Introduction: Some History

Quantum Chromodynamics is the fundamental theory of the strong interactions. In everyday life, the strong interactions have the important task of binding protons and neutrons inside the nucleus. Back in 1932, immediately after the neutron had been discovered, it was realized that the forces keeping together the nucleons in nuclei have nothing to do with the electromagnetic forces, which are much weaker. The official birthdate of QCD is 1974 — it took therefore more than 40 years to understand what is the fundamental theory describing the dynamics of the nuclear forces and to explain the immense bulk of the accelerator data accumulated by that time.

The gradual progress in our understanding of the dynamics of nucleons and other strongly interacting particles, called *hadrons*,[†] cannot be visually represented by a reasonably straight line. It may be better described by a kind of fractal which, starting from Ignorance and passing through innumerable Fallacies, ends up eventually at the predestined final point of Truth.

Maybe we would still be far from the truth were it not for the constant pressure exerted by experimental data. QCD was created while constantly profiting from genuinely physical feedback. Unfortunately, this was probably the last example of such an efficient cooperation between theorists and experimentalists. Those fundamental laws of Nature which are still not known refer to a scale of energies currently not accessible to experiment. As for the most intriguing question of what is going on at the Planck scale $\sim 10^{19}$GeV — it presumably will *never* be answered experimentally. We

---

[†] This term coined by L. Okun is derived from the Greek *hadros*, which means "bulky" as opposed to *leptos* meaning "small".

will have to guess the fundamental Theory of Everything without much experimental feedback. The situation is difficult but, of course, not hopeless. Everybody knows an example when such a purely theoretical quest was successful: the (classical) theory of gravitation (or general relativity). It was created by Einstein in 1916 by *fiat*, confirmed by experiment soon thereafter, and, since then, there has been little essential progress in further understanding gravity.

Going back to the "fractal" history of QCD, this fractal smoothens out in our perception as time passes by and the details which are no longer relevant are forgotten. It is still wiggly enough, but its length seems to be finite and the main line of development can (and will) be described in the few pages that follow.

The first theory of strong interactions was constructed by Yukawa in 1935. By analogy with QED, where the interaction between charged particles is mediated by electromagnetic fields whose quanta are photons, Yukawa assumed that the interaction of the nucleons is mediated by some new field whose quanta (the *mesons*) represent a new kind of particle. This new theory differed from QED in three points:

- The electromagnetic field is vectorial in nature. This leads to the repulsion of same-sign charges, and only charges of opposite signs attract. However, the nucleus involves only one kind of particle (in two isotopic forms) and the interaction is attractive. That can only be realized if the quanta of the mediating field have even spin. The simplest possibility is that the spin is zero.
- Photons are massless and the Coulomb interaction is long-range. But the nuclear forces have the short range of order 1 fm. That means that the "meson" should be massive, with mass μ being of order $(1\text{ fm})^{-1} \sim 200$ MeV.
- We are dealing with *strong* interactions, and this means that the meson–nucleon coupling constant (the analog of electric charge) is large.

The theory of Yukawa was brilliantly confirmed in 1947 when the meson with the required properties was discovered in cosmic rays.[‡] That was, of course, the π meson. Three charged states of the π meson (π^+, π^0, and π^-)

[‡] Before that the muon was discovered and the attempts to associate it with the Yukawa mesons lead to considerable confusion. This is one of the "wiggles" we disregard here.

were found which meant that it is an isovector rather than an isoscalar (the apparatus of the isotopic symmetry was already well developped by that time). Also, it turned out to be pseudoscalar rather than scalar. Bearing all this in mind, the Lagrangian of the meson–nucleon theory takes the form

$$\mathcal{L}_{\text{meson}} = i\bar{N}\not{\partial}N - m\bar{N}N + \frac{1}{2}(\partial_\mu\phi)^2 - \frac{\mu^2}{2}\phi^2 - ig\bar{N}\gamma^5\tau_i N\phi_i \,, \qquad (0.1)$$

where $\tau_{1,2,3}$ are the Pauli matrices.

Soon after that, it was understood that, even though the theory (0.1) certainly correctly describes some features of nuclear interactions, it cannot be the final truth. First of all, after the discovery of the π mesons many more mesons and baryons were discovered, first in cosmic rays and then in accelerator experiments. The simple Lagrangian (0.1) had no place for them.

Secondly, the theory (0.1) had severe problems of a purely theoretical nature. The main problem was that the coupling constant g was rather big ($g^2/4\pi \sim 14$ which is ~ 1000 times larger than $e^2/4\pi \equiv \alpha$). As a result, practically no calculations were possible. Indeed, at that time (we are in the fifties now) the only known way to work with quantum field theory was to use the perturbative expansions. This works marvellously well for QED, where the expansion parameter is small, but fails completely in the case of the meson theory. Furthermore, at the beginning of the fifties, the so called zero charge phenomenon was discovered by Landau, Abrikosov, and Khalatnikov. It was found that the effective charge grows with the characteristic momentum. This means that the "continuum limit", where the ultraviolet cutoff is sent to infinity, cannot be reached: if the physical charge is kept fixed, we run into the *Landau pole* and the bare coupling constant becomes infinite at some finite Λ_{ultr}. On the other hand, if we keep g^2_{bare} finite, the physical charge goes to zero.

Thus, it became clear that a theory like (0.1), as well as QED, is sick and has no meaning beyond perturbation theory. In QED the latter works and, if we are not afraid of the inevitable trouble which we would meet if we tried to explore the unattainably high energies $E \sim M_{\text{pl}}$, we can obtain nontrivial theoretical predictions and compare them with experiment. In the meson theory, we cannot do that and, since the theory just makes no sense when the coupling is big, nonperturbative methods do not work either. This means that the Lagrangian (0.1) should be put into the waste basket.

Actually, the prevailing attitude of physicists for almost 20 years, from the mid-fifties to the mid-seventies, was that not only *that* Lagrangian, but also all other Lagrangians, the entire Lagrangian method of field theory should be discarded. It was assumed that the correct theory should be founded on some other principles. The idea of the *bootstrap* (that there are no fundamental particles and no fundamental fields whatsoever, and everything depends on itself in some self-consistent way) was popular. The theory of the S-matrix was being intensely developped. We now understand that the purely kinematical requirements of unitarity, causality, space–time and internal symmetries are not restrictive enough to determine the form of the S-matrix in a four-dimensional theory, but that *was* the hope. One of the positive developments in this "medieval" period was Regge theory. It is perceived now as good phenomenological effective theory describing hadron scattering at high energies and small tranverse momenta. But, in old days, this was considered *the* theory, the most refined and ingenious product of kinematical methods.

Thus, physicists, leaving aside the ambition to reveal a fundamental field theory, occupied themselves with a more pragmatic task — to handle the rapidly growing zoo of new "elementary" particles and resonances, if not in a refined theoretical fashion, then at least in some phenomenological way. It was basically the same job as had been done by Mendeleev for the zoo of chemical elements a century earlier. The "Mendeleev Table" of hadrons had been constructed by the early sixties by the efforts of many people, of which M. Gell-Mann probably deserves most credit. It was found that all known hadrons can be grouped into some octets and decuplets representing multiplets of $SU(3)$. This $SU(3)$ symmetry [now called *flavor* $SU(3)$] was not exact, but slightly broken in a controllable way. The model resulted in a lot of non-trivial predictions. The most famous one was the prediction of the existence of the Ω^--hyperon, including an estimate of its mass; this prediction was soon confirmed by experiment. This discovery played the same role as the discovery of missing chemical elements which filled out the empty cells in Mendeleev's Table. One could recall as well the classic story of the discovery of the planet Neptune.

We observe in experiment the octets and the decuplets, but where are the triplets and anti-triplets, the particles belonging to the fundamental representation of $SU(3)$ and its conjugate? These mysterious unobserved particles were called *quarks*.[§] The constituent quark model for hadrons was created. From a purely algebraic result concerning tensor products of several fundamental representations: $\mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{8} + \mathbf{1}$ and $\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{10} + \mathbf{8} + \mathbf{8} + \mathbf{1}$, it was inferred that mesons (grouped in octets) represent bound states of a quark and an antiquark, while the baryons (grouped in octets and decuplets) consist of three quarks. Assuming that the quarks are fermions and ascribing them the proper quantum numbers: the fractional electric charges $e_u = 2/3, \quad e_d = e_s = -1/3$, the strangeness -1 and the isospin 0 for the strange quark and the strangeness 0 and the isospin 1/2 for the u and d quarks, and assuming that the hadrons are composed of quarks in roughly the same manner as the nuclei are composed of nucleons, the quantum numbers of all observed hadrons were correctly reproduced.

That was, of course, an important success, but the dynamics of the forces binding together the quarks in hadrons remained completely unknown. Even if one did not try to answer this question in precise terms, there were two points of trouble. First, it was absolutely unclear why the quarks are always bound and not present in Nature in free unbound form. (A *lot* of experimental efforts were devoted to the search of fractionally charged particles in accelerators, in cosmic rays, in meteorites and in the ice of Greenland: to no avail.) The second difficulty was more technical. Take e.g. the Δ^{++}-isobar which, according to the constituent quark model, is made of three u quarks in S wave with the same spin orientation. But the Pauli principle forbids it (by the same token, it forbids the existence of all other baryons including nucleons).

The remedy for the second trouble was found rather soon. It was assumed that the quark of each flavor (u, d, and s) is found in Nature in three *color*[¶] forms. Thus, on top of $SU(3)$ flavor, we have $SU(3)$ color, which is exact. Pauli's principle is no longer violated because we can write the

---

[§] At first, there were three of them: "up", "down", and "strange" quarks (or just u, d, and s quarks). Now we know three more: "charmed", "bottom" (or "beautiful"), and "top" quarks. The experimental values for the quark masses are

$$\begin{gathered} m_u \approx 4 \text{ MeV} \ , \quad m_d \approx 7 \text{ MeV} \ , \quad m_s \approx 150 \text{ MeV} \ , \\ m_c \approx 1.35 \text{ GeV} \ , \quad m_b \approx 4.8 \text{ GeV} \ , \quad m_t \approx 170 \text{ GeV} \ . \end{gathered} \tag{0.2}$$

[¶] This is not the first time the word "color" appeared in this book. It is actually present right in its title in a disguised form. The Greek word for "color" is *chroma* and *chromo*dynamics means the dynamics of color.

wave function for the baryons in the required antisymmetric form $\epsilon^{ijk}q^i q^j q^k$ $(i, j, k = 1, 2, 3$ are the color indices).

It was postulated that by some miraculous reason only the color–singlet states are allowed to exist in Nature in unbound form. The quarks are color–triplets and are therefore doomed to stay confined within the hadrons. The word *confinement* was coined, but no satisfactory explanation for this phenomenon was suggested. In the middle of the sixties (and even much later) many theorists found this picture so crazy that they did not believe in the existence of quarks. Their position was pretty much the same as the position of Copernicus who said that the astronomical data could be explained more easily if adopting a mathematical model that the Earth and other planets go around the Sun. He never said it *is* the case, indeed (and that is why Copernicus in contrast to Galilei was never bothered by the Church).

This viewpoint actually prevailed until 1969 when experiments on deep inelastic electron scattering were performed at SLAC. It was shown that the total cross section of the deep inelastic scattering of virtual photons on nucleons (recall that the deep inelastic scattering is the inclusive process $e + N \to e +$ anything or, which is the same, $\gamma^* N \to$ anything, where γ^* is the virtual photon emitted by an electron) behaved in the same way as the cross section of the scattering of γ^* on a charged pointlike particle. That meant that there *are* some pointlike constituents within the proton. They were at first called *partons*, but it soon became clear that everything works wonderfully well if one assumes that these partons have the quantum numbers of quarks. Thus, quarks surpassed the status of an amusing mathematical model and displayed themselves as dynamical entities which interact and, hence, are really there (*interago ergo sum*).

But what about field theory? Was it indeed true that theorists, being frustrated by the Landau pole problem, abandoned all field theory studies? Not, of course. The latter were no longer streamline, but the nature of science is such that one can always find some people standing aside the crowd, doing unpopular things. Sometimes, this happens to be the right thing to do.

The field theory on whose basis QCD was eventually built, Yang–Mills theory, had actually been invented a long time ago, in 1954. Yang and Mills generalized the principle of local gauge invariance to the non-Abelian gauge group $SU(2)$ and wrote the celebrated Lagrangian $\mathcal{L}_{YM} = -(1/4)F^a_{\mu\nu}F^{\mu\nu\, a}$.

The spectrum of this theory involves massless self-interacting vector particles belonging to the triplet of $SU(2)$. Yang and Mills added the mass term $-m^2(A_\mu^a)^2/2$ to the Lagrangian and tried to associate the resulting massive vector particles with the ρ mesons. This attempt did not work. As a result, Yang–Mills theory was not considered to be a realistic physical theory during the following 10–15 years.

But the model as such was interesting and many theorists continued to play with it. It turned out that the quantization of such a theory is far from being easy. Proceeding in a naïve way, one gets nonsensical results, including violation of unitarity. At the beginning of the sixties, Feynman tried to resolve this problem (he was interested in Yang–Mills theory as a model for quantum gravity: the nonlinear interactions of gauge bosons in the former carry some features of the nonlinear interactions of gravitons in the latter). Feynman did not succeed by himself, but he had some important insights, which helped Faddeev and Popov to develop their *ghost* method in 1967. Including the ghosts made perturbation theory self-consistent and it became possible to calculate scattering amplitudes and other quantities at, in principle, any order in the coupling constant.

While some people tried to construct a consistent theoretical scheme for the quantization of Yang–Mills theory, some other researchers contemplated on its possible phenomenological applications. Nobody was too much in a hurry.

When the idea to describe ρ mesons as Yang–Mills quanta was abandonned, people did not further think about the possibility of applying this theory to the physics of strong interactions, but rather turned to the weak interactions. Fermi theory, involving four-fermion vertices with dimensionful coupling, was sick: nonrenormalizable and nonunitary. The idea was to consider it as the effective low-energy Lagrangian of some other theory involving only dimensionless couplings of fermions with massive charged vector bosons. Charged fundamental vector particles are most naturally described as Yang–Mills particles (ρ mesons are not well described in this way, but neither are they fundamental). Of course, in pure Yang–Mills theory the quanta are massless and one should have guessed how to give them mass in some benign way in order not to spoil renormalizability and unitarity of the massless theory (which, as a matter of fact, were not yet proven at that time!)

An almost correct model of the weak interactions was constructed along these lines by Glashow back in 1961 and the fully correct one, which unites weak and electromagnetic interactions, by Weinberg and Salam in 1967. The $W^\pm$ and Z^0 bosons were endowed with masses by the non-Abelian Higgs mechanism. The renormalizability and unitarity of this model were proven later by 't Hooft and Veltman.

The idea that Yang–Mills theory describes not only the physics of weak interactions, but also the strong interaction dynamics was put forward in 1973 in three independent papers by Pati and Salam; Fritzsch, Gell-Mann, and Leutwyler; and Weinberg. They suggested that the quarks interact with each other by the exchange of *gluons*, massless vector particles belonging to the octet representation of $SU(3)$. The gluons also interact between themselves, as is dictated by the Yang–Mills Lagrangian with the $SU(3)$ color gauge group. A real breakthrough occurred when the charge renormalization in this theory was calculated by Gross, Wilczek, and Politzer, and the phenomenon of asymptotic freedom was discovered: the effective charge does not grow with energy as is the case for Landau-pole-plagued theories like QED, but rather decreases.

History is rather dramatic at this point. The first person who correctly calculated the renormalization of the coupling constant in pure Yang–Mills theory was Khriplovich as early as 1969. He calculated the 1-loop gauge boson polarization operator $\Pi_{\mu\nu}$ in Coulomb gauge, which does not involve ghosts. In Coulomb gauge, $\Pi_{\mu\nu}$ is the *only* source of the charge renormalization; the contribution of the vertices is irrelevant. Khriplovich did not understand, however, the meaning of his result, and did not think of applying Yang–Mills theories to physics of strong interaction. There was an even earlier paper in 1965 by Terent'ev and Vanyashin, who found indications that charge may decrease with energy, but they did the calculation in a nonrenormalizable theory (without ghosts, which were not yet known at that time) and ascribed the strange behavior of the charge to this. Finally, 't Hooft also calculated it about a year before Gross, Wilczek, and Politzer, but did not, at first, understand the overwhelming significance of his result and did not publish it.[‖]

At any rate, by the end of 1974 the dust settled, and the picture became clear. The asymptotic freedom of the theory:

(1) Made it self-consistent: there *is* a continuum limit where the bare charge falls off as the ultraviolet cutoff increases in such a way that

---

[‖] The interested reader can read about all that in more details in a historical review article by M. Shifman written for the volume [11].

the effective charge defined at a given energy scale is kept fixed.

(2) Explained SLAC deep inelastic data. When the energy is large, the interaction between quarks is relatively weak, we can treat them as free and that is what was seen in experiment. Moreover, a quantitative theory of "scaling violations", i.e. the theory of deviations from the free quark–parton model due to interaction between quarks was developped by Politzer in 1974. These deviations turned out to be rather weak. Now experiments are precise enough to see them, and they perfectly well agree with the theory.

(3) Explained why strong interactions are strong. If the coupling falls off at small distances, it grows at large distances, becoming roughly of order 1 at the distances of order 1 fm, which are relevant to the nuclear physics.

(4) Did not explain the phenomenon of confinement (to derive it from the first principles of QCD is still a major unresolved task), but made it probable. If the interaction is strong at large distances, who knows how it behaves there? It is quite conceivable that the spectrum of asymptotic physical states does not resemble the set of the fundamental fields and, in particular, does not include colored states. Some modern ideas on the possible mechanisms of confinement will be discussed at the end of this book.

Since that time, many other experimental tests of QCD were carried out, and there is not a slightest doubt anymore that QCD is correct. Probably, the most spectacular confirmation comes from jet physics. Consider the process of e^+e^- annihilation with creation of a bunch of hadrons in the final state. If the energy of colliding electron and positron is large, the process goes, according to the theory, in two stages. At the first stage, the fundamental particles, the quarks and gluons are produced. This fundamental process can be $e^+e^- \to q\bar{q}$, $e^+e^- \to q\bar{q}g$, $e^+e^- \to q\bar{q}q\bar{q}$, etc. The cross section of these processes can be calculated in the framework of QCD perturbation theory.

But since quarks and gluons do not exist in free form, we observe only hadrons in the final state. Thus, at the second stage each energetic quark or gluon creates a *jet* of hadrons going roughly in the same direction as the initial colored particles (see Fig. 0.1).

Such jet processes are indeed seen. The most common is the process when only two jets are created. The underlying fundamental process is

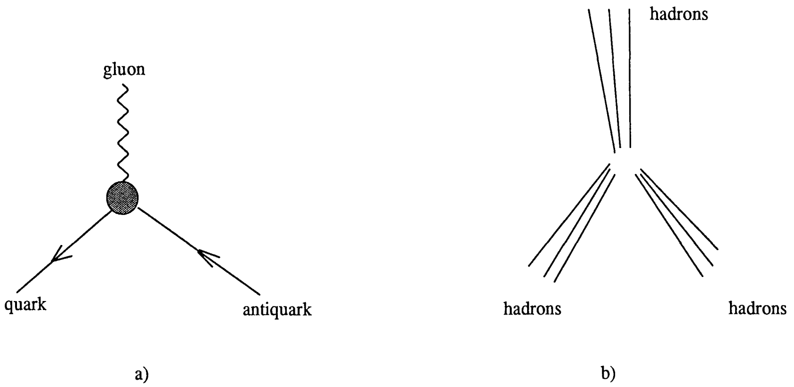

Fig. 0.1 Production of three jets in e^+e^- annihilation. *a)* quark and gluon stage, *b)* hadron stage.

$e^+e^- \to q\bar{q}$. The three-jet process (corresponding to the QCD process $e^+e^- \to q\bar{q}g$) is less probable because the cross section involves an extra power of the strong coupling constant α_s (which is small at high energies). An event where four jets are created (corresponding to the QCD processes $e^+e^- \to q\bar{q}gg$ and $e^+e^- \to q\bar{q}q\bar{q}$) is still less probable, etc. Experimentalists can measure jet cross sections, *viz.* the probabilities of producing, say, 3 jets with given total energies (E_{jet} is the sum of the energies of all particles forming the jet) and forming given angles. There is quantitative agreement between such data and the theoretical result for the differential cross section of the process $e^+e^- \to \bar{q}qg$.

There are many other tests of QCD: a series of tests for heavy quarkonium systems, for the decays of W and Z bosons, for the masses, widths, static and dynamical characteristics of low lying hadron resonances (they were calculated in QCD numerically by doing the functional integrals on the lattice and semianalytically using the ingenious *ITEP sum rules* technique) as well as many other things. The voices of the last "heretics" who have still expressed doubts whether QCD is the true theory of strong interactions, faded away by the end of the seventies.