Relativity and Cosmology:

The development of the modern cosmological worldview

Waseda University, SILS,
History of Modern Physical Sciences

The End of "Classical Physics"

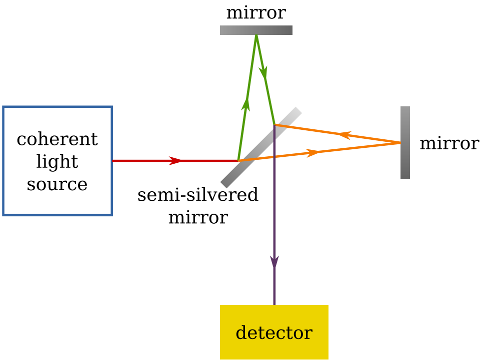
At the end of the century, L. Boltzmann gave a lecture arguing that the Laplacian-Newtonian program had been undermined by the 19th century theories, and that these new theories were in turn under threat.

Boltzmann, Lecture, 1899

"Many may have thought at the time of Lessing, Schiller and Goethe, that by constant further development of the ideal modes of poetry practiced by these masters dramatic literature would be provided for in perpetuity, whereas today one seeks quite different methods of dramatic poetry and the proper one may well not have been found yet. Just so, the old school of painting is confronted with impressionism, secessionism, pleinairism, and classical music with music of the future. Is not this last already out-of-date in turn? We therefore will cease to be amazed that theoretical physics is no exception to this general law of development."

The Michelson-Morley Experiments

From the late 1880s, Michelson and Morley, and then Miller, carried out a series of precise measurements of the speed of two beams of light traveling at right angles to one another. If the light were moving in a medium, there should have been detectible time differences in the two directions. The experiment was generally understood to have given a null result. Miller continued to carry out the experiment into the 1930s.



Non-Classical Mechanics

There were a number of papers in relativity before Einstein published – for example, Lorentz and FitzGerald had worked out the famous transformation equations to explain the null result of the Michelson-Morley experiment.

In 1902, H. Poincaré (1854–1912) stated that there is no absolute space, no absolute time, and no direct intuition of simultaneity. Moreover, that it might be possible to enunciate mechanical facts with reference to non-Euclidean space.

Poincaré, 1904

It should be possible to "construct a whole new mechanics, of which we only succeed in catching a glimpse, where inertia increasing with velocity, the velocity of light would become an impassible limit. The ordinary mechanics, more simple, would remain a first approximation, since it would be true for velocities not too great, so that."

The Special Theory of Relativity

In 1905, while still working as a technical clerk in the Bern patent office, A. Einstein (1879–1955) set out a new theory of the electromagnetic field, which was based on a fundamental reconceptualization of time and space.

The special theory was based on two fundamental principles:

- "Light is always propagated in empty space with a definite velocity, c, which is independent of the state of motion of the emitting body."
- The laws of electrodynamics and optics will be the same for all inertial reference frames.

Hence, distances and times have to be determined on the basis of the constant $c \approx 300,000$ km/s (2.9×10^5). In order to do this, the observers use theoretical *rigid yardsticks* and *light clocks*.

Time, Space, and Simultaneity

Although the constancy of the speed of light was also held in traditional, ether-based, classical physics, the two principles together implied that concepts such as time, space and mass must be understood in terms of a given frame of reference – and will change relative to a change of frame.

Specifically, in reference frames moving relative to the observer, time slows down and distances contract, by a factor involving $(1 - (v^2/c^2))^{1/2}$. Moreover, sets of events that are simultaneous in one inertial frame will not appear to be simultaneous to observers in another frame that is moving relative to the initial frame.

Einstein produced a set of equations that could be used to transform between "events" – time-space points – in different frames. This was generalized, in 1908, by the mathematician H. Minkowski (1864–1909), who showed that time-space could be described as a 4-dimensional manifold, and the *time-space interval* between any two events could be stated with a metric.

The General Theory of Relativity

For the next 10 years, Einstein and a number of others worked on extending the theory of relativity so that gravity could be described as equivalent to the effect of a frame moving at a varying velocity. The basic principles were that the laws of physics are the same in every reference frame, and that we have no way of deciding on a privileged frame – so that, for example, we cannot tell the difference between gravity and acceleration.

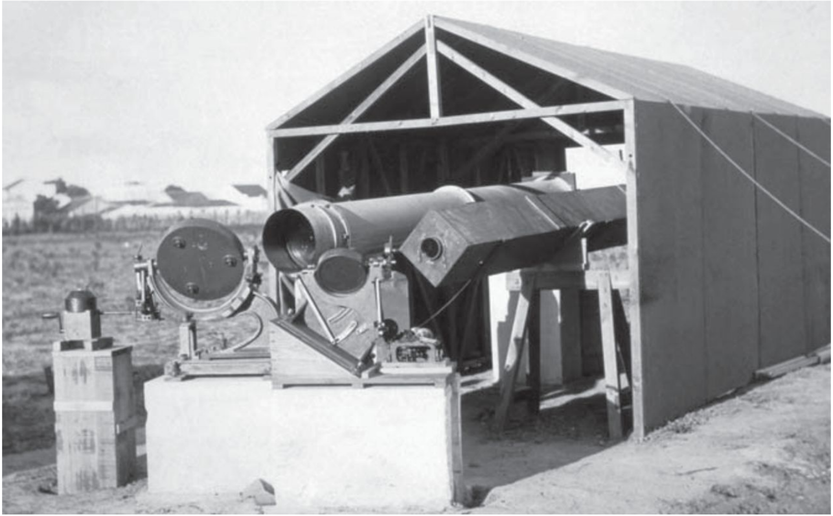
On this basis, in 1915, Einstein published his gravitational field equations, which described how the geometry of space-time relates to the matter and energy of the field. This was a tensor field theory in which gravitation was the result of the space-time field, similar to the way that electromagnetic force was the result of the electromagnetic field. In his 1915 paper, he showed how his gravitational theory could solve the problem of the perihelion of Mercury. He then went on to predict gravitational red-shifting for light leaving a massive object, and the deflection of light in a gravitational field in the vicinity of a massive object.

Testing the Theory

The early reception of the theory was purely technical – only mathematically minded physicists payed it any attention. This changed following an expedition to observe the total solar eclipse of 1919, organized by Arthur Eddington (1882–1944).

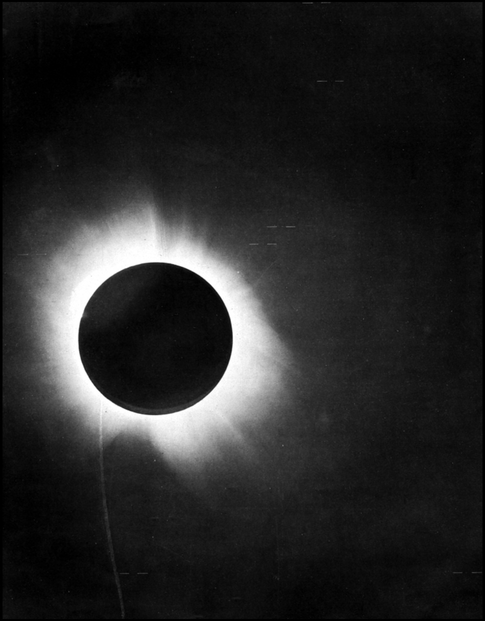
Einstein calculated that under Newton's theory stars near the sun would be deflected 0.8 seconds of arc ($0.0000\bar{2}^\circ$), while under his theory they would be deflected by 1.7 seconds ($0.00047\bar{2}^\circ$).

The solar eclipse was observed in two different sites – Brazil and West Africa. Photographs were taken of a field of stars before and during the solar eclipse, and the locations of the stars around the sun were measured and compared. There were three different telescopes which took 28 photos, most of which were of poor quality. The majority of the plates were discarded due to the suspicion that there were systematic errors with the telescope. The eight good plates from Brazil produced a mean displacement of 1.98 seconds of arc.



16-inch telescope and large coelostat, left, and smaller 4-inch telescope in a square box with its much smaller coelostat, right, Sobral, Brazil

“In Plate 1 is given a half-tone reproduction of one of the negatives taken with the 4-inch lens at Sobral. This shows the position of the stars, and, as far as possible in a reproduction of this kind, the character of the images, as there has been no retouching. A number of photographic prints have been made and applications for these from astronomers, who wish to assure themselves of the quality of the photographs, will be considered and as far as possible acceded to.”



Relativistic Cosmology, Static Universes

Because the general theory of relativity offered a new, and sophisticated, mathematical theory of gravity, there were almost immediately attempts to see how this would play out in terms of cosmology.

In 1917, Einstein published his "Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie," in which he proposed a static universe that has a homogeneous and symmetrical distribution of matter, in a closed and unbounded space of constant positive curvature. In order to keep the universe stable, he added a term for an unknown force that acted contrary to gravity that he called λ, and which he later abandoned. (λ is related to what we call the *cosmological constant*.)

A few months later, the Dutch astronomer Willem de Sitter (1872–1934) proposed a solution to the equations for a closed static universe with *no mass* – again involving λ. In the de Sitter universe, the spectra of distant light sources is red-shifted.

A Dynamic Universe

In the early 1920s, Alexander Friedmann (1888–1925), a Russian meteorologist, expanded on the work of Einstein and de Sitter, and began to explore the mathematics of solutions to the cosmological equations that involved a non-static universe.

After showing that the Einstein and de Sitter models were the only possible solutions for a static universe, he then began to explore the possibilities of solutions that have curvatures that change as a function of time – considering both positive closed spaces, as well as negative open spaces. In this regard, he introduced a scale function, $R(t)$, such that if at some time, t_0, the distance between two points is 1, at some other time, t, it will be $R(t)$.

He was particularly interested in the case of an expanding universe, which led to an equation relating $R(t)$ to the parameter λ, and to a specification of a beginning of time – when all of space was a single point.

Apparent Doppler Effect, and the Primeval Atom

Throughout the 1920s, there was growing observational evidence that distant objects – which many thought were probably different galaxies – were more and more redshifted.

In 1927, the Belgian priest and physicist George Lemaître (1894–1966) produced an independent study of the relativistic equations, postulating an expanding, closed universe of positive curvature, and explicitly working to explain the redshift observed for distant objects.

Lemaître, "A Homogenous Universe of Constant Mass…," 1927

"The receding velocities of extra-galactic nebulae are a cosmological effect of the expansion of the universe."

That is, he considered that the redshifts were caused by the motion of the extra-galactic objects with the expansion of space itself from an original "primeval atom."

Albert Einstein, Hideki Yukawa, John Wheeler and Homi J. Bhabha hiking in the woods, Institute for Advanced Studies, Princeton, 1947

Our Place in the Universe

In the first part of the 20th century, observational astronomers developed techniques for determining the absolute distances of stars and nebulae, based on the discovery of what came to be called *standard candles*. This allowed the development of what we call the *cosmic distance ladder*.

By building up a larger and larger database of the distances of different objects it became possible to say something precise about the overall structure and size of the Milky Way, and to locate our system – the Solar System – within it.

Finally, by extending the distance ladder to more distant objects, it became clear that many nebulae are outside the Milky Way – extragalactic objects. This, finally, established empirically that the Milky Way is just one of many galaxies, and that most of the other galaxies appear to be moving away from us.

Standard Candles

In 1908, Henrietta Leavitt (1868–1921), of Harvard College Observatory, noticed that 16 variable stars in the Magellanic Cloud exhibit a fixed relationship between their periods of variation and luminosity. On the assumption that they were all at roughly the same distance, this allowed her to establish their *relative luminosity*. These became known as *cepheid variables* and by finding such variables at distances that could be measured by parallax, their *absolute luminosity* could be determined and it became possible to find the radial distance of a large number of stars and clusters.

This was one of the first examples of measuring distance using what astronomers call a *standard candle* – that is, an object whose absolute luminosity is known. In the following decades more details of the cepheid variables allowed for greater distances to be measured, and eventually Ia supernovae (white dwarf supernovae) were used as standard candles for very distant galaxies.

The Shape and Size of the Milky Way

The method of cepheid variables was used to determine the distances to a number of distant objects, and in 1918, Harlow Shapley (1885–1972), at Harvard College Observatory, produced a formula for computing distances using cepheids.

On this basis, Shapley compiled a large data set of cepheids and began to develop a picture of the shape of the Milky Way as a flattened, spiral galaxy. By a series of bold conjectures about the distance ladder, he calculated the distance of a number of remote clusters as 200,000 light years, and the diameter of the whole Milky Way as 300,000 light years. (Now 150–200kly.)

This was much larger than most people had imagined and it lead Shapley and his colleague Adriaan van Maanen (1884–1946), at Mount Wilson Observatory, to conclude that the Milky Way made up the entire universe.

In 1912, Vesto Slipher (1875–1969), at the Lowell Observatory, used the method of cepheid variables to determine the Doppler shift of M31, the Andromeda nebula, and a number of other nearby nebulae. Although M31 was blueshifted, the majority of nebuae seemed to be redshifted.

In 1923, Edwin Hubble (1889–1953), found a new cepheid in M31 that allowed him to compute the absolute distance of the system. The value he came up with was around 900kly, which was much to far away for it to be part of the Milky Way. (Now around 2.5mly.)

Although not everyone was convinced, Hubble and many others came to the conclusion that M31, the Andromeda galaxy, was a *galaxy* of the same scale as the Milky Way.

The Hubble Law

Following the work of Slipher, Hubble and his colleague Milton Humason (1891–1972) used the 100-inch telescope at Mount Wilson to add to the set of known redshifts for distant galaxies.

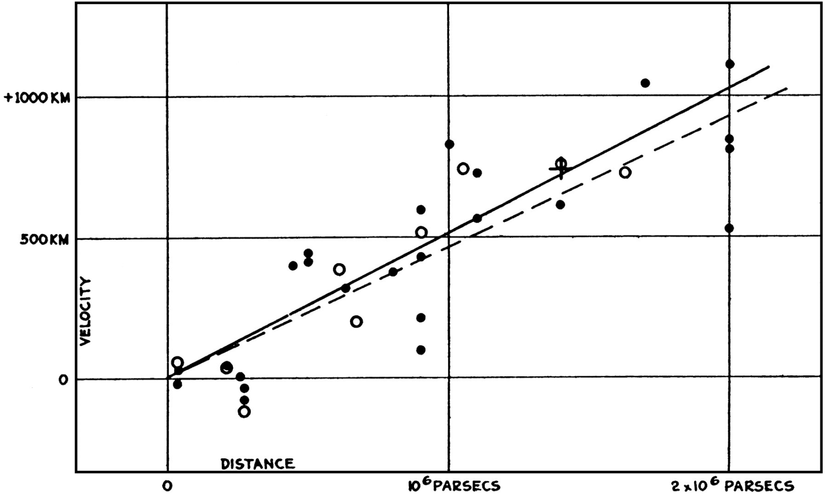
Hubble, "A Relation between Distance and Radial Velocity…," 1929

"The results establish a roughly linear relation between velocities and the distances among nebulae for which velocities have been previously established, and the relation appears to dominate the distribution of velocities."

Where v is velocity, d distance, and H some empirical parameter, the law was expressed by others as

$$v = Hd.$$

Although Hubble stated the linear law connecting redshifts and distances, he balked at an assertion that the universe is actually expanding and referred to the redshifts as "apparent velocities."



Hubble's velocity-distance diagram of 1929

Edwin Hubble stands by the 48-inch telescope at the Palomar Observatory, San Diego, California

An Expanding Universe

Following Hubble's paper, theorists began to put more weight on relativistic cosmological models that involve mass, a cosmological constant, and an expanding universe. By the early 1930s, Eddington, de Sitter, and many others had begun to take seriously the various models predicted by the Friedmann-Lemaître equations.

Overview studies, by de Sitter and others, showed that there were a number of possible solutions to the equations that predicted expansion – from a point, from a volume, increasing indefinitely, increasing to a maximum, returning to a point, and so on.

Eddington, *The Expanding Universe*, 1933

"I shall speak of the theoretical work of Einstein of Germany, de Sitter of Holland, Lamaître of Belgium. For observational data I turn to the Americans, Slipher, Hubble, Humason recalling however that the vitally important datum of distance is found by a method which we owe to Hertzsprung of Denmark... My subject disperses the galaxies, but it unites the earth."

The History of the Cosmos

The view of the cosmos as developing from an initial singularity went through a number of changes in the 20th century. The first theories, in the 1910s and 20s, were relativistic and geometrical, but with little in the way of nuclear physics. The early period of the work on the physics of the early universe took place in the 1930s and 40s, and included attempts to explain the distribution of elements in the known universe.

In the 1950s, there was a major challenge to this perspective, known as the steady-state theory, which was based on a number of important principles, and in the context of which physicists worked out a theory of the production of heavy elements in the stars.

Finally, in the 1960s, a new synthesis was worked out involving a hot big bang and the production of heavy elements in stars. This theory predicted that there would be a background electromagnetic radiation, detectable in all directions. This cosmic background radiation was discovered independently by radio astronomers.

The Hot Big Bang

The early theories of a primeval atom had been geometrical and relativistic, but had not considered any of the physical consequences of pressing everything into such a small space. This began to change in the and 1940s and 50s with the work of the Russian-American George Gamow (Gamov Georgiy, 1904–1968) and his colleagues R. Alpher (1921–2007) and R. Herman (1914–1997).

They started by working on stellar nucleosynthesis and the question of whether or not the abundances of elements in the universe could be explained on the basis of the nuclear physics of stars. Gamow proposed thinking of the elements as formed in the early universe.

Gamow and Flemming, "'Report on…," 1942

"It seems … probable that the elements originated in a process of explosive character, which took place at the 'beginning of time' and resulted in the present expansion of the universe."

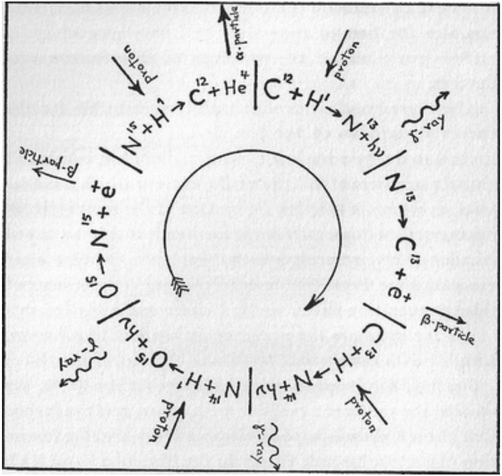
The Alchemy of Stars

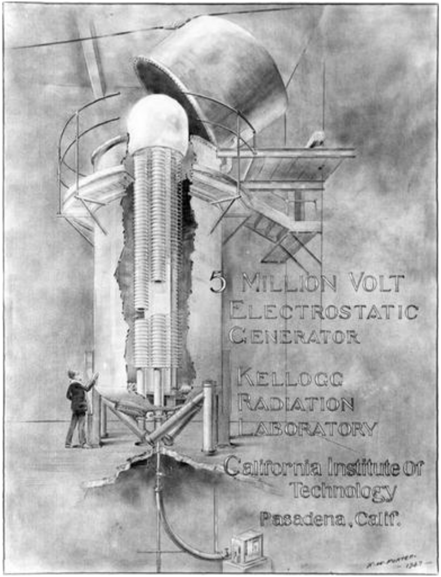
In 1938, C.F. von Weizsäcker (1912–2007) published a theory on the nuclear fusion powering stars. He proposed that stars of less than 1.5 solar masses were driven by mostly proton-proton reactions, while larger stars would involve a cycle of chain reactions in which carbon (C), nitrogen (N), and oxygen (O) act as catalysts.

Following this, Hans Bethe (1906–2005), along with S. Chandrasekhar (1910–1995) and Bengt Strömgren (1908–1987), worked out the details of the so-called CNO cycle in much more quantitative detail.

Their theory started from the assumption of large quantities of hydrogen acting under gravitational pressure, and considered the nuclear reactions of a series of elements when bombarded with protons. Bethe's highly polished work was well received, but both Weizsäcker and Bethe agreed that the theory could not explain the abundances of heavy elements.

Gamow's illustration of the Bethe-Weizäker CNO cycle, from *The Birth and the Death of the Sun*, 1940. It shows a series of proton captures, starting at C^{12}.





Design drawing of the electrostatic accelerator at the Kellogg Institute for Radiation Physics at Caltech, Pasadena, California. The machine was used to study the nuclear reactions in the CNO cycle. Russell Porter, 1947.

The experimental work of this machine was interrupted by the war, when most nuclear physicists in the US were recruited into the Manhattan Project. After the war, with a much deeper basis in nuclear physics, a number of researchers worked out the details of the nuclear reactions.

The Steady State Challenge

In the 1950s, especially in Britain, a number of theorists challenged the concept of a big bang universe. In the 1930s, Edward Authur Milne (1896–1950) had endorsed the concept of universe expanding into Euclidean space, and he rejected general, but not special, relativity. Although Milne's own ideas were not much adopted, his general approach influenced the next generation.

Three Cambridge physicists, Herman Bondi (1919-2005), Thomas Gold (1920–2004) and Fred Hoyle (1915–2001), proposed a pair of theories involving an eternal universe, in which galaxies are receding from one another, with a steady creation of matter – say hydrogen atoms – in the void, *ex nihilo*. They could derive the density of matter, the metric of space, and the rate of expansion – but they had to deny the 2nd law of thermodynamics.

Methodological Considerations

The steady-state theorists were motivated by a number of methodological issues. For example, they were concerned with what was called the *age paradox* – which was the fact that at that time the age of the universe, as calculated from the Hubble parameter known at that time, was much less the age of the earth, as calculated by geologists from radiation dating.

Another concern was that the relativistic equations were so general that solutions could be found for nearly any set of observations, and hence failed to meet Karl Popper's criteria for being scientific.

Finally, they found it suspicious that although from our perspective space is homogenous and isotropic – the *cosmological principle* – according to the big bang theory, we are in a very special position in time. Against this they proposed the *perfect cosmological principle* – which is the claim that we are neither at a privileged position nor in a privileged epoch.

The Construction of Heavy Elements

One of the major concerns of the steady-state theorists was to explain the production of heavy elements in stellar nucleosynthesis. For this reason, Hoyle turned his attention to these processes.

Hoyle entered into a collaboration with the nuclear physicist William Fowler (1911–1995) who was experimentally studying nucleosynthesis at CalTech. In this way, they were able to confirm experimentally a number of the processes that Hoyle was proposing from a theoretical perspective. In 1957, along with the English-American astronomers Margaret Burbidge (1919–2020) and Geoffrey Burbidge (1925–2010), Hoyle and Fowler produced a monograph-length study called "Synthesis of the Elements in Stars," which became a classic in the field and is now known as the B^2FH paper. Their theory showed the importance of carbon in making possible the synthesis of heavier elements.

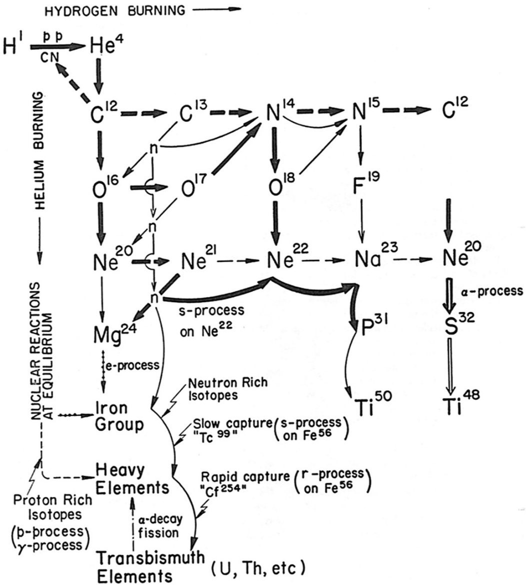


Diagram showing the production of elements in stars in the B^{2}HF paper, 1957

Gamow and his Wife Spoof the B^{2}HF Paper.

Barbara and George Gamow, "New Genesis"

"In the beginning, God created radiation and ylem, without shape, into which nucleons were rushing.

And then, let there be mass 1, 2, 3 … up to 92 elements.

But when he looked back, he had missed calling for mass five, and, naturally, no heavier elements could be formed.

And then he said, let there be Hoyle, and he told him to make heavy elements in any way he pleased.

And Hoyle decided to make heavy elements in stars and to spread them around by supernova explosion."

But it was difficult to explain the production of heavy elements any other way and the results of the B^{2}HF paper were incorporated into the next generation of big bang theories, and became a key piece of the argument for the big bang.

From the 1960s, cosmology became accepted as a scientific discipline and its practitioners began to call themselves cosmologists. A number of groups, independently, rederived the the hot big bang theory, now in more computational and physical detail, and with more precise predictions.

In the early 60s, two separate groups independently made the prediction of background radiation from the big bang. In 1962, the Soviet astrophysicists A. Doroshkevich (1937–) and A. Novikov (1935–) predicted the radiation and even suggested that the Holmdel Horn antenna could be used to find it. In 1964 the Princeton the physicist R. Dicke (1916–1997) suggested that his former student P.J.E. Peebles (1935–) calculate the background radiation from the big bang. Peebles and his team produced an estimate, and began an observational project to see if it could be detected.

Background Radiation

Although cosmic background radiation had been predicted by a number of theorists, the actual discovery of the effect was accidental, insofar as Arno Penzias (1933–) and Robert Wilson (1936–) did not set out with this goal.

In 1963, Penzias and Wilson were working with the Holmdel Horn Antenna at Bell Laboratories that had been built for communication, and which they now wished to repurpose for astronomy. What they found, to their dismay was that there seemed to be a slight radio hiss coming from every direction and even empty spaces in the night sky. After trying systematically to reduce interference, they continued to find background radiation of around 3.5° K (−269.7° C).

After consulting various colleagues, including Dicke, they came to understand that this effect was predicted by the hot big bang models, and they published their findings in 1965.

Arno Penzias (right) and Robert Wilson (left) in front of the Bell Laboratories horn radio antenna, Crawford Hill, New Jersey.

Black Holes

Although “dark stars” were postulated in the 18th century, the earlist mathematical theories of these objects were developed from general relativity by Karl Schwarzchild (1843–1916) and S. Chandrasekhar. The former showed, in 1916, that if the radius of an object is smaller than $r_s = 2GM/c^2$, then no radiation would be able to escape from it. The latter argued, in 1930, that if a star is larger than about 1.4 times the size of the sun, when it runs out of nuclear fuel it will collapse into something smaller than a white dwarf – that is, as we, now say, either a neutron star or a black hole.

In 1939, J.R. Oppenheimer (1904–1967) and H. Snyder (1913–1962) showed that a sufficiently large star would collapse to produce a singularity cut off from us by an *event horizon*. The term “black hole” was introduced by John Wheeler (1911–2008) in 1968 in part of a general attempt to rejuvenate these objects as a more serious topic of physical investigation. (The terminology was originally rejected by the editors of *Physical Review*, as being too “obscene.”)

Supermassive Black Holes

In the 1980s, there were speculations that there might be supermassive black holes at the centers of some galaxies, with masses up to millions of times that of the sun. Radio astronomers began to study the radiation coming from the centers of objects such as M87, in Virgo, and M31, the Andromeda Galaxy.

In 1988, C. Townes (1915–2015) and R. Genzel (1852–), at UC Berkeley, suggested that there was a supermassive black hole at the center of our galaxy, and used the infrared spectra to study the motion of ionized hydrogen clouds around the center of the Milky Way.

Confirmation for this position came in 2002, when Genzel published some 10 years of his observations of the rapid and highly eccentric orbits of stars close to the galactic center. It was possible to calculate that these stars were orbiting a body of some 3,000,000 times the solar mass.

Dark Matter

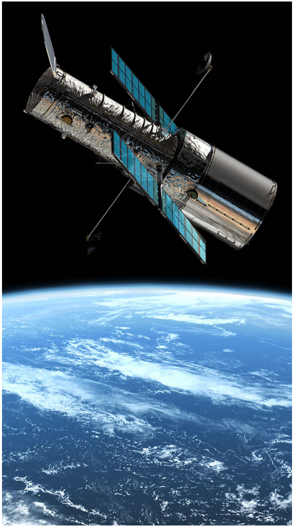
In the 1930s, Fritz Zwicky (1898–1974), at the Mount Wilson Observatory, had shown that the visible matter in the Coma cluster is much less than the matter predicted gravitationally by the motions. Both the observations and computations were confirmed, but there was little work in this area. In the 1970s, Vera Rubin (1928–2016), William Ford (1931–), and others at the Carnegie Institution, showed that the visible matter in galaxies is rotating as though it is being carried around in much more mass. Evidence from a large number of different observations indicates that there is a huge amount of matter that does not emit electromagnetic radiation, but which interacts gravitationally with visible matter.

This matter is generally called "dark matter," because it does not register in any electromagnetic instruments. It is currently believed that there is much more of it than visible, or baryonic, matter, but there is still no established position on what it is.

The Hubble Space Telescope

The Hubble was launched in 1990, but the mirror had the wrong shape. In 1993, astronauts repaired the telescope by installing, essentially, corrective lenses. Since that time, it has been in continuous operation, having been repaired and updated by astronauts a number of times.

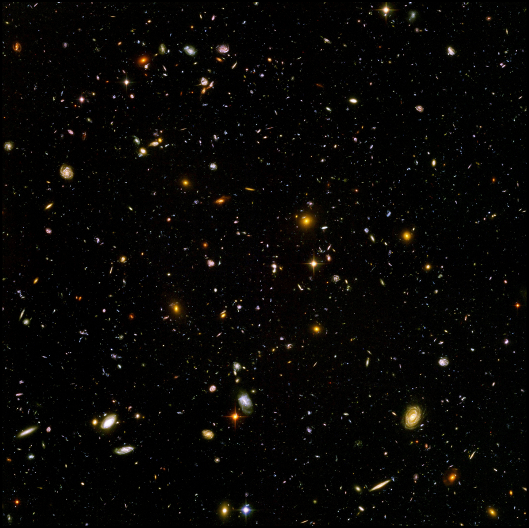
It is arguably one of the most important single pieces of scientific equipment ever deployed and is a good example of what can be done with Big Science. Almost all of the images that we have become accustomed to seeing of deep space were produced by the Hubble. (And now from the James Web Telescope.)



The first Hubble Deep Field, produced in 1995, collected light for 10 days from an empty dot of sky.

The Hubble Ultra Deep Field, 2004, collected light for 114 days from an otherwise blank section of sky located southwest of Orion in the southern-hemisphere constellation Fornax.

It revealed around 10,000 galaxies from 13 billion (1.3×10^{10}) years ago.



Hubble Ultra Deep Field

The Hubble eXtreme Deep Field was published in 2012. It collected light for about 23 days over the course of 10 years.

It contains around 5,500 galaxies from 13.2 billion (1.32×10^{10}) years ago.

Hubble eXtreme Deep Field

Dark Energy

In the 1990s, two teams of astronomers at Harvard University, and UC Berkeley used the Hubble Telescope to determine the Hubble parameter, H, for galaxies many times more distant than those observed by Hubble himself. In 1998, by using Ia supernovae as a standard candle to determine the *distance*, $v = Hd$ was used to show that the observed redshift for distant galaxies was less than that predicted by the law. This implies that the expansion of the universe is accelerating as one moves forward in time, or closer to us. This was an unexpected, and shocking, realization. At this time the standard thinking had been that the expansion was slowing.

One response to this was a renewal of the relativistic models with a positive cosmological constant, λ. This was now interpreted as a non-radiating – that is, dark – energy that propels the expansion of the universe. In this view, λ is a quantum, zero-point, vacuum energy that is very weak, but as the density of the universe dropped, it became relatively stronger and began to drive the expansion.

Overview

The theories of relativity transformed the way that we think about such basic concepts as time, space, simultaneity and mass. The general theory of relativity provided a new theory of gravity, in which the gravitational force was understood as an effect of the curvature of the space-time field.

The modern conception of cosmology – a universe expanding from a hot big bang – is based on the discovery of the changing redshifting of distant galaxies and the cosmic microwave radiation. But both of these were recognized as discoveries in processes in which theory was as important as observation.

Throughout the 20th century, cosmologists became increasingly convinced that there is far more non-radiating matter than radiating matter, and that the energy of the vacuum of space has come to play an important role in the development of the universe. There is still no generally accepted theory explaining dark matter and dark energy.