

Chapter 1

The Fundamental Interactions

The four fundamental interactions that we recognize today are the *gravitational, electromagnetic, weak and strong nuclear forces*. So far they have been sufficient to describe most of the observed properties of matter, but it is not impossible that some other force will assume such fundamental role in the future. However, not everyone agrees that there are compelling evidences for such assumption, at least within a simple and understandable argument in the Occam sense.¹ Indeed, recent cosmological evidences show that the remaining 96% of the known universe may be filled with something that we do not quite understand, and this is very appropriately called the dark component of the universe, composed of dark matter showing attractive gravitation and dark energy showing something like a repulsive gravitation. They are dark because they cannot be seen, like ordinary matter.

The missing matter in galaxies and clusters was noted in 1933 by Fritz Zwicky, when he was observing the rotation of stars in galaxies using Newton's gravitational law. The name dark matter is credited to Vera Rubin in 1970 [1, 2]. Dark energy is far more recent, appearing in 1998 as a possible explanation for the accelerated expansion of the universe, observed by search teams looking at very distant type Ia supernovae [3]. An alternative current thinking about dark energy is that it is the quantum vacuum, something like the void proposed by Thomas Bradwardine in another dark age [4], but a void endowed with some energy. Such new cosmology is an integrated science involving not only optical ground-based and space telescopes and radio telescopes operating in a wide range of frequencies up to x-rays but also gravitational wave detectors, cosmic ray detectors, deep underground neutrino experiments and high energy particle colliders, and geometry and mathematical analysis.

On the theoretical side of these fascinating and hard to ignore facts, there is a strong competition between explanations, not all of them following Occam's maxim or the existence of a classical or quantum void. The ideas vary from the supposed existence of new and exotic forms of matter; or that space-time may have more

¹ The maxim that the true explanation for some natural phenomenon is also the simplest one is generally attributed to William de Occam, a Franciscan friar and philosopher from the middle ages period.

than the usual four dimensions; or that space–time may have a foam-like structure formed by “atoms” of space–time; or that space–time is as smooth as possible; or that the universe exists because we are here; or several combinations of these. Here we present just a glimpse of these concepts that are part of well-established theories. We will not discuss other theories which are not proven consistent either theoretically or experimentally. We start at the beginning, with the ideas of geometry and topology of Riemann which emerged around 1850 [5, 6].

Since the advent of Riemann’s concepts of curvature of a manifold, it has become the main tool behind all the modern theories of fundamental interactions. In the following sections we will discuss the essential ideas of Riemann, Lie, Weyl, Einstein, and others and how they have established such stronghold for modern science.

As we shall see, the Riemann curvature depends on the preliminary notion of an *affine connection* or equivalently of a covariant derivative. On the other hand, as a consequence of Noether’s theorem, the curvature appears as observables in nature, responsible for the fundamental interactions. Therefore, we may pause for a reflection on the possibility that geometry and analysis are ultimately dependent on the underlying physics of the fundamental interactions [7].

Einstein’s general relativity of 1916 set the geometrical paradigm that we use today based on the notion of curvature set by Riemann. Thus, we no longer think of gravitation as a force, but rather as a curvature of space–time, as compared with the idealized Minkowski’s flat space–time. It was only after 1954, with the works of C. N. Yang and R. Mills, that it was understood that the other three fundamental interactions, known as *gauge interactions*, also have the same curvature meaning. However, this latter development is not intuitive and it took a long time to mature [8].

The development of gauge theory started in 1919, with two independent ideas. The first one was the proposal of Hermann Weyl to describe a geometrical theory of the electromagnetic field [9]. The second one was the development by Emmy Noether of very general theorems concerning the construction of the observables of a physical theory, starting from the knowledge of its symmetries [10].

Weyl’s original idea was to generalize Einstein’s gravitational theory by incorporating the electromagnetic field as part of the space–time geometry: He reasoned that in the same way as the gravitational field is defined by the quadratic form defined in space–time

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu$$

where the coefficients $g_{\mu\nu}$ are solutions of Einstein’s equations, the electromagnetic field would be defined by the coefficients of a linear form

$$dA = A_\mu dx^\mu$$

where A_μ are the components of the electromagnetic four-vector potential. To achieve such geometric unification of gravitation and electromagnetism, Weyl

modified the metric affine connection condition $g_{\mu\nu;\rho} = 0$ (the so-called *metricity condition* of Riemann's geometry), where the semicolon denotes Riemann's covariant derivative, by the more general condition [9]

$$g_{\mu\nu;\rho} = -2A_\rho g_{\mu\nu} \quad (1.1)$$

In this way Weyl hoped to obtain the electromagnetic field described by the Maxwell tensor

$$F_{\mu\nu} = \frac{\partial A_\mu}{\partial x^\nu} - \frac{\partial A_\nu}{\partial x^\mu}$$

satisfying Maxwell's equations

$$F^{\mu\nu}{}_{;\nu} = J^\mu \quad (1.2)$$

$$F_{\mu\nu;\rho} + F_{\rho\mu;\nu} + F_{\nu\rho;\mu} = 0 \quad (1.3)$$

where the partial derivatives (∂) are replaced by the covariant derivative ($;$) defined by Weyl's connection satisfying condition (1.1) and where J^μ are the components of the four-dimensional current density (see e.g., [11]).

Weyl's proposal did not succeed essentially because in translating Maxwell's equations to his new geometry, the Poincaré symmetry was lost and so also the compatibility with the gauge transformations of the electromagnetic potential. Indeed, the above expressions for $F_{\mu\nu}$ are covariant (that is, they keep the same form on both sides) under the Poincaré transformations in space-time and also under the transformations of the electromagnetic potential given by

$$A'_\mu = A_\mu + \partial_\mu\theta(x) \quad (1.4)$$

where, as indicated, the parameter $\theta(x^\mu)$ is a function of the space-time coordinates. These transformations are not a consequence of the Poincaré transformations of coordinates, but they form a group by their own properties, acting in the space of the potentials, in such a way that they are compatible with the Poincaré coordinate symmetry. Therefore, when Weyl tried to write the gauge transformations (1.4) of the electromagnetic field in a curved space-time, the Poincaré symmetry was replaced by the group of diffeomorphisms of coordinate transformations of the curved space-time. In doing so, the mentioned compatibility between the two symmetries was lost. This result implied that at each point of the curved space-time, the potential A_μ would not behave as the known electromagnetic potential (which can be calibrated by all observers in the orderly way dictated by (1.4)). In the Weyl proposal, the two symmetries would lead to unpredictable results for the electromagnetic field. In view of such inconsistency, Weyl abandoned his theory.

The solution to *Weyl's gauge inconsistency* appeared only after the development of quantum theory. In 1927 Vladimir Fock and Fritz London suggested that Weyl's

idea could in principle make sense in quantum mechanics, where the gauge transformation (1.4) would be replaced by a *unitary gauge transformation* like [12, 13]

$$A'_\mu = e^{i\Theta(x)} A_\mu$$

Since in quantum theory only the norm $\|A_\mu\|$ is an observable, the above unitary transformation would not affect the observable potentials themselves, regardless of the type of coordinate transformations. Consequently, the Fock–London suggestion would apply only to a quantum version of the electromagnetic theory, with the above unitary gauge transformations.

With this new interpretation, Weyl reconsidered his theory in 1929, when he introduced the concept of *gauge transformations* in the sense that it is used today, specially including the “local gauge transformations” in which the parameters are dependent on coordinates. Such unitary gauge transformation would be understood as an intrinsic property of the quantum electromagnetic field, retaining its Poincaré invariance in the classical limit [14]. In this case, the diffeomorphism invariance of his theory would not interfere with the unitary transformations of that quantum gauge transformations.

However, little was known in 1929 about the quantum behavior of electrodynamics. With the lack of experimental support to the quantum interpretation, Weyl’s theory entered into a second dormant period lasting to about 1945, when new properties of fields and elementary particles would become more evident. The resulting theory called *quantum electrodynamics* (or QED) associated with the one-parameter unitary gauge group $U(1)$ was described by J. Schwinger around 1951 [15]. Instead of describing just the already known classical electromagnetic interaction between charged particles, in QED the interaction between charged particles was intermediated by photons.

The second important contribution to gauge theory from the period 1918 to 1919 was the theorem by Emmy Noether showing how to construct the observables of a physical theory, starting from the knowledge of its Lagrangian and its symmetries [10]. Although the motivations and results of Weyl and Noether were independent, they met at the point where Noether introduced a matrix-vector quantity (a vector whose components are matrices) to obtain the divergence theorem in the case where the parameters depended on the coordinates. Later on it was understood that Noether’s matrix-vector defined the same gauge potential when it is written in the *adjoint representation* of the *Lie algebra* of the gauge group. Then the properties of the adjoint representations of the Lie algebras of local symmetry groups became central to the development of gauge theory. More than that, Noether’s theorem made it possible to predict new results from gauge theories. It also meant that the gauge potentials are observables and not just mathematical corrections to derivatives and to the divergence theorem.

In 1956 Yakir Aharonov and David Bohm suggested an experiment to find an observable effect associated with the magnetic potential vector \mathbf{A} (and not by the magnetic field \mathbf{B}) itself, on the electron deviation in a double slot experiment. The proposed experiment used a long spiral coil parallel to the slots, in such a way

that the magnetic field being confined to the center of the coil would not interfere directly on the electron beam [16]: The experiment was realized in 1960, showing that indeed there was shift on the phase of the wave function, given explicitly by the vector potential

$$\Psi' = e^{-\frac{ie}{\hbar} \oint_c \langle \mathbf{A}, d\ell \rangle} \Psi \quad (1.5)$$

where the integral is calculated on the closed path c formed by the electron beams in Fig. 1.1. Since this is essentially the consequence of a phase transformation depending on the local coordinates, this experiment can also be seen as a confirmation of the Fock and London interpretation.

The rest of the history of development of gauge theory represents an exuberant mixture of theory and experiment, mainly because it depended on the development of nuclear theory and indeed on the whole phenomenology of particle physics. The reader may find details in, e.g., [17–19] among other fine reviews.

In 1932 Werner Heisenberg had already proposed that protons and neutrons could be described as being distinct states of one same particle, the nucleon [20]. This nucleon was consistently described by a new quantum number, the isotopic spin (or *isospin*), mathematically described by the $SU(2)$ group. This is formally similar to the orbital spin, but here it is regarded as an *internal symmetry*. However, a major difference with the QED gauge theory is that the $SU(2)$ isospin symmetry is a *global symmetry*, in the sense that its parameters do not depend of the coordinates [21].

In 1954 Yang and Mills proposed a generalization of electrodynamics, where the $U(1)$ group was replaced by the *local* $SU(2)$ group. In this case, instead of a vector potential A_μ , the proposed theory has a non-Abelian 2×2 matrix-vector potential \mathbf{A}_μ , whose components are defined in the Lie algebra of the new symmetry [22], similar to the matrix-vector invented by Noether. However, the physical interpretation of the local $SU(2)$ required further developments, like the Weinberg–Salam theory.

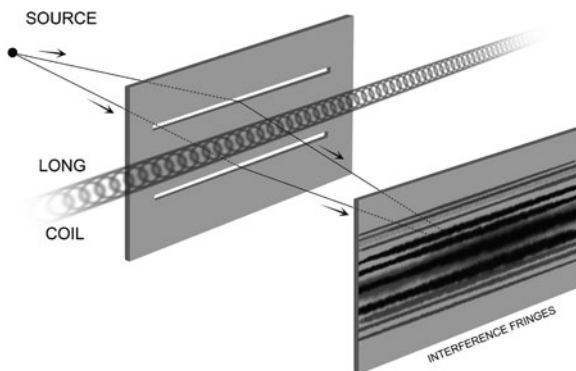


Fig. 1.1 The Aharonov–Bohm experiment

In the period 1967–1968, Weinberg and Salam independently proposed a unification of the electrodynamics with the weak nuclear force called the *electroweak theory* based on the symmetry $SU(2) \times U(1)$ [23–25]. This theory predicted three intermediate particles with integer spins, the W^+ , W^- , and Z^0 bosons, whose existence was confirmed experimentally at CERN in 1983, producing also the experimental evidence of the $SU(2)$ gauge theory.

The existence of sub-nuclear particles called quarks (the word quark was extracted from Finnegan’s Wake [26] by Gell-Mann in 1964). In 1961 Gell-Mann and independently Yuval Ne’eman had formulated a particle classification scheme called the eightfold way, which after much hard work resumed in the $SU(3)$ symmetry [27, 28]. In this scheme quarks were held together by gluons in a more general model of strong nuclear interactions, today understood as a Yang–Mills theory with eight local parameters organized in a local $SU(3)$ gauge symmetry. The $SU(3)$ group contains $SU(2)$ and $U(1)$ as subgroups associated with the isospin and hypercharge, respectively. The four remaining parameters would describe the components of the strong nuclear force called gluons, represented in the Lie algebra of $SU(3)$ as the bounding force between quarks. The result is a theory of strong interactions nicknamed quantum chromodynamics (QCD). At the time of this writing, quarks were never observed as free particles but always bounded to another by gluons. Summarizing, the original works of Weyl and Noether took shape with the name of gauge field theory (or Yang–Mills theory), involving three of the four fundamental interactions, associated with the local gauge symmetries $U(1)$, $SU(2)$, and $SU(3)$, respectively.

Those three groups can be seen as parts or subgroups of a larger symmetry group, the *combined symmetry group*. The simplest combined symmetry is just the Cartesian product $U(1) \times SU(2) \times SU(3)$, which forms the basic or *standard model* of particle interactions. It was soon found that the standard model is not sufficient to describe other aspects of the structure of particle physics, like, for example, their organization into families. Thus, a more general group of symmetries was and still is sought for, which can eventually lead to a grand unification theory (GUT), involving the three gauge interactions. Suggested candidates are $SU(5)$, $SU(6)$, $SO(10)$, and products of these and other symmetries. More recently the exceptional groups such as E_7 or E_8 have emerged as a necessary component of such scheme [29].

Since all possible combinations of gauge symmetries are relativistic, they should also combine with the Poincaré group as the symmetry of Minkowski’s space–time. This is necessary because the experimental basis of particle physics is constructed with the representations of that group [30]. This fact opened another problem: In 1964 O’Raifeartaigh showed that an arbitrary combination between gauge symmetries and the Poincaré group implied that all particles belonging to the same spin multiplet would have the same mass, which is of course not correct. In 1967 O’Raifeartaigh’s theorem was generalized by Coleman and Mandula, with the same conclusion: The combined gauge–Poincaré symmetry is not compatible with the experimental facts at the level of energies where the particle masses are evaluated.

This mass splitting problem became known as the *no-go theorem* for the compatibility between particle physics and gauge theory.

A more detailed analysis of these theorems shows that the difficulty lies in the translational subgroup of the Poincaré group. The conclusion is that either the Poincaré translations are left aside, or else the Lie algebra structure should change, or finally that the combined symmetry would not hold at the level of measurement of the particle masses [31–33]. Clearly, such fundamental issues required a radical solution if the whole scheme of gauge theories was to succeed.

Among these proposed solutions, one suggested the replacement of the Poincaré group by the deSitter group, which have the same number of parameters as the Poincaré group [34]. This choice would be naturally justified by the presence of the cosmological constant in Einstein's equations, which forbids the Poincaré symmetry in favor of the deSitter group. The currently observed acceleration of the universe finds in the cosmological constant a simple explanation, provided the cosmological constant problem can be explained.

In 1972 Roger Penrose, with different motivations, suggested that the translational symmetry could be hidden by use of the conformal group which is also a symmetry of Maxwell's equations [35]. The violation of the causality was perhaps the main restriction imposed on the use of the conformal group as a fundamental symmetry of physics. Further considerations on the conformal symmetry emerged again in 1998 for a possible mechanism to conciliate particle physics and gravitation. This was codenamed the ADS/CFT correspondence: Conformal invariant field theories can set in correspondence with isometric invariant fields in the five-dimensional anti-deSitter space. Since all known gauge fields are quantized, the ADS/CFT correspondence can be used to define quantum theory in a gravitational background [36] sometimes together with supersymmetric theories. Supersymmetry was introduced in 1974 by J. Wess and B. Zumino, as a modification of the Poincaré Lie algebra structure such that particles with half and integer spins would be interchangeable [37]. Since the generators of infinitesimal transformations of this new symmetry do not close as a Lie algebra, the resulting "graded Lie algebra" in principle would solve the mass splitting problem.

However, we cannot give up the Poincaré symmetry at the cost of having to define a new particle physics, based either on the deSitter group or on a supersymmetric group. As it was found later on, supersymmetry has to be broken at the lower levels of energies where the standard model of particle interactions applies. To the present, none of the new particles predicted by supersymmetric theory were found.

The presently adopted option to solve the mentioned no-go theorem is the so-called Higgs mechanism for spontaneous breaking of symmetries, proposed by Higgs in 1964. Essentially, the Higgs field is a postulated scalar field required to break the combined symmetry, so that the observed masses of the particle multiplets become distinct [38, 39].

Gravitation remains a great mystery. Assuming that the standard theory of gravitation is Einstein's general relativity, it still cannot explain the motion of stars in galaxies and clusters; the early inflation and the currently observed accelerated

expansion of the universe. It has also resisted all attempts to be compatible with quantum mechanics even after nearly a century of hard work on its quantization.

The earliest consideration on quantum gravity was made by Planck in 1907, when he attempted to define a natural system of physical units, in which Newton's gravitational constant G , the speed of light c , and Planck's reduced constant \hbar have value 1, and everything else would be measured in centimeters. The result is that quantum gravity would exist only at the energy level of 10^{19} Gevs, at the small length of 10^{-33} cm, which defines the so-called *Planck regime* [40]. This regime created the *hierarchy problem of the fundamental interactions*, because all other fundamental interactions exist at $\approx 10^3$ GeVs and nothing happens in between these limits. Another problem associated with the Planck regime is that it holds only in the border between three theories, Newtonian gravity, special relativity, and general relativity, each one with different symmetries and therefore with different observers. As a way to maintain a special system of units, the physics at such triple border is difficult to understand. Yet, there are over 20 theories of quantum gravity proposed up to the present ranging from the ADM program from the early 1960s, to string theory and loop quantum gravity, to massive gravity, all depending on the Planck regime.

In 1971 'tHooft showed that all gauge theories are finite when quantized by perturbative processes [41]. Therefore, if gravitation could be written as a gauge theory, then in principle we could apply 'tHooft's result to obtain a quantum theory of gravity. However, in spite of the many efforts made to write a gauge theory of gravity, it is not yet clear what is the appropriate gravitational gauge symmetry. We will return to this topic in the last chapter.

We can say that the history of the fundamental interactions is a monument to the human effort to understand nature, written in a rich mathematical language. Our objective is to discuss the various concepts involved, such as manifolds, space-times, basic field theory, symmetry, Noether's theorem, connections, culminating with our central theme, *the Riemann curvature*.