Introduction

In this chapter we continue our study of nuclear physics, discussing biological effects of ionizing radiation, nuclear fission, nuclear fusion, and nuclear reactors.

Next we go further inwards, discussing the physics of "elementary particles," also known as high-energy physics.

Then we'll zoom way out to discuss the universe at large, its origin and evolution: cosmology. As part of this discussion we'll also introduce Einstein's theory of gravity, general relativity.

We'll conclude the course by providing an overview/summary of modern physics, including quantum field theory, the theme of unification, and the essential role of symmetry in modern physics.

Biological Effects of Ionizing Radiation

In every-day language we tend to reserve the word "radiation" to mean alpha-rays, beta-rays, and gamma-rays, which were studied in the previous chapter. Physicists use the word more generally to mean something that spreads out; thus, we use "electromagnetic radiation" to mean electromagnetic waves. The nomenclature is further complicated because gamma-rays fit into both categories.

In an attempt to clarify the nomenclature confusion, the phrase "ionizing radiation" has been coined to distinguish radiation coming from atomic nuclei (alpha, beta, and gamma) from other kinds of radiation. The typical ionization energy of atoms in molecules is on the order of 1 eV or
10 eV, whereas the typical energies of alpha, beta, and gamma radiation are on the order of 1 MeV. Thus, each particle of nuclear radiation can potentially ionize many, many atoms, rattling around like a ball in a pin-ball machine. Nuclear radiation is therefore potentially very damaging if it is incident on a human body.

However, the damage potential of nuclear radiation is difficult to estimate, and depends on many factors. For example, alpha particles are typically stopped after travelling a few centimetres in air, so an alpha-particle emitter that lies elsewhere in the same room may be of no danger. However, some radioactive substances evaporate and float about in the air; if you were to inhale such a substance, then it would decay while in your lungs, and this might be extremely dangerous.

We won't get into quantitative measures of radioactive dose and biological effects; if you're interested in digging into this more deeply, a good next step is to enroll in PHYS 2P02, where this is discussed in detail, along with therapeutic uses of radiation, the workings of medical imaging devices, and much more.

**Induced Nuclear Reactions**

Nuclear radiation, as we learned in the previous chapter, is a natural phenomenon. It's unpredictable, in the sense that it's not possible to know when exactly a particular radioactive nucleus will transform into another one with the consequent emission of radiation; for this reason, the word "spontaneous" is used to describe radioactivity. However, it is possible to know with extreme precision which proportion of a sample of identical radioactive nuclei will transform in a given time period. The half-life of a radioactive species can also therefore be known with a very high precision.

Although radioactivity is a set of natural, spontaneous phenomena, is it possible to control this in any way? That is, is it possible to "cause" a nucleus to transform by disturbing it in some way? The answer is yes, it is
possible to "induce" a **nuclear reaction** under the right circumstances by firing a particle into a nucleus. Such nuclear reactions were often called "nuclear transmutations" in the early days of nuclear research.

The first nuclear reaction produced in a laboratory was due to Ernest Rutherford in 1919; he fired alpha particles into nitrogen nuclei and produced oxygen and ejected protons:

\[
^4_2\text{He} + ^{14}_7\text{N} \rightarrow ^{17}_8\text{O} + ^1_1\text{P}
\]

This was another of Rutherford's revolutionary discoveries, aided by Patrick Blackett, who devoted five years to photographing particle tracks in cloud chambers to support Rutherford's discovery. In 1925, Blackett published the results of his work, which involved analyzing over 400,000 particle tracks in more than 20,000 photographs. Eight (8 !!) of these tracks were forked, which showed that in these cases an alpha particle had momentarily merged with a nitrogen atom to form a fluorine atom, which then almost immediately decayed into an oxygen atom and a proton.

This work was part of Rutherford's larger program of analyzing scattering experiments, and showed that usually scattering of incoming alpha particles resulted in each particle retaining its identity. It was only in a minute fraction of cases that a nuclear reaction occurred.

Induced nuclear reactions have been used to produce chemical elements not found in nature, following a suggestion by Fermi in 1934. Here's an example:
Two important examples of induced nuclear reactions are nuclear fission and nuclear fusion, which we'll discuss next.

**Nuclear Fission**

Otto Hahn and Lise Meitner worked together studying radioactivity since 1912, and they were joined in 1929 in Berlin by Fritz Strassman. The discovery of the neutron in 1932 gave them another tool in their study; they could now fire neutrons at targets and study the scattering patterns and the products of any transformations produced. They were expecting to produce small quantities of elements close in atomic number to the target element. By working with uranium as the target nucleus, they were hoping to produce trans-uranic elements; that is, they were hoping to discover elements heavier than uranium that had never been produced before.

By 1938, the situation in Germany had become so dangerous for Jews that Meitner, who was Jewish, had to flee Germany. She stayed in Germany far longer than was wise; German officials were alert that she...
might leave and were watching the borders for her. In the end she managed to make it to the Dutch border and into the Netherlands only thanks to the help of two Dutch physicists, Dirk Coster and Adriaan Fokker. She travelled under cover, and Coster was somehow able to convince German border guards that she did indeed have permission to leave Germany and enter the Netherlands. Meitner had no possessions and no job, and a possible job in the Netherlands fell through. She eventually found work in Stockholm, and continued research there in collaboration with her nephew Otto Frisch, who was working at the Bohr institute in Denmark at the time.

Meanwhile, Hahn and Strassmann reported that their experiments of firing neutrons at uranium had produced barium, an element with much smaller atomic mass. This was very surprising, and an important discovery. Evidently firing neutrons at uranium could induce nuclear fission; that is, uranium nuclei were splitting into several large fragments. They immediately communicated this discovery to Meitner, and she and Frisch worked out the theory of the reaction based on Bohr's "liquid-drop" model of the nucleus.

Several different fission reactions are possible when neutrons are fired at uranium-235 nuclei. Here's one possibility:
This fission reaction is net-exothermic, in the sense that the products carry a lot more kinetic energy than the reactants had. Thus, once you supply a little bit of energy in the form of the incoming neutrons, you get a lot more back out. The initial nuclear and electrostatic potential energy decreases when the fragments are produced; the difference is converted to kinetic energy in the fragments. About 200 MeV is produced per fission, which is about 100 billion times the amount of energy produced per molecular reaction in a normal combustion reaction (such as the chemical burning of a fuel); nuclear fission is therefore a potentially very valuable process for "energy production," and indeed we use this process in nuclear reactors to produce electricity, as we'll describe in the next section.

Strassmann and his wife concealed a Jewish friend in their home for months during WW II, risking their own lives to do so. Strassman had resigned in 1933 from the Society of German Chemists when it came under Nazi control, and felt extremely lucky to find a job with Hahn and Meitner, who were able to find a position for him at half-pay.

The discovery of nuclear fission made nuclear bombs possible, and although he never did any research on bomb production, Hahn felt guilty
that his discovery eventually led to the deaths of hundreds of thousands of people when the United States dropped nuclear bombs on Japan in 1945. Although Hahn was on the brink of suicide, he was helped through his despair by friends, including Max von Laue, and after the end of WW II, he devoted an enormous amount of time to the advancement of science and social responsibility. Among the many, many awards he received in his lifetime, it's notable that he was inducted as an Honorary Officer of the Order of the British Empire in 1957 by his country's wartime enemy.

The publications of Hahn and Strassmann, and of Meitner and Frisch, both in 1939, electrified the scientific community, both for their intrinsic scientific interest, and also for their possible dangerous consequences. It was quickly realized that nuclear fission might be used to create bombs, and there was alarm that Hitler's Germany might be working on a nuclear bomb. Einstein was the most famous scientist in the world at this time, and he had escaped Germany in 1933, settling in Princeton, New Jersey. American scientists Szilard, Teller, and Wigner persuaded Einstein to write to the U.S. president of the time, F.D. Roosevelt, warning him of the potential danger and urging him to take action. As a result, the American "Manhattan Project" was initiated, and nuclear weapons were engineered in short order after an enormous effort from an enormous number of people.

Science is a double-edged sword. On one hand, it is a search for truth and understanding of the beautiful world we live in. However, the new understandings we obtain ought to be put to good use, for the benefit of all, not their detriment, but this does not always happen.

**Nuclear Reactors**

It was known since the time of Hahn, Meitner, and Strassmann that neutrons need to be slow to effectively induce nuclear fission. Fast neutrons don't do the job nearly as well, because the probability of their capture by a nucleus is much less.
Thermal neutrons have kinetic energies of about 0.04 eV or less, and they are effective at inducing fission. For uranium-235, thermal neutrons are about 500 times more effective than fast (kinetic energy of about 1 MeV) neutrons at inducing fission.

Referring to the previous figure of a fission reaction, each uranium-235 fission of this type "produces" 3 neutrons. (That is, some of the neutrons that were "buried" inside a uranium nucleus are freed and escape, instead of remaining buried inside one of the major fragments.) Another major type of uranium-235 fission reaction produces 2 neutrons, and they are about equally likely, so one could say that the average uranium-235 fission produces 2.5 neutrons. In any case, each uranium-235 fission produces more free neutrons than were present before the fission.

This raises the possibility of a "chain reaction;" what if each of the 2 or 3 neutrons produced by a fission goes on to produce another fission, each of which releases 2 or 3 more neutrons, and so on. The number of neutrons flying around rapidly increases, and so the fission rate might also rapidly increase. Each fission is exothermic, so the energy production might also increase rapidly. Such a rapid increase in energy production amounts to a bomb, and this is indeed how nuclear fission bombs work.
If you can figure out a way to safely absorb some of the neutrons, then you might be able to control nuclear fission so that you don't get an explosive amount of energy released, but rather get a useful amount of energy released gradually. Fermi led the team that designed and built the first nuclear reactor in Chicago; they ran the first controlled nuclear fission reaction there in December 1942. (Fermi won the 1938 Nobel Prize for physics, went to Stockholm with his family to receive the prize, and then went straight to the U.S. and applied for immigration; his wife was Jewish, and would be in danger in his native Italy, which was allied with Nazi Germany at the time.)
Fermi used graphite bricks as neutron absorbers; graphite also slows neutrons, and that is the other problem in creating a sustained nuclear chain reaction. Each fission produces fast neutrons, and they must be slowed down to make them effective at producing additional fissions. Graphite is good at slowing down neutrons (it's a neutron "moderator"), and so Fermi's design was successful.

Modern nuclear reactors use a variety of different types of neutron absorbers (such as cadmium and indium), and various types of moderators (water, graphite, and "heavy water"). The following diagrams give schematic representations of a nuclear power reactor.
Nuclear power has its positive and negative aspects. Some of the positive aspects include safety, sustainability, and low carbon emissions when calculated over the lifetime of a reactor. Some of the negative aspects are that serious accidents are potentially very dangerous (witness Chernobyl in 1986 and Fukushima in 2011), building nuclear power stations is extremely expensive, the problem of nuclear waste disposal
has not really been solved yet, and there is the danger of further proliferation of nuclear weapons.

Canada uses CANDU nuclear reactors, which are moderated by heavy water. Heavy water (deuterium dioxide) absorbs fewer neutrons than ordinary water, which allows the use of unenriched uranium fuel. This is cheaper (no enrichment plants needed), and safer, because unenriched fuel is unusable "as is" in a bomb. However, a negative aspect of heavy water moderators is that it requires a greater thickness of heavy water to effectively moderate neutrons. Another problem with heavy water moderators is the tritium problem.

**Nuclear Fusion**

The primary way that stars produce light is nuclear fusion. Nuclear fission is a high-atomic-mass nucleus splitting up into two or more fairly large fragments and a few other bits. Nuclear fusion is the opposite, the joining up of two (or more) low-atomic-mass nuclei to form a larger nucleus.

Recall the binding energy curve for atomic nuclei:
By sketching horizontal and vertical lines on the graph above, you should be able to convince yourself that the fission of a uranium nucleus releases about 1 MeV of energy per nucleon, whereas the fusion of deuterium into helium releases about 6 MeV per nucleon. It seems that one could get more "bang for the buck" from fusion than from fission. This is verified by experiments.

Nuclear fusion is the reason that stars shine, but conditions in the core of a star (with temperatures of 10 million K and up, and correspondingly enormous pressures) are so extreme that it's difficult to imagine how one could reproduce them in a laboratory. Nevertheless, researchers have been trying since at least the early 1950s to develop practical nuclear fusion reactors. (We've had nuclear fusion bombs since the early 1950s as well, but no reactors yet.) The problems in controlling nuclear fusion reactors are mainly related to how to confine the fuel. You need a very hot gas at high pressure to induce fusion, but what kind of material can contain such a substance at such extreme conditions? The two main ideas being pursued right now are confinement using magnetic fields (the fusion reactants are first charged, to form what is called a plasma, and then one can imagine using a magnetic field to act on the moving charged particles), and using lasers. So far, no luck. Do you have any good ideas?

The idea of using fusion reactors instead of fission reactors to generate electricity is that the reaction products from fusing hydrogen are helium, which is absolutely benign. Fission reactors produce highly radioactive, and therefore highly dangerous, reaction products, and what to do with them is a nightmarish problem.

Our basic idea for producing electricity, which is based on Faraday's law of induction, is to turn a coil of wire within a magnetic field. The turning typically comes from boiling water to create steam, and then directing jets of steam at high pressure against a turbine, on which are attached the coils of wire. Is there a way of using the products of fusion or fission reactors directly, without going to the intermediary of boiling water? In a
nuclear fission reactor, is there a way of using neutrons directly to produce power? Nobody has thought of a way of doing so yet.

Elementary Particles (High Energy Physics)

Continuing inwards, one notes that atoms are made up of protons, neutrons, and electrons. And what about the latter particles; are they made up of still smaller particles? If so, are the constituents made up of even smaller particles? Where does it all end? Is there any such thing as a truly "elementary" particle? These notes summarize very briefly what we know nowadays, and how we came to know it.

Particle accelerators through time

There are (historically) four main sources of elementary particles used for study:

- Cosmic rays
- Radioactive elements
- Nuclear reactors
- Particle accelerators

Cosmic rays are high-energy particles that hit Earth from sources outside it. About 90% of them are protons, most of the rest are alpha particles, and the rest are electrons and other subatomic particles, and occasionally include other atomic nuclei. (The word "ray" is usually used for some kind of electromagnetic radiation, which cosmic rays are not, but once a word is used for long enough it's very hard to change it even if it's incorrect.)

Who knows where cosmic rays come from? It is thought that they originate in the supernovae explosions of giant stars and perhaps also in active galactic nuclei, as they are the only known ways to produce such
extremely energetic particles. Cosmic rays should not be confused with the solar wind, which are energetic particles that "boil" off the Sun. Although they are energetic, solar wind particles typically have energies that are much lower than typical cosmic ray energies.

The peak of the energy distribution of cosmic rays occurs at about 300 MeV, but some rare cosmic rays have energies up to about 300 million TeV, which is about 40 million times the highest energies achieved in particle accelerators on Earth! When cosmic rays smash into air molecules in the upper atmosphere, they create showers of other particles, which can then be detected and studied.

In early studies of elementary particles (starting about 100 years ago), radioactive substances were used as a source of particles that could then be directed at targets. When nuclear reactors began to be developed in the 1950s, one could produce a greater variety of beams of particles.

Using radioactive elements or nuclear reactors to produce beams of particles, one did not have control over the energies of the particles. Neither did one have control over the energies of cosmic rays; one just waited for the cosmic rays to strike the atmosphere, and tried to record the resulting reactions. With the development of particle accelerators, one began to have control over the energies of the particles produced.

Cockroft-Walton generator (1932):
Ernest O. Lawrence's first cyclotron was built in 1931, based on an early linear accelerator of Rolf Wideroe, and was about 4 inches in diameter! The advantage of Lawrence's design is that it could fit in his small laboratory, and did not require very high voltages, which were difficult to work with.
Cyclotron (Ernest O. Lawrence, 1932; the "Dees" are 27 inches in diameter):

Lawrence's cyclotron at UC Berkeley, 1939 (with 60-inch Dees):
The earliest cyclotrons could be held in a person's hands; currently the world's biggest cyclotron has Dees 18 m wide, is located at TRIUMF in Vancouver, BC, and can produce 500 MeV particle beams.

Linear accelerators: The largest in the world is currently at SLAC in California, is 3.2 km long, and can produce electrons and positrons with energies of up to 50 GeV. Accelerating light particles in circular accelerators is problematic because they emit so much electromagnetic radiation that they decelerate significantly while on a circular track.

Here's part of the SLAC detector room:
And here's part of the SLAC beamline:

The DESY complex in Germany was opened in 1964, and was at that time the most powerful particle accelerator in the world, with beam energies of about 7 GeV.

Fermilab, near Chicago, had two circular tunnels, each of circumference 6.3 km, and a maximum collision energy of nearly 2 TeV:
CERN, Switzerland: The Large Hadron Collider (LHC) at CERN is tunnel 100 m below ground and has a radius of 27 km, the longest particle collider in the world. The projected collision energy at maximum is 14 TeV.

The biggest discovery so far at the LHC is the Higgs boson in 2012/2013.
CMS detector at the LHC:
Part of the ATLAS detector at the LHC:
Some particle accelerators produce information in the form of photographs of the tracks that particles make, either in bubble chambers, or cloud chambers, or in some other ways. The detectors are often placed in magnetic fields, so that charged particles leave curved tracks, and neutral particles leave straight tracks. Here are some examples:
Knowing which particles were fired at the target, and knowing the composition of the target, the task is to analyze the spray of particles whose paths are recorded in the photograph and determine what their properties are (mass, charge, etc.). It's a complex task.

One of the side effects of the fact that particle accelerators became larger and larger, and therefore involved larger and larger teams from many countries, is the development of the World Wide Web. Tim Berners-Lee of CERN wrote a proposal in 1989 to facilitate communications amongst the growing teams of scientists and engineers, based on a number of software projects that he had been developing in the previous decade. Berners-Lee and Robert Cailliau suggested in 1990 that the project be expanded to facilitate communication more widely, and in December of 1990 Berners-Lee completed the world's very first web page and first web browser. Berners-Lee and Cailliau publicized their ideas and work, and their movement quickly grew.

**Brief history of some discoveries in particle physics**

Atoms were first postulated as far back as the times of the ancient Greek philosophers, with Democritus one of the notable proponents. They thought of atoms as indivisible, as did scientists relatively recently in more modern times. The chemist John Dalton thought of atoms as indivisible in the early 1800s, and made a lot of progress understanding chemical reactions on this basis.

However, by the end of the 1800s, those who believed in atoms (and there was not widespread belief) began to realize that atoms were not indivisible, because they were composed of positive and negative parts. Emil Wiechert, William Kaufmann, and J.J. Thomson did experiments on cathode rays in 1897, and following a series of such experiments it was
clear that a new particle had been identified: the electron. It was clear that electrons were found inside atoms, and so it was also clear that atoms are not indivisible; they had to have positive parts as well as the negative electrons.

The Geiger-Marsden experiment, and Rutherford's analysis of the results, made it clear that the positive parts of the atom were concentrated in a minute volume at its centre; the atomic nucleus had been discovered. By 1920, Rutherford named the lightest known nucleus, that of hydrogen, the proton, and by 1932 James Chadwick (one of Rutherford's students) had discovered the neutron.

Meanwhile, Einstein had proposed (based on earlier work by Planck) that electromagnetic radiation exists as particles, which were later called photons, and which he used to neatly explain the photoelectric effect. Although Einstein's photon hypothesis was met with strong opposition in its first two decades, experimental evidence by Compton in 1923 provided strong support for the photon hypothesis.

So at this time, we had a fairly simple list of fundamental particles: proton, neutron, electron, photon, and that is all. All matter and radiation was thought to be composed of these four particles and nothing else.

However, there were already puzzles that challenged this simple picture. In studies of beta decay, it was found that ejected electrons did not carry away all of the energy that was expected. Some energy was missing. Meitner assumed that the energy was lost when the ejected electrons scattered off other electrons in nearby atoms. This was disputed by Charles Ellis and William Wooster, who performed beta-decay experiments inside a calorimeter, so that they could measure the temperature increases as a result of the decay. If Meitner was right about the energy being transferred by electrons scattering from atoms, the temperature should increase by a certain definite amount. Ellis and Wooster didn't find the expected temperature increase, and so they
concluded that something was amiss; there was still missing energy unaccounted for.

Meitner did not agree with the results of Ellis and Wooster, although she agreed that their experimental setup was clever. So she engaged the help of Wilhelm Orthmann, who worked in the lab of Walther Nernst, the world's foremost expert in calorimetry, and they repeated the experiment of Ellis and Wooster, but with greater precision. Their results agreed with those of Ellis and Wooster, so it was now clear that the missing energy was a puzzle.

As it became possible to measure the total spins of nuclei, it also became clear that there were anomalies in beta-decay. The inferred spins of nuclei undergoing beta-decay didn't match up with their spins inferred from other experiments. There was also "missing spin."

This led Wolfgang Pauli to propose the existence of the neutrino in 1930. With this hypothesis, Pauli was able to resolve both the missing energy problem and the missing spin problem simultaneously, provided that the neutrino was very light and had spin-1/2, like the electron and the nucleons.

There was some confusion about this in the following years when Chadwick discovered the neutron in 1932. Fermi clarified this by emphasizing that Pauli's neutral particle ought to be called neutrino, not neutron, the suffix "-ino" in Italian meaning "little one."

The neutrino interacts hardly at all with ordinary matter, and so it eluded detection for some time. Experiments by Cowan and Reines in 1956 finally provided convincing evidence of the neutrino's existence. Because the probability of interaction of a neutrino with matter is so small, one needs a huge flux of neutrinos and one needs to patiently wait for a long time before one can collect sufficient data to be convincing. With the development of nuclear reactors, it became possible in the 1950s to produce intense beams of neutrinos that could then be studied.
Mesons

Meanwhile, more particles continued to be discovered. The next characters in our story are a family of particles called mesons.

In the late 1920s and 1930s, the first steps in the development of quantum field theories were being made, and the paradigm of a quantum field theory is that each of the four fundamental forces is transmitted by an exchange of "messenger particles." For the electromagnetic field the messenger particle is the photon; what is the messenger particle for the strong force?

In 1934 Yukawa created an early theory of the strong force, and used it to calculate the mass of the messenger particle, and came up with a value that is about 300 times the electron mass; because its mass was intermediate between the electron mass and the mass of a nucleon, Yukawa called it a meson.

In 1937 two separate teams, Anderson and Neddermeyer, and Street and Stevenson, identified a new particle by detecting the decay products of cosmic rays smashing into the atmosphere, and Oppenheimer suggested that this newly discovered particle was Yukawa's postulated meson.

However, the newly discovered particle didn't seem to interact very much with ordinary matter, which was puzzling. (It's supposed to transmit the strong force, so it ought to interact very readily with protons and neutrons in ordinary matter.) Further studies showed up discrepancies in the mass and lifetime of these particles. The puzzle was resolved in 1947 by Powell's group, who determined that what was thought to be one type of particle was actually two different species of particle, which became known as pions and muons. (Marshak and Bethe, and others, had suggested as much based on theoretical arguments.) The pion has a much shorter lifetime than the muon, and is now known to be made up of quarks, whereas the muon is a lepton, and so not made up of
quarks, and therefore not a meson. (Nowadays particles are not classified by mass, but in other ways. Baryons and mesons are made up of quarks, whereas leptons are not.) Powell's group took some of their detectors to mountain-tops, and so were able to detect pions before they decayed. (The muons, being longer-lived, were surviving to reach sea level in much greater numbers.)

The discovery of mesons was followed by the discovery of many other particles. As particle accelerators became more powerful, more particles were being discovered.

Anti-Particles

A notable development on the theoretical side was Dirac's theory of the electron, which he published in 1927. The history of the Dirac equation has parallels with the history of the Schrödinger equation. Schrödinger first tried to "guess" an equation that was consistent with special relativity, but his attempts were beset with a seemingly fatal problem, the existence of a series of negative-energy states that had no lowest-energy member. Because physical systems naturally move into states of lowest energy, the existence of such a series is problematic, because a particle could continue to move to lower and lower energy levels, all the while emitting energy, so that an infinite amount of energy could be emitted. Schrödinger couldn't immediately see a way around this problem, so he gave up this line of inquiry, and rather attempted to guess an equation based on Newton's second law of motion; in this he was successful, and the Schrödinger equation was born. (Later, Klein and Gordon re-invented Schrödinger's relativistic attempt, which is now known as the Klein-Gordon equation.)

Dirac attempted to guess a quantum equation that would apply to a particle with spin, such as an electron, and that is consistent with special relativity. With great ingenuity, Dirac succeeded, but the Dirac equation has the same problem that Schrödinger's original relativistic equation had: the existence of an infinite number of negative-energy states, that
had no lowest energy. However, rather than give up on this attempt, because the equation was so beautiful, Dirac tried a work-around. He said, OK, we've got this infinite number of negative energy states; what if they are all occupied with electrons? In this case, there is no problem with a particle cascading down indefinitely and then giving off an infinite amount of energy. There is nowhere to cascade down to, because all negative-energy states are already occupied with electrons.

This collection of electrons, all in negative-energy states, became known as the "Dirac sea." Because these hypothetical electrons were uniform, homogeneous, they were considered unobservable, and would have no observable effects on any real particles. However, Dirac explained that if a photon should be absorbed by one of these negative-energy electrons, it could be promoted to a positive-energy state; a sort of "unreal" electron could somehow become real. This would leave a "hole" in the Dirac sea of electrons. Dirac carefully (and mathematically) analyzed the properties of such a hole, and concluded that it would have all the properties of a real, positively-charged particle. In fact, he initially hoped that he could explain the existence of protons in this way, that a proton is nothing more than a hole in a Dirac sea of electrons. Alas, this proved impossible, because the effective mass of a hole in the Dirac sea of electrons would have to be the same as the mass of an electron.

This seemed (to some) to be the death of Dirac's beautiful electron theory, but in 1931 Carl Anderson discovered positively-charged particles that had the same mass as electrons! These were the holes in Dirac's sea!

Nowadays we don't speak of the Dirac sea because there is no need to do so. The negative-energy states in the Dirac equation can be interpreted as positive-energy states for an associated type of particle: the anti-particle of the electron. The anti-electron is also called the positron. Far from being a problem, the negative-energy states of the Dirac equation are a prediction of an anti-particle for the electron.

With the work of Ernst Stueckelberg and Richard Feynman in the 1940s,
It was realized that this phenomenon is a generic for all quantum field theories; thus, for each fundamental particle, there is an associated anti-particle. Besides the discovery of positrons by Anderson in 1931, anti-protons were discovered by Emilio Segre and Owen Chamberlain in 1955 and anti-neutrons were discovered by Bruce Cork in 1956.

It turns out that some particles are their own anti-particles; photons, for example are their own anti-particle. In general, particles and anti-particles share many of the same properties, but have some of their quantum numbers reversed. For example, the neutron and anti-neutron are both neutral, but neutrons have baryon number +1 and anti-neutrons have baryon number -1. (Also, although the neutron is neutral overall, its charge distribution is positive near the centre and towards the surface, but negative in between, whereas the anti-neutron has the opposite charge distribution.)

Can you make atoms out of anti-particles? You'd have to be clever to isolate such anti-atoms for a long enough time to study them if you wished to determine their properties. Atoms consisting of an anti-proton and an anti-electron (anti-hydrogen) have been produced and studied for about the past 20 years at CERN. Lately, they've been getting good at decelerating anti-hydrogen atoms, trapping them, cooling them, and studying them. Any differences between the properties of anti-hydrogen and normal hydrogen might provide deeper insight into elementary particles and the structure of atoms.

**Strange particles**

Starting in 1947, a great quantity of strange subatomic particles began to be discovered. The rate of discovery increased greatly following the construction of the Brookhaven Cosmotron in 1952, the first particle accelerator to operate in the GeV range (individual protons were accelerated up to 3 GeV).

As more and more particles were discovered, trying to piece them all
together into a coherent picture became more and more difficult. However, imaginative physicists (both theoretical and experimental physicists) managed to create a beautiful theory that makes sense out of the bewildering proliferation of particles.

One of the ways that physicists made sense of the enormous flux of new particle data was to notice that certain pairs of particles seemed to be created together, but other pairs seemed never to be created together in certain particle reactions. This led physicists to invent new quantum numbers to describe this; if they postulated that the new quantum number satisfied a conservation law, then this explained why certain pairs were never seen. It was quite a difficult guessing game, which explains why only a certain few were able to play the game successfully.

For example, in the 1950s, Murray Gell-Mann and Kazuhiko Nishijima independently invented the concept of strangeness (this word was used because so many of the new particles were very puzzling), and postulated that it is a conserved quantity in all interactions involving the strong force.

This, and other developments, helped to bring some order to the growing particle menagerie, but particle physics was still in an wild, unruly state. It's reminiscent of the situation in atomic physics in the mid-1800s, where there was a bewildering array of chemical elements, and seemingly no way to understand the mess. Mendeleev brought order to a messy situation by creating the periodic table of the chemical elements, and by bringing order he noticed "missing" elements, and was able to predict their properties based on regularities in the table. Once these "missing" elements were actually discovered, this brought credibility to the table, and strong support for Mendeleev's train of thought. However, there was still no underlying reasons for the ordering of the elements; the physical explanation behind the table came later, with the development of quantum mechanics starting in 1926.

**Symmetry**
Picking up another thread of this beautiful tapestry, the theory of groups and group representations had been developed by mathematicians, starting with work by Lagrange and Vandermonde, both in 1770. They studied groups of permutations towards developing a theory of polynomial equations and their solutions. This part of the story was brought to a dramatic high point in 1829 when Galois developed the theory of groups to a sufficient extent that he was able to prove that there can be no general formula for the solution of a polynomial equation of degree 5 or higher ("the insolvability of the quintic").

Towards the latter part of the 1800s, various mathematicians began to develop group theory applied to geometrical transformations, stimulated by the foundational work of Felix Klein in 1872 and Sophus Lie in 1884. By the 20th century, it had become clear that the theory of group representations is an important and fundamental branch of mathematics, and by now great swaths of mathematics can be understood in terms of group representation theory.

Hermann Weyl made important developments to group representation theory in the 1920s, and he was one of the pioneers (along with Eugene Wigner and John von Neumann) in applying group representation theory to the newly developed quantum mechanics in the late 1920s.

It turns out that groups and representations of groups are a powerful way to describe symmetry, and symmetry is intimately linked to conservation principles in physics (look up "Noether's theorem" to learn more), and since conservation principles are so fundamental in physics, it's then no surprise that symmetry (and its related mathematics, which is representation theory) is so important in fundamental physics.

For an introduction into some of the mathematics of symmetry, see the appendix at the end of these notes.

**Representation theory applied to particle physics**
In 1961, Gell-Mann (and, independently, Yuval Ne'eman) discovered that eight of the known spin-1/2 baryons and eight of the known spin-0 mesons could be placed in similar hexagonal diagrams, which became known as the "eight-fold way." The first figure shows spin-1/2 baryons:

The next figure shows spin-0 mesons:
The next figure shows spin-3/2 baryons:

Just as Mendeleev did, Gell-Mann noticed missing particles in some of his
diagrams, and so he predicted the existence of particles with certain very specific properties (lifetime, mass, charge, and other quantum numbers). For example, in 1962 Gell-Mann predicted the existence of the \( \Omega^- \) particle at the lower vertex of the previous diagram; in 1964, a group at Brookhaven led by Nicholas Samios discovered the particle and confirmed that its properties were just as Gell-Mann predicted (to within experimental uncertainty). The discovery provided solid support for Gell-Mann's classification scheme, and every other subsequent particle discovery fit neatly into the some geometrical pattern in Gell-Mann's scheme.

Gell-Mann and others were quite successfully at extracting insightful patterns from the enormous and rapidly growing amount of particle data that was available at that time. But is there any deeper understanding available for this enormous mess of particles?

In 1964, Gell-Mann (and, independently, George Zweig) proposed that all hadrons were composed of three fundamental particles, which Gell-Mann called quarks. He labelled the three quarks u, d, and s, for "up quark," "down quark," and "strange quark." Gell-Mann got the whimsical name for the new particles from the James Joyce novel, *Finnegan's Wake*. Nowadays we know that there are six fundamental quarks.

### Quark flavor properties

<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol</th>
<th>Mass (MeV/c^2)</th>
<th>J</th>
<th>B</th>
<th>Q</th>
<th>I^3</th>
<th>C</th>
<th>S</th>
<th>T</th>
<th>B'</th>
<th>Antiparticle symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First generation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Up</td>
<td>u</td>
<td>1.7 to 3.1</td>
<td>( \frac{1}{2} )</td>
<td>+( \frac{1}{2} )</td>
<td>+( \frac{3}{2} )</td>
<td>+( \frac{3}{2} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Antiup ( \bar{u} )</td>
</tr>
<tr>
<td>Down</td>
<td>d</td>
<td>4.1 to 5.7</td>
<td>( \frac{1}{2} )</td>
<td>+( \frac{1}{2} )</td>
<td>+( \frac{3}{2} )</td>
<td>-( \frac{1}{2} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Antidown ( \bar{d} )</td>
</tr>
<tr>
<td><strong>Second generation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charm</td>
<td>c</td>
<td>1.290 ±0.50</td>
<td>( \frac{1}{2} )</td>
<td>+( \frac{1}{2} )</td>
<td>+( \frac{3}{2} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Anticharm ( \bar{c} )</td>
<td></td>
</tr>
<tr>
<td>Strange</td>
<td>s</td>
<td>1.00 ±0.30</td>
<td>( \frac{1}{2} )</td>
<td>+( \frac{1}{2} )</td>
<td>-( \frac{1}{2} )</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>Antistrange ( \bar{s} )</td>
<td></td>
</tr>
<tr>
<td><strong>Third generation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top</td>
<td>t</td>
<td>172 900 ±600 ±900</td>
<td>( \frac{1}{2} )</td>
<td>+( \frac{1}{2} )</td>
<td>+( \frac{3}{2} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Antitop ( \bar{t} )</td>
<td></td>
</tr>
<tr>
<td>Bottom</td>
<td>b</td>
<td>4 190 ±180/0</td>
<td>( \frac{1}{2} )</td>
<td>+( \frac{1}{2} )</td>
<td>+( \frac{3}{2} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>Antibottom ( \bar{b} )</td>
<td></td>
</tr>
</tbody>
</table>

Gell-Mann and Zweig's original three quarks fit neatly into their own "eightfold-way" diagram:

\[
\begin{align*}
S &= 0 & Q &= \frac{2}{3} \\
S &= -1 & Q &= -\frac{1}{3}
\end{align*}
\]

The three corresponding anti-quarks fit into their own diagram:

\[
\begin{align*}
S &= 1 & \bar{Q} &= -\frac{2}{3} \\
S &= 0 & \bar{Q} &= \frac{1}{3}
\end{align*}
\]

The search was on for quarks! However, nobody has yet been able to isolate quarks so that they can be studied independently (although in the past few years RHIC at Brookhaven claims to have produced a quark-gluon plasma, so stay tuned for further developments), and this was cause for severe doubts about the quark model in the late 1960s and early 1970s. However, there were already experimental results hinting at structure within protons and neutrons. In 1967, Richard Taylor, Jerome Friedman, Henry Kendall and their group provided excellent evidence for quarks within protons and neutrons by firing high-energy electrons at them and analyzing the paths of the scattered electrons, in much the
same way as Rutherford analyzed the scattering of alpha-particles from gold nuclei. Their results were clear: There are quite definitely three very small particles residing within protons and neutrons. Follow-up experiments since then have strengthened and extended these conclusions, so that quarks and the quark model are firmly established.

All you linear algebra fans might like to read a little more about the mathematics behind the quark model, so before we push on I'll say a bit about this right now. The basic idea dates back to 1932, when the neutron was discovered. Heisenberg proposed that the proton and neutron, which have nearly the same mass, are not distinct particles at all, but merely different states of the same particle, which he called the nucleon.

This may seem strange at first, because we're used to thinking of the neutron and the proton as different particles. However, think about a hydrogen atom. If the electron in the atom gets bumped up to a higher energy state, we wouldn't think of this as a different type of atom, would we? No, we would still call it a hydrogen atom, just in a higher energy state. This is Heisenberg's basic idea in thinking of the proton and neutron as different states of the same particle.

There is more to Heisenberg's idea. He treated the two basic states of the nucleon as basis states in the linear algebra sense; that is, he defined an abstract two-dimensional complex vector space, and thought of the proton and neutron as the simplest basis vectors for this space. Furthermore, he defined a set of linear transformations on this space, one of which transformed a proton to a neutron, and another of which did the opposite.

To satisfy the general properties of quantum mechanics, these transformations had to be members of the Lie group SU(2), which can be thought of as the group of 2 by 2 matrices with complex number entries that have determinant equal to 1 and for which the inverse of each matrix is equal to the transpose of its complex conjugate. Furthermore,
the Schrodinger equation describing the strong force acting on a nucleon is invariant with respect to any of these SU(2) transformations, so we say that a system of nucleons described in this way has an SU(2) symmetry.

Gell-Mann used the same idea to describe the symmetry of quarks, only more operators were necessary because there were three quarks. Each quark can be considered to be a basis vector in a certain abstract complex vector space, and then one can define operators that turn each quark into the other. This is done by first defining a set of "creation operators" and "annihilation operators" of various types. For example:

- \( U^\dagger \) creates an up-quark and \( U \) annihilates an up-quark
- \( D^\dagger \) creates a down-quark and \( D \) annihilates a down-quark
- \( S^\dagger \) creates a strange-quark and \( S \) annihilates a strange-quark

Thus, the operator \( U^\dagger D \) transforms a down-quark into an up-quark, because when it operates on the basis vector \( d \), the down-quark is annihilated and then an up-quark is created. In symbols, this can be expressed as:

\[
(U^\dagger D)|d\rangle = U^\dagger (D|d\rangle) = U^\dagger |0\rangle = |u\rangle
\]

The vector labels are enclosed by a strange looking half-bracket symbol (physics convention for writing a vector in quantum mechanics, called "Dirac notation"), and the linear transformations are written using upper-case letters in these notes. Note that a "0" in a half-bracket symbol stands for "the vacuum," which means no particle is present.

Similarly, the operator \( D^\dagger U \) transforms an up-quark into a down-quark, because

\[
(D^\dagger U)|u\rangle = D^\dagger (U|u\rangle) = D^\dagger |0\rangle = |d\rangle
\]

Similarly, the operator \( U^\dagger S \) transforms a strange-quark into an up-quark, the operator \( S^\dagger U \) transforms an up-quark into a strange quark, the
operator $D^+ S$ transforms a strange-quark into a down-quark, and the operator $S^+ D$ transforms a down-quark into a strange-quark.

Graphically, here's how it looks:

Based on these basic transformations, if I tell you that an $\Omega^-$ particle is composed of three strange-quarks, can you determine the quark content of all the other particles in the following baryon decuplet?
Fun, eh?

Thus, the baryon decuplet illustrated above is described by an abstract 10-dimensional complex vector space, where each of the ten particles is a basis vector in the space, and the indicated transformations act on the basis vectors to produce other basis vectors. By analyzing the properties of the transformations, one can determine that they belong to the Lie group SU(3), so one says that quarks are described by an SU(3) symmetry.

For fun, and as a challenge, figure out the quark content of each of the eight members of the baryon octet illustrated earlier in the notes.

We'll leave our little foray into advanced mathematics there for now.

Continuing with the story, there is a problem with the analysis presented so far. Three of the particles in the baryon decuplet just illustrated appear to violate the Pauli exclusion principle, since they are made up of three identical particles in the identical quantum state. Recall that when
you "fill" the energy levels of an atom with electrons, you can only fit a maximum of two electrons into each energy state, because electrons are spin-1/2 fermions and the Pauli exclusion principle applies to them. The same is true of quarks, which are also spin-1/2 fermions. This would have been a death blow to the quark model, unless a work-around were found, and in 1964 O.W. Greenberg suggested a way to resolve this difficulty. He suggested that each quark comes in three "colours," which he called red, blue, and green. He did not mean for these to be thought of as real colours, but just as new, whimsical, abstract properties. He might have just as well called these new quantum numbers "Greenberg-ness," "Redbergness," and "Bluebergness," but he chose not to do so. To be clear, a down-quark that is also red has 1 unit of "redness," 0 units of "blueness," and 0 units of "greenness." An anti-quark that is also red has -1 units of "redness," 0 units of "blueness," and 0 units of "greenness." And so on.

Anyway, Greenberg's suggestion has a useful analogy with real colours. If you combine equal intensities of red, blue, and green light, it is perceived by us humans as white light. Greenberg considered "white" to be "colourless" for his analogy, and proposed that only colourless particles could be observed. This "explains" why individual quarks are never observed, because they all carry a certain colour. However, a baryon is made up of three quarks, and the three quarks are always of complementary colours (i.e., one red, one blue, and one green) so that the "total" colour is always colourless. Similarly, a meson is made up of a quark and an anti-quark, one with a certain colour, and the other with the matching anti-colour, so that again the combination is colourless.

And Greenberg's proposal also neatly solves the Pauli exclusion principle problem, because each quark in a three-quark particle carries a different colour. The addition of the new quantum numbers allows each quark to be in a different quantum state within the three-quark particle.

Furthermore, colour is the key to understanding the strong force that acts between quarks. It turns out that the messenger particles that
transmit the strong force between quarks are called gluons, and they also carry colour. (Contrast with electromagnetism, where the messenger particles are photons, which do not carry electric charge.) The exchange of a gluon results in the changing of the colours of the quarks doing the exchange. For example, if a blue up-quark sends a blue/anti-red gluon to a red strange-quark, then the strange-quark will become blue and the up-quark will become red. The exchange of the gluon causes the two quarks to feel the attractive strong force.

In a similar way, all pairs of quarks feel the attractive strong force by exchanging the gluon with the appropriate colour/anti-colour combination.

The mathematical details of such interactions have been worked out to a high degree, the result being a theory of the strong force called quantum chromodynamics (QCD). But much work remains to be done to clarify many details of the theory. So far, so good: Many of the predictions of the theory have been verified very nicely by experiments. However, QCD has not been verified to anywhere near the extent that QED has been, so much remains to be done.

The symmetry of QCD also turns out to be SU(3), so quarks are associated with two distinct SU(3) symmetries: flavour SU(3), related to the transformations between various flavours (u, d, s, etc.) of quarks, and colour SU(3), related to the transformations between various colours (r, b, g) of quarks.

The colour hypothesis is consistent with the fact that no particles composed of four or five quarks has ever been observed, nor one with two quarks (only mesons that contain a quark and the corresponding anti-quark have been observed). But it's not much of an explanation for why quarks are confined within particles; that had to wait until the 1970s with the development of the concept of "asymptotic freedom."

Classification of elementary particles
Where do we stand today?

Scattering experiments at higher and higher energies had results that were consistent with the fact that protons and neutrons have three "lumps" in them, so this provided support for the quark model. Further support came in 1974, when teams led by Samuel Ting at Brookhaven and Burton Richter at SLAC independently discovered a new particle, the $J/\psi$, which could be naturally interpreted in terms of a fourth quark. It turns out that just such a fourth quark had already been proposed by Bjorken and Glashow in 1964, and argued for by Glashow, Iliopoulos, and Maiani in 1970.

In subsequent years, new particles continued to be discovered, and as they were all successfully explained in terms of quarks, the quark model gained support.

A big breakthrough occurred in 1983, when a group led by Rubbia at CERN observed three new particles, the $W^+$, $W^-$, and $Z^0$. All three of these particles had been predicted independently by Glashow, Weinberg, and Salam, and before their discovery their masses had been predicted very precisely. The experiments confirmed the predicted masses very accurately, which provided tremendous support to the new picture of fundamental particle physics.

By this time, views in fundamental particle physics had converged on what became known as the standard model of particle physics. The model included a listing of particles considered to be fundamental, and a picture of the universe as consisting of quantum fields. Particles are considered to be excitations of the fields. Forces are transmitted between matter particles by the exchange of "messenger" particles. Each type of force field has its own characteristic messenger particle: the electromagnetic field has the photon, the weak force has three messenger particles, the $W^+$, $W^-$, and $Z^0$, and the strong force has eight gluons. Each of these three fundamental forces is described by a
detailed, well-established quantum field theory: Quantum electrodynamics (QED) for electromagnetism (with a U(1) symmetry), the theory of weak interactions (with SU(2) symmetry), and quantum chromodynamics (QCD) for strong interactions (with SU(3) symmetry).

The current "periodic table" of fundamental particles is as follows:

There are three "generations" of fundamental matter particles, one in each of the first three columns of the diagram. Ordinary matter particles that we observe in nature are composed of fundamental particles in the first generation. Other particles, such as ones that are produced in giant particle accelerators, are composed of fundamental particles in the second and third generations.

As far as we know, these particles are fundamental, in the sense that they have no further internal structure; but who knows? We can only probe so deeply, and perhaps one day we will be able to probe yet more deeply and we'll discover more structure deep down. (Story about
Bertrand Russell and the turtles.

The fourth column in the diagram above contains "messenger particles" that transmit fundamental forces. Because QED has U(1) symmetry, and U(1) is a one-dimensional group, there is only one messenger particle in QED, the photon. The photon is its own anti-particle. The weak force has SU(2) symmetry, and SU(2) is three-dimensional, so there are three messenger particles for the weak force, the Z⁰, which is its own anti-particle, and the W⁺ and the W⁻ particles, each of which is the anti-particle of the other. The strong force has SU(3) symmetry, and SU(3) is eight-dimensional, so there are eight messenger particles for the strong force, called gluons. The gluons come in four particle/anti-particle pairs.

It's interesting to note that what we normally consider the basic matter particles (quarks and the hadrons (baryons and mesons) that are made up of quarks, and also leptons) are all fermions, whereas the messenger particles are all bosons. The simplest quantum field theories describing the interactions of such particles all worked only if the messenger bosons are all massless. This is problematic, because although the photon and the gluons are massless, the vector bosons that mediate the weak force (the W⁺, W⁻, and Z⁰) are not massless. In 1962, Philip Anderson suggested a work-around for this problem by postulating the existence of another field that would interact with the vector bosons and give them mass. A number of people worked on this idea in short order and successfully made the idea consistent with relativity, thus patching up this problem in the standard model. Although a number of people worked on this idea, Peter Higgs got a little more credit than the others in that this field is now called the Higgs field. And, as we know, every quantum field has a particle associated with its excitations, and so this particle became known as the Higgs boson.

Although this particle has been theorized for about 50 years, it was discovered only recently at the LHC, with first news coming out in 2012 and sufficient data accumulated and analyzed for confirmation by 2013. The discovery of the Higgs particle filled in the last gap in the standard model.
model of particle physics, and so it's now a very mature and successful theory.

However, this is not to imply that the standard model of particle physics is complete and written in stone. There are still many puzzles to work out, and undoubtedly new discoveries await enterprising explorers (one of you, perhaps?).

First and foremost of the annoyances in the standard model is that gravity is not included in this picture. Well, perhaps not everyone would say this is the foremost problem, because gravity is so weak that it normally doesn't play a role in particle physics, nuclear physics, atomic physics, chemistry, etc., but it is annoying to many people because it messes with the way we view the world.

The basic view of the world assumed by quantum field theory is that the world is made up of quantum fields. Each wave or particle that we observe is an excitation of one of the quantum fields. For example, the electromagnetic field is a quantum field, and an excitation of the electromagnetic field is a photon. Yes, we have a good classical theory of electromagnetism, Maxwell's theory, but this is considered to be an approximation to the deeper, more fundamental quantum field theory of electromagnetic phenomena, called quantum electrodynamics (QED).

QED was the first quantum field theory, and it remains the most accurately verified physical theory. Subsequently, quantum field theories for two of the other fundamental forces were developed, one for the theory of the weak force, and the other (QCD) for the theory of the strong force.

Each and every fundamental particle is viewed as an excitation of a quantum field. For example, every electron is viewed as an excitation of the electron field. Every quark and neutrino, every single fundamental particle is an excitation of a certain kind of quantum field. (By the way, this gives us a new way of thinking of a "vacuum" as well; in the past it
was thought that a vacuum is a void, totally empty of anything. In the QFT picture, even a vacuum is a very exciting place, with all kinds of quantum fields doing their thing, and virtual particles blinking in and out of existence.)

The problem with gravity is that nobody has been able to come up with a quantum theory of gravity. We have good classical theories of gravity (Newton's theory, and its successor Einstein's theory, which we'll learn more about next week), but no quantum theory of gravity.

And this troubles people. When you think of everything in the world as deriving from a quantum field, and then there is no quantum theory of gravity, what are you supposed to think? Maybe someone will cook up a nice quantum theory of gravity next month, and everyone will be happy and the problem will be solved.

But maybe you are wrong about everything? If there is no quantum theory of gravity, then maybe we've got the wrong way of looking at the world. Or not wrong, exactly, because quantum field theories are pretty amazingly good, and their predictions have been verified extremely accurately. But maybe there is a better, more insightful way of looking at the world, and so there is a sense of unease.

Many people are working on developing a theory of quantum gravity, and maybe one of these days someone will score a breakthrough. It's a hard problem, though, so it may be a while.

Gravity does not fit into the standard model of particle physics, but the other three fundamental forces do.

There are other puzzles, open problems, and desiderata with the standard model of particle physics. Some of these are:

- Neutrinos seem to have mass, but what exactly are their masses? Measuring their masses is difficult because they interact so
sporadically with other matter, but this needs to be done. The standard model has to be slightly modified to accommodate neutrinos with non-zero mass, and this needs to be tidied up.

• Why are there only three generations of fundamental matter particles? There is cosmological evidence that there should be no more than three generations of fundamental matter particles, but this needs to be clarified.

• Is there a way to predict some of the properties of fundamental matter particles based on the properties of others? That is, a theory becomes more elegant if the number of its "free parameters" is reduced, and one wonders whether this can be done for the standard model.

• Why is there more matter than anti-matter in the universe?

• What is dark matter? (We'll explore evidence for dark matter next week.)

• Can the standard model of particle physics be extended to include gravity?

There are many other open questions in particle physics, but this should give you a flavour of some of the problems being actively researched.

**The theme of unification in fundamental physics theory**

As we've seen in the course, one of the persistent themes in physical theory is unification. We understand more deeply when we realize that two phenomena that were originally thought to be distinct are actually aspects of the same basic thing.

It was previously thought that electricity, magnetism, and optics were separate branches of physics, but with the work of Maxwell, it was realized that they are all united as aspects of the same basic phenomenon, electromagnetism. Einstein made this connection even deeper with his theory of special relativity.

In the same way, work in quantum field theory has helped us to see that electromagnetism and the weak force are aspects of the same basic
force, as shown by the electroweak theory of Glashow, Weinberg, and Salam, which was supported by many experiments since the 1970s.

Is there a theory that would unify electroweak forces and strong forces? After intensive research for many years, nobody is quite sure about this, and work continues. To learn more about such attempts, search for "grand unified theories."

And of course, gravity is still left out; attempts to include gravity in unified theories are listed under the heading "theories of everything."

One currently popular class of theories purporting to be theories of everything are called string theories, or superstring theories. These are still theories in formation, and have not developed to the extent that they can make definite predictions, so they are not quite theories yet, but they are working hard to get there. The problems faced by workers in this field are enormously difficult, and a lot of very bright people are working on such theories, but so far there is no definitive theory. Nevertheless, the partial results have been enticing, and so people have continued to be encouraged to work in this field.

One of the attempts to go beyond the standard model involves a concept called *supersymmetry*, which proposes that each of the fermions in the standard model has a boson partner, and each of the bosons in the standard model has a fermion partner. These supersymmetric partners are given whimsical names, such as selectron (short for "supersymmetric electron"), squark, neutralino, wino, zino, gluino (see photographs below), and so on.

There is currently absolutely no evidence whatsoever for supersymmetry, and so this must be listed as a very speculative idea. (The LHC has already ruled out most of the "parameter space" for supersymmetry, and future research there may snuff out supersymmetry's few remaining breaths.)
But one never knows how things will pan out; it's best if many researchers try out many *different* ideas rather than everyone treading down the same path that might lead nowhere. Time will tell.

Fundamental particles in the grocery store:

Startling discovery of new fundamental particles in a local grocery store:
Appendix: Symmetry in Fundamental Physics

(This section to be completed.)
The idea of a group

Groups of numbers

Groups of transformations

Discrete groups

Continuous groups

Permutation groups

Applications of groups in physics; the connection between symmetry and conservation laws

Noether's theorem; symmetry in classical systems

Symmetry and conserved quantities in quantum systems

Standard model symmetries

2: Cosmology

Einstein's theory of gravity: General relativity

In 1905, Einstein made an amazing number of revolutionary discoveries,
so much so that this year is widely considered an *annus mirabilis*, his miracle year. His first remarkable work was his paper on the photoelectric effect, which contained the photon hypothesis, and explained an effect using the photon hypothesis that nobody else could by using the wave model of light. Next was his Ph.D. thesis, which provided powerful arguments for the existence of atoms. Although it was only published in 1906, the work was completed in April 1905 and he followed up less than two weeks later with a published paper on Brownian motion, which provided additional support for the existence of atoms. Then came his paper on the foundations of special relativity, and finally a subsequent relativity paper on $E = mc^2$.

It was quite a year for a complete unknown who was living very modestly and working full-time as a humble patent clerk, third-class.

As you may recall from first semester, the special theory of relativity places consistency restrictions on all physical theories; essentially, the speed of light is an absolute "speed limit," and so each physical theory must incorporate the fact that no material object can be accelerated to this speed limit or beyond. In his first paper Einstein already analyzed Newtonian mechanics and found that it failed the relativity test, and began to develop a new mechanical theory to replace Newtonian mechanics. We now call this "relativistic mechanics," even if the name is inappropriate; after all, Newtonian mechanics is also relativistic, it just satisfies Galilean relativity instead of Einstein's theory of relativity. Maxwell's theory of electromagnetism already satisfies Einstein's relativity, and so needed no modification.

Then Einstein turned to Newton's theory of gravity, and soon determined that it did not satisfy his new relativity principle, and so he set himself the task of constructing a new theory of gravity that does satisfy special relativity. The result, after ten years of extremely hard work, is Einstein's theory of gravity, also called the general theory of relativity.

**Minkowski spacetime**
Einstein's former mathematics professor, Hermann Minkowski, had also come to the same conclusion, and he did important work that laid some of the foundations for Einstein's theory of gravity. You may recall from first semester that Einstein was not a regular attendee of his undergraduate lectures, preferring to work intensively on the latest theories of physics on his own. As a result, his professors did not form a very high opinion of Einstein, and Minkowski in particular said that Einstein would never amount to anything. But Minkowski deserves credit for immediately recognizing Einstein's special relativity papers as outstanding work, and Minkowski began to work on mathematical aspects of relativity and made important contributions.

In the few years after 1905, only a very few physicists recognized the value of Einstein's theory of special relativity; most just ignored it because they didn't understand it and didn't think it was important. Planck was the first to recognize the value of Einstein's relativity, and he began to correspond with Einstein, began working on the theory and publishing papers on it, and publicized the theory by giving talks and even courses on it.

Minkowski's contribution was to recast Einstein's theory in a more geometrical form, and to express it in the brand-new mathematical language of tensor analysis, which he did in a paper published in 1908. Einstein wasn't thrilled with this at first, saying sarcastically that now that the mathematicians had gotten hold of his theory, not even he understood it anymore. He initially figured that Minkowski's work was "superficially learned" (see Pais, Subtle is the Lord ..., Page 152). But, as we shall see, Einstein came around to adopt Minkowski's perspective, and it allowed him to create his theory of gravity.

Here's the beginning of a famous address Minkowski gave in Koln in 1908:

"The views of space and time which I wish to lay before you have
sprung from the soil of experimental physics, and therein lies their strength. They are radical. Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality."

After Minkowski's address, research activity in Einstein's theory of relativity jumped dramatically, and it's probable that his talk had a lot to do with this. Certainly the response to Minkowski's talk was strong, both positive and negative. Many physicists were shocked at Minkowski's suggestion that Euclidean geometry was no longer applicable to the relativistic world, and expressed their disapproval most vigorously. However, others were very enthusiastic. Besides those already named, Arnold Sommerfeld and Max Born were so positive about Minkowski's perspective that they proclaimed it as the future of physics. Max von Laue, who worked at Sommerfeld's Munich institute, prepared the first textbook on relativity in 1911, and it propagated the geometric approach to relativity. However, von Laue was careful to downplay Minkowski's strong language, and to separate to some extent the new mathematical formalism of Sommerfeld from the philosophical position of Minkowski. This continued throughout the many editions of von Laue's influential work, even in the 50th anniversary edition of 1955!

One gets the sense from this how difficult it is for us humans to change the way we think, and to adapt to new ideas. A flexible few are able to do it easily, but the rest of us find it difficult. Typically it takes a generation or two of physicists to die with their old ideas before new ideas take hold and flourish! In the case of relativity, by about 1911 the many of the top physicists had understood the theory and realized its importance. Indeed, by 1908 Einstein was becoming very well-known in the physics community, and he was even nominated for the Nobel Prize by Wien, for relativity, in 1912.

Minkowski himself died at the young age of 44 in 1909, and therefore did not live to see the full flourishing of his radical ideas in Einstein's theory of gravity, which was completed in 1915.
Newton's theory of gravity is inconsistent with special relativity

But why doesn't Newton's gravity fit nicely into the framework of special relativity, as Minkowski and Einstein both realized? Well, in special relativity one defines inertial reference frames much as one does in Newtonian mechanics; once one finds one inertial reference frame, any other reference frame moving at constant velocity relative to it is also inertial. Furthermore, each reference frame is assumed to extend as widely as one likes.

With gravitation involved, one cannot have an inertial reference frame extending indefinitely. Consider the following diagram:

Particle A falls freely towards the centre of the Earth, and so does Particle B. It follows that if you have an inertial reference frame attached to particle A, then another reference frame moving at a constant velocity relative to it, say along the dashed red line attached to B, ought to also be inertial according to the principle of special relativity. However, experiments done in closed compartments at A and B will reveal differences, so this is problematic.

A second problem with Newton's law of gravity, which also makes it inconsistent with special relativity, is that it assumes instantaneous
action at a distance. Compare Newton's law of gravity with Coulomb's law; the two laws have the same form, so why is the latter consistent with special relativity whereas the former is not? One way to see this is to note that if charged particles are moving, then Coulomb's law is modified in Maxwell's equations. Magnetic forces are also present for moving charges, and there is a very specific prescription for the transmission of changes in the electromagnetic field for moving charges; the changes are transmitted at the speed of light by electromagnetic waves. There is no such provision in Newton's law of gravity; where are the gravitational waves that transmit changes in the gravitational field for a moving particle? They don't exist in Newton's theory of gravity, and that is a problem; instantaneous changes in the gravitational field are inconsistent with the universal speed limit in special relativity.

The happiest thought in Einstein's life

Another line of thought that Einstein pursued was to generalize special relativity to include frames of reference that were in relative acceleration. Special relativity deals with reference frames that are moving at a constant velocity with respect to each other.

Pondering these two problems (how to generalize special relativity to include accelerated reference frames, and how to generalize Newton's law of gravity to make it consistent with special relativity), led Einstein to the remarkable insight that the two problems are the same!

Einstein thought about a person falling freely, from a roof perhaps, and suddenly realized that while falling freely, a person would not feel gravity. This he described as the happiest thought of his life, as quoted here from Pais's, *Subtle is the Lord* ... (Page 178):

"When, in 1907, I was working on a comprehensive paper on the special theory of relativity for the *Jahrbuch der Radioaktivität und Electronik*, I had also to attempt to modify the Newtonian theory of gravitation in such a way that its laws would fit in the [special
relativity] theory. Attempts in this direction did show that this could be done, but did not satisfy me because they were based on physically unfounded hypotheses.

"Then there occurred to me the happiest thought of my life, in the following form. The gravitational field has only a relative existence in a way similar to the electric field generated by magnetoelectric induction. **Because for an observer falling freely from the roof of a house there exists** — at least in his immediate surroundings — **no gravitational field** [italics in the original]. Indeed, if the observer drops some bodies then these remain relative to him in a state of rest or uniform motion, independent of their particular chemical or physical nature (in this consideration the air resistance is, of course, ignored). The observer therefore has the right to interpret his state as 'at rest.'

"Because of this idea, the uncommonly peculiar experimental law that in the gravitational field all bodies fall with the same acceleration attained at once a deep physical meaning. Namely, if there were to exist just one single object that falls in the gravitational field in a way different from all others, then with its help the observer could realize that he is in a gravitational field and is falling in it. If such an object does not exist, however — as experience as shown with great accuracy — then the observer lacks any objective means of perceiving himself as falling in a gravitational field. Rather he has the right to consider his state as one of rest and his environment as field-free relative to gravitation.

"The experimentally known matter independence of the acceleration of fall is therefore a powerful argument for the fact that the relativity postulate has to be extended to coordinate systems which, relative to each other, are in non-uniform motion."

Nowadays, Einstein's "happiest thought" is called the equivalence principle; formally, one can consider this the principle that a free fall due
to a gravitational field is equivalent to constant acceleration in a region of space where there is no net gravitational field.

Einstein was careful to say that this is true only "locally." The Earth out on the prairies looks flat, but only from the ground ("locally"); from out in space, the Earth definitely looks curved. Similarly, a smooth curve is extremely well approximated by its tangent line (i.e., the curve looks flat), but only close to the point of tangency. The same is true of a laboratory in free fall in a gravitational field; you can't tell the difference between it and a laboratory moving far from any gravitational fields with a constant acceleration, but only if the laboratory is very small. If the free-falling laboratory is large, then the "high" end of the laboratory will experience a weaker gravitational field than the "low" end. Similarly, two particles at opposite sides of the falling laboratory will tend to move closer to each other as the laboratory falls. Consider the following diagrams:

In a "large" laboratory, you'd be able to tell the difference between a gravitational field due to (say) the Earth, and an accelerating laboratory far from any massive bodies. Thus, the equivalence principle is a local principle; but that's OK, that's the way it works in calculus, derivatives tell us local rates of change, and similarly the equivalence principle tells us
about a local equivalence.

It can be shown that the equivalence principle is the same as the principle that gravitational and inertial mass are equivalent. This principle was recognized by Newton, and he did experiments to verify this principle. Modern experiments verifying the equivalence principle are discussed later in these notes.

**Spacetime is curved**

Particles that start moving parallel don't stay parallel in a gravitational field. If you drop two objects in your free-falling laboratory parallel to the "vertical" walls, the falling objects don't stay parallel; instead, they start curving towards each other. But if we want to base physics on geometry, how can we model this? Where else do we see paths that start parallel but don't stay parallel? On curved surfaces! Could it be? Could the universe be curved?
But what does it mean for the universe to be curved? According to Einstein's theory of special relativity, the arena for doing physics is Minkowski's *spacetime*; that's right, not just space, but space and time. If the universe is curved, it would have to be space and time together, as a unity, that is curved.

Note that there is no way to *derive* this conclusion. The curvature of spacetime is a hypothesis, a creation if you like. The test of its validity is whether it turns out to be valuable in making predictions that can be verified by experiments. It turns out that Einstein's theory of gravity based on the curvature of spacetime makes very definite predictions that can be accurately calculated and that have been verified by precision experiments over and over again over the past 100 years. It's a battle-tested theory and it has withstood all the tests of time so far.

When Einstein had the brainstorm that perhaps spacetime is curved, and that this curvature is the origin of gravity, he turned to his friend Marcel Grossmann, who was a mathematician. (You may recall from first semester that Grossmann was the fellow that helped Einstein make it
through his undergraduate degree by lending his detailed notes so that Einstein could cram for his exams.) Einstein needed to learn the mathematics of curved spaces, and Grossmann came through again, helping Einstein quickly get up to speed on the latest relevant mathematics: Differential geometry and tensor analysis.

**Einstein's gravitational field equations**

Newton's law of gravity allows one to calculate the strength of the force acting between two objects given their masses and the distance between them. Later researchers cast Newton's theory into the form a field theory, once fields and potentials were invented.

The perspective of Einstein's theory is that gravity is caused by the curvature of spacetime. Einstein's equations provide a prescription for calculating how much curvature is produced by a given distribution of mass. In other words, mass is a source of spacetime curvature. However, there must be other sources, too, because according to Einstein's theory of special relativity, energy can be converted to mass, and vice versa. Thus, energy density must also be a source of spacetime curvature.

It turns out that there are other sources of spacetime curvature as well, such as pressure. It's a bit complicated.

And how does one describe spacetime curvature mathematically? This is complicated. For all you calculus fans, think about how you would describe the curvature of a curve. Curvature has something to do with how fast the slope of the graph changes as you move along it, so curvature has something to do with the second derivative of the function describing the curve. However, if you keep the curve rigid, like a piece of wire, and rotate it, then the describing function changes, whereas we wouldn't want to say that the curvature of the curve changes. It turns out that this means that the expression for the curvature of a curve is a combination of first and second derivatives of the defining function. You can look up the expression towards the end of your calculus textbook, in
the part that is normally studied in second-year calculus. Here it is:

\[ K = \frac{y''}{\left[1 + (y')^2\right]^{3/2}} \]

Some of you calculus fans out there might enjoy playing with this formula and trying to develop an intuitive feel for it. Try it out on your favourite curves (polynomials, trig, etc.) and see if the values you end up with correspond with your sense for how "curvy" the graph is, thinking of the graph as a road that you might drive your car (or motorcycle, or bicycle) on. Have fun!

The next step would be to bump up the formula to deal with curves that are defined "parametrically" rather than as graphs of functions. In this way, you can also deal smoothly with curves in three-dimensional space.

So much for curves; how do you describe the curvature of a surface? Ah, this is much more complicated, because it depends on the direction of your motion along the surface. It turns out that you need a two-by-two matrix of expressions involving second derivatives to describe the curvature of a two-dimensional surface. Similarly, you need a four-by-four matrix of expressions involving second derivatives to describe the curvature of a four-dimensional spacetime. We shall leave the mathematical details to a higher-level course (search for "differential geometry of surfaces" when you're ready to dig into this more deeply).

Once you figure out how to describe the curvature of four-dimensional spaces (or, rather, spacetimes, because that's what's relevant here), you are then ready for the mathematical statement of Einstein's theory of gravity. In words (these are approximately the words of John Wheeler), Einstein's theory of gravity says that:

**Matter (and energy, etc.) tell spacetime how to curve. Curved spacetime tells particles how to move.**
The greater the density of mass (and energy, etc.), the greater is the curvature of spacetime, and therefore the greater is the distortion of a particle's path from a straight line. But deviations from straight-line paths is just what we'd interpret as being due to a force! In Einstein's perspective, there are no forces, really. Particles just try to do their best to move in straight lines, but they can't do so in locations where spacetime is very curved, so they move in paths that are as straight as possible. Such paths are called geodesics.

The statement that "curved spacetime tells particles how to move" can be said in other words as "particles move in geodesics."

On a flat piece of paper, a geodesic is a straight line. On the surface of a sphere, a geodesic is a "great circle."
For the curious, here's what Einstein's equations look like, at least in their original form:

\[ G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \]

The left side of the equation above is the Einstein tensor, and it describes the curvature of spacetime. It can be thought of as a 4 by 4 matrix, and each of the 16 entries in the matrix has quantities, and perhaps their partial derivatives, and perhaps their second-order partial derivatives. On the right side of the equations, Newton's constant \( G \) appears, as well as the speed of light \( c \). The other gadget on the right side is called the stress-energy tensor (in some books it's called the energy-momentum tensor).
stress-energy tensor (in some books it's called the energy-momentum tensor), and it too can be thought of as a 4 by 4 matrix; it contains all of the information about the distribution of the sources of the gravitational field, which are mass, energy, pressure, momentum flux, stresses (due to forces other than gravity; remember, in this perspective, gravity manifests as geometry, not as a force), and so on.

The tensors on the left and right side of the previous display are both symmetric, so they each have a maximum of 10 independent components, not 16. If you wrote each component as a separate equation, you would then have 10 coupled second-order partial differential equations. These are called the Einstein equations.

To make matters worse, these equations are typically (except for the simplest situations) nonlinear. This means that there is no superposition principle in Einstein's theory of gravity, as there is in quantum mechanics and in Maxwell's theory of electromagnetism. This makes solving Einstein's equations a more challenging technical problem than if they were linear. Oh well, that's the way it goes.

The first thing Einstein did was to apply his equations to the very simplest situations he could think of. (Physicists credo: Start with the simplest situations.) In applying his equations to an empty universe (setting the right side of his equations to zero), he found that there were no static solutions! This troubled him, because it was widely assumed at the time that the universe overall was static (though, of course, there were plenty of motions within it). However, he was able to patch things up by adding a term to his basic equations, as follows:

\[ G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \]

Ch32L Page 66
In the updated equations, the new term includes a constant, $\Lambda$, which became known as "the cosmological constant," and the metric tensor describes "distance" relations between nearby points in spacetime, so it has something to do with the curvature of spacetime as well. Einstein interpreted the cosmological constant as representing a mysterious overall repulsion in the universe, that served to balance the attractive forces of gravity, leading to overall balance and a static universe on large scales.

Einstein dreamed up these equations in 1915, after a period of years of very intensive work. He applied the equations to the very simplest situations, and then essentially collapsed in exhaustion, and needed to be nursed back to health for a few months. In the meantime, others took up his equations and applied them to other, more complex situations. Work is ongoing.

**Experimental evidence for the equivalence principle**

It is very typical of Einstein to found his theories very solidly on well-established experimental evidence. Remember that the equivalence principle as used by Einstein as a foundation for general relativity (the local equivalence of free fall in a gravitational field and uniform acceleration far from any gravitational fields) is itself equivalent to the equality of inertial and gravitational mass.

Evidence for the equivalence principle goes way back to Newton. Newton was well aware of the *logical* difference between the concepts of gravitational mass (which appears in his formula for the gravitational force) and inertial mass (which appears in Newton's second law). Despite the logical difference between the two concepts, they are numerically identical, and Newton verified this by swinging various materials, all with the same "mass," in a pendulum. Perhaps different substances with the same gravitational mass had different inertial masses. To make sure that
air resistance was as equal as possible for each material, he used a small wooden box with a door for the pendulum "bob," and placed each substance into the box. Then he displaced the pendulum and measured its motion. His results were that each different substance had the same ratio of inertial mass to gravitational mass, so it's consistent to take them as equal.

Newton's experiment was improved by Bessel in the early 1800s, and further improved most dramatically in a series of experiments performed by Roland von Eötvös from 1885 until 1909! Eötvös used a torsion balance instead of a pendulum; that is, he hung two objects with the same mass but made of different materials from a thin wire. The two objects are affected by both the Earth's gravity and the Earth's rotation, but one acts through the gravitational mass and the other acts through the inertial mass. If there were a difference in gravitational mass and inertial mass, then a balance that were originally at rest would spontaneously start to rotate. Any potential rotation was detected by mounting a small mirror on the wire and observing it through a small telescope.

(Source: [http://tudtor.kfki.hu/eotvos1/onehund.html](http://tudtor.kfki.hu/eotvos1/onehund.html))
Eötvös produced stunning accuracy; to within one part in $10^{12}$ he found no difference between gravitational mass and inertial mass. Later experimenters (for example, Roll, Krotkov, and Dicke in 1964; Braginsky and Panov in 1971; Su et al in 1994) pushed the accuracy to one part in $10^{13}$. A similar experiment was performed by shining laser light on a special reflector placed on its surface by astronauts in the 1960s; in this experiment, the Earth and the Moon are the two objects accelerated by the Sun's gravity!

Here's the torsion balance used in a recent repetition of this experiment:
So, there's an enormous quantity of experimental support for the principle on which Einstein founded general relativity; this doesn't support the theory itself, but it does support its fundamental postulate. For experimental support of the theory itself, one has to use the theory to make predictions, and then the predictions themselves have to be tested experimentally. That's where we'll turn next.

**Some predictions of Einstein's theory of gravity**

1: Mercury's orbit

One of the long-standing problems in solar system celestial mechanics involved extremely slight anomalies in the orbit of Mercury. This kind of thing had been seen before; in the 1700s, careful measurements detected anomalies in the orbit of Uranus. That is, using Newton's law of gravity and Newton's laws of motion, it was possible to predict very accurately the future positions of all the planets. Of course, this was a complex matter, and one had to account first and foremost for the
gravitational force exerted on a planet by the Sun. However, to make the prediction more accurate, one then had to also account for the small perturbations due to gravitational pulls from other planets. When this is carefully done, it all works well and all makes sense.

Except for Uranus, the predictions were not coming out quite right. The future positions of the planet, as carefully measured, were not quite matching up with the predictions. By 1845 Urbain Le Verrier and, independently, John Adams, had calculated that there must be another planet beyond Uranus causing the wobble in its orbit. In 1846, Johann Galle and Heinrich d'Arrest found Neptune just where it was expected. This was a great triumph for Newtonian mechanics!

Pluto was discovered in 1930 by Clyde Tombaugh, by a similar process of careful observations of anomalies in the orbit of Neptune.

But the anomaly in Mercury's orbit was a real puzzle, because there did not seem to be any place that another planet could be, seeing how close to the Sun Mercury is. To be sure, the anomaly in Mercury's orbit is very small indeed. It doesn't move in a perfect ellipse, as predicted by Newton's law of gravity; instead, the entire ellipse rotates extremely slowly, throwing off its calculated position by a mere 43 seconds of arc per *century*. (A minute of arc is 1/60th of a degree, and a second of arc is 1/3600th of a degree!)

The precession of Mercury's orbit was first noted by Le Verrier (the same Le Verrier who predicted Neptune) in 1859, after studying 150 years of data collected on Mercury's orbit. (Science is truly a collective effort; how many generations of astronomers were responsible for the collection of so much data on Mercury. And what about all the other planets and stars? One begins to appreciate the truly enormous effort required to bring us to our present understanding of the universe.)

Einstein used his field equations to calculate the orbit of Mercury based on his theory instead of Newton's theory of gravity, and what do you
know: There popped out the missing 43 seconds of arc per century, and Einstein's theory of gravity had its first success.

The orbits of other planets also precess, but at much smaller rates, and so it was difficult in Einstein's time to test his theory on them. However, modern measurements verify Einstein's theory for the other planets as well. The same effect is observed in some binary star systems (in others it's too difficult to observe), and in each case Einstein's theory is confirmed.

2: Bending of light by gravitational fields

In Newton's theory of gravity, it's unclear whether light should be affected by a gravitational field. After all, only particles with mass experience a force, according to Newton's law of gravity, and by Einstein's time it was clear that photons were massless.

In Einstein's theory of gravity, there is no ambiguity about this. All particles travel along geodesics in curved spacetime, so light is indeed affected by gravity.

Einstein first realized that light is affected by gravity in 1907, as discussed earlier in these notes ("the happiest thought"). By 1911 he realized that light from a distant star grazing the surface of the Sun would be deflected, and he calculated a value for the deflection angle, obtaining a little less than 1 second of arc. However, he didn't have his full theory available yet, and so his result was off by a factor of 2. By 1912, Einstein realized that spacetime is curved, and by 1915, when Einstein had his field equations in hand, he was able to determine the deflection angle more accurately, and obtained 1.74 arc seconds.

Here's a schematic diagram of the effect (much exaggerated):
Here's a copy of a letter Einstein wrote to astronomer George Hale in 1913, asking him to look for the deflection:
Einstein had good luck here. It was not possible to observe eclipses of the Sun for a while (partly because of World War I), so there was a delay in making the critical observations. Had observations been done earlier, Einstein would have been proven wrong, and his later correction would have been seen as a case of the boy crying wolf. However, by 1915 he had the correct result, and so the confirmation, when it came in 1919, made him front-page news, and quite a celebrity, the most famous scientist in the world.

Arthur Eddington was one of the early relativity enthusiasts, and he wrote books to popularize science in general, and relativity in particular. His views were not very popular in World War I-era Britain, because Einstein was German and therefore belonged to an enemy nation. But
Eddington was a Quaker and a pacifist, and his attitudes display science at its best: An international, cooperative pursuit.

Eddington was secretary of the Royal Astronomical Society, and Frank Dyson was Astronomer Royal in Britain, and they managed to obtain funding for an expedition to two locations (one in Brazil, the other in western Africa) to view the solar eclipse of 1919. Their measurements confirmed Einstein's predictions and made the latter famous. The whole expedition was rather remarkable, coming so soon after WWI ended in 1918.

Of course, Eddington's measurements were disputed; that's the way science works. However, they have been repeated many times over the years, with increasing accuracy, and by the 1960s there was general agreement that this prediction of Einstein's theory was amply confirmed.

Such measurements also confirm that spacetime really is curved. For example, one could claim that the bending of starlight as it grazes the Sun takes place in a flat spacetime, in much the same way that a car goes around a curve on a flat road. In 1976 (reported by Shapiro et al in 1977) a radar beam from Earth was sent to a spacecraft orbiting Mars along a
route that grazed the Sun. One can calculate the time delay for the signal to reflect from the spacecraft and return to Earth assuming that spacetime is flat, and then again assuming that spacetime is curved according to Einstein's field equations.

The time delays in the two cases differ by a factor of about 7000 (200 microseconds vs. 30 nanoseconds), so the experiment was easily able to confirm the prediction of general relativity: Spacetime really is curved. This type of experiment has been done many times in the 1970s and since then, with results always supporting Einstein's theory (see http://www.physics.ucsd.edu/~tmurphy/apollo/doc/Will_1.pdf for a review).

3: Gravitational redshift

The third of the classical predictions of Einstein's theory of gravity is gravitational redshift. In 1907, well before he had his field equations and the full theory of general relativity, Einstein gave his first argument that light moving in a gravitational field should experience a shift in wavelength. This effect is not the Doppler effect, which occurs because of relative motion between the emitter and observer of light, but rather for a different reason, that we'll now discuss.
Einstein, the master of the thought experiment, argued as follows. Suppose a particle of mass \( m \) falls from a height \( y \) above the surface of the Earth. Then when it reaches the Earth, its total energy will be

\[ mc^2 + mgy \]

Suppose that when the particle reaches the surface of the Earth it is somehow (remember, this is 1907) converted to a photon. Then, using the Planck relation, the energy of the photon will be

\[ hf_E \]

and

\[ mc^2 + mgy = hf_E \]

If the photon then rises back up to a height \( y \), and is then converted back to a particle with mass, then the photon must lose energy on the way up. Otherwise, we have just built a machine that will produce an unlimited amount of energy from nothing; in total, an amount of energy equal to \( mgy \) is created for each cycle of the machine. Einstein did not believe that this is possible (nobody does), and so the resolution is that the photon must lose energy equal to \( mgy \) on the way up, and it does this by a reduction in its frequency.

But there is nothing special about this photon; if this photon experiences a shift in frequency, then so does any photon "climbing" away from a massive object. Because the frequency decreases, which shifts the wavelength of a visible photon towards the red end of the spectrum, this phenomenon is called gravitational redshift.

The magnitude of the redshift can be calculated as follows: Let \( f_R \) represent the frequency of the photon when it has climbed back up to height \( y \). Then, using the principle of conservation of energy,
\[ mc^2 + mgy = hf_E = mgy + hf_R \]  \hspace{1cm} (1)

Using the first and third terms of equation (1), we obtain an expression for the mass of the particle in terms of the frequency of the photon received back at the top:

\[ mc^2 + mgy = mgy + hf_R \]

\[ mc^2 = hf_R \]

\[ m = \frac{hf_R}{c^2} \]  \hspace{1cm} (2)

Now take the expression for the mass of the particle from equation (2) and insert it into the second two terms of equation (1) to obtain an expression for the frequency change:

\[ hf_E = mgy + hf_R \]

\[ hf_E = (hf_R/c^2)gy + hf_R \]

\[ f_E = (f_R/c^2)gy + f_R \]

\[ \frac{f_E - f_R}{f_R} = \frac{gy}{c^2} \]

A more accurate version of this equation can be obtained by using the correct Newtonian expression for gravitational potential energy

\[ -\frac{GMm}{r} \]

instead of the approximation \( mgy \), where \( M \) is the mass of the central object (Sun, Earth, whatever) and \( r \) is the distance to the centre of the central object. The result is \( (R \) is the radius of the central object)
\[
\frac{f_E - f_R}{f_R} = \frac{GM}{c^2}\left(\frac{1}{R} - \frac{1}{R+y}\right)
\]

If \(y\) is much larger than \(R\) (for example, for a photon emitted from the surface of the Sun and received on Earth), we can omit the last term, which is very small relative to the first term in the parenthesis, to obtain

\[
\frac{f_E - f_R}{f_R} = \frac{GM}{c^2R}
\]

Using \(c = f\lambda\), one can also express the previous equation in terms of the wavelengths of the photons:

\[
\frac{\lambda_R - \lambda_E}{\lambda_E} = \frac{GM}{c^2R}
\]

This summarizes the results of Einstein's calculation from 1907 (he obtained the same result in 1911 using a different argument), before he had his finished theory of general relativity. Once he had his full theory in 1915, he was able to make this result even more precise by (in essence) using the relativistic expression for the gravitational potential energy rather than the Newtonian expression used in this rough calculation. The more accurate result is (based on the approximation that the receiving point is far from the emission point, relative to the radius of the emitting body)

\[
\frac{f_R}{f_E} = \sqrt{1 - \frac{2GM}{c^2R}}
\]

(It will be an interesting mathematical exercise for some of you (binomial
theorem!) to show that the more accurate expression reduces to the approximation as long as the mass and radius of the emitting object combine to make the last term in the previous equation small relative to 1.)

An early test of Einstein's calculation of gravitational redshift was performed by Pound and Rebka at Harvard in 1960 (improved to 1% accuracy by Pound and Snider in 1964), a notable achievement as this was the first terrestrial experiment confirming Einstein's theory of gravity. There is an internal tower at Jefferson Hall at Harvard that allowed them to study the motion of a photon over a vertical distance of 22.5 m (they started in the basement). The experiment was made possible by the recently discovered Mossbauer effect, which allowed them to make extremely accurate wavelength measurements.

By now there have been numerous such measurements, all in accord with the predictions of Einstein's theory of gravity.

Other consequences of Einstein's theory of gravity

1: Gravitational time dilation

If Alice lives on the 10th floor of a building and Basil lives in the basement, and they try to synchronize their clocks, they will judge that their clocks do not move at the same rate. For example, if they use light signals to achieve the synchronization, there will be a problem, because light signals experience a frequency shift when they move up or down in a gravitational field. By the equivalence principle, the same will be true no matter which method of synchronization is used, so this is an intrinsic property of clocks in gravitational fields, and not an artifact of the method used to achieve the synchronization.

To summarize: The stronger the gravitational potential, the slower that clocks tick.
So gravitational redshift and the fact that clocks in gravitational fields have different rates depending on their gravitational potential are two effects with the same root cause.

Using Einstein's theory of gravitation, one can calculate the effect precisely. How would one go about measuring such an effect? A classic experiment was performed in 1971 by Joseph Hafele and Richard Keating. They took four of the most precise clocks available at that time ("atomic clocks") on two round-the-world journeys, one going roughly East and the other going roughly West. Afterwards they compared the clocks that had journeyed with some stay-at-home clocks, and noted the differences in elapsed time. The differences agreed with the predictions of both special relativity and general relativity to within experimental uncertainty. A similar experiment was performed more accurately in 1975 by a group at the University of Maryland, and similar experiments were carried out with increasing accuracy by the U.S. National Physical Laboratory in 1996 and again in 2010. All of the experiments confirmed the special and general theories of relativity.

In 2010, Chou et al reported on experiments demonstrating the gravitational time dilation effect for clocks whose altitudes differed by less than a metre!
Global Positioning System (GPS)

The global positioning system consists of 24 satellites orbiting the Earth in 6 distinct planes (4 per plane), all with an orbital period of 12 hours. Each GPS satellite carries a clock that is accurate to within a few parts in $10^{13}$ per month, and is updated several times per day to prevent long-term drift in accuracy.
--- schematic description of how the system works

--- the system's accuracy depends on incorporating both special relativity (because the satellites are moving relative to the Earth) and general relativity (because they are at a lower gravitational potential, they experience gravitational time dilation relative to the Earth's surface)

It can be calculated that the errors introduced into GPS if relativity were not accounted for amount to about 0.84 parts in $10^{10}$ for special relativity.
and about twice that amount for general relativity, and the effects have the opposite signs, so the net effect is about 0.8 parts in $10^{10}$. Thus, if relativistic effects were not accounted for, after 12 s passed there would be an error of about 1 ns, after 1 minute passed the error would grow to 5 ns, and after 1 hour passed the error would grow to 300 ns. To achieve accuracy to within 2 m (which is a standard that is required in some applications), one needs a timing accuracy of about 6 ns, since light travels about 1 foot per ns. You can see that the whole system would quickly become useless unless relativity is taken into account.

This example shows us that relativity is very practical in our modern world, and its applications are not restricted to exotic locales such as particle accelerators and black holes.

Speaking of which:

2: Black holes

While Einstein was recovering from exhaustion after years of intensive work on general relativity, others began to apply his theory to the simplest physical systems. Just about the simplest physical system that contains mass consists of a single point mass. Karl Schwarzschild tackled this problem and solved Einstein's equations for a point mass, which he communicated to Einstein by letter in December of 2015. Einstein was pleasantly surprised that his very complex equations could admit exact solutions, as he anticipated that only approximate solutions could be obtained.

We won't go into the details of Schwarzschild's solution of Einstein's equations; it suffices to say that they specify the curvature of spacetime for all points in a space that consists of a single point mass. Going into the details requires considerably more mathematics than we have available at the moment. Schwarzschild's solution is remarkable because there is a certain distance from the point mass that represents a kind of boundary. Now known as the Schwarzschild radius, its value is
The boundary around the point mass having the Schwarzschild radius is called the point mass's event horizon. Any particle or photon trajectory that begins inside the event horizon does not pass outside the event horizon; not even light can escape from within the event horizon. We have a black hole.

Black holes were first imagined by Laplace way back in 1796; while studying escape velocity, he speculated that there might be stars so massive that not even light could escape them. With Schwarzschild's work, we now had a specific criterion for their existence. (The name "black hole" was coined by John Wheeler.)

Is a black hole just an imaginary mathematical thing, or do they really exist in nature? If they do really exist, how might they form?

Stars are normally in a state of hydrostatic equilibrium; the inward gravitational force that tends to compress them is balanced by an outward force due to gas pressure caused by the extremely dense, hot gas in their cores. However, once they "run out" of nuclear fuel, the balance is disrupted and they can collapse violently. Depending on their initial masses, various things might happen.

Our Sun, and stars with similar mass, will eventually grow into red giants, then blow off their outer layers and their cores will contract into very compact, dense, hot stars called white dwarfs. In a white dwarf, the gravitational attraction is balanced by a quantum effect called electron degeneracy pressure (the Pauli principle again!). However, if the mass is too great, it won't be possible to maintain this balance, and the star will pass right through the white dwarf stage and collapse into a neutron star. A neutron star has had most of its electrons and protons jammed together into neutrons, and so it's a bit like a giant atomic nucleus.
Our Sun's diameter is about 100 times Earth's diameter. White dwarfs are typically the size of the Earth, so you can imagine how dense they are. Neutron stars have diameters on the order of 10 km! Their density is enormous.

Neutron stars are kept in balance against gravity by neutron degeneracy pressure. However, if they are too massive, collapsing stellar cores pass right through the neutron star stage and collapse into ... well, it's not exactly clear what happens here, but it is widely believed that they collapse into black holes.

The Chandrasekhar limit is the maximum mass of a white dwarf, and its value (about 1.4 solar masses) was worked out by him in 1930. The Oppenheimer-Volkoff limit is the maximum mass of a neutron star, and its value was originally worked out by them in 1939; however, its value is less certain, because we don't understand the properties of such dense nuclear matter very well. Its value is thought to be about 2 solar masses.

Very massive stars explode their outer layers ("supernovae") before their cores collapse, and they lose a significant amount of mass in this way. But surely many supermassive stars will have collapsed cores that have more than 2 solar masses. What happens to them? It is thought that they collapse right through the neutron star stage and become black holes.

So physicists think that black holes are real objects, not just mathematical abstractions. But what is the observational evidence for black holes? Before we do this, let's calculate the Schwarzshild radius for objects of various sizes and masses just to get a feel for how large a black hole event horizon is likely to be. Remember, though, that these calculations are meant to apply to point masses; thus, if we get a Schwarzshild radius that is smaller than the size of the actual object, then an event horizon doesn't actually exist for the object, because it couldn't masquerade as a point mass inside its event horizon.
Example: A white dwarf of 1 solar mass and radius equal to Earth's radius.

\[ M = 2 \times 10^{30} \text{ kg} \]

\[
\frac{2GM}{c^2} = \frac{2(6.67 \times 10^{-11})(2 \times 10^{30})}{(3 \times 10^8)^2} = 3000 \text{ m} = 3 \text{ km}
\]

Conclusion: This white dwarf does not form an event horizon.

Example: A neutron star of 3 solar masses and radius 10 km.

\[ M = 6 \times 10^{30} \text{ kg} \]

\[
\frac{2GM}{c^2} = \frac{2(6.67 \times 10^{-11})(6 \times 10^{30})}{(3 \times 10^8)^2} = 9000 \text{ m} = 9 \text{ km}
\]

Conclusion: This neutron star does not form an event horizon either, but it's getting close.

So we're getting a sense for the typical sizes of event horizons, for star-sized masses. In the neutron star example just above, if somehow the neutron star managed to get compressed to a size that was just a little smaller, it would then be entirely inside its Schwarzschild radius. As long as the neutron star was spherically symmetrical, and non-rotating, its gravitational field outside itself would be indistinguishable from the gravitational field of a point mass having the same mass. (In the Newtonian case, this is Gauss's law for gravitation; in the relativistic case, this is called Birkhoff's theorem.)

So this is a realistic process by which a black hole could form: Gravitational collapse of a supermassive star. As long as the mass is large enough that the star eventually collapses to within its Schwarzschild radius, we expect that an event horizon would form, and that anything
produced by the star (particles, light, whatever) would be unable to escape from inside the event horizon.

But do we have any observational evidence for black holes? Yes, but necessarily it must all be indirect.

The first class of evidence we have for black holes is the existence of powerful X-ray sources in binary star systems. For example, some binary systems consist of a normal star and an X-ray source. The two objects revolve around each other, as can be determined by the periodic variation in light reaching us from the ordinary star (it sometimes passes behind the other object) and also by the periodic Doppler shift in the light from the ordinary star (sometimes it moves towards us, sometimes away from us). Some of the material from the ordinary star is pulled towards the black hole, and accumulates in what astronomers call an accretion disk. The material in the accretion disk becomes compressed and heated so much that it emits intense X-rays.

If the mass of the X-ray source is greater than a few solar masses (i.e., greater than the Oppenheimer-Volkoff limit) then it is presumed to be a black hole.

The other main class of evidence for black holes is the existence of supermassive compact objects at the centres of galaxies. There is reasonably good evidence that there is a supermassive black hole at the centre of many galaxies, including our own. In the case of our own galaxy, it is thought that there is a black hole of mass about 3 million solar masses. In our galaxy, the motions of stars near the centre have been accurately observed for decades, and their motions fitted by ellipses. The mass of the central object can then be determined by calculations based on Newtonian mechanics (Kepler's third law of planetary motion). The Schwarzschild radius of the black hole at the centre of our galaxy is about 9 million km, which is nothing compared to the size of the galaxy's central bulge. Because we can see objects moving in the galactic centre that are fairly close to the black hole's...
Schwarzschild radius, we are fairly confident that the object is a black hole, and not some other concentration of matter.

Although the evidence is indirect, there is an enormous amount of overlapping evidence pointing in the direction of yes, these are black holes, so although we can't be sure, astronomers have a lot of confidence that they are really there.

Finally, some sources have propagated a misconception that black holes are somehow dangerous vacuum cleaners, that suck up matter in a way that a star, for example, would not. This is not the case; gravitationally, black holes act like any other concentration of mass. Just as the Sun might suck in matter that is close to it, so might a black hole. And just as the planets exist in stable orbits around the Sun, it's quite possible for astronomical objects to orbit black holes in nice stable ways. (The temperature might be different, though!)

**Further experimental tests of Einstein's theory of gravity**

The three classical tests of general relativity are the perihelion shift of Mercury's orbit, gravitational deflection of starlight grazing the edge of the Sun, and gravitational redshift. There are many (many, many) other ways in which Einstein's theory of gravity has been tested, including measurements on pulsars (rapidly rotating neutron stars that send periodic pulses of energy our way, much like the rotating light at the top of a lighthouse), gravitational lensing and other cosmological measurements, and many measurements inside our own solar system (often involving measurements of round trips of radar signals).

**Gravitational waves**

One of the expectations in general relativity is that moving matter should create gravitational waves. Matter (and energy, pressure, etc.) is the source of curvature of spacetime; if matter moves, then spacetime curvature should also change, and these changes (ripples in the fabric of
spacetime, if you wish) should propagate with the speed of light.

There are lots of places in the universe where matter in motion should be creating plenty of gravitational waves that propagate throughout space, much as electromagnetic waves do, and so we should be able to detect them. The problem is that gravitational waves are expected to be extremely weak, and therefore difficult to detect.

The effect of gravitational radiation on the orbits of binary pulsars has been detected, but gravitational waves themselves have not been directly detected. A number of very sensitive detectors (for example, LIGO) have attempted to detect gravitational waves, but so far there have been no detections. A more advanced, sensitive version of LIGO is scheduled to open later this year, and holds out hope that gravitational waves will be detected.

Should physicists become successful, gravitational wave observatories would give us another tool for studying the universe. We shall see.

Currently we have no quantum theory of gravity, and this is still a very active area of research. An aesthetic desire of many physicists is to marry Einstein's theory of gravity with quantum theory to produce a quantum theory of gravity. However, this has proved to be an extremely difficult problem, and so far nobody has done it. Once again, perhaps, we shall see.

Finally, it's worth mentioning that over the years there have been DOZENS of rival theories of gravity, but they have all been discarded based on the results of numerous experiments. It's remarkable that there have been so many rival theories, and that Einstein's theory of gravity has stood the test of time so robustly.

**Modern cosmology**

Besides applying Einstein's theory of gravity to star-like objects (point
mass/black hole, etc.), other physicists also thought about the origin and evolution of the universe as a whole: cosmology. Two of the early researchers who solved Einstein's equations in a cosmological context were Alexandre Friedmann in 1922 and George Lemaitre in 1927. They worked independently, and apparently each did not hear about the work of the other. In 1928, H.P. Robertson independently did very similar work, which was extended in 1936 by A.G. Walker, and nowadays all four share credit for the Friedmann-Lemaitre-Robertson-Walker cosmological models.

Before we describe their models, let's discuss a momentous discovery made by Edwin Hubble in 1929. Using the work of Vesto Slipher, who carefully measured redshifts for a number of nearby galaxies, Hubble and Milton Humason set about measuring the distances to these galaxies. What he found astounded him: There seemed to be a linear relation between the distance of a galaxy and its recession speed; that is, the recession speed of a galaxy is proportional to its distance from us.

The astounding conclusion is that the universe is expanding!
The subsequent collection of data reinforced this conclusion. It's perhaps difficult to appreciate what a momentous discovery this was. The prevailing view since ancient times was that the Earth was at the centre of the universe, and only in Renaissance times did this view begin to be transformed to the Sun being at the centre. But what they thought of as the universe was little more than the solar system. As late as 1920 there was a famous public debate between prominent astronomers Harlow Shapley and Heber Curtis about whether or not faint nebulae were part of our galaxy or were their own "island universes." The shape and size of our own galaxy was still a matter of much discussion and research.

Hubble was firmly on the side of those who believed that faint nebulae were their own galaxies and existed outside our galaxy. When he showed convincingly that the universe was expanding it was quite a bombshell, suddenly disorienting people who had not yet gotten used to the fact that our galaxy was not the entire universe.

At this point, Einstein commented that his introduction of the cosmological constant in his equations was the biggest blunder of his life, for he could have predicted the expansion of the universe 14 years before Hubble's discovery. But really it was not a blunder at all, and the possibility of an expanding universe was contained in the cosmological models of Friedmann, Lemaître, Robertson, and Walker. (However, this early work was mainly ignored as purely abstract speculation, because "everyone knew" that the universe is static.) In fact, the cosmological models fit neatly into three families, based on the possible values of the overall density of matter and energy in the universe:
Each class of model universes is associated with a sign of the overall curvature of spacetime, either positive, zero, or negative.

The Big Bang Theory

You'll notice that all of the models shown in the diagram above have a beginning, what is now called the Big Bang. This class of models is now standard, although they fought it out hard with a class of models called Steady State until the mid 1960s. (Steady State theories (also called continuous creation models) were first proposed by James Jeans in the 1920s, and later championed by Fred Hoyle, Geoffrey Burbidge, and Jayant Narlikar.) The decisive observation that convinced most astronomers of the Big Bang theory of the universe was the discovery of cosmic microwave background radiation by Arno Penzias and Robert Wilson in 1964, confirming an important prediction of the Big Bang theory.
The Big Bang theory is nowadays mainstream cosmology, and the theory has been developed to a very high degree mathematically, and intersects with elementary particle physics and nuclear physics. The theory predicts that the early universe was in a very dense, hot, compact state, and that in the early universe the only elements created were hydrogen, helium, and lithium, in very specific proportions. These predictions, and other predictions of the theory, are very well confirmed by observations.

Olbers's paradox: Another piece of evidence for the Big Bang theory.

Recent observations suggest that the expansion of the universe is accelerating, and this has caused astronomers and cosmologists to rethink the details of their models. This is currently a very active area of research, and new discoveries are being made on an ongoing basis.

One of the important areas of research is to search for detail in the cosmic microwave background radiation. For example, is it isotropic? Are there any details that would help us learn more about the early universe? Several dedicated telescopes are working on this: COBE, WMAP, and Planck are three of note.

Some cosmological puzzles

1: Dark matter

Galactic rotation curves
The "expected" curve is based on Newtonian mechanics (Kepler's third law)

Candidates for dark matter: abandoned TV sets, MACHOs, WIMPs, etc.

2: Dark energy

The overall balance of "forces" in the universe: gravitation is attractive, but the cosmological constant is repulsive; which will win in the end? The acceleration of the universe's expansion suggests that the cosmological constant is winning.

Nowadays the value of the cosmological constant is said to derive from "dark energy." What is its nature? What is its origin? Nobody is sure, but it's currently a very active area of research.

A natural connection is to relate the cosmological constant to dark energy, but not everyone does this. In quantum field theory, one can calculate the "vacuum energy" and associate this with the origin of dark energy (and therefore the value of the cosmological constant), but all such calculations predict an enormous value for the cosmological
constant, and measured values are quite close to zero. So far nobody has figured out how to reconcile this extreme disagreement between prediction and observation, so stay tuned (or, better yet, get involved!).

These problems certainly highlight a very nice aspect of cosmological research, in that it combines understanding of very many fields of physics: special and general relativity, classical mechanics, quantum mechanics, thermal physics, fluid mechanics, electricity and magnetism, nuclear physics, particle physics, even aspects of chemistry, biology, and philosophy; you name it, it's in there! This makes astronomy, astrophysics, and cosmology difficult subjects, but also very stimulating and far-reaching.

**Unsolved problems in modern cosmology**

(To be completed)

Mysteries: quasars, active galaxies

Mysteries: fine-tuning puzzles

Inflationary cosmological models

Dark matter

Dark energy

The ultimate fate of the universe

3: *The landscape of modern physics*

(To be completed.)
Review of the fundamental themes of modern physics

Repetition of the Griffiths diagram

The four forces revisited

Gauge theories, QFT

Symmetry

The theme of unification revisited

String Theory, M-Theory, etc.

The active areas of research nowadays

Lots of unsolved problems