
БУДОВА РЕЧОВИНИ

Stephen W. Hawking

Elementary Particles and the Forces of Nature

1. Aristotle believed that all the matter in the universe was made up of four basic elements – earth, air, fire, and water. These elements were acted on by two forces: gravity, the tendency for earth and water to sink, and levity, the tendency for air and fire to rise. This division of the contents of the universe into matter and forces is still used today. Aristotle believed that matter was continuous, that is, one could divide a piece of matter into smaller and smaller bits without any limit: one never came up against a grain of matter that could not be divided further. A few Greeks, however, such as Democritus, held that matter was inherently grainy and that everything was made up of large numbers of various different kinds of atoms. (The word *atom* means “indivisible” in Greek.) For centuries the argument continued without any real evidence on either side, but in 1803 the British chemist and physicist John Dalton pointed out that the fact that chemical compounds always combined in certain proportions could be explained by the grouping together of atoms to form units called molecules. However, the argument between the two schools of thought was not finally settled in favor of the atomists until the early years of this century. One of the important pieces of physical evidence was provided by Einstein. In a paper written in 1905, a few weeks before the famous paper on special relativity, Einstein pointed out that what was called Brownian motion – the irregular, random motion of small particles of dust suspended in a liquid – could be explained as the effect of atoms of the liquid colliding with the dust particles.

2. By this time there were already suspicions that these atoms were not, after all, indivisible. Several years previously a fellow of Trinity College, Cambridge, J. J. Thomson, had demonstrated the existence of a particle of matter, called the electron, that had a mass less than one thousandth of that of the lightest atom. He used a setup rather like a modern TV picture tube: a red-hot metal filament gave off the electrons, and because these have a negative electric charge, an electric field could be used to accelerate them toward a phosphor-coated screen. When they hit the screen, flashes of light were generated. Soon it was realized that these electrons must be coming from within the atoms themselves, and in 1911 the New Zealand physicist Ernest Rutherford finally showed that the atoms of matter do have internal structure:

they are made up of an extremely tiny, positively charged nucleus, around which a number of electrons orbit. He deduced this by analyzing the way in which alpha-particles, which are positively charged particles given off by radioactive atoms, are deflected when they collide with atoms.

3. At first it was thought that the nucleus of the atom was made up of electrons and different numbers of a positively charged particle called the proton, from the Greek word meaning “first,” because it was believed to be the fundamental unit from which matter was made. However, in 1932 a colleague of Rutherford’s at Cambridge, James Chadwick, discovered that the nucleus contained another particle, called the neutron, which had almost the same mass as a proton but no electrical charge. Chadwick received the Nobel Prize for his discovery.

4. Up to about thirty years ago, it was thought that protons and neutrons were “elementary” particles, but experiments in which protons were collided with other protons or electrons at high speeds indicated that they were in fact made up of smaller particles. These particles were named quarks by the Caltech physicist Murray Gell-Mann, who won the Nobel Prize in 1969 for his work on them. The origin of the name is an enigmatic quotation from James Joyce: “Three quarks for Muster Mark!” The word *quark* is supposed to be pronounced like *quart*, but with a *k* at the end instead of a *t*, but is usually pronounced to rhyme with *lark*.

5. There are a number of different varieties of quarks: there are six “flavors,” which we call up, down, strange, charmed, bottom, and top. The first three flavors had been known since the 1960s but the charmed quark was discovered only in 1974, the bottom in 1977, and the top in 1995. Each flavor comes in three “colors,” red, green, and blue. (It should be emphasized that these terms are just labels: quarks are much smaller than the wavelength of visible light and so do not have any color in the normal sense. It is just that modern physicists seem to have more imaginative ways of naming new particles and phenomena – they no longer restrict themselves to Greek!) A proton or neutron is made up of three quarks, one of each color. A proton contains two up quarks and one down quark; a neutron contains two down and one up. We can create particles made up of the other quarks (strange, charmed, bottom, and top),

but these all have a much greater mass and decay very rapidly into protons and neutrons.

6. We now know that neither the atoms nor the protons and neutrons within them are indivisible. So the question is: what are the truly elementary particles, the basic building blocks from which everything is made? Since the wavelength of light is much larger than the size of an atom, we cannot hope to “look” at the parts of an atom in the ordinary way. We need to use something with a much smaller wavelength. Quantum mechanics tells us that all particles are in fact waves, and that the higher the energy of a particle, the smaller the wavelength of the corresponding wave. So the best answer we can give to our question depends on how high a particle energy we have at our disposal, because this determines on how small a length scale we can look. These particle energies are usually measured in units called electron volts. (In Thomson’s experiments with electrons, we saw that he used an electric field to accelerate the electrons. The energy that an electron gains from an electric field of one volt is what is known as an electron volt.) In the nineteenth century, when the only particle energies that people knew how to use were the low energies of a few electron volts generated by chemical reactions such as burning, it was thought that atoms were the smallest unit. In Rutherford’s experiment, the alpha-particles had energies of millions of electron volts. More recently, we have learned how to use electromagnetic fields to give particles energies of at first millions and then thousands of millions of electron volts. And so we know that particles that were thought to be “elementary” thirty years ago are, in fact, made up of smaller particles. May these, as we go to still higher energies, in turn be found to be made from still smaller particles? This is certainly possible, but we do have some theoretical reasons for believing that we have, or are very near to, a knowledge of the ultimate building blocks of nature.

Every-thing in the universe, including light and gravity, can be described in terms of particles. These particles have a property called spin.

One way of thinking of spin is to imagine the particles as little tops spinning about an axis. However, this can be misleading, because quantum mechanics tells us that the particles do not have any well-defined axis. What the spin of a particle really tells us is what the particle looks like from different directions. A particle of spin 0 is like a dot: it looks the same from every direction. On the other hand, a particle of spin 1 is like an arrow: it looks different from different directions. Only if one turns it round a complete revolution (360 degrees) does the particle look the same. A particle of spin 2 is like a double-headed arrow: it looks the same if one turns it round half a revolution (180 degrees). Similarly, higher spin particles look the same if one turns them through smaller fractions of a complete revolution. All this seems fairly straightforward, but the remark-able fact

is that there are particles that do not look the same if one turns them through just one revolution: you have to turn them through two complete revolutions! Such particles are said to have spin $\frac{1}{2}$.

All the known particles in the universe can be divided into two groups: particles of spin $\frac{1}{2}$, which make up the matter in the universe, and particles of spin 0, 1, and 2, which, as we shall see, give rise to forces between the matter particles. The matter particles obey what is called Pauli’s exclusion principle. This was discovered in 1925 by an Austrian physicist, Wolfgang Pauli – for which he received the Nobel Prize in 1945. He was the archetypal theoretical physicist: it was said of him that even his presence in the same town would make experiments go wrong! Pauli’s exclusion principle says that two similar particles can-not exist in the same state; that is, they cannot have both the same position and the same velocity, within the limits given by the uncertainty principle. The exclusion principle is crucial because it explains why matter particles do not collapse to a state of very high density under the influence of the forces produced by the particles of spin 0, 1, and 2: if the matter particles have very nearly the same positions, they must have different velocities, which means that they will not stay in the same position for long. If the world had been created without the exclusion principle, quarks would not form separate, well-defined protons and neutrons. Nor would these, together with electrons, form separate, well-defined atoms. They would all collapse to form a roughly uniform, dense “soup.”

A proper understanding of the electron and other spin- $\frac{1}{2}$ particles did not come until 1928, when a theory was proposed by Paul Dirac. Dirac’s theory was the first of its kind that was consistent with both quantum mechanics and the special theory of relativity. It explained mathematically why the electron had spin- $\frac{1}{2}$; that is, why it didn’t look the same if you turned it through only one complete revolution, but did if you turned it through two revolutions. It also predicted that the electron should have a partner: an anti-electron, or positron. The discovery of the positron in 1932 confirmed Dirac’s theory and led to his being awarded the Nobel Prize for physics in 1933. We now know that every particle has an antiparticle, with which it can annihilate. (In the case of the force-carrying particles, the antiparticles are the same as the particles themselves.) There could be whole antiworlds and antipeople made out of antiparticles. However, if you meet your antiself, don’t shake hands! You would both vanish in a great flash of light. The question of why there seem to be so many more particles than antiparticles around us is extremely important.

In quantum mechanics, the forces or interactions between matter particles are all supposed to be carried by particles of integer spin – 0, 1, or 2. What happens is that a matter particle, such as an electron or a quark, emits a force-carrying particle. The recoil from this

emission changes the velocity of the matter particle. The force-carrying particle then collides with another matter particle and is absorbed. This collision changes the velocity of the second particle, just as if there had been a force between the two matter particles. It is an important property of the force-carrying particles that they do not obey the exclusion principle. This means that there is no limit to the number that can be exchanged, and so they can give rise to a strong force. However, if the force-carrying particles have a high mass, it will be difficult to produce and exchange them over a large distance. So the forces that they carry will have only a short range. On the other hand, if the force-carrying particles have no mass of their own, the forces will be long range. The force-carrying particles exchanged between matter particles are said to be virtual particles because, unlike "real" particles, they cannot be directly detected by a particle detector. We know they exist, however, because they do have a measurable effect: they give rise to forces between matter particles. Particles of spin 0, 1, or 2 do also exist in some circumstances as real particles, when they can be directly detected. They then appear to us as what a classical physicist would call waves, such as waves of light or gravitational waves. They may sometimes be emitted when matter particles interact with each other by exchanging virtual force-carrying particles. (For example, the electric repulsive force between two electrons is due to the exchange of virtual photons, which can never be directly detected; but if one electron moves past another, real photons may be given off, which we detect as light waves.)

Force-carrying particles can be grouped into four categories according to the strength of the force that they carry and the particles with which they interact. It should be emphasized that this division into four classes is man-made; it is convenient for the construction of partial theories, but it may not correspond to anything deeper. Ultimately, most physicists hope to find a unified theory that will explain all four forces as different aspects of a single force. Indeed, many would say this is the prime goal of physics today. Recently, successful attempts have been made to unify three of the four categories of force.

The first category is the gravitational force. This force is universal, that is, every particle feels the force of gravity, according to its mass or energy. Gravity is the weakest of the four forces by a long way; it is so weak that we would not notice it at all were it not for two special properties that it has: it can act over large distances, and it is always attractive. This means that the very weak gravitational forces between the individual particles in two large bodies, such as the earth and the sun, can all add up to produce a significant force. The other three forces are either short range, or are sometimes attractive and sometimes repulsive, so they tend to cancel out. In the quantum mechanical way of looking at the gravitational field, the force between two matter particles is pictured as being carried by a particle of

spin 2 called the graviton. This has no mass of its own, so the force that it carries is long range. The gravitational force between the sun and the earth is ascribed to the exchange of gravitons between the particles that make up these two bodies. Although the exchanged particles are virtual, they certainly do produce a measurable effect – they make the earth orbit the sun! Real gravitons make up what classical physicists would call gravitational waves, which are very weak – and so difficult to detect that they have not yet been observed.

The next category is the electromagnetic force, which interacts with electrically charged particles like electrons and quarks, but not with uncharged particles such as gravitons. It is much stronger than the gravitational force: the electromagnetic force between two electrons is about a million million million million million million million million (1 with forty-two zeros after it) times bigger than the gravitational force. However, there are two kinds of electric charge, positive and negative. The force between two positive charges is repulsive, as is the force between two negative charges, but the force is attractive between a positive and a negative charge. A large body, such as the earth or the sun, contains nearly equal numbers of positive and negative charges. Thus the attractive and repulsive forces between the individual particles nearly cancel each other out, and there is very little net electromagnetic force. However, on the small scales of atoms and molecules, electromagnetic forces dominate. The electromagnetic attraction between negatively charged electrons and positively charged protons in the nucleus causes the electrons to orbit the nucleus of the atom, just as gravitational attraction causes the earth to orbit the sun. The electromagnetic attraction is pictured as being caused by the exchange of large numbers of virtual massless particles of spin 1, called photons. Again, the photons that are exchanged are virtual particles. However, when an electron changes from one allowed orbit to another one nearer to the nucleus, energy is released and a real photon is emitted – which can be observed as visible light by the human eye, if it has the right wave-length, or by a photon detector such as photographic film. Equally, if a real photon collides with an atom, it may move an electron from an orbit nearer the nucleus to one farther away. This uses up the energy of the photon, so it is absorbed.

The third category is called the weak nuclear force, which is responsible for radioactivity and which acts on all matter particles of spin- $\frac{1}{2}$, but not on particles of spin 0, 1, or 2, such as photons and gravitons. The weak nuclear force was not well understood until 1967, when Abdus Salam at Imperial College, London, and Steven Weinberg at Harvard both proposed theories that unified this interaction with the electromagnetic force, just as Maxwell had unified electricity and magnetism about a hundred years earlier. They suggested that in addition to the photon, there were three other spin-1 particles, known

collectively as massive vector bosons, that carried the weak force. These were called W^+ (pronounced W plus), W^- (pronounced W minus), and Z^0 (pronounced Z naught), and each had a mass of around 100 GeV (GeV stands for gigaelectron-volt, or one thousand million electron volts). The Weinberg-Salam theory exhibits a property known as spontaneous symmetry breaking. This means that what appear to be a number of completely different particles at low energies are in fact found to be all the same type of particle, only in different states. At high energies all these particles behave similarly. The effect is rather like the behavior of a roulette ball on a roulette wheel. At high energies (when the wheel is spun quickly) the ball behaves in essentially only one way – it rolls round and round. But as the wheel slows, the energy of the ball decreases, and eventually the ball drops into one of the thirty-seven slots in the wheel. In other words, at low energies there are thirty-seven different states in which the ball can exist. If, for some reason, we could only observe the ball at low energies, we would then think that there were thirty-seven different types of ball!

In the Weinberg-Salam theory, at energies much greater than 100 GeV, the three new particles and the photon would all behave in a similar manner. But at the lower particle energies that occur in most normal situations, this symmetry between the particles would be broken. W^+ , W^- , and Z^0 would acquire large masses, making the forces they carry have a very short range. At the time that Salam and Weinberg proposed their theory, few people believed them, and particle accelerators were not powerful enough to reach the energies of 100 GeV required to produce real W^+ , W^- , or Z^0 particles. However, over the next ten years or so, the other predictions of the theory at lower energies agreed so well with experiment that, in 1979, Salam and Weinberg were awarded the Nobel Prize for physics, together with Sheldon Glashow, also at Harvard, who had suggested similar unified theories of the electromagnetic and weak nuclear forces. The Nobel committee was spared the embarrassment of having made a mistake by the discovery in 1983 at CERN (European Centre for Nuclear Research) of the three massive partners of the photon, with the correct predicted masses and other properties. Carlo Rubbia, who led the team of several hundred physicists that made the discovery, received the Nobel Prize in 1984, along with Simon van der Meer, the CERN engineer who developed the antimatter storage system employed.

The fourth category is the strong nuclear force, which holds the quarks together in the proton and neutron, and holds the protons and neutrons together in the nucleus of an atom. It is believed that this force is carried by another spin-1 particle, called the gluon, which interacts only with itself and with the quarks. The strong nuclear force has a curious property called confinement: it always binds particles together into combinations that have no color. One cannot have a

single quark on its own because it would have a color (red, green, or blue). Instead, a red quark has to be joined to a green and a blue quark by a “string” of gluons (red + green + blue = white). Such a triplet constitutes a proton or a neutron. Another possibility is a pair consisting of a quark and an antiquark (red + antired, or green + antigreen, or blue + antiblue = white). Such combinations make up the particles known as mesons, which are unstable because the quark and antiquark can annihilate each other, producing electrons and other particles. Similarly, confinement prevents one having a single gluon on its own, because gluons also have color. Instead, one has to have a collection of gluons whose colors add up to white. Such a collection forms an unstable particle called a glueball.

The fact that confinement prevents one from observing an isolated quark or gluon might seem to make the whole notion of quarks and gluons as particles somewhat metaphysical. However, there is another property of the strong nuclear force, called asymptotic freedom, that makes the concept of quarks and gluons well defined. At normal energies, the strong nuclear force is indeed strong, and it binds the quarks tightly together. However, experiments with large particle accelerators indicate that at high energies the strong force becomes much weaker, and the quarks and gluons behave almost like free particles.

This title is rather an exaggeration: the resultant theories are not all that grand, nor are they fully unified, as they do not include gravity. Nor are they really complete theories, because they contain a number of parameters whose values cannot be predicted from the theory but have to be chosen to fit in with experiment. Nevertheless, they may be a step toward a complete, fully unified theory. The basic idea of GUTs is as follows: as was mentioned above, the strong nuclear force gets weaker at high energies. On the other hand, the electromagnetic and weak forces, which are not asymptotically free, get stronger at high energies. At some very high energy, called the grand unification energy, these three forces would all have the same strength and so could just be different aspects of a single force. The GUTs also predict that at this energy the different spin- $\frac{1}{2}$ matter particles, like quarks and electrons, would also all be essentially the same, thus achieving another unification.

The value of the grand unification energy is not very well known, but it would probably have to be at least a thousand million million GeV. The present generation of particle accelerators can collide particles at energies of about one hundred GeV, and machines are planned that would raise this to a few thousand GeV. But a machine that was powerful enough to accelerate particles to the grand unification energy would have to be as big as the Solar System – and would be unlikely to be funded in the present economic climate. Thus it is impossible to test grand unified theories directly in the laboratory. However,

just as in the case of the electromagnetic and weak unified theory, there are low-energy consequences of the theory that can be tested.

The most interesting of these is the prediction that protons, which make up much of the mass of ordinary matter, can spontaneously decay into lighter particles such as antielectrons. The reason this is possible is that at the grand unification energy there is no essential difference between a quark and an antielectron. The three quarks inside a proton normally do not have enough energy to change into antielectrons, but very occasionally one of them may acquire sufficient energy to make the transition because the uncertainty principle means that the energy of the quarks inside the proton cannot be fixed exactly. The proton would then decay. The probability of a quark gaining sufficient energy is so low that one is likely to have to wait at least a million million million million million years (1 followed by thirty zeros). This is much longer than the time since the big bang, which is a mere ten thousand million years or so (1 followed by ten zeros). Thus one might think that the possibility of spontaneous proton decay could not be tested experimentally. However, one can increase one's chances of detecting a decay by observing a large amount of matter containing a very large number of protons. (If, for example, one observed a number of protons equal to 1 followed by thirty-one zeros for a period of one year, one would expect, according to the simplest GUT, to observe more than one proton decay.)

A number of such experiments have been carried out, but none have yielded definite evidence of proton or neutron decay. One experiment used eight thousand tons of water and was performed in the Morton Salt Mine in Ohio (to avoid other events taking place, caused by cosmic rays, that might be confused with proton decay). Since no spontaneous proton decay had been observed during the experiment, one can calculate that the probable life of the proton must be greater than ten million million million million million years (1 with thirty-one zeros). This is longer than the lifetime predicted by the simplest grand unified theory, but there are more elaborate theories in which the predicted lifetimes are longer. Still more sensitive experiments involving even larger quantities of matter will be needed to test them.

Even though it is very difficult to observe spontaneous proton decay, it may be that our very existence is a consequence of the reverse process, the production of protons, or more simply, of quarks, from an initial situation in which there were no more quarks than antiquarks, which is the most natural way to imagine the universe starting out. Matter on the earth is made up mainly of protons and neutrons, which in turn are made up of quarks. There are no antiprotons or antineutrons, made up from antiquarks, except for a few that physicists produce in large particle accelerators. We have evidence from cosmic

rays that the same is true for all the matter in our galaxy: there are no antiprotons or antineutrons apart from a small number that are produced as particle/antiparticle pairs in high-energy collisions. If there were large regions of antimatter in our galaxy, we would expect to observe large quantities of radiation from the borders between the regions of matter and antimatter, where many particles would be colliding with their anti-particles, annihilating each other and giving off high-energy radiation.

We have no direct evidence as to whether the matter in other galaxies is made up of protons and neutrons or antiprotons and anti-neutrons, but it must be one or the other: there cannot be a mixture in a single galaxy because in that case we would again observe a lot of radiation from annihilations. We therefore believe that all galaxies are composed of quarks rather than antiquarks; it seems implausible that some galaxies should be matter and some antimatter.

Why should there be so many more quarks than antiquarks? Why are there not equal numbers of each? It is certainly fortunate for us that the numbers are unequal because, if they had been the same, nearly all the quarks and antiquarks would have annihilated each other in the early universe and left a universe filled with radiation but hardly any matter. There would then have been no galaxies, stars, or planets on which human life could have developed. Luckily, grand unified theories may provide an explanation of why the universe should now contain more quarks than antiquarks, even if it started out with equal numbers of each. As we have seen, GUTs allow quarks to change into antielectrons at high energy. They also allow the reverse processes, antiquarks turning into electrons, and electrons and antielectrons turning into antiquarks and quarks. There was a time in the very early universe when it was so hot that the particle energies would have been high enough for these transformations to take place. But why should that lead to more quarks than antiquarks? The reason is that the laws of physics are not quite the same for particles and antiparticles.

Up to 1956 it was believed that the laws of physics obeyed each of three separate symmetries called C, P, and T. The symmetry C means that the laws are the same for particles and antiparticles. The symmetry P means that the laws are the same for any situation and its mirror image (the mirror image of a particle spinning in a right-handed direction is one spinning in a left-handed direction). The symmetry T means that if you reverse the direction of motion of all particles and antiparticles, the system should go back to what it was at earlier times; in other words, the laws are the same in the forward and backward directions of time. In 1956 two American physicists, Tsung-Dao Lee and Chen Ning Yang, suggested that the weak force does not in fact obey the symmetry P. In other words, the weak force would make the universe develop in a different way from the way in which the mirror image of the universe would

develop. The same year, a colleague, Chien-Shiung Wu, proved their prediction correct. She did this by lining up the nuclei of radioactive atoms in a magnetic field, so that they were all spinning in the same direction, and showed that the electrons were given off more in one direction than another. The following year, Lee and Yang received the Nobel Prize for their idea. It was also found that the weak force did not obey the symmetry C. That is, it would cause a universe composed of antiparticles to behave differently from our universe. Nevertheless, it seemed that the weak force did obey the combined symmetry CP. That is, the universe would develop in the same way as its mirror image if, in addition, every particle was swapped with its antiparticle! However, in 1964 two more Americans, J. W. Cronin and Val Fitch, discovered that even the CP symmetry was not obeyed in the decay of certain particles called K-mesons. Cronin and Fitch eventually received the Nobel Prize for their work in 1980. (A lot of prizes have been awarded for showing that the universe is not as simple as we might have thought!).

There is a mathematical theorem that says that any theory that obeys quantum mechanics and relativity must always obey the combined symmetry CPT. In other words, the universe would have to behave the same if one replaced particles by antiparticles, took the mirror image, and also reversed the direction of time. But Cronin and Fitch showed that if one replaces particles by antiparticles and takes the mirror image, but does not reverse the direction of time, then the universe does *not* behave the same. The laws of physics, therefore, must change if one reverses the direction of time – they do not obey the symmetry T.

Certainly the early universe does not obey the symmetry T: as time runs forward the universe expands – if it ran backward, the universe would be contracting. And since there are forces that do not obey the symmetry T, it follows that as the universe expands, these forces could cause more antielectrons to turn into quarks than electrons into antiquarks. Then, as the universe expanded and cooled, the antiquarks would annihilate with the quarks, but since there would be more quarks than antiquarks, a small excess of quarks would remain. It is these that make up the matter we see today and out of which we ourselves are made. Thus our very existence could be regarded as a confirmation of grand unified theories, though a qualitative one only; the uncertainties are such that one cannot predict the numbers of quarks that will be left after the annihilation, or even whether it would be quarks or antiquarks that would remain. (Had it been an excess of antiquarks, however, we would simply have named antiquarks quarks, and quarks antiquarks.)

Grand unified theories do not include the force of gravity. This does not matter too much, because gravity is such a weak force that its effects can usually be neglected when we are dealing with elementary particles or atoms. However, the fact that it is both long range and always attractive means that its effects all add up. So for a sufficiently large number of matter particles, gravitational forces can dominate over all other forces. This is why it is gravity that determines the evolution of the universe. Even for objects the size of stars, the attractive force of gravity can win over all the other forces and cause the star to collapse.

Literature

1. Stephen W. Hawking, *A Brief History of Time From the Big Bang to Black Holes*, 1988.
2. Краткая история времени. От большого взрыва до черных дыр (Амфора, Санкт-Петербург, 2001).

Stephen W. Hawking, who was born in 1942 on the anniversary of Galileo's death, holds Isaac Newton's chair as Lucasian Professor of Mathematics at the University of Cambridge. Widely regarded as the most brilliant theoretical physicist since Einstein, he is also the author of *Black Holes and Baby Universes*, published in 1993, as well as numerous scientific papers and books.