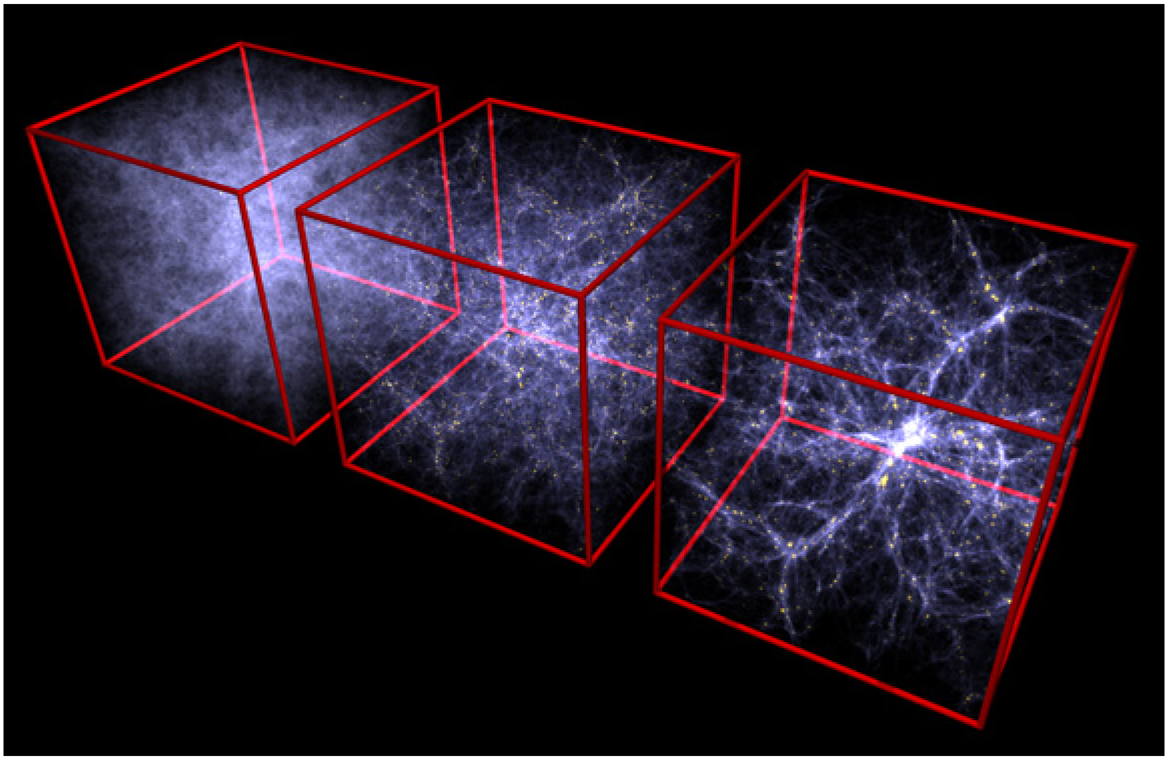


Figure 28: Small fluctuations in density in the far left box collapse into large structures on the right in this computer simulation of the universe.
Source: © Courtesy of V. Springel, Max-Planck-Institute for Astrophysics, Germany.

One source of fluctuations is quantum mechanics itself. The roll of the inflaton down its potential hill cannot be the same at all points in space, because small quantum fluctuations will cause tiny differences in the inflaton trajectories at distinct points. But because the inflaton's potential energy dominates the energy density of the universe during inflation, these tiny differences in trajectory will translate to small differences in local energy densities. When the inflaton decays, the different regions will then reheat the Standard Model particles to slightly different temperatures.

Who caused the inflation?

This leaves our present-day understanding of inflation with the feel of a murder mystery. We've found the body—decisive evidence for what has happened through the nearly uniform CMB radiation and numerous other sources. We have an overall framework for what must have caused the events, but we don't know precisely which suspect is guilty; at our present level of knowledge, many candidates had opportunity and were equally well motivated.

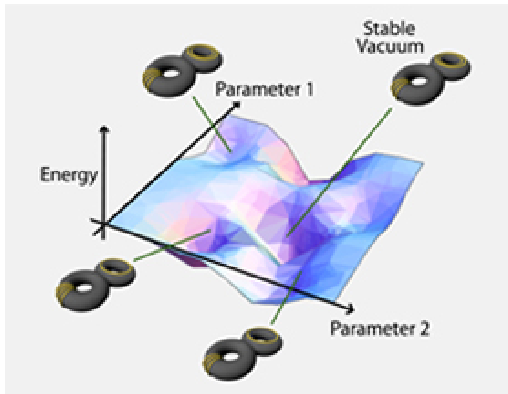


Figure 29: Typical string theories or supersymmetric field theories have many candidate scalar field inflatons.
Source:

In inflationary theory, we try to develop a watertight case to convict the single inflaton that was relevant for our cosmological patch. However, the suspect list is a long one, and grows every day. Theories of inflation simply require a scalar field with a suitable potential and some good motivation for the existence of that scalar and some rigorous derivation of that potential. At a more refined level, perhaps they should also explain why the initial conditions for the field were just right to engender the explosive inflationary expansion. Modern supersymmetric theories of particle physics, and their more complete embeddings into unified frameworks like string theory, typically provide abundant scalar fields.

While inflationary expansion is simple to explain, it is not simple to derive the theories that produce it. In particular, inflation involves the interplay of a scalar field's energy density with the gravitational field. When one says one wishes for the potential to be flat, or for the region over which it is flat to be large, the mathematical version of those statements involves M_{Planck} in a crucial way: Both criteria depend on M_{Planck}^2 multiplied by a function of the potential. Normally, we don't need to worry so much about terms in the potential divided by powers of M_{Planck} because the Planck mass is so large that these terms will be small enough to neglect. However, this is no longer true if we multiply by M_{Planck}^2. Without going into mathematical details, one can see then that even terms in the potential energy suppressed by a few powers of M_{Planck} can qualitatively change inflation, or even destroy it. In the analogy with the rolling ball on the hill, it is as if we need to make sure that the hilltop is perfectly flat with well-mown grass, and with no gophers or field mice to perturb its flat perfection with minute tunnels or hills, if the ball is to follow the inflation-causing trajectory on the hill that we need it to follow.

Figure 30: The LHC creates the highest-energy collisions on Earth, but these are irrelevant to Planck-scale physics.
Source: © CERN.

In particle physics we will probe the physics of the TeV scale at the LHC. There, we will be just barely sensitive to a few terms in the potential suppressed by a few TeV. Terms in the potential that are suppressed by M_{Planck} are, for the most part, completely irrelevant in particle physics at the TeV scale. If we use the cosmic accelerator provided by the Big Bang, in contrast, we get from inflation a predictive class of theories that are crucially sensitive to quantum gravity or string theory corrections to the dynamics. This is, of course, because cosmic inflation involves the delicate interplay of the inflaton with the gravitational field.

Section 9: *Inflation in String Theory*

Because inflation is sensitive to M_{Planck}-suppressed corrections, physicists must either make strong assumptions about Planck-scale physics or propose and compute with models of inflation in theories where they can calculate such gravity effects. String theory provides one class of theories of quantum gravity well developed enough to offer concrete and testable models of inflation—and sometimes additional correlated observational consequences. A string compactification from 10D to 4D often introduces interesting scalar fields. Some of those fields provide intriguing inflationary candidates.

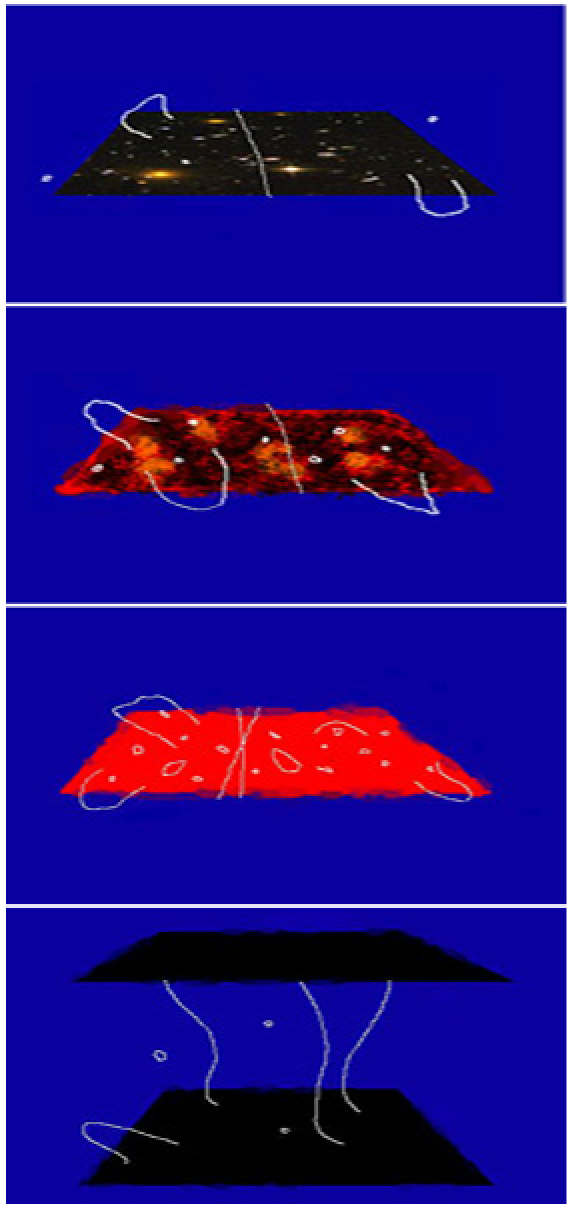


Figure 31: A brane and an anti-brane moving toward one another and colliding could have caused inflation and the Big Bang.
Source: © S.-H. Henry Tye Laboratory at Cornell University.

Perhaps the best-studied classes of models involve p-branes of the sort we described earlier in this unit. So just for concreteness, we will briefly describe this class of models. Imagine the cosmology of a universe that involves string theory compactification, curling up six of the extra dimensions to yield a 4D world. Just as we believe that the hot gas of the Big Bang in the earliest times contained particles and anti-particles, we also believe that both branes and anti-branes may have existed in a string setting; both are p-dimensional hyperplanes on which strings can end, but they carry opposite charges under some higher-dimensional analog of electromagnetism.

The most easily visualized case involves a 3-brane and an anti-3-brane filling our 4D spacetime but located at different points in the other six dimensions. Just as an electron and a positron attract one another, the brane and anti-brane attract one another via gravitational forces as well as the other force under which they are charged. However, the force law is not exactly a familiar one. In the simplest case, the force falls off as $1/r^4$, where r is the distance separating the brane and anti-brane in the 6D compactified space.

Models with sufficiently interesting geometries for the compactified dimensions can produce slow-roll inflation when the brane and anti-brane slowly fall together, under the influence of the attractive force. The inflaton field is the mode that controls the separation between the brane and the anti-brane. Each of the branes, as a material object filling our 4D space, has a tension that provides an energy density filling all of space. So, a more accurate expression for the inter-brane potential would be $V(r) \sim 2T_3 - 1/r^4$, where T_3 is the brane tension. For sufficiently large r and slowly rolling branes, the term $2T_3$ dominates the energy density of the universe and serves as the effective cosmological constant that drives inflation.

As the branes approach each other and $r \sim \ell_{string}$, this picture breaks down. This is because certain open strings, now with one end on each of the branes as opposed to both ends on a single brane, can become light. In contrast, when $r >> \ell_{string}$, such open strings must stretch a long distance and are quite heavy. Remember that the energy or mass of a string scales with its length. In the regime where r is very small, and the open strings become light, the picture in terms of moving branes breaks down. Instead, some of the light open strings mediate an instability of the brane configuration. In the crudest approximation, the brane and anti-brane simply annihilate (just as an electron and anti-electron would), releasing all of the energy density stored in the brane tensions in the form of closed-string radiation. In this type of model, the Big Bang is related to the annihilation of a brane with an anti-brane in the early universe.

Other consequences of inflation

Any well-specified model of cosmic inflation has a full list of consequences that can include observables beyond just the density fluctuations in the microwave background that result from inflation. Here, we mention some of the most spectacular possible consequences.

Quantum jiggles: We do not know the energy scale of inflation directly from data. In many of the simplest theories, however, this energy scale is very high, close to the Grand Unified Theory scale of 10^{16} GeV. It is therefore quite possible that inflation is probing energies 13 orders of magnitude higher than we'll see at the LHC.

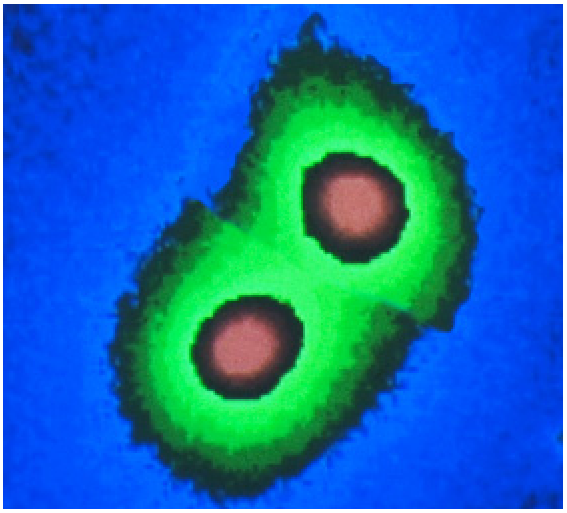


Figure 32: A cosmic string could produce a double image of a galaxy behind it.
Source: © NASA/ESA Hubble Space Telescope.

If the scale of inflation is high enough, we may see further corroborating evidence beyond the solution of the horizon problem and the explanation of density fluctuations. The density fluctuations we discussed in the previous section came from the quantum jiggles of the inflaton field itself. But during inflation, quantum jiggles also originate in the other fields present, including the gravitational field. Future cosmological experiments exploring those phenomena could pin down the scale of inflation to just a few orders of magnitude shy of the Planck scale.

Cosmic strings: Very particular models often come with their own smoking-gun signatures. Take, for example, the class of speculative models we discussed earlier based on the slow attraction and eventual annihilation of a 3-brane and an anti-3-brane. The annihilation process involves the dynamics of open strings that stretch between the 3-brane and its partner, and that eventually "condense." This condensation process creates cosmic strings as the branes annihilate, which can be thought of as 1-branes or even fundamental strings that thread our 4D spacetime, and have grown to macroscopic size. If these tension-bearing cosmic strings really were created at the end of inflation, they should be present in the universe today, with tell-tale signatures in experiments that study the distribution of matter through gravitational lensing. Future experiments should rule out the presence of such strings or detect them for a wide range of values of the possible scale of inflation.

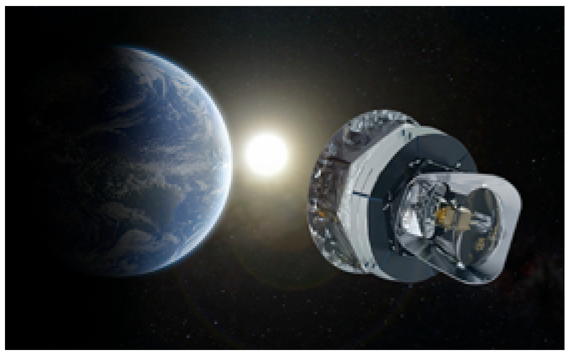


Figure 33: The Planck satellite will make precise measurements of fluctuations in the CMB.
Source: © ESA.

Density fluctuations: The slow-roll approach to inflation, with an inflaton field moving on a flat potential, is the class of models we focused on in the last section, but is not the only model of inflation. In some more modern theories, again partially inspired by branes in superstring theory, the inflaton undergoes rapid motion. Instead of the flatness of the potential, a delicate interplay between the potential and the complicated structure of the inflation kinetic terms produces the inflation. If any such model captures a grain of truth, then the pattern of density fluctuations would bear tell-tale structural signatures. Measurements by the Planck satellite that the European Space Agency launched in May 2009 should put constraints on the validity of those models.

Section 10: *Fundamental Questions of Gravity*

Figure 34: Some of Einstein's great insights began in thought experiments.
Source: © Marcelo Gleiser.

Quantum gravity is not a mystery that is, as yet, open to experiment. With a few exceptions such as inflationary cosmology and, quite possibly, the correct interpretation of dark energy (see Unit 11), thoughts about quantum gravity remain in the theoretical realm. This does not mean that theories of quantum gravity cannot be important and testable in some circumstances. In a given model, for instance, string theory models of particle physics make very definite statements about the origins of the mass hierarchy of fermions or the number of generations of Standard Model particles. But these problems may also have other solutions, insensitive to the structure of gravity at short distances; only in very few cases do we suspect that quantum gravity must be a part of the solution to a problem.

Here, we discuss some of these issues that are intrinsically gravitational. They have not yet confronted experiment directly. But we should remember that Einstein formulated special relativity by reconciling different thought experiments, and that therefore even thought experiments about quantum gravity may eventually be useful.

Black holes and entropy

Black holes are objects so dense that the escape velocity from their surface exceeds the speed of light, c. Because of that, one would think that in a relativistic theory, outside observers performing classical experiments can never see their surfaces. As a rough-and-ready definition, we will call the surface defining the region where light itself can no longer escape from the gravitational attraction of a black hole, the event horizon. Nothing, in a classical theory, can be emitted from this horizon, though many things can fall through.

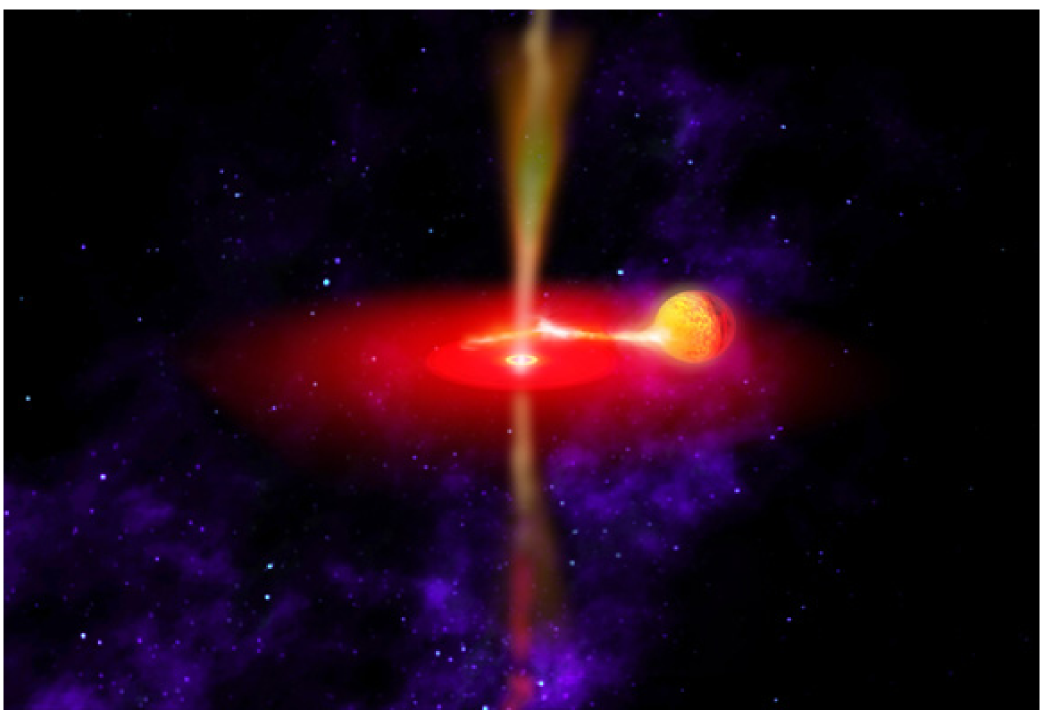


Figure 35: Artist's conception of a black hole accreting matter from a companion star.
Source: © Dana Berry (CfA, NASA).

Careful consideration of the theory of black holes in classical general relativity in the early 1970s led Jacob Bekenstein, Stephen Hawking, and others to a striking set of conclusions. They found that as a chargeless, non-rotating black hole accretes matter, its mass grows by an amount proportional to the strength of gravity at the black hole's surface and the change in its surface area. Also, the black hole's surface area (defined by its event horizon) cannot decrease under any circumstances, and usually increases in time.

At a heuristic level, Bekenstein and Hawking's laws for black holes seem reminiscent of the laws of thermodynamics and statistical mechanics: The change in energy is proportional to the change in entropy and the entropy (a measure of disorder) of a system can only increase. This is no coincidence. The results of general relativity imply what they seem to: A black hole does carry an entropy proportional to its surface area, and, of course, it has an energy that grows with its mass.

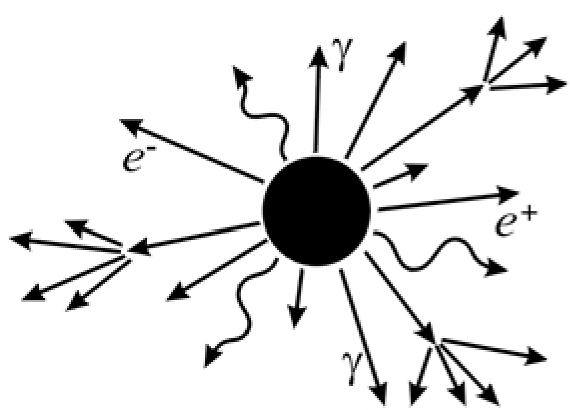


Figure 36: Black holes radiate by a quantum mechanical process.
Source: © Reprinted with permission from Nova Science Publishers, Inc. from: Sabine Hossenfelder, "What Black Holes Can Teach Us," in Focus on Black Hole Research, ed. Paul V. Kreitler (New York: Nova Publishers Inc., 2006), 121-58.

One mystery remains, however. In thermodynamics, the change in energy is proportional to the temperature times the change in entropy; and hot bodies radiate. Even though there is an analogous quantity—the surface gravity—in the black hole mechanics, no classical process can bring radiation through the horizon of a black hole. In a brilliant calculation in 1974, Stephen Hawking showed that, nevertheless, black holes radiate by a *quantum* process. This quantum effect occurs at just the right level to make the analogy between black hole thermodynamics and normal thermodynamics work perfectly.

Hawking's calculation reinforces our belief that a black hole's entropy should be proportional to its surface area. This is a bit confusing because most theories that govern the interactions of matter and force-carrying particles in the absence of gravity posit that entropy grows in proportion to the *volume* of the system. But in a gravity theory also containing these other degrees of freedom, if one tries to fill a region with enough particles so that their entropy exceeds the area bounding the region, one instead finds gravitational collapse into a black hole, whose entropy is proportional to its surface area. This means that at least in gravity theories, our naive idea that the entropy that can be contained in a space should scale with its volume must be incorrect.

Holography, multiple dimensions, and beyond

This concept that in every theory of quantum gravity, the full entropy is proportional only to the area of some suitably chosen boundary or "holographic screen" in the system, and not the full volume, carries a further implication: that we may be able to formulate a theory of gravity in D +1 spacetime dimensions in just D dimensions—but in terms of a purely non-gravitational quantum field theory. Dutch theorist Gerard 't Hooft and his American colleague Leonard Susskind put forward this loose idea, called holography, in the early 1990s. The idea, as stated, is a bit vague. It begs questions such as: *On which "bounding surface" do we try to formulate the physics? Which quantum field theory is used to capture which quantum gravity theory in the "bulk" of the volume?*

In the late 1990s, through the work of Argentine theorist Juan Maldacena and many others, this idea received its first very concrete realization. We will focus on the case of gravity in four dimensions; for different reasons, much of the actual research has been focused on theories of gravity in five dimensions. This work gives, for the first time, a concrete non-perturbative formulation of quantum gravity in the 4D Anti de Sitter spacetime—the simplest solution of Einstein's theory with a negative cosmological constant.

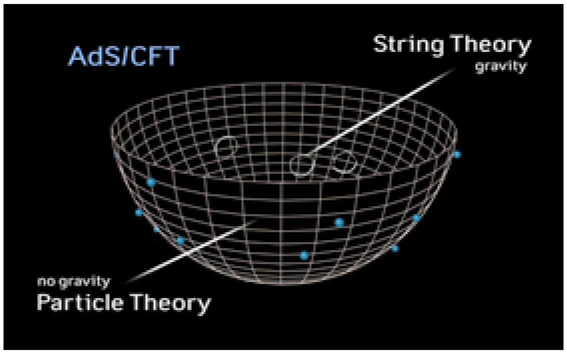


Figure 37: The AdS/CFT duality relates a theory on the boundary of a region to a theory with gravity in the interior.
Source:

It turns out that the symmetries of the 4D space—that is, in 3+1 dimensions—match those of a quantum field theory of a very particular sort, called a "conformal field theory," in 2+1 spacetime dimensions. However, we also expect that quantum gravity in 3+1 dimensions should have the same behavior of its thermodynamic quantities (and in particular, its entropy) as a theory in 2+1 dimensions, without gravity. In fact, these two observations coincide in a beautiful story called the Anti de Sitter/Conformal Field Theory (AdS/CFT) correspondence. Certain classes of quantum field theories in 2+1 dimensions are exactly equivalent to quantum gravity theories in 3+1 dimensional Anti de Sitter space. Physicists say that these two theories are dual to one another.

The precise examples of this duality that we recognize come from string theory. Compactifications of the theory to four dimensions on different compact spaces give rise to different examples of AdS_4 gravity and to different dual field theories. While we do not yet know the gravity dual of every 2+1 dimensional conformal field theory or the field theory dual of every gravity theory in AdS_4, we do have an infinite set of explicit examples derived from string theory.

The value of duality

This duality has a particularly interesting, useful, and, on reflection, necessary aspect. The 2+1 dimensional field theories analogous to electromagnetism have coupling constants g analogous to the electron charge e. The gravity theory also has a natural coupling constant, given by the ratio of the curvature radius of space to the Planck length, which we will call L. In the known examples of the duality between AdS space and quantum field theories, large values of L, for which the gravity theory is weakly curved, and hence involves only weak gravity, correspond to very large values of g for the quantum field theory. Conversely, when the quantum field theory is weakly coupled (at small values of g), the gravity theory has very small L; it is strongly coupled in the sense that quantum gravity corrections (which are very hard to compute, even in string theory) are important.

This kind of duality, between a strongly coupled theory on the one hand and a weakly coupled theory on the other, is actually a common (though remarkable and beautiful) feature in physical theories. The extra shock here is that one of the theories involves quantum gravity in a different dimension.

This duality has had two kinds of uses to date. One obvious use is that it provides a definition of quantum gravity in terms of a normal field theory for certain kinds of gravitational backgrounds. Another use, however, that has so far proved more fruitful, is that it gives a practical way to compute in classes of strongly coupled quantum field theories: You can use their weakly curved gravitational dual and compute the dual quantities in the gravity theory there.

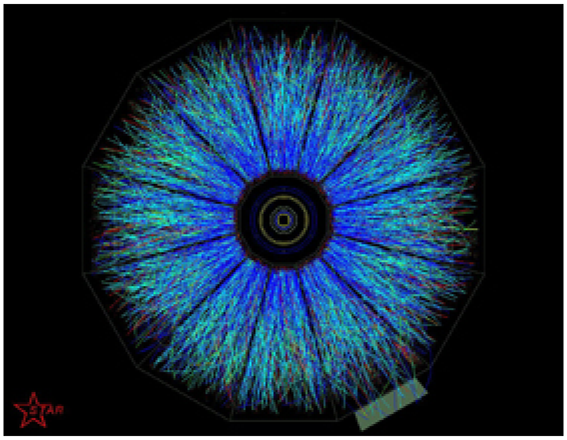


Figure 38: The aftermath of a heavy ion collision at RHIC, the Relativistic Heavy Ion Collider.
Source: © Courtesy of Brookhaven National Laboratory.

Physicists have high hopes that such explorations of very strongly coupled quantum field theories based on gravity duals may provide insight into many of the central problems in strongly coupled quantum field theory; these include a proper understanding of quark confinement in QCD, the ability to compute transport in the strongly coupled QCD plasma created at present-day accelerators like Brookhaven National Laboratory's Relativistic Heavy Ion Collider (RHIC) that smash together heavy ions, the ability to solve for quantities such as conductivity in strongly correlated electron systems in condensed matter physics, and an understanding of numerous zero-temperature quantum phase transitions in such systems. However, we must note that, while this new approach is orthogonal to old ones and promises to shed new light in various toy models of those systems, it has not yet helped to solve any of the central problems in those subjects.

The initial singularity

Even more intriguing is the question of the initial cosmological singularity. In general relativity, one can prove powerful theorems showing that any expanding cosmology (of the sort we inhabit) must have arisen in the distant past from a point of singularity in which the energy density and curvature are very large and the classical theory of relativity is expected to break down. Intuitively, one should just run the cosmological expansion backwards; then the matter we currently see flying apart would grow into an ever-denser and more highly curved state.

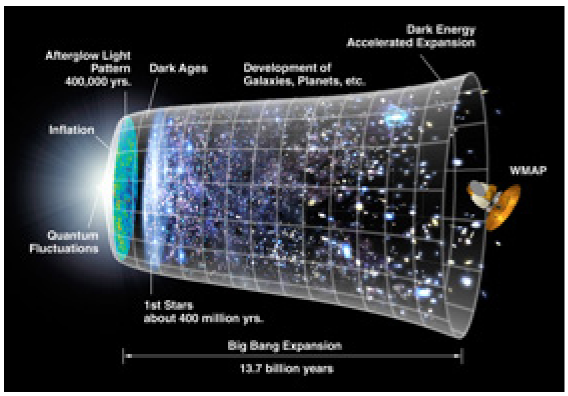


Figure 39: If we run time backwards, the entire universe collapses into a single, infinitely dense point.
Source: © NASA, WMAP Science Team.

Our current picture of the Big Bang, including the successes of Big Bang nucleosynthesis and the CMB, gives us confidence that we can safely extrapolate our current cosmology back to temperatures of order MeV. In most inflationary theories, the Big Bang engendered temperatures far above the TeV scale. But at some still higher energy scale, probing the start of inflation and beyond, we do not know what happened; we do not even have any ideas that provide hints of testable predictions.

What is the origin of our observable universe? Is it one of many, coming from quantum tunneling events out of regions of space with different macroscopic laws of physics? Or is it a unique state, arising from some unknown initial condition of quantum gravity that we have yet to unravel? And, how does this physics eventually deal with the singularity theorems of general relativity, which assure us that extrapolation backwards into the past will lead to a state of high density and curvature, where little can be reliably calculated? These mysteries remain at the forefront of modern research in quantum gravity.

Section 11: Further Reading

- Brian Greene, "The Elegant Universe," Vintage Books, 2000.
- Juan Maldacena, "The Illusion of Gravity," *Scientific American*, November, 2005, p. 56–63.
- Lisa Randall, "Warped Passages: Unraveling Mysteries of the Universe's Hidden Dimensions," *Harper Perennial*, 2006.
- Barton Zwiebach, "A First Course in String Theory," *Cambridge University Press*, 2009.

Glossary

Anti-de Sitter/Conformal Field Theory (AdS/CFT): AdS/CFT is a mathematical relationship between two separate descriptions of the same physics. According to AdS/CFT, a string theory in a region of Anti-de Sitter (AdS) space is equivalent to a conformal field theory (CFT) on the boundary of that region. Anti-de Sitter space has negative curvature (a two-dimensional plane is curved in a saddle shape rather than flat), and is one of the simplest geometries in which the equations of general relativity can be solved. A conformal field theory is the type of field theory used in the Standard Model. Although AdS/CFT describes an artificially simple situation—we appear to live in flat space, not Anti-de Sitter space—the mathematical correspondence between the two descriptions of physics has allowed relatively straightforward field theory calculations to shed light on problems associated with the quantum mechanics of black holes. AdS/CFT has also been used the other way, with black hole calculations providing insight into complicated particle collisions and condensed matter systems that are difficult to understand with the conventional field theory approach.

blackbody: A blackbody is an object that absorbs all incident electromagnetic radiation and re-radiates it after reaching thermal equilibrium. The spectrum of light emitted by a blackbody is smooth and continuous, and depends on the blackbody's temperature. The peak of the spectrum is higher and at a shorter wavelength as the temperature increases.

black hole: A black hole is a region of space where gravity is so strong that nothing can escape its pull. Black holes have been detected through their gravitational influence on nearby stars and through observations of hot gas from surrounding regions accelerating toward them. These black holes are thought to have formed when massive stars reached the end of their cycle of evolution and collapsed under the influence of gravity. If a small volume of space contains enough mass, general relativity predicts that spacetime will become so highly curved that a black hole will form.

brane, p-brane: In string theory, branes are fundamental objects that exist in a specific number of spatial dimensions. The "p" in p-brane stands for the number of dimensions that brane has. For example, a string is a 0-brane, a membrane is a 2-brane, and we could live on a 3-brane.

closed string: In string theory, a closed string forms a loop. Unlike open strings, closed strings are not attached to other objects; however, a closed string can be broken apart to form an open string.

Open strings and closed strings have different properties, and give rise to different sets of fundamental particles.

compactification: In string theory, the term compactification refers to how an extra dimension is made small enough that we cannot perceive it. The three spatial dimensions we are familiar with from daily life are essentially infinite, while compactified dimensions are curled up, and have a finite size that ranges from a few microns (10^{-6} m) down to the Planck length.

cosmic microwave background: The cosmic microwave background (CMB) radiation is electromagnetic radiation left over from when atoms first formed in the early universe, according to our standard model of cosmology. Prior to that time, photons and the fundamental building blocks of matter formed a hot, dense soup, constantly interacting with one another. As the universe expanded and cooled, protons and neutrons formed atomic nuclei, which then combined with electrons to form neutral atoms. At this point, the photons effectively stopped interacting with them. These photons, which have stretched as the universe expanded, form the CMB. First observed by Penzias and Wilson in 1965, the CMB remains the focus of increasingly precise observations intended to provide insight into the composition and evolution of the universe.

cosmic string: A cosmic string is a one-dimensional topological defect stretching across the universe, essentially an extremely thin, extremely dense line in space that would deform spacetime around it according to general relativity. Cosmic strings have been predicted by various theories, but never detected. It is possible that if the period of inflation in the early universe ended in a collision between a brane and an anti-brane, cosmic strings were produced in the process.

cosmological constant: The cosmological constant is a constant term that Einstein originally included in his formulation of general relativity. It has the physical effect of pushing the universe apart. Einstein's intent was to make his equations describe a static universe. After astronomical evidence clearly indicated that the size of the universe is changing, Einstein abandoned the cosmological constant though other astrophysicists, such as Georges Lemaître and Sir Arthur Stanley Eddington, thought it might be the source of cosmic expansion. The cosmological constant is a simple explanation of dark energy consistent with the observations; however, it is not the only possible explanation, and the value of the cosmological constant consistent with observation is over 60 orders of magnitude different from what theory predicts.

event horizon: A black hole's event horizon is the point of no return for matter falling toward the black hole. Once matter enters the event horizon, it is gravitationally bound to the black hole and cannot escape. However, an external observer will not see the matter enter the black hole. Instead, the gravitational redshift due to the black hole's strong gravitational field causes the object to appear to approach the horizon increasingly slowly without ever going beyond it. Within the event horizon, the black hole's gravitational field warps spacetime so much that even light cannot escape.

general relativity: General relativity is the theory Einstein developed to reconcile gravity with special relativity. While special relativity accurately describes the laws of physics in inertial reference frames, it does not describe what happens in an accelerated reference frame or gravitational field. Since acceleration and gravity are important parts of our physical world, Einstein recognized that special relativity was an incomplete description and spent the years between 1905 and 1915 developing general relativity. In general relativity, we inhabit a four-dimensional spacetime with a curvature determined by the distribution of matter and energy in space. General relativity makes unique, testable predictions that have been upheld by experimental measurements, including the precession of Mercury's orbit, gravitational lensing, and gravitational time dilation. Other predictions of general relativity, including gravitational waves, have not yet been verified. While there is no direct experimental evidence that conflicts with general relativity, the accepted view is that general relativity is an approximation to a more fundamental theory of gravity that will unify it with the Standard Model. See: gravitational lensing. gravitational time dilation, gravitational wave, precession, spacetime, special relativity, Standard Model.

gravitational lensing: Gravitational lensing occurs when light travels past a very massive object. According to Einstein's theory of general relativity, mass shapes spacetime and space is curved by massive objects. Light traveling past a massive object follows a "straight" path in the curved space, and is deflected as if it had passed through a lens. Strong gravitational lensing can cause stars to appear as rings as their light travels in a curved path past a massive object along the line of sight. We observe microlensing when an object such as a MACHO moves between the Earth and a star. The gravitational lens associated with the MACHO focuses the star' light, so we observe the star grow brighter then dimmer as the MACHO moves across our line of sight to the star.

graviton: The graviton is the postulated force carrier of the gravitational force in quantum theories of gravity that are analogous to the Standard Model. Gravitons have never been detected, nor is there a viable theory of quantum gravity, so gravitons are not on the same experimental or theoretical footing as the other force carrier particles.

ground state: The ground state of a physical system is the lowest energy state it can occupy. For example, a hydrogen atom is in its ground state when its electron occupies the lowest available energy level.

hierarchy problem: The hierarchy problem in theoretical physics is the fact that there appear to be two distinctly different energy scales in the universe for reasons that are not understood. The first energy scale, called the "electroweak scale," is associated with everything except gravity. The electroweak scale is set by the mass of the W and Z bosons at around 100 GeV, and determines the strength of the strong, electromagnetic, and weak interactions. The second is the Planck scale, at 10^{19} GeV, which is associated with gravitational interactions. Another way of stating the hierarchy problem is to ask why gravity is 39 orders of magnitude weaker than the other fundamental forces of nature.

inflation: Inflation is a period of exponential expansion thought to have occurred around 10^{-36} seconds after the universe began. During this period, which lasted for a few million Planck times, the universe expanded by a factor of at least 10^{25}, smoothing out temperature and density fluctuations to produce the nearly uniform universe we observe today. Although the mechanism driving inflation is still not understood, evidence from the cosmic microwave background supports its existence.

inflaton: The inflaton is a hypothetical scalar field that could drive the period of inflation that took place in the early universe.

nucleosynthesis: The term "nucleosynthesis" refers either to the process of forming atomic nuclei from pre-existing protons and neutrons or to the process of adding nucleons to an existing atomic nucleus to form a heavier element. Nucleosynthesis occurs naturally inside stars and when stars explode as supernovae. In our standard model of cosmology, the first atomic nuclei formed minutes after the Big Bang, in the process termed "Big Bang nucleosynthesis."

open string: In string theory, an open string has two distinct ends. Open strings can have one end attached to another object like a brane, and the two ends of an open string can connect to form a closed string. Open strings and closed strings have different properties, and give rise to different sets of fundamental particles.

brane, p-brane: In string theory, branes are fundamental objects that exist in a specific number of spatial dimensions. The "p" in p-brane stands for the number of dimensions that brane has. For example, a string is a 0-brane, a membrane is a 2-brane, and we could live on a 3-brane.

Planck length: The Planck length is the fundamental unit of length used in high energy physics, and is a combination of Planck's constant, Newton's constant of universal gravitation, and the speed of light. The Planck length is approximately 1.6×10^{-35} m.

Planck mass: The Planck mass is the fundamental unit of mass used in high energy physics, and is a combination of Planck's constant, Newton's constant of universal gravitation, and the speed of light. The Planck mass is approximately 2.2×10^{-8} kg.

Planck time: The Planck time is the time it takes light to travel one Planck length, and is considered the fundamental unit of time in high energy physics. The Planck time is approximately 5.4×10^{-44} seconds.

potential energy: Potential energy is energy stored within a physical system. A mass held above the surface of the Earth has gravitational potential energy, two atoms bound in a molecule have chemical potential energy, and two electric charges separated by some distance have electric potential energy. Potential energy can be converted into other forms of energy. If you release the mass, its gravitational potential energy will be converted into kinetic energy as the mass accelerates downward. In the process, the gravitational force will do work on the mass. The force is proportional to the rate at which the potential energy changes. It is common practice to write physical theories in terms of potential energy, and derive forces and interactions from the potential.

quantum chromodynamics: Quantum chromodynamics, or QCD, is the theory that describes the strong nuclear force. It is a quantum field theory in which quarks interact with one another by exchanging force-carrying particles called "gluons." It has two striking features that distinguish it from the weak and electromagnetic forces. First, the force between two quarks remains constant as the quarks are pulled apart. This explains why single quarks have never been found in nature. Second, quarks and gluons interact very weakly at high energies. QCD is an essential part of the Standard Model and is well tested experimentally; however, calculations in QCD can be very difficult and are often performed using approximations and computer simulations rather than solved directly.

recombination: In the context of cosmology, the term recombination refers to electrons combining with atomic nuclei to form atoms. In our standard model of cosmology, this took place around 390,000 years after the Big Bang. Prior to the time of recombination, the universe was filled with a plasma of electrically charged particles. Afterward, it was full of neutral atoms.

scalar field: A scalar field is a smoothly varying mathematical function that assigns a value to every point in space. An example of a scalar field in classical physics is the gravitational field that describes the gravitational potential of a massive object. In meteorology, the temperature and pressure distributions are scalar fields. In quantum field theory, scalar fields are associated with spin-zero particles. All of the force-carrying particles as well as the Higgs boson are generated by scalar fields.

singularity: Singularity is a mathematical term that refers to a point at which a mathematical object is undefined, either because it is infinite or degenerate. A simple example is the function 1/x. This function has a singularity at x = 0 because the fraction 1/0 is undefined. Another example is the center of a black hole, which has infinite density. In our standard model of cosmology, the universe we live in began as a spacetime singularity with infinite temperature and density.

special relativity: Einstein developed his theory of special relativity in 1905, 10 years before general relativity. Special relativity is predicated on two postulates. First, the speed of light is assumed to be constant in all inertial frames. Second, the laws of physics are assumed to be the same in all inertial frames. An inertial frame, in this context, is defined as a reference frame that is not accelerating or in a gravitational field. Starting from these two postulates, Einstein derived a number of counterintuitive consequences that were later verified by experiment. Among them are time dilation (a moving clock will run slower than a stationary clock), length contraction (a moving ruler will be shorter than a stationary ruler), the equivalence of mass and energy, and that nothing can move faster than the speed of light. See: general relativity, spacetime.

strong interaction: The strong interaction, or strong nuclear force, is one of the four fundamental forces of nature. It acts on quarks, binding them together into mesons. Unlike the other forces, the strong force between two particles remains constant as the distance between them grows, but actually gets weaker when the particles get close enough together. This unique feature ensures that single quarks are not found in nature. True to its name, the strong force is a few orders of magnitude stronger than the electromagnetic and weak interactions, and many orders of magnitude stronger than gravity.

topology: Topology is the mathematical study of what happens to objects when they are stretched, twisted, or deformed. Objects that have the same topology can be morphed into one another smoothly, without any tearing. For example, a donut and a coffee cup have the same topology, while a beach ball is in a different topological category.

winding mode: In string theory, a winding mode is a distinct way in which a string can wrap around a compactified extra dimension. If we imagine a single extra dimension compactified into a circle, the simplest winding mode is for the string to wind once around the circle.