The making of the standard model

Gerard 't Hooft

A seemingly temporary solution to almost a century of questions has become one of physics' greatest successes.

The standard model of particle physics is more than a model. It is a detailed theory that encompasses nearly all that is known about the subatomic particles and forces in a concise set of principles and equations. The extensive research that culminated in this model includes numerous small and large triumphs. Extremely delicate experiments, as well as tedious theoretical calculations — demanding the utmost of human ingenuity — have been essential to achieve this success.

Prehistory

The beginning of the twentieth century was marked by the advent of two new theories in physics[1]. First, Albert Einstein had the remarkable insight that the laws of mechanics can be adjusted to reflect the principle of relativity of motion, despite the fact that light is transmitted at a finite speed. His theoretical construction was called the special theory of relativity, and for the first time, it was evident that purely theoretical, logical arguments can revolutionize our view of nature. The second theory originated from attempts to subject the laws found by James Maxwell for the continuum of electric and magnetic fields to the laws of statistical mechanics. It was Max Planck who first understood how to solve this: the only way to understand how heat can generate radiation is to assume that energy must be quantized. This theory became known as quantum mechanics.

At first, it was thought that quantum mechanics would apply only to atoms and the radiation emitted by their electrons. But, gradually, it became clear that the laws of quantum mechanics had to be completely universal to make sense. This idea of universality was in common with Einstein's theories of relativity. In particular, quantum mechanics had to apply not only to electrons but also to the particles that reside in atomic nuclei.

It was clear, right from the beginning, that the two new theoretical constructions would need to be combined into one. The vast amounts of energy found to inhabit atomic nuclei implied that 'relativistic quantum mechanics' had to apply to atomic nuclei in particular. Thus, a new problem became evident and soon garnered worldwide attention: how is quantum mechanics reconciled with special relativity? This question kept physicists busy for most of the rest of the century, and it was not completely answered until the standard model saw the light of day.

The early days

By 1969, the reconciliation of quantum mechanics with special relativity was still a central issue[2], but much more had been discovered through experimental observation[3]. Matter particles (see page 270) had been divided into leptons and hadrons. The known leptons were the electron, the muon and their two neutrinos (these last assumed to be massless); hadrons, such as protons and pions, obeyed the conservation laws of quantum numbers known as 'strangeness' and 'isospin'. Hadrons are divided into mesons, which can be described loosely as an association of a quark and an antiquark, and baryons, which can be simply depicted as being made up of either three quarks or three antiquarks. The symmetry of strong interactions between subatomic particles was known to be approximated by the 'eightfold way' (Fig. 1). And it seemed that all hadrons had infinite series of excited states, in which angular momentum was bounded by the square of the mass measured in units of ~1 gigaelectronvolt (Fig. 2). This feature of all hadrons was telling us something important about strong interactions, but the first attempts to understand it consisted of rather abstract formalisms.

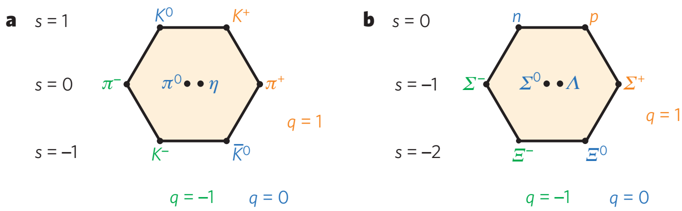


Figure 1 | The eightfold way. Spin-zero mesons (**a**) and spin-half baryons (**b**) can be grouped according to their electric charge, *q*, and strangeness, *s*, to form octets (which are now understood to represent the flavour symmetries between the quark constituents of both mesons and baryons).

It was also known that there are weak forces and electromagnetic forces, to which subatomic particles owe some of their properties. However, only the electromagnetic force was understood in sufficient detail for extremely precise calculations to be checked against accurate experimental observations. Theorists had tried to devise methods to subject not only the electromagnetic force but also other forces to the laws of quantum mechanics and special relativity. Despite their efforts over nearly 50 years, attempts to improve this 'quantum field theory' to include weak interactions failed bitterly. And describing the strong interactions between mesons and baryons drove them to despair.

The theorists at that time therefore concluded that quantum field theory should be dismissed as a possible way of addressing the dynamics of particle interactions. We now know that this was a misjudgement. Their mistrust of quantum fields was, however, understandable: in all known quantum field systems, there were divergences in the high-energy domain, making these systems unsuitable for describing strong interactions. Yet it was clear that strong interactions, such as those that hold a nucleus together, do exist. The error made by the theorists was that this 'bad' high-energy behaviour was thought to be an unavoidable, universal feature of all quantum field theories[4].

Because of this widespread objection to quantum field theories, few theorists ventured to investigate field theoretical methods. They should have realized that their objections could be swept away when the forces are weak. Indeed, the weak force was the first subatomic force to be formulated using the new 'gauge theories'[2]. Such theories had been proposed in 1954 by Chen Ning Yang and Robert Mills (Fig. 3), who were inspired by the fact that the two basic forces of nature that were well understood, gravity and electromagnetism, were both based on the principle of local gauge invariance: that is, that symmetry transformations can be performed in one region of space-time without affecting what happens in another. This beautiful idea got off to a slow start, even after Peter Higgs, François Englert and Robert Brout realized in 1964 how the structure of the vacuum can be modified by the field of a scalar (spin-zero) particle, which came to be called the Higgs particle. With the inclusion of the Higgs particle, the Yang–Mills field equations could

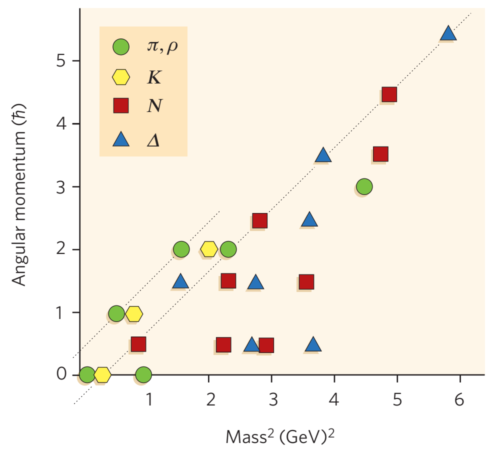


Figure 2 | A hint at the nature of the strong force. All strongly interacting particles and their excited states seem to have an angular momentum (in units of the reduced Planck constant $\hbar$) that is less or about equal to the square of their mass (measured in gigaelectronvolts, GeV). The limits for various particle species form lines that seem to be straight and parallel. *N*, nucleon (which includes neutrons and protons).

now be used to describe the weak force accurately; the force would be carried by the quanta of the Yang–Mills field, which had gained mass by this 'Brout–Englert–Higgs mechanism'. Reasonably realistic models in which exactly this happens were proposed by Abdus Salam, Sheldon Glashow and Steven Weinberg in the 1960s.

The 1970s

In 1971, Martinus Veltman and I demonstrated that it is exactly these theories (in which the mass of the Yang–Mills gauge quanta is attributed to the field of a Higgs particle) that are 'renormalizable', and it seems that this was all that was needed for a full rehabilitation of quantum field theory to begin[4]. Renormalization is the mathematical description of the requirement for distinguishing, at a fundamental level, the algebraic mass terms and coupling terms in the equations from the actual physical masses and charges of the particles. The choice of values for these algebraic parameters depends crucially on the smallest distance scales taken into account in the theory. So, if it were insisted that all particles are truly point-like — that is, the smallest distance scale should be zero — then these algebraic parameters would need to be infinite. The infinite interactions were needed to cancel the infinitely strong self-interactions of particles that the equations inevitably lead to. But the mathematical procedure of cancelling infinite forces against one another needed to be performed with considerable care. Many theorists did not understand what was going on and aired their strong suspicions that 'all this' had to be 'rubbish'.

We were learning not only how to construct realistic and logically coherent models but also how to study the behaviour of these theories over short distances by using the concept of the 'renormalization group'. Introduced by Ernst Stückelberg and André Petermann in 1953, this mathematical procedure allows one to go from one distance scale to another. It is used both in condensed-matter theory and in elementary particle physics, for which it was pioneered by Curtis Callan and Kurt Symanzik. A function that can be computed for every theory, named β-function by Callan and Symanzik, determines what might happen: if β is positive, the strengths of the couplings are increased at shorter distances; if β is negative, they are weakened. The error that I mentioned earlier was that all quantum field theories were thought to have positive β-functions. Indeed, it was claimed that this could be proved. Owing to various miscommunications, earlier calculations that yielded negative β-functions (including calculations by me) were systematically ignored, until in 1973, David Politzer, David Gross and Frank Wilczek published their findings that, for Yang–Mills theories, β is generally negative. Therefore, the strength of interactions would be reduced at short distances, making them controllable — a property that was named asymptotic freedom. Until this point, Yang–Mills theories had been understood to describe only electromagnetic and weak interactions. But the discovery of asymptotic freedom immediately turned Yang–Mills theory into a prime candidate for describing strong interactions as well.

In fact, experimental observations had been pointing in the same direction. A Yang–Mills structure not only fitted beautifully with the algebraic symmetries that had been established for the strong force (such as the eightfold way) but also could be deduced from observations made at the Stanford Linear Accelerator Center (SLAC), in California, where strong interactions seemed to show scaling behaviour, as though their strength diminished at short distances (known as Bjorken scaling)[4]. Indeed, theorists had concluded that no quantum field theory would be suitable for the strong force — until the asymptotic freedom of Yang–Mills fields was uncovered.

Basically, Yang–Mills fields are a generalization of the electromagnetic field, for which Maxwell had determined the equations a century earlier. Particles carry a generalized type of electric charge, allowing them not only to be accelerated by the Yang–Mills fields but also to be transmuted into other kinds of particle under the influence of these fields. Thus, electrons can transform into neutrinos, protons into neutrons and so on, as a result of the weak force. The strong force is understood as a new kind of field acting on quarks, which are the building blocks of protons and neutrons inside the atomic nuclei. In addition to ordinary electric charges, quarks also carry a threefold charge, which is reminiscent of colour vision (and hence they are usually called red, green and blue). For this reason, the Yang–Mills theory for the strong force is called quantum chromodynamics, the Greek word *chromos* meaning colour.

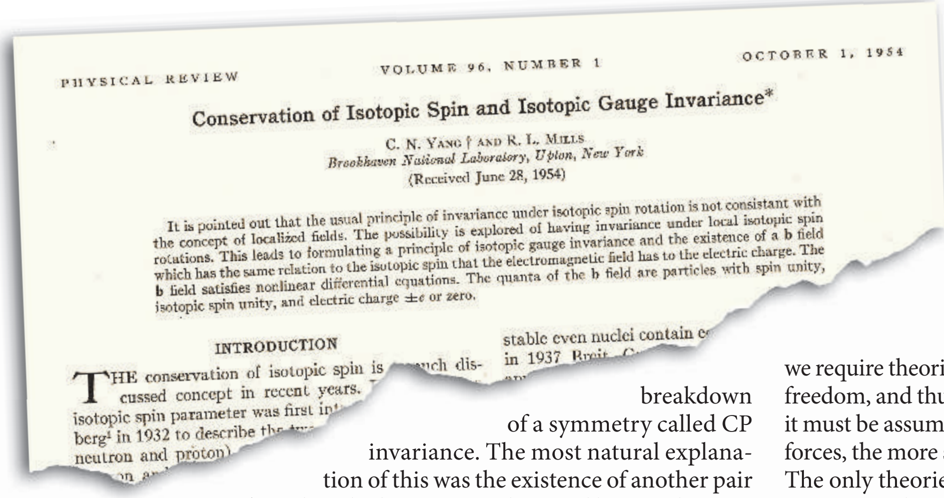
Getting the details right

For the first time, all known particles and all known forces between them could be cast in a single model. This model described three closely related Yang–Mills systems for the three major forces (strong, weak and electromagnetic), one Higgs field and several matter fields. These matter fields were Dirac fields, describing the four known leptons and the three known quarks (up, down and strange), all of which have half a unit of spin. According to this theory, the Dirac particles cannot interact directly with one another but only by exchanging energy quanta of the Yang–Mills field. The interactions between Yang–Mills fields and matter fields are identical for all particle types; only the Higgs field couples differently to the different matter fields. And only in this way is differentiation brought about between the various kinds of particle according to this new insight. By breaking the symmetry of the vacuum, the Higgs field could also give masses to the Yang–Mills quanta. But even the Higgs field is allowed to have only a limited number of interaction coefficients, so this model had only a small number of adjustable parameters: the masses of the quarks and leptons and a handful of 'mixing parameters'. The gravitational force, being excessively weak when acting between individual particles, could be included only to the extent that it acts classically.

The early versions of this model had other deficiencies. One of these was the remarkable absence of interactions due to the exchange of the neutral component, the Z boson, of the weak Yang–Mills quanta (the charged components being the W^+ and W^- bosons). These 'neutral current interactions' were detected for electrons and neutrinos in pivotal experiments at CERN in 1973 (Fig. 4). But they should also have caused strangeness-changing interactions among hadrons, and the existence of these was excluded by experimental observations. A possible remedy to this problem had already been proposed by Glashow, John Iliopoulos and Luciano Maiani in 1969, but this required a drastic revision of the model: the addition of a fourth quark, which was named charm.

The discovery of a series of new particles in 1974, beginning with the J/ψ particle at SLAC and at Brookhaven National Laboratory (the Alternating Gradient Synchrotron, in Upton, New York), marked a revolution of sorts. These new particles contained the elusive charm quark. Furthermore, their properties dramatically confirmed quantum chromodynamics and asymptotic freedom.

More details were then added. A rare type of transition observed in a special type of K meson called a K_L meson seemed to imply


PHYSICAL REVIEW VOLUME 96, NUMBER 1 OCTOBER 1, 1954

Conservation of Isotopic Spin and Isotopic Gauge Invariance*

C. N. Yang† and R. L. Mills
Brookhaven National Laboratory, Upton, New York
(Received June 28, 1954)

It is pointed out that the usual principle of invariance under isotopic spin rotation is not consistant with the concept of localized fields. The possibility is explored of having invariance under local isotopic spin rotations. This leads to formulating a principle of isotopic gauge invariance and the existence of a **b** field which has the same relation to the isotopic spin that the electromagnetic field has to the electric charge. The **b** field satisfies nonlinear differential equations. The quanta of the **b** field are particles with spin unity, isotopic spin unity, and electric charge $\pm e$ or zero.

INTRODUCTION

THE conservation of isotopic spin is ... much dis- cussed concept in recent years. ... isotopic spin parameter was first in... berg[1] in 1932 to describe the two ... neutron and proton) ...

stable even nuclei contain ... in 1937 Breit, C...

Figure 3 | Yang–Mills gauge theory. The field equations introduced by Chen Ning Yang and Robert Mills in 1954 became the basis for the three forces of the standard model — electromagnetic, weak and strong. Image reproduced, with permission, from ref. 7.

breakdown of a symmetry called CP invariance. The most natural explanation of this was the existence of another pair of quarks, which were named top and bottom, because only the delicate interplay of at least six quarks with the Higgs field could give rise to the observed CP breakdown. Mathematical consistency of the scheme also required the existence of more leptons. The tau lepton, and its associated neutrino, were discovered and confirmed around 1978. The bottom quark (in 1977) and, finally, the top quark (in 1995) were also proved to exist.

Thus, a picture emerged of three generations, each containing two species of lepton and two species of quark. All of these particles interact with three types of Yang–Mills field and one Higgs field and, of course, with gravity. This picture was subsequently referred to as the standard model. In the 1970s, it was generally thought that this standard model would merely be a stepping stone. Amazingly, however, no improvements seemed necessary to explain the subsequent series of experimental observations. The standard model became a 'standard theory' — an accurate and realistic description of all of the particles and forces that could be detected.

One further detail did need to be added. The standard model was originally designed to accommodate only strictly massless neutrinos, but there was one anomaly — the neutrino flux from the Sun[5]. Pivotal observations announced in 1998, made using the Kamiokande detector, in Japan, showed that neutrinos can mix and therefore must have mass. Adding neutrino mass terms to the standard model was, however, only a minor repair and not totally unexpected, although it did add more parameters to the model. The earlier version had 20 fundamentally freely adjustable constants (parameters) in it; now, this number would need to be increased to at least 26.

Super theories

By the 1980s, it was understood that quantum field theories are perfect frameworks for the detailed modelling of all known particles. Indeed, if we require theories with only a limited number of elementary degrees of freedom, and thus a finite number of freely adjustable parameters, then it must be assumed that all forces are renormalizable. But, for all strong forces, the more stringent condition of asymptotic freedom is required. The only theories with these desired properties are theories in which Dirac particles interact exclusively with Yang–Mills fields and (where needed) with Higgs fields. This is now regarded as the answer to that problem of more than half a century ago — how to reconcile quantum mechanics with special relativity.

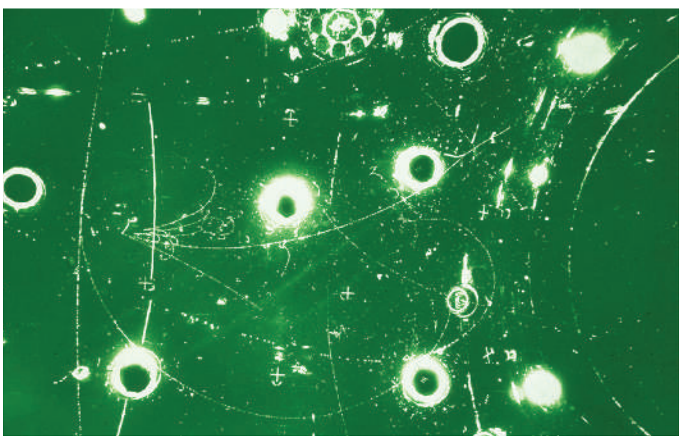


Figure 4 | The neutral current. An image from the heavy-liquid bubble chamber Gargamelle, at CERN, in 1973. The curling tracks reveal the interaction of a neutrino with a nucleon through the neutral current of Z exchange. Image reproduced with permission from CERN.

The mere fact, however, that these three Yang–Mills field systems are based on exactly the same general gauge principle, acting on the same sets of Dirac particles, has inspired many researchers not to stop here but to search for more common denominators. Can we find a completely unified field theory? Such theories have been sought before, notably by Einstein and by Werner Heisenberg in their later years, but their efforts were bound to fail because the Yang–Mills theories were then unknown. Now, it seems that we have the key to doing a much better job.

Indeed, we do have clues towards constructing a unified field theory. Despite its stunning successes, there are weaknesses in the standard model. Mathematically, the model is nearly, but not quite, perfect. Also, from a physics point of view, there are problematic features. One is the occurrence of gigantic differences in scale: some particles are extremely heavy, whereas others are extremely light. At distance scales that are short compared with the Compton wavelength of the heaviest particles — the cut-off scale below which field theories become important for these particles — there seems to be a crucial 'fine-tuning' among the effective couplings. And, most importantly, the quantum effects of the gravitational force are not included. These issues are the focus of new generations of theoretical proposals[6]. Might there be a new symmetry — a 'supersymmetry' — between Dirac particles and the force-carrying particles? Might particles turn out to be string-like rather than point-like? Or will a new generation of particle accelerators reveal that quarks and leptons are composites?

In the strongest possible terms, as theorists, we now urge our friends in experimental science to do whatever they can to obtain further information on the properties of nature's building blocks at the tiniest possible scales. In our business, this means reaching for the highest attainable energies: the Large Hadron Collider will make such a step. We can hardly wait. ■

Gerard 't Hooft is at the Institute for Theoretical Physics, Utrecht University and the Spinoza Institute, Post Office Box 80.195, 3508 TD Utrecht, The Netherlands.

1. Pais, A. *Niels Bohr's Times, in Physics, Philosophy, and Polity* (Clarendon, Oxford, 1991).
2. Crease, R. P. & Mann, C. C. *The Second Creation: Makers of the Revolution in Twentieth-Century Physics* (Macmillan, New York, 1986).
3. Källén, G. *Elementary Particle Physics* (Addison-Wesley, Reading, Massachusetts, 1964).
4. Hoddeson, L., Brown, L. M., Riordan, M. & Dresden, M. (eds) *The Rise of the Standard Model: a History of Particle Physics from 1964 to 1979* (Cambridge Univ. Press, Cambridge, 1997).
5. Bahcall, J. N. *Neutrino Astrophysics* (Cambridge Univ. Press, Cambridge, 1989).
6. Ross, G. G. *Grand Unified Theories* (Perseus, Reading, Massachusetts, 2003).
7. Yang, C. N. & Mills, R. L. Conservation of isotopic spin and isotopic gauge variance. *Phys. Rev.* **96,** 191–195 (1954).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence should be addressed to the author (g.thooft@phys.uu.nl).