Quarks, partons
and
Quantum Chromodynamics

A course in the phenomenology of the quark-parton model and Quantum Chromodynamics [1]

Jiří Chýla

Institute of Physics, Academy of Sciences of the Czech Republic

Abstract

The elements of the quark-parton model and its incorporation within the framework of perturbative Quantum Chromodynamics are laid out. The development of the basic concept of quarks as fundamental constituents of matter is traced in considerable detail, emphasizing the interplay of experimental observations and their theoretical interpretation. The relation of quarks to partons, invented by Feynman to explain striking phenomena in deep inelastic scattering of electrons on nucleons, is explained. This is followed by the introduction into the phenomenology of the parton model, which provides the basic means of communication between experimentalists and theorists.

No systematic exposition of the formalism of quantized fields is given, but the essential differences of Quantum Chromodynamics and the more familiar Quantum Electrodynamics are pointed out. The ideas of asymptotic freedom and quark confinement are introduced and the central role played in their derivation by the color degree of freedom are analyzed.

Considerable attention is paid to the physical interpretation of ultraviolet and infrared singularities in QCD, which lead to crucial concepts of the "running" coupling constant and the jet, respectively. Basic equations of QCD–improved quark–parton model are derived and the main features of their solutions discussed. It is argued that the "pictures" taken by modern detectors provide a compelling evidence for the interpretation of jets as traces of underlying quark–gluon dynamics. Finally, the role of jets as tools in searching for new phenomena is illustrated on several examples of recent important discoveries.

---

[1] Available at http://www-hep.fzu.cz/ chyla/lectures/text.pdf

Contents

Prolog

This text is an extended version of the lectures read in a one-semester course to students at the Faculty of Mathematics and Physics of the Charles University in Prague. Some parts of it go, however, beyond the introductory level and could be used by postgraduate students. These more advanced parts concern primarily questions related to various aspects of the renormalization procedure.

The reader is expected to be familiar with the technique of Feynman diagrams, but no special ability in carrying out actual calculations is assumed. On many places I emphasize the intuitive interpretation of these diagrams in terms of the "virtual particles" (or, more appropriately, "virtual states"). I am convinced that this interpretation helps in deriving various formulae and approximations and provides deeper physical insight into the meaning of perturbation theory.

Throughout the text the system of units conventional in elementary particle physics is used. In these units $\hbar = c = 1$, masses, momenta and energies are measured in GeV and the lengths are given in GeV^{-1}. The conversion factor reads: 1 fm$=10^{-13}$ cm $\doteq 5$ GeV^{-1}.

There is a number of books that should be consulted for more in-depth exposition of formal aspects of the elements of quantum field theory. Out of the textbooks on foundations of quantum field theory the recently published three–volume

S. Weinberg: The Quantum Theory of Fields [1]

by Steven Weinberg has rapidly become a modern classic. Beside formal aspects of quantum field theory it contains in the second volume numerous modern applications as well. Despite its age and, consequently, absence of any reference to quarks and partons, I still recommend any beginner to look into the good old

J.D. Bjorken & S.D. Drell: Relativistic Quantum Theory [2]

for a very physical introduction to the Feynman diagram technique. To get a feeling of the power of the modern language of path integrals in their derivation I recommend a small, but marvelous book

R. Feynman: QED, strange theory of light and matter [3].

The basics of the quark-parton model can be found in many books, but the original text

R. Feynman: Photon-hadron Interactions [4],

as well as another classic book

F. Close: An Introduction to Quarks and Partons [5],

are still highly recommended. The chapter on group theory is based on an excellent book

H. Georgi: Lie Algebras in Particle Physics [6],

tailored specifically to the needs of an active particle physicist. To get a proper historical perspective of the developments of subatomic physics up to early fifties the reader should consult the book

S. Weinberg: Discovery of Subatomic Particles [7],

written in the same spirit as the well known *First three minutes* by the same author. Detailed and instructive description of the crucial experiments on deep inelastic scattering of electrons on nucleons, carried out at SLAC in late sixties, can be found in 1992 Nobel lectures of J. Friedman, W. Kendall and R. Taylor [8].

There is a number of other good books or lecture notes covering parts of this course, among them

Ta-Pei Cheng, Ling-Fong Li: Gauge Theory of Elementary Particle Physics [9],
F. Yndurain: Quantum Chromodynamics [10],
Y. Dokshitzer, V. Khoze, A. Mueller and S. Troyan: Basics of QCD [11],
G. Altarelli: Partons in Quantum Chromodynamics [12],
R. Field: Quantum Chromodynamics [13],
K. Huang: Quarks and leptons [14],
F. Halzen, A. Martin: Quarks and leptons [15],
C.Quigg: Gauge theories of strong, weak and electromagnetic interactions [16].

Although this text concerns primarily the strong interactions, some working knowledge of the current electroweak theory, going usually under the name *Standard Model* (SM), is necessary. I have therefore included elements of this theory on several appropriate places, but the reader is recommended to consult the new book of Jiří Hořejší *Fundamentals of electroweak theory* [17] for excellent pedagogical exposition of the foundations of modern gauge theory of electroweak interactions. This book also gives the reader some feeling how nontrivial and twisted was the development of its basic concepts and how vital was the permanent confrontation between emerging theory and experiment.

The text has been written with two types of students in mind. First, a beginner in the field would probably like to concentrate on mastering the quark-parton model and have only limited aspirations as far as QCD is concerned. An eager experimentalist could be a typical example. On the other hand, for theoretically oriented or postgraduate students more emphasis on the QCD would be appropriate. To reconcile both of these needs the text is written at two levels: elementary and advanced ones. Parts of the text contain the advanced topics are marked in the Contents as well as in the text by an asterisk.

Helped by comments and suggestions of a number of my students, I have tried to correct misprints and errors, but more are likely to remain. I will therefore be grateful to the reader for alerting me [2] to them or for any other comments or suggestions concerning the content of this text

Beside minor corrections and modifications of the previous versions, the present one contains one important extension, concerning the the emphasis on historical aspects of major experimental discoveries and their influence on the origins of basic theoretical concepts. The more I know about this point, the more I am convinced that without the knowledge of historical circumstances, with all its turns, of the development of the Standard Model, we cannot appreciate the complicated path to its current status and the nontrivial nature of its basic structure. I have therefore added to most of the Chapters introductory Section containing basic information on the history relevant for the subject dealt with in that Chapter. On a number of places, particularly in the Chapter on deep inelastic scattering and the origin of the Parton Model, some of such remarks are inserted directly into the text as well. In these historical digressions I have greatly benefited from several books of Abraham Pais, particularly his remarkable historical treatise *Inward bound* [18], which contains wealth of information unmatched by any other source. The Czech reader may find an excellent description of historical development of the Standard Model with emphasis on the electroweak sector in [19].

Many important topics that might be relevant for students coming to the rapidly expanding field of subnuclear physics had to be left out, either because of my incompetence, or the lack of space. In a some future version I intend to include introductory exposition concerning the basic ideas and techniques of the so called **Monte-Carlo event generators**. These programs, based on current understanding of the Standard Model and generating, via the Monte Carlo techniques, the whole complicated event structure, have become indispensable tools not only in the theory, but even more in experiment. All modern particle detectors use them heavily at various stages of the extraction of meaningful information from raw experimental data.

I also plan to include discussion of basic theoretical ideas on the structure of the photon and a review of relevant experimental data. Though not the main topic, hard $\gamma\gamma$ collisions will be studied at LHC and may bring interesting information that otherwise will have to wait for future International Linear Collider.

In writing up this text I have benefited from innumerable discussions with my colleague and friend P. Kolář, which helped shape my understanding of the subjects covered in these lecture notes. The section on dimensional regularization is based to large extent on his notes. I am grateful to Jiří Hořejší for initiating my interest in historical aspects of the development of the Standard model in general and QCD in particular. I have enjoyed greatly many interesting discussions we have had over the last year concerning the birth of fundamental concepts of this theory as well as of the crucial discoveries leading to current picture of subnuclear structure and forces. We share the point of view that good understanding of these aspects of the Standard Model is vital for its further development.

Prague, February 2009

Chapter 1

Introduction

1.1 What are these lectures all about?

Before going into the systematic exposition of the quark-parton model and Quantum Chromodynamics, I will briefly review the basic concepts and overall structure of the **Standard model** (SM), the theoretical framework that encompasses the current status of our understanding of strong and electroweak interactions. Clear and pedagogical introduction into the electroweak sector of the Standard Model can be found, for instance, in [17]).

The fundamental objects of the SM are **fermions** with spin 1/2, divided into two distinct groups: **quarks** and **leptons**, forming three **generations** of identical structure. Each of the 6 quarks carries a distinct internal quantum number, called **flavor**, which is *conserved* in strong (but not electroweak) interactions. The corresponding antiparticles are grouped in a similar pattern. The interactions between these fundamental fermions are mediated by the **intermediate vector bosons** ($W^\pm$,Z,γ) in the case of electroweak interactions and by **gluons** in the case of strong interactions. The full set of fundamental particles of the SM is completed by the so far elusive **Higgs boson**. Gravitational effects are neglected. In this text we shall be interested primarily in the dynamics of strong interactions, but leptons turn out to be very useful tools for the study of the structure of **hadrons**, particles composed out of quarks and gluons.

The major conceptual problem in QCD is related to the fact that although we talk about quarks and gluons as basic building blocs of matter, they have so far not been observed in the nature as free particles. This fundamental property of QCD, called **quark and gluon confinement** makes it fundamentally different from the familiar Quantum Electrodynamics. I shall discuss our current understanding of this phenomenon in the part on hadronization, but it is clear that it has also far-reaching philosophical implications.

	charge	Generations		
		1.	2.	3.
Quarks	2/3	u(up)	c(charm)	t(top)
	-1/3	d(down)	s(strange)	b(bottom)
Leptons	0	ν_e	ν_μ	ν_τ
	-1	e^-	μ^-	τ^-

Table 1.1: Fundamental fermions of the Standard model

standing of this phenomenon in the part on hadronization, but it is clear that it has also far-reaching philosophical implications.

The reverse side of the quark (or better, color, confinement) is another novel phenomenon, called **asymptotic freedom**. It turns out, and will be discussed extensively later on, that as quarks and gluons get closer to each other, the strength of their interaction, measured by the effective color charge, progressively *decreases* until it vanishes at zero distance. Both of these surprising phenomena, i.e. quark confinement as well as asymptotic freedom, turn out to be related to the **nonabelian** structure of QCD, or, in simpler terms, to the fact that *gluons carry color charge* (contrary to photons, which are electrically neutral). [1] With color charge introduced, the basic set of quarks in Table 1.1 should actually be tripled as each of them exists in three different color states.

In order to understand the way the current knowledge of the structure of matter has in the past century been discovered by colliding various particles, it is vital to understand in detail the setup and reasoning

---

[1]This relation is, however, sensitive to several aspects of the theory, like the numbers of colors and flavors and the dimensionality of the space-time. For instance, in 2+1 dimensions the confinement is the property of **classical theory** already, while even in nonabelian gauge theories asymptotic freedom may be lost in the presence of large number of quark flavors.

used by Rutherford in 1911 in his discovery of atomic nucleus. In fact, all the major discoveries concerning the structure of matter, which are the subject of these lectures, repeat the same basic strategy pioneered by Rutherford and differ only in conclusions drawn from experimental findings. I will therefore start by recalling the essence of his experiment and then discuss some important aspects of the scattering process, first in classical and then in nonrelativistic quantum mechanics.

1.2 Rutherford experiment

The schematic layout of the Rutherford experiment is sketched in Fig. 1.1. Narrow beam of α-particles emanating from the natural radioactive source inside the lead box and formed by means of a diaphragm had been directed towards a thin golden scattering foil. Scattered α-particles were observed as scintillations on the screen made of ZnS. Counting the number of such oscillations per unit of time in a given interval of scattering angles allowed Rutherford and his collaborators Geiger and Marsden to measure the angular distribution of the scattered particles. According to the "pudding" model of the atom, popular at the beginning of the last
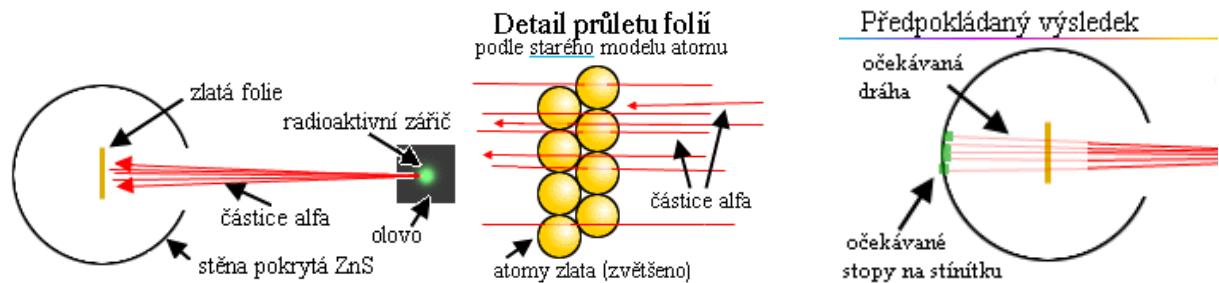
Figure 1.1: Layout of the Rutheford experiment and prediction of the "pudding" model of the atom.

century, the positive electric charge was assumed to be distributed more or less uniformly over the whole atom. The electrons, known since 1897 to be part of the atom, were imagined to resemble raisins in the pudding. Using the classical mechanics to calculate the angular distribution of α-particles scattered in such a medium (for details of this calculation see next Section) lead to the expectation, sketched in Fig. 1.1c, that almost all the particles scatter at very small angles, whereas the number of particles scattering at large angles should be entirely negligible. Such distribution is in marked contrast to what we expect for the scattering off a pointlike charge of the same magnitude, described by the Coulomb potential $V(r) = \alpha/r$. The latter leads to angular distribution proportional to $1/\sin^4(\vartheta/2)$, which peaks at small angles, but is nonnegligible even at very large angles including those corresponding to backscattering.

Surprisingly, the observations made by Geiger and Marsden and represented in a schematic way in Fig. 1.2a, showed convincingly that the large angle scattering is quite frequent, in complete disagreement with the expectations of the "pudding" model of atom. Moreover, quantitative analysis of the angular distribution of scattered α-particles suggested that it is proportional to $1/\sin^4(\vartheta/2)$, the form expected for the scattering off a pointlike electric charge. The conclusion drawn by Rutherford (see next Section for details of the reasoning)
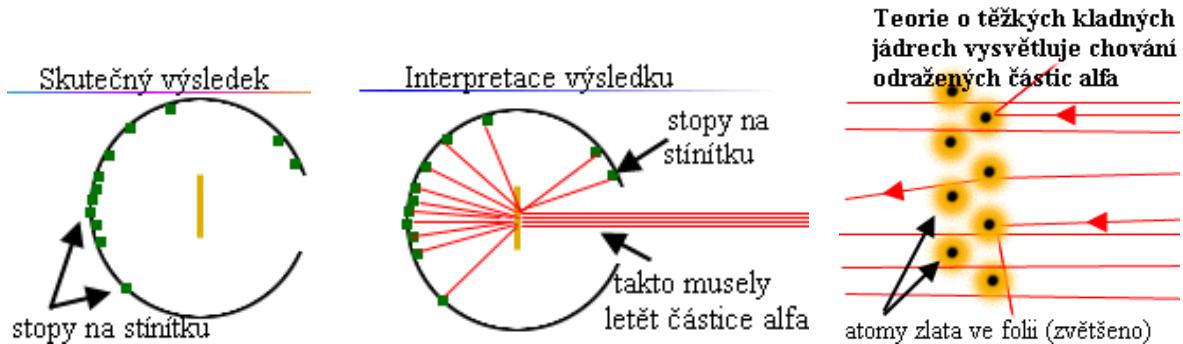
Figure 1.2: Results of the Rutheford experiment and their interpretation as evidence that positive of the atom charge must be concentrated in a small spatial region.
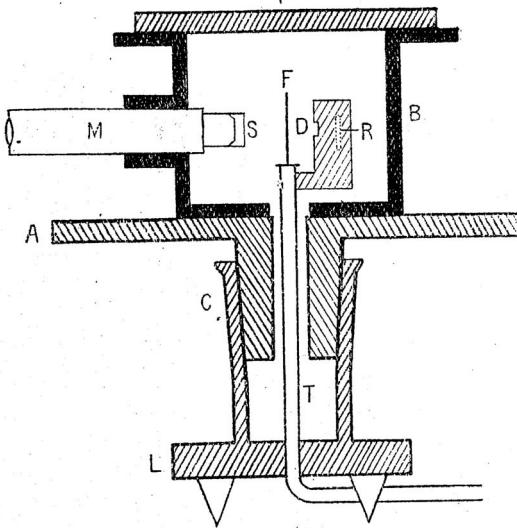
I. Angle of deflexion, ϕ.	II. $\dfrac{1}{\sin^4\phi/2}$	III. SILVER. Number of scintillations, N.	IV. $\dfrac{N}{\sin^4\phi/2}$	V. GOLD. Number of scintillations, N.	VI. $\dfrac{N}{\sin^4\phi/2}$
150	1·15	22·2	19·3	33·1	28·8
135	1·38	27·4	19·8	43·0	31·2
120	1·79	33·0	18·4	51·9	29·0
105	2·53	47·3	18·7	69·5	27·5
75	7·25	136	18·8	211	29·1
60	16·0	320	20·0	477	29·8
45	46·6	989	21·2	1435	30·8
37·5	93·7	1760	18·8	3300	35·3
30	223	5260	23·6	7800	35·0
22·5	690	20300	29·4	27300	39·6
15	3445	105400	30·6	132000	38·4
30	223	5·3	0·024	3·1	0·014
22·5	690	16·6	0·024	8·4	0·012
15	3445	93·0	0·027	48·2	0·014
10	17330	508	0·029	200	0·0115
7·5	54650	1710	0·031	607	0·011
5	276300	...	...	3320	0·012

Figure 1.3: Schematic layout of the apparatus used by Geiger and Marsden (left) and the Table containing the results of their measurements.

was that the positive electric charge is concentrated in atom inside the region that is much smaller than the atom itself, see Fig. 1.2c. Since the experiments at SLAC in 1950-1960, to be mentioned in Chapter 4, we know that the nucleus is actually not pointlike, but has a finite radius, which is about 100000 times smaller than that of the hydrogen atom. The concept of atomic nucleus was born.

In view of the pioneering character of the Rutherford experiment it is perhaps useful to recall two important circumstances. First, the "Rutherford experiment" was actually a series of observations carried out by Geiger and Marsden between 1906 and 1913, which had lead Rutherford to his model of atom. In Fig. 1.3, taken from the original paper of Geiger and Marsden [20], the vertical cross section of their apparatus is shown. Its main components were a cylindrical metal box B the size of a present day desktop PC, which contained the source R of α-particles, the scattering foil F, and a microscope M, to which the ZnS screen S was rigidly attached. The microscope M and box B, fastened to a platform A, could be rotated around the foil and radioactive source, which remained in position. The beam of α-particles from the source R formed by the diaphragm D was directed onto the scattering foil F. By rotating the microscope the α-particles scattered at different angles could be observed on the screen S. The distances between the source R and the foil F and this foil and the screen S were both typically 2 cm. Observations were taken at scattering angles between 5° and 150°. Geiger and Marsden measured the angular dependence of a number of scintillations observed during their experiments and compared it to the form $1/\sin^4(\vartheta/2)$ expected for the pointlike positive electric charge. The arrangement in which the microscope rotates around the stationary beam and target, invented by Geiger and Marsden, had since been used in modern variants in many other experiments, most notably at SLAC in fifties and sixties (see Chapter 5 for details), and usually called "single arm spectrometer". s

In the paper [20], which includes the comprehensive analysis of all technical aspects of their experiments Geiger and Marsden presented their results in a form of the Table, reproduced in Fig. 1.3, which lead them to conclude "considering the enormous variation in the numbers of scattered particles, from 1 to 250 000, the deviations from constancy of the ratio are probably well within the experimental error. The experiments, therefore, prove that the number of α-particles scattered in a finite direction varies as $1/\sin^4(\vartheta/2)$." The Summary of [20] starts with the words: "The experiments described in the foregoing paper $\cdots$ there exists at the center of the atom an intense highly concentrated electrical charge" as proposed by Rutherford. All scattering experiments perfomed ever since are in essence nothing but more or less sophisticated versions of the Rutherford experiment. What has, however, changed quite dramatically, are the characteristics of fours basic ingredients of any scattering experiment:

- source of scattering particles,

- nature and form of the target,

- technique for detecting the scattered particles,

- framework for theoretical predictions to be compared to data.

Instead of simple natural radioactive source physicists now use powerful accelerators, where protons, antiprotons, electrons and positron are given energies bilion times bigger that those in Ruthherford experiment. The role of a simple golden foil has been taken over by sophisticated arrangements of the target material and the task of observing and recording the position of the scattered particle, entrusted to the eyes and hands of Geiger and Marsden, requires now huge detectors the size of a big family house and weighing thousands of tons. Last, but not least, also the complexity of theoretical calculations has increased enormously. Instead of a simple analytic expression, like the formula (1.4 in the next Section), theoretical predictions that can be compared to experimental data employ in most cases complicated computer programs.

1.3 Bohr model of atom

Rutherford's discovery of the atomic nucleus lead directly to the formulation of the so called planetary model of atoms, in which negatively charged electrons orbited positive, practically pointlike nuclei, in much the same way as planets orbit the Sun. Since the very beginning this model faced serious problems to explain the basic features of atoms, in particular

- the existence of only discrete sets of allowed classical orbits characteristic for each element,

- the stability of the lowest energy (ground) states and

- the identity of all atoms of a given element.

By Maxwell laws of classical electrodynamics, negatively charged electrons orbiting positive nuclei should copiously radiate electromagnetic waves and thereby rapidly collapse, which, fortunately, does not happen. Equally incomprihensible apeared the fact that only well defined, discrete classical orbits were allowed to reproduce the obseved spectrum of deexcitation lines. Nothing like that holds for planetary systems hold together by gravity. A collision with an asteroid would cause arbitrarily small perturbation of an Earth's orbit, kicking the Earth into a new orbit, where it would live until the next encounter.

The impossibility to reconcile the laws of classical electrodynamics with the basic properties of atoms lead Bohr to the formulation of his model of atomic structure. This model attempted to explain the above meantioned properties of atoms by adopting several **quantum postulates**. and employing the hypothesis of energy quantization, laid out by Planck in 1900 in order to explain the spectrum of blackbody radiation. The resulting "Old Quantum Theory" was essentially a desperate attempt to apply the old classical concepts of particle tracectory for the description of subatomic world. It was developed into a rather sophisticated framework which successfully reproduced the above basic properties of atoms, by Bohr, Sommerfeld and others. It eventually failed but to understand the reasons for this failure is essential for understanding the fundamental inevitability of abandoning the classical physics for the description of the microworld. There is no space in this text to elaborate on the Bohr Model, but it is still worthwhile to read the original Bohr papers [22, 23] where its foundations are laid out.

1.4 Scattering experiment and the concept of cross-section

As most of the measurements in the world of elementary particles are based on the scattering experiments let me first recall some basic relevant facts. They are simple but may get lost in the flood of mathematical formalism on one side and practical realization of such experiments on the other.

Consider first the scattering of a pointlike particle by a (for simplicity spherically symmetric) potential $V(r)$, first in nonrelativistic classical mechanics. The familiar example of how physical laws can be inferred from experimental observations is the story of Johannes Kepler, who by observing the movement of celestial objects was able to formulate his famous laws. This procedure can be formalized as follows. The beam of particles is sent on the force center we want to study and by observing the resulting deflection we can, if the beam is sent under all impact parameters b, (see Fig. 1.4) deduce most of the information on the potential. I used the term "most" as there are certain features, concerning the properties of bound states and resonances in quantum physics, which can't be determined in this way. Moreover, as we shall see in the next Section, to probe the form of the potential at all distances requires, in classical physics strictly and in quantum physics practically, infinite energies. Instead of the whole classical path it, in fact, suffices to measure the distribution of the scattering angle $\vartheta(b)$ as a function of the impact parameter b.

Experimentally, the above algorithm can be realized by sending a uniform flux of test particles, defined by their number J per unit area perpendicular to the incoming beam, onto the target, and measure the angular distribution $N(\vartheta)\mathrm{d}\cos\vartheta$, the number of particles scattered into the angular interval $(\vartheta, \vartheta + \mathrm{d}\cos\vartheta)$. In this approach the impact parameter b does not appear in the final expression, but we have to ensure that the flux of incoming particles is uniform *over the whole plane* perpendicular to its direction. This leads us to the definition of the **cross-section**, here differential with respect to the scattering angle ϑ

$$\mathrm{d}\sigma(\vartheta, \phi) \equiv \frac{N(\vartheta, \phi)\mathrm{d}\Omega}{J} \Rightarrow \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \frac{N(\Omega)}{J} = \frac{N(\vartheta)}{2\pi J}. \tag{1.1}$$
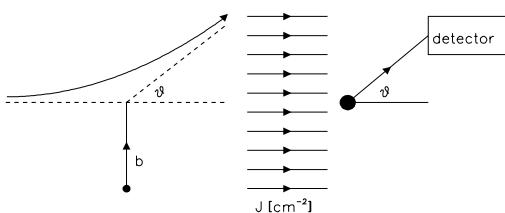
Figure 1.4: Layout of the scattering experiment.

Provided there is one-to-one correspondence between the impact parameter b and scattering angle ϑ the differential cross section (1.1) can expressed in terms of $b(\vartheta)$ as follows

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = b(\vartheta)\frac{\mathrm{d}b(\vartheta)}{\mathrm{d}\cos\vartheta} \tag{1.2}$$

where $\mathrm{d}\Omega = \mathrm{d}\cos\vartheta\mathrm{d}\phi$ with ϑ, ϕ being the polar and azimuthal angles respectively. Note, that although the scattering angle ϑ is a function of the impact parameter b the opposite is not necessarily the case as more impact parameters may lead to the same scattering angle. This fact will be crucial in the analysis, performed in the next Section, of mathematical formalism behind the interpretation of the Rutherford experiment.

For the scattering of a unit test charge (in multiples of positron charge) on the Coulomb potential $V(r) = \alpha/r$ of a unit infinitely heavy target charge, the relation between the impact parameter b and the scattering angle ϑ reads

$$b(\vartheta) = \frac{1}{\kappa\tan(\vartheta/2)}, \ \kappa \equiv \frac{p^2}{m\alpha} = \frac{2E}{\alpha} \iff \vartheta(b) = 2\arctan\left(\frac{1}{\kappa b}\right) \doteq \frac{2\alpha m}{p^2 b} = \frac{\alpha}{Eb}, \tag{1.3}$$

the latter approximation holding for large product Eb. Combining (1.3) and (1.1) leads to the Rutherford formula for differential cross section

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \frac{1}{4\kappa^2\sin^4(\vartheta/2)} = \frac{\alpha^2 m^2}{4p^4\sin^4(\vartheta/2)} = \frac{\alpha^2}{16E^2\sin^4(\vartheta/2)} = \frac{4\alpha^2 m^2}{q^4}, \tag{1.4}$$

where m, p are the mass and momentum of the beam particle and $q^2 = 4p^2\sin^2\vartheta/2$ is square of the momentum transfer $\vec{q} \equiv \vec{p} - \vec{p'}$.

The derivation of (1.3-1.4), or in general of the differential cross section for scattering in a spherically symmetric potential is quite simple. As the particle scattering on a spherically symmetric potential $V(r)$ moves in a plane it suffices to work in two dimensions. Writing the corresponding lagrangian in spherical coordinates we get

$$\mathcal{L} = \frac{1}{2}m\left[(\dot{r})^2 + r^2(\dot{\phi})^2\right] - V(r) \tag{1.5}$$

where the dot stands for time derivative. This lagrangian implies two constants of motion: orbital momentum M and energy E

$$M \equiv mr^2\dot{\phi}, \qquad E \equiv \frac{1}{2}m(\dot{r})^2 + U(r), \qquad U(r) \equiv \frac{M^2}{2mr^2} + V(r), \tag{1.6}$$

which can be expressed in terms of initial momentum p and impact parameter b as $M = pb$, $E = p^2/2m$. The scattering angle ϑ equals $\vartheta = \pi - 2\phi_0$ where ϕ_0 is given as

$$\phi_0 \equiv \int_{r_0}^{\infty}\frac{\mathrm{d}\phi}{\mathrm{d}r}\mathrm{d}r = \int_{r_0}^{\infty}\frac{\dot{\phi}}{\dot{r}}\mathrm{d}r = \int_{r_0}^{\infty}\frac{M}{mr^2}\frac{\mathrm{d}r}{\sqrt{2(E - U(r))/m}} \tag{1.7}$$

where we have simply substituted for the time derivatives the expressions following from (1.6), and r_0, denoting the smallest distance of the trajectory from the scattering center, characterized by the vanishing

of time derivative $\dot{r} = 0$, which implies that it is given as the solution of equation $E = M^2/2mr^2 + V(r)$. Solving the equation for r_0 and evaluating (1.7) for the Coulomb potential $V(r) = \alpha/r$ yields (1.3) and

$$r_0 = \frac{1}{\kappa}\left(1 + \sqrt{1 + \kappa^2 b^2}\right) = \frac{1}{\kappa}\left(1 + \sqrt{1 + \cot^2(\vartheta/2)}\right) \geq r_{\min} \equiv \frac{2}{\kappa} = \frac{\alpha}{E}. \tag{1.8}$$

To penetrate close to the origin requires large angle scattering, but the angular dependence $r_0(\vartheta)$, plotted in Fig. 1.5, shows that above roughly $\vartheta \simeq 120°$ the values of r_0 is insensitive to ϑ and equals approximately $2/\kappa$. The angular distribution of scattered α-particles as measured by Marsden and Geiger [20], reproduced in Fig. 1.5, showed excellent agreement with the $1/\sin^4(\vartheta/2)$ behavior expected of a pointlike electric charge throughout the covered region, which extended up to 150 degrees. Taking into account the energy of α-particles from radioactive decay of radium was 7.68 MeV, its electric charge (2), as well as that of gold (79) gave $r_{\min} \doteq 30$ fm, which is about 10000 times smaller than the radius of atom.
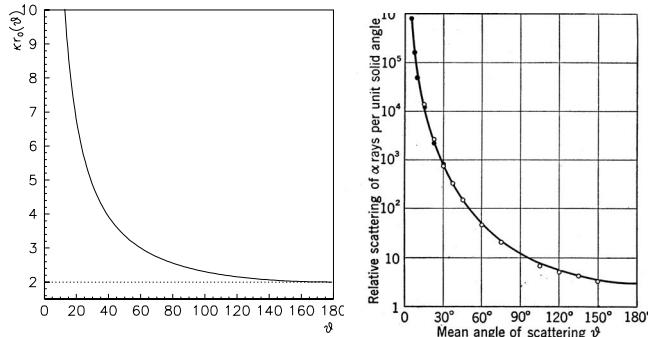
Figure 1.5: $\kappa r_0(\vartheta)$ (left) and the angular distribution of α-particles as measured by Geiger and Marsden (right).

One feature of the formula (1.4) is worth noting. Although (1.4) decreases for a fixed scattering angle ϑ with increasing momentum p of the incoming particle, the form of this distribution remains the same for all momenta p. This follows immediately from the relation between ϑ and b: to get the same scattering angle ϑ as p increases, one must simply shoot the incoming particle at appropriately smaller impact parameter! This form invariance of the differential cross section does not, however, hold generally. It is obvious that for a finite incoming momentum p the angular distribution is sensitive to the form of the potential $V(r)$ at distances $r \geq r_0$ only. The value of r_0 depends on p and m as well as details of the potential $V(r)$ in the vicinity of $r = 0$. To probe $V(r)$ at infinitely small distances requires infinitely large energies of incoming particles. This statement holds in classical, but as we shall see below, not in quantum mechanics.

The integrated (over the angles) cross-section diverges due the behavior of (1.4) at small angles. In classical physics and as far as integrated cross-sections are concerned, there is no principal difference, between the Coulomb potential α/r and, for instance, the familiar Yukawa potential $(\alpha/r)\exp(-\mu r)$ as both are infinite. In quantum physics the situation is different as there the Yukawa, but not the Coulomb, potential has finite integrated cross-section.

In classical physics the concept of cross section is useful, but may be dispensed with as more direct methods, based on the classical equations of motion for particle trajectories, are available. In quantum physics, on the other hand, this concept plays an indispensable role, as the latter methods inapplicable. We can not trace down exact trajectories of quantum particles (indeed they have no good meaning at all) neither can we determine impact parameter of each incoming particle individually.

1.5 Scattering in nonrelativistic classical physics

Let me now discuss in some detail the scattering of a pointlike test particle of unit charge on a repulsive potential $V(r)$ generated by the spherically symmetric distribution $\rho(r)$ of unit electric charge inside the sphere of radius R. As no charge is located outside this sphere, the Gauss theorem tells us that the force (and thus the intensity of the electric field $\vec{E}$) outside the sphere coincide with that generated by unit pointlike charge. For the corresponding potential this implies the following relation

$$r \geq R: \quad V(r) = \frac{\alpha}{r}, \tag{1.9}$$

$$r \leq R: \quad V(r) = e(r)\frac{\alpha}{r} + 4\pi\alpha\int_r^\infty z\rho(z)\mathrm{d}z, \quad e(r) \equiv 4\pi\int_0^r z^2\rho(z)\mathrm{d}z, \quad \frac{\mathrm{d}V(r)}{\mathrm{d}r} = -\frac{e(r)\alpha}{r^2} \tag{1.10}$$

where $e(r)$, the effective charge at distance r, is just the total electric charge located inside the sphere of radius r. Clearly, $e(R) = 1$.

Imagine we want to find out the form of $\rho(r)$ from the results of a scattering experiment by measuring the angular distribution of scattered particles. It is obvious that to do that the test particle must penetrate closer to the source of the potential than R, otherwise it will not be influenced by the form of $\rho(r)$ at all and will feel the total charge e_0 only. This requires that $\kappa \equiv p^2/m\alpha > \kappa_{min} = 2/R$. Even then, however, not all particles will pass through the region $r < R$, but only those whose impact parameter satisfies the inequality $b \leq b_0 \equiv R\sqrt{1 - 2/\kappa R} \leq R$. The rest of incoming particles will generate exactly the same angular distribution as the pointlike potential $V(r) \equiv \alpha/r$.

I will now consider beside the pointlike charge three different forms of extended charge distribution $\rho(r)$ that will illustrate the relation between the structure of the target and the corresponding angular distribution of scattered particles. First two of them coincide at $r \geq R$ with the potential α/r induced by the pointlike charge distribution and differ only inside the sphere $r \leq R$. The third distribution has no sharp edge but the parameter R characterizes the rate of exponentially decreasing factor $\exp(-r/R)$. We find ($x \equiv r/R$)

$$\rho_1(R,r) = \frac{\delta(1-x)}{4\pi R^3}, \qquad e_1(x) = \theta(x-1), \qquad V_1(r) = \frac{\alpha}{r}\left[\theta(x-1) + x\theta(1-x)\right] \qquad (1.11)$$

$$\rho_2(R,r) = \frac{3}{4\pi R^3}\theta(1-x), e_2(x) = x^3\theta(1-x) + \theta(x-1), V_2(r) = \frac{\alpha}{R}\left[\frac{3}{2} - \frac{x^2}{2}\right]\theta(1-x) + \frac{\alpha}{r}\theta(x-1) (1.12)$$

$$\rho_3(R,r) = \frac{1}{8\pi R^3}e^{-x}, \qquad e_3(x) = 1 - e^{-x}(1 + x + x^2/2), V_3(r) = \frac{\alpha}{R}\left[\frac{e_3(x)}{x} + e^{-x}(1+x)\right] \qquad (1.13)$$

Note that the first distribution $\rho_1(R,r)$ corresponds to putting the total (in our case unit) charge uniformly on the surface of the sphere with the radius R, whereas $\rho_2(R,r)$ corresponds to constant volume charge density inside that sphere. The former will play important role in quantum field theory when introducing the concept of renormalized coupling constants.

The form of the potentials $V_i(r)$, shown in Fig. 1.7, inside the sphere leads to the modification of the relation (1.3) that holds for the pointlike source. For the potential $V_1(r)$ we find after elementary calculation

$$\vartheta(R,\kappa,b) = \pi - 2\arccos\left(\frac{1}{\sqrt{1+\kappa^2 b^2}}\right), \qquad b > b_0 = \frac{1}{\kappa}\sqrt{(R\kappa - 1)^2 - 1}, \quad (1.14)$$

$$\vartheta(R,\kappa,b) = \pi - 2\left[\arccos\left(\frac{1}{\sqrt{1+\kappa^2 b^2}}\right) + \arccos\left(\frac{b}{b_0}\right) - \arccos\left(\frac{1 + \kappa b^2/R}{\sqrt{1+\kappa^2 b^2}}\right)\right], \quad b \leq b_0. \quad (1.15)$$
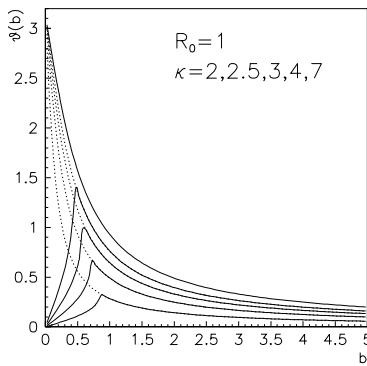
Figure 1.6: a) $\vartheta(R,\kappa,b)$ as a function of b for the distribution $\rho_1(1,r)$ and several values of κ (solid curves), from above in decreasing order. The dotted curves correspond to $R = 0$.

Graphical representation of this relation is shown in Fig. 1.6 for $R = 1$ (in arbitrary units) and five values of the parameter κ, one of them, $\kappa = 2 = \kappa_{min}$, corresponding to the threshold case that the scattered particles do not enter the interior of the sphere with radius R and are thus not influenced by the potential inside it. Several conclusions can be drawn from this figure.

First, for $\kappa > \kappa_{min}$ there is a maximal scattering angle $\vartheta_{max}(R,\kappa) = \vartheta(R,\kappa,b_0)$ corresponding to $b = b_0$, which is a decreasing function of κ. This behaviour simply reflects the fact that large scattering angles require strong repulsive potential, which, however, is not the property of the potential $V_i(r)$. The fact that the angular distribution vanishes for $\vartheta > \vartheta_{max}$ is a direct consequence of the sharp edge of the distribution $\rho_1(r)$. For smoothly vanishing $\rho(r)$ the resulting angular will not vanish at large angle but will merely be strongly suppressed there.

Secondly, to calculate the angular distribution from the known dependences $\vartheta(R,\kappa,b)$ we cannot simply invert then and use (1.1) as there are two impact parameters for each scattering angle ϑ. We must therefore use (1.1) separately in the two regions $b < b_0$ and $b > b_0$ and then sum the results. As a result, the angular distribution will differ from that of the pointlike charge distribution not only for $\vartheta > \vartheta_{max}(R,\kappa)$, where it vanishes, but also in the region $\vartheta < \vartheta_{max}$.

Third, as for fixed κ the maximal angle $\vartheta_{max}(R,\kappa)$ is a decreasing function of the radius R, we have to go to large angles, i.e large momentum transfers, to probe the charge distribution inside the sphere of that radius.

13
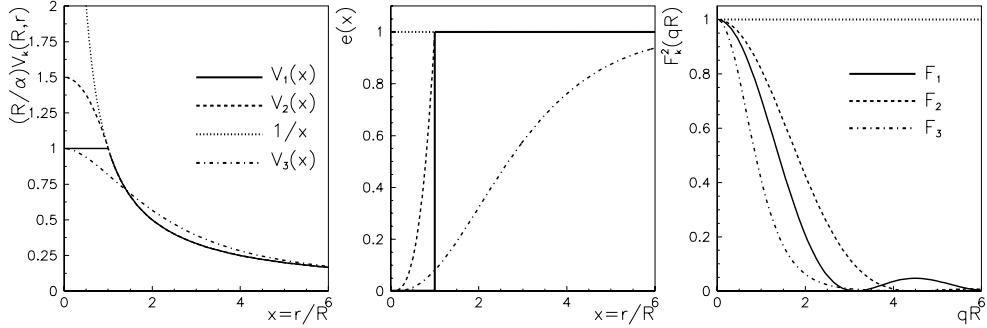
Figure 1.7: The form of potentials $V_k(r)$ defined in (1.11-1.13) and multiplied by R/α (a), corresponding to the effective charges $e_k(r/R)$ plotted in (b) and the square of the formfactors F_k. Dotted curves correspond to the pointlike charge at the origin.

We have thus seen that by measuring the angular distribution sufficiently precisely we can determine the form of the potential and thus the corresponding charge distribution inside the sphere. In principle we don't have to go to large angles to do as the differences at large and small angles are related. We have, however, use sufficiently energetic primary particles in order to satisfy the condition $\kappa = p^2/m\alpha > \kappa_{\min} = 2/R$. In quantum theory, as we shall see below, the first statement holds as well, but the second does not.

So far we have treated the distributions $\rho(r)$ of the target charge as fixed, whereas in realistic conditions of the Rutherford and other experiments the corresponding carriers were actually other particles. This is expected to be a good approximation, if the rest mass of these particles is much larger than the momentum transfer q, which, indeed was the case in the Rutherford experiment.

1.6 Scattering in nonrelativistic quantum physics

In nonrelativistic quantum mechanics the differential cross section $d\sigma/d\cos\vartheta$ for elastic scattering of a spinless pointlike particle with mass m and a unit charge [2] on a spherically symmetric potential $V(r)$ is given (in suitable normalziation) as square of the corresponding scattering amplitude $f(\vartheta)$. In the lowest order perturbation theory, called usually the Born approximation, $f(\vartheta)$ is just the Fourier transform of the potential acting on the scattering particle:

$$\frac{d\sigma}{d\Omega} = \mid f(\vartheta) \mid^2, \ \ f(\vartheta) = -\frac{m}{2\pi} \int \exp(i\vec{q}\,\vec{r})V(r)d\vec{r}, \quad \vec{q} \equiv \vec{p} - \vec{p}\,', \quad q^2 = 4p^2 \sin^2 \vartheta/2. \tag{1.16}$$

For the Coulomb potential of the form $V(r) = \alpha/r$ we easily find

$$f(\vartheta) = f(q^2) = \frac{2\alpha m}{q^2} \ \Rightarrow \frac{d\sigma^{Born}}{d\Omega} = \frac{4\alpha^2 m^2}{q^4} = \frac{\alpha^2 m^2}{4p^4 \sin^4(\vartheta/2)} \tag{1.17}$$

The form (1.17) is often taken as defining the pointlike nature of the target particle, generating the potential $V(r)$. Remarkably, the Born approximation in quantum mechanics coincides for the Coulomb potential with exact result of the classical calculation. This coincidence does not, however, have a more general validity. For Yukawa potential $\alpha \exp(-r/R)/r$ we similarly find

$$f(\vartheta) = f(q^2) = \frac{2\alpha m}{(q^2 + \mu^2)} \ \Rightarrow \frac{d\sigma^{Born}}{d\Omega} = \frac{4\alpha^2 m^2}{(q^2 + \mu^2)^2} = \frac{4\alpha^2 m^2}{(4p^2 \sin^2(\vartheta/2) + \mu^2)^2} \tag{1.18}$$

where the presence of $\mu \equiv 1/R$ in the numerator of (1.18) screens off the singularity of the Coulomb amplitude (1.17) in forward direction so that the integrated cross-section becomes *finite*.

$$\sigma_{integ}^{Born} = \frac{16\pi\alpha^2 m^2}{4p^2 + \mu^2} \frac{1}{\mu^2} \tag{1.19}$$

---

[2]For beam particle with charge e the results can be obtained by the substitution $\alpha \to \alpha e$.

For the case the potential $V(r)$ is generated by the (spherically symmetric) distribution $\rho(r)$ of elementary sources, each of them having the same form $V(r)$, i.e

$$V_\rho(r) \equiv \int \mathrm{d}\vec{z}\rho(z)V(\vec{r} - \vec{z}) \tag{1.20}$$

Born approximation for the scattering amplitude $f_\rho(q^2)$ yields, upon exchanging the order of integrations,

$$f_\rho(q^2) = F(q^2)f(q^2), \quad F(q^2) = \int \exp(i\vec{q}\,\vec{r})\rho(\vec{r})\mathrm{d}\vec{r}, \tag{1.21}$$

where the elementary scattering amplitude $f(q^2)$ is given in (1.16). For the distributions $\rho_k(r)$ corresponding to the potentials $V_k(r)$ defined in the previous Section, the form of their formfactors $F_k(q^2)$ is shown in Fig. 1.7. The rapidly dying oscillations at large angles (i.e. large values of qR) for the case of charge distributions ρ_1 and ρ_2 reflect their sharp edges. No such oscillations are, on the other hand, present for the smoothly decreasing distribution ρ_3. Note that all the formfactors $F_k(qR)$ deviate from unity, which corresponds to the pointlike charge distribution, at all qR. This implies that contrary to classical mechanics, there is no strict lower limit on energy E of the incoming particle necessary for probing the charge distribution inside the sphere R. On the other hand, to access the region where the formfactors F_k differ appreciably from unity, requires $qR \gtrsim 1$. This in turn implies $pR \gtrsim 1$ and thus sets the lower limit on p.

1.7 Scattering on bound systems

So far we have discussed scattering off a given distribution of elementary scatterers described by some continuous function $\rho(r)$ normalized to the unity. To come closer to real situation let us now consider a bound system of such elementary scatterers, described by potentials $\mathcal{V}_i(\vec{r} - \vec{r}_i) = \alpha_i V(\vec{r} - \vec{r}_i)$, still fixed at points $\vec{r}_i$ and differing only by the magnitude of their "charges" α_i, which sum up to the total charge of the target $\sum_i \alpha_i \equiv \alpha$, which can be without loss of generality set to unity. Let us moreover assume the existence of a minimal distance Δ^{min} between any pair of them. In the Born approximation the expression for the scattering amplitude $F(\vartheta)$ on such a composite system reads

$$F(\vartheta) = -\frac{m}{2\pi} \int \exp(i\vec{q}\,\vec{r}) \sum_i \alpha_i V(\vec{r} - \vec{r}_i)\mathrm{d}\vec{r} = \left(\sum_i \alpha_i e^{i\vec{q}\vec{r}_i}\right) \underbrace{\frac{-m}{2\pi} \int \exp(i\vec{q}\,\vec{r})V(r)\mathrm{d}\vec{r}}_{f(\vartheta)} \tag{1.22}$$

where the last integral is just the elementary amplitude for the scattering on a pointlike centre with unit charge. The differential cross-section for the scattering on such a composite target is then given as

$$\frac{\mathrm{d}\sigma^{comp}}{\mathrm{d}\Omega} = \frac{\mathrm{d}\sigma^{point}}{\mathrm{d}\Omega} \left|\sum_i \alpha_i \exp(i\vec{q}\,\vec{r}_i)\right|^2 = \frac{\mathrm{d}\sigma^{point}}{\mathrm{d}\Omega} \left(\sum_i \alpha_i^2 + \underbrace{2\mathrm{Re}\sum_{i<j} \alpha_i\alpha_j \exp(i\vec{q}(\vec{r}_i - \vec{r}_j))}_{Z}\right) \tag{1.23}$$

This expression has simple forms in two different regions:

- $q^2 \to 0$: where $Z \to \sum_{i<j} 2\alpha_i\alpha_j$ and the differential cross-section on the composite target behaves like that on the pointlike target with the unit total charge α. For small momentum transfers, the beam particle "sees" merely the total charge of the target.

- $q^2 \to \infty$: where $Z \to 0$ (due to nonzero Δ^{min}!) and we get

$$\frac{\mathrm{d}\sigma^{comp}}{\mathrm{d}\Omega} \to \frac{\mathrm{d}\sigma^{point}}{\mathrm{d}\Omega} \sum_i \alpha_i^2 = \sum_i \frac{\mathrm{d}\sigma^i}{\mathrm{d}\Omega} \tag{1.24}$$

In this case the beam particle "sees" the individual charges and measures the **incoherent** sum of individual cross-sections, proportional to squares of the charges. Assuming there are N scattering centers each carrying the same charge $1/N$, we conclude that the differential cross section (1.24) decreases at large q^2 with N like $1/N$!

Realistic systems, like atoms or nuclei, their constituents are not fixed in space, but bound by some, usually two-body, forces and therefore we have to take into account these forces when calculating scattering on individual scattering constituents. Provided the system is weakly bound, i.e. the characteristic binding energy is finite and small with respect to momentum transfer q of the collision, this binding can to first approximation be neglected and we can again use (1.24), although the kinematics must take into account the recoil of the target constituent. The main difference with respect to fixed scattering centers will be in the identity of the final state. It is obvious that a violent scattering on a bound system will usually break it up and thus lead to <u>inelastic</u> collision. The probability that after such violent collision the constituents will recombine into the original system, though nonzero, is very small. So to see inside the bound system we thus have to break it up into the individual constituents!

The above reasoning is straightforward and almost trivial for weakly bound systems, but can under certain circumstances be used even for the strongly bound ones. Imagine, for instance the system of particles bound pairwise by the linearly rising potentials $V(r) = \omega r$, which prevents the separation of individual particles with any finite energy. On the other hand, provided that $q^2 \gg \omega$, which guarantees that the interparticle distances remain small during the interaction time, estimated by $1/q$, compared to the distances where the linearly rising potential becomes comparable to q, we can separate the process into two stages:

Hard scattering: describing the scattering of incoming particles on individual scatterers treated as free particles, which is sensitive to their charges,

Recombination: nontrivial process in which the excited system de-excites by producing the original or other strongly bound systems.

The essential feature of this space-time separation of the hard inelastic collision is that the recombination stage determines the final states of the collision, but <u>not</u> the angular distribution of the scattered test particles!

1.8 Exercises

1. Solve Gauss' law $\Delta \phi(\vec{r}) = \kappa \delta(r)$ in 1+1 and 2+1 dimensions and discuss the results.

2. Derive (1.2).

3. Derive (1.4).

4. Derive (1.18) and (1.19).

5. Evaluate the formfactor (1.21) as well as the potential (1.10) for the distribution $\rho_Y(r) = (\mu^3/8\pi) \exp(-\mu r)/\kappa$.

Chapter 2

Elements of groups and algebras

2.1 Definitions and basic facts

Only a brief reminder of the very basics of group theory is given in this chapter. The emphasis is put on practical abilities to carry out simple calculations that will be needed in the following chapters when discussing the quark-parton model and QCD.

Group: a set $\mathbf{G}$ with binary operation "$\bullet$" satisfying the following conditions:

1. $\forall x, y \in \mathbf{G} : x \bullet y \in \mathbf{G}$

2. $\exists e \in \mathbf{G} : \forall x \in \mathbf{G}, e \bullet x = x \bullet e = x$ (existence of a unit element e)

3. $\forall x \in \mathbf{G} \ \exists x^{-1} \in \mathbf{G} : \ x \bullet x^{-1} = x^{-1} \bullet x = e$ (existence of an inverse element)

4. $\forall x, y, z \in \mathbf{G} : x \bullet (y \bullet z) = (x \bullet y) \bullet z$ (associativity of $\bullet$)

Groups enter physics basically because they correspond to various symmetries. The concept crucial for the description of transformations of physical quantities under these symmetry transformations is that of the **group representation**.

Group representation: the mapping $\mathbf{D}$: $\mathbf{G} \Rightarrow \mathcal{L}_{\mathcal{H}}$ of the group $\mathbf{G}$ onto the space $\mathcal{L}_{\mathcal{H}}$ of linear operators on Hilbert space $\mathcal{H}$, which preserves the property of group multiplication "$\bullet$"

$$\forall x, y, z \in \mathbf{G} : x \bullet y = z \Rightarrow D(x)D(y) = D(z) \tag{2.1}$$

where the multiplication on the r.h.s. of (2.1) is defined in the space $\mathcal{L}_{\mathcal{H}}$ of linear operators on $\mathcal{H}$. The representations will in general be denoted by boldface capital $\mathbf{D}$ with possible subscripts or superscripts, or as is common for SU(3) group, by a pair of nonnegative integers (i,j), specifying the so called *highest weight* of the representation (see Section 2.6). The image of a group element $g \in \mathbf{G}$ in such a representation will be denoted simply as $D(g)$.

Examples:

- $\mathbf{G}$=R, the set of real numbers with conventional addition as the group binary operation $\bullet$, $D(x) = \exp(i\alpha x)$ where α is a fixed real number. This representation plays a crucial role in the construction of abelian gauge theories, like QED.

- P_3, permutation group of three objects. Define (1,2) as permutation of 1,2 etc., (1,2,3) as simultaneous transpositions: $1 \rightarrow 2$, $2 \rightarrow 3$, $3 \rightarrow 1$ and e as a unit element. One of its representations is given by the following six matrices, with normal matrix multiplication defining the product of linear operators:

$$D(12) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad D(13) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad D(23) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

$$D(123) = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad D(321) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad D(e) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

Realization of operators by matrices: the action of any linear operator $\hat{O}$ on vectors $\mid j\rangle$ from a given normalized basis of $\mathcal{H}$ can be represented by means of the matrix O_{ij} defined as

$$O_{ji} \equiv \langle j \mid \hat{O} \mid i\rangle \Rightarrow \hat{O} \mid i\rangle = O_{ji} \mid j\rangle, \tag{2.2}$$

where the summation over the repeating indices is understood as usual. In all the following applications I shall, when talking about the group representation, always have in mind the above matrix representation.

Equivalence of representations: takes into account the fact that the matrices O_{ij} depend on the chosen basis of $\mathcal{H}$. Two representations $\mathbf{D_1}$ and $\mathbf{D_2}$ are said to be equivalent if there exists unitary operator $S \in \mathcal{L}_{\mathcal{H}}$ such that

$$\forall x \in \mathbf{G}, \ D_1(x) = SD_2(x)S^{-1}. \tag{2.3}$$

The transformation of $D(x)$ can be interpreted as resulting from the change of the basis of $\mathcal{H}$ induced by S.

Direct sum of representations: assuming $\mathbf{D}_i$ act on corresponding Hilbert space $\mathcal{H}_i$ of dimension n_i, $i = 1, 2$, the direct sum $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$ is the Hilbert space of dimension $n_1 + n_2$ and its $n_1 + n_2$ dimensional basis spanned by the set of vectors

$$\underbrace{\mid e_1^1\rangle, \mid e_2^1\rangle, \cdots, \mid e_{n_1}^1\rangle}_{\text{basis of } \mathcal{H}_1}, \underbrace{\mid e_1^2\rangle, \mid e_2^2\rangle, \cdots, \mid e_{n_2}^2\rangle}_{\text{basis of } \mathcal{H}_2}. \tag{2.4}$$

The corresponding direct sum of representations $\mathbf{D} = \mathbf{D_1} \oplus \mathbf{D_2}$ is made up from matrices $(n_1 + n_2) \times (n_1 + n_2)$ of the block diagonal form

$$\begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}. \tag{2.5}$$

Direct product of representations: a more complicated but simultaneously more interesting construction. Using the same notation as above, the basis of the direct product $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ is formed by $n_1 \times n_2$ pairs of vectors of the form $\mid e_i^1\rangle \mid e_j^2\rangle$ where $\mid e_i^j\rangle \in \mathcal{H}_j, j = 1, 2$. The direct product of representations $\mathbf{D_1}, \mathbf{D_2}, \mathbf{D} = \mathbf{D_1} \otimes \mathbf{D_2}$ is formed by the matrices which act on these basis vectors as follows ($g \in \mathbf{G}$ is an element of group $\mathbf{G}$)

$$\underbrace{D_1 \otimes D_2(g)}_{\in \mathbf{D_1} \otimes \mathbf{D_2}} \mid e_i^1\rangle \mid e_j^2\rangle = (\underbrace{D_1(g)}_{\in \mathbf{D_1}} \mid e_i^1\rangle)(\underbrace{D_2(g)}_{\in \mathbf{D_2}} \mid e_j^2\rangle). \tag{2.6}$$

The matrices corresponding to (2.6) are $(n_1 \times n_2) \times (n_1 \times n_2)$ dimensional and can be written as

$$[(D_1 \otimes D_2)(g)]_{ij,i'j'} = [D_1(g)]_{ii'} [D_2(g)]_{jj'}. \tag{2.7}$$

Reducibility of a representation: the possibility to transform all elements of a given representation $\mathbf{D}$, acting on the Hilbert space $\mathcal{H}$, by means of a unitary operator S to block–diagonal form:

$$\exists S \in \mathcal{L}_{\mathcal{H}} : \forall x \in \mathbf{G}, SD(x)S^{-1} = \begin{pmatrix} D_1(x) & 0 \\ 0 & D_2(x) \end{pmatrix}, \tag{2.8}$$

where the matrices $D_1(x), D_2(x)$ act on Hilbert spaces $\mathcal{H}_1$ and H_2 respectively and $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$. If such a unitary operator S does exist the representation $\mathbf{D}$ is called **reducible** and can be written as a direct sum $\mathbf{D} = \mathbf{D_1} \oplus \mathbf{D_2}$. If not $\mathbf{D}$ is said to be **irreducible**.

Further important concepts:

- **Abelian** group: the group multiplication • is **commutative**.

- **Finite and infinite groups:** according to the number of independent elements

- **Compact and noncompact groups:** roughly speaking compact groups have finite "volume" as measured by some measure on the group space, while the noncompact have infinite volume. Example: the group of phase factors introduced above is compact, as the unit circle in complex plane is the finite.

2.2 Lie groups and algebras

I shall not attempt to give an exact definition of this important concept for the case of a general Lie group, as it would require introducing too much of heavier mathematical arsenal, which we shall actually not need in our practical applications. Instead let me emphasize the basic attributes of this class of groups. Lie groups are characterized by the fact that their elements are labelled by a set of continuous parameters, which can be imagined as coordinates in a subset of $\mathbf{R}^n$, n-dimensional Cartesian space with conventionally defined metric and topological properties. These properties are necessary for the formulation of the concept of continuous mapping and continuous functions, which are central to mathematically rigorous definition of Lie groups. The exact meaning of the words "labelled by a set of continuous parameters" is also more involved, but roughly speaking there exists a *local* one-to-one mapping between the group elements and points in some subset of $\mathbf{R}^n$ which is **continuous** in both ways [1] and which allows us to translate all the operations and questions from the group space to analogous operations and questions in $\mathbf{R}^n$ where we know what to do. For instance for a continuous group to be a Lie group the group multiplication $x \bullet y$ must be a continuous function of both x and y and the operation of taking the inverse x^{-1} must be continuous function of x. We shall not go into more details of this abstract construction as for us the group elements will always be matrices, parameterized by a few real numbers and for the vector space of matrices the topology is easily defined via the norm of a matrix.

For Lie groups, and in particular their matrix representations there is also intuitive understanding of the group "volume" and thus of the difference between the compact and noncompact groups. I shall now present, without proofs, several propositions that will be useful in further considerations.

Proposition 2.1 *Any element of a* **compact** *Lie group can be written in the form*

$$\forall g \in \mathbf{G}, g = g(\alpha_a) = \exp(i\alpha_a X_a),$$

where the operators X_a, called **generators** *of the Lie group $\mathbf{G}$, form the basis of a vector space $\mathbf{X}$ (over the field of complex mumbers) of dimension m with the operation of "adding", denoted as "+".*

The classification and properties of representations are simplest for the class of compact Lie groups, where all those needed in formulating the Standard Model do belong.

Proposition 2.2 *All irreducible representations of compact Lie groups are finite dimensional.*

Proposition 2.3 *Finite dimensional representations of a compact Lie group $\mathbf{G}$ are equivalent to representations by* unitary *operators, i.e. the generators X_a are* hermitian *operators.*

In the case of matrix groups, or representations, both the elements g of the group $\mathbf{G}$ and the generators X_a are again *matrices*. As a result, also the generators of any irreducible representation of a compact Lie group can be represented by finite dimensional hermitian matrices. Note that the generators of a given Lie group are not determined by this group *uniquely*, as any change of the basis vectors of the associated Hilbert space implies the change of these generators.

Algebra: A quite general concept, we shall restrict our discussion to algebras over the field $\mathcal{C}$ of complex numbers. The $\mathcal{C}$-algebra is a vector space $\mathcal{A}$ over the field of complex numbers equipped with a bilinear binary operation (here denoted as simple multiplication) $\mathcal{A} \times \mathcal{A} \to \mathcal{A}$, which means that $\forall x, y, z \in \mathcal{A}, a, b \in \mathcal{C}$

- (x+y)z=xz+yz

- x(y+z)=xy+xz

- (ax)(by)=(ab)(xy)

In the special case of **Lie algebra**, the binary operation, called "commutator", is *antisymmetric* and satisfies the Jacobi identity (2.12).

Structure constants: Due to the fact that the product of elements of a Lie group $\mathbf{G}$ like

$$\exp(i\lambda X_a)\exp(i\lambda X_b)\exp(-i\lambda X_a)\exp(-i\lambda X_b)$$

---

[1]The continuity, being defined by means of open sets, needs topology on the group space as well as in $\mathbf{R^n}$. For mathematically rigorous treatment of this topic consult [24].

is also an element of this group, it must be expressible as $\exp(i\beta_c X_c)$. Using the Taylor expansion on both sides we find

$$[X_a, X_b] \equiv X_a X_b - X_b X_a = i f_{abc} X_c; \quad \beta_c = -\lambda^2 f_{abc}, \tag{2.9}$$

where f_{abc} are real numbers, called the **structure constants** of the Lie group **G**. They are by definition *antisymmetric* in first two indices. Similarly to generators they are, however, *not unique* and do depend on the choice of the basis in **X**. They can be used to express the product of two elements of the group as follows:

$$\exp(i\alpha_a X_a)\exp(i\beta_b X_b) = \exp(i\gamma_c X_c) \Rightarrow \gamma_c = \alpha_c + \beta_c - \frac{1}{2} f_{abc}\alpha_a\beta_b + \cdots \tag{2.10}$$

Proposition 2.4 *The structure constants f_{abc} satisfy the following Jacobi identity:*

$$f_{ade}f_{bcd} + f_{cde}f_{abd} + f_{bde}f_{cad} = 0. \tag{2.11}$$

Proof: (2.11) is a simple consequence of the following relation between the commutators of generators:

$$[X_a, [X_b, X_c]] + [X_c, [X_a, X_b]] + [X_b, [X_c, X_a]] = 0, \tag{2.12}$$

which is straightforward to verify. In this case the vector space is that of the generators X_a and the "commutator" is defined by the conventional commutator of matrices, representing the generators. Note that the elements of an algebra spanned on the generators of a compact Lie group contain also operators that are *not* hermitian!

Analogously as in the case of groups we can define representation of an algebra as a mapping of its elements to the operators on some Hilbert space, which conserves the commutator (on a Hilbert space of operators commutator of its elements A, B is simply $AB - BA$) and introduce the concepts of reducibility, irreducibility etc..

Generators of the direct product $\mathbf{D_1} \otimes \mathbf{D_2}$ of n_1 dimensional representation $\mathbf{D_1}$ and n_2 dimensional representation $\mathbf{D_2}$ are $(n_1 \times n_2) \times (n_1 \times n_2)$ dimensional matrices of the form

$$\left(X^{\mathbf{D_1}\otimes\mathbf{D_2}}\right)_{ij,i'j'} = \underbrace{\left(X^{\mathbf{D_1}}\right)_{ii'}}_{\text{acts on } \mathcal{H}_1 \text{ only}} \delta_{jj'} + \delta_{ii'} \underbrace{\left(X^{\mathbf{D_2}}\right)_{jj'}}_{\text{acts on } \mathcal{H}_2 \text{ only}} . \tag{2.13}$$

Adjoint representation: On the basis of the Jacobi identity it is straightforward to verify that the following $n \times n$ matrices, where n is the dimension of the vector space **X**,

$$(T_a)_{bc} \equiv -i f_{abc} \tag{2.14}$$

do satisfy the same commutation relations as X_a themselves:

$$[T_a, T_b] = i f_{abc} T_c \tag{2.15}$$

and therefore also form a representation of **G**. This representation is called **adjoint** representation and will play an important role in the description of particle multiplets in quark model and of gluons in QCD.

Although we can choose any basis in **X**, one which is particularly suitable is defined by imposing the following normalization conditions:

$$\text{Tr}(X_a X_b) = \lambda \delta_{ab}, \tag{2.16}$$

where the constant λ is *positive* for compact Lie groups. [2] In this normalization we have

$$f_{abc} = -\frac{i}{\lambda}\text{Tr}([X_a, X_b]X_c) = -\frac{i}{\lambda}\text{Tr}([X_c, X_a]X_b) = -\frac{i}{\lambda}\text{Tr}([X_b, X_c]X_a), \tag{2.17}$$

which implies that in this normalization f_{abc} is *fully antisymmetric* tensor.

Hilbert space of adjoint representation: The Hilbert space associated to the adjoint representation can be constructed from the *vector* space **X** spanned on the generators of **G** by defining the binary

---

[2]Note that since for compact Lie groups $\text{Tr}(X_a X_b)$ is real symmetric tensor, it can always be diagonalized.

operation of a "scalar" product of any two vectors $\mid X_a \rangle, \mid X_b \rangle$, associated to the generators X_a, X_b, in the following way:

$$\langle X_b \mid X_a \rangle \equiv \frac{1}{\lambda} \text{Tr} \left(X_b^+ X_a \right), \tag{2.18}$$

where "$+$" denotes the hermitian conjugation of the operator and $\langle a \mid, \mid a \rangle$ are the usual Dirac "bra" and "ket" vectors [24]. For the adjoint representation we have the following chain of equalities

$$T_a \mid T_b \rangle = (T_a)_{cb} \mid T_c \rangle = -i f_{acb} \mid T_c \rangle = \mid i f_{abc} T_c \rangle = \mid [T_a, T_b] \rangle, \tag{2.19}$$

where the first equality is a consequence of the definition of action of the matrix T_a on the ket vector $\mid T_b \rangle$, the second uses just the definition (2.14), the third is trivial and the last one returns to (2.15). In other words the action of the operator (for us matrix) T_a on the ket vector $\mid T_b \rangle$ produces the vector, associated to the commutator of the matrices T_a, T_b!

Further important concepts:

- **Abelian** algebra: Algebra with the property that all its elements *commute* with each other.

- **Invariant subalgebra** $\mathcal{S}$ of algebra $\mathcal{A}$: $\forall a \in \mathcal{S}, \forall x \in \mathcal{A} : [a, x] \in \mathcal{S}$.

- **Simple algebra:** algebra which contains no nontrivial invariant subalgebra.

- **Semisimple algebra:** algebra which contains no *abelian* invariant subalgebra. This type of groups has great physical relevance in theories unifying various kinds of interactions, like the electroweak theory within the SM.

- **Rank** of the algebra: the maximal number of mutually commuting generators. Crucial characteristics of nonabelian algebras. Rank determines the number of independent quantum numbers, which uniquely characterize each state within a given irreducible representation, or, as is common to say, multiplet.

2.3   SU(2) group and algebra

The simplest of nonabelian Lie groups, with plenty of applications in particle physics is the SU(2) group. Moreover, most of the techniques useful for the more complicated case of SU(3) and other groups can be generalized from the technically simpler case of SU(2).

Definition 2.1 *SU(2) is formed by unitary 2×2 matrices with unit determinant. The associated Lie algebra is made out of traceless hermitian matrices 2×2.*

There are 3 independent generators J_1, J_2, J_3 of SU(2), which form the basis of SU(2) albegra and which satisfy the well-known commutation relations

$$[J_i, J_j] = i \varepsilon_{ijk} J_k. \tag{2.20}$$

Remark 2.1 *In fact this algebra is equivalent to that of SO(3) group, indicating that the relation between the group and the associated algebra is not unique. This has to do with the fact that albegras express only the local properties of the groups, but don't describe the global ones.*

Representations of SU(2): can be constructed by means of various techniques. The one which proved to be the most useful for generalization to SU(3) is called the method *of the highest weight*. We shall be interested in the construction of **irreducible** representations only.

Construction: proceeds in several steps:

1. SU(2) is of rank 1, i.e. there is only one operator, which fully characterizes the state within a given multiplet. Let us choose J_3 for that purpose and denote the state with the highest weight j as $\mid j \rangle$:

$$J_3 \mid j \rangle = j \mid j \rangle; \quad \langle i \mid j \rangle = \delta_{ij}. \tag{2.21}$$

2. define the lowering and rising operators

$$J^{\pm} \equiv \frac{J_1 \pm \mathrm{i}J_2}{\sqrt{2}} \quad \Rightarrow [J_3, J^{\pm}] = \pm J^{\pm}, \quad [J^+, J^-] = J_3. \tag{2.22}$$

3. As a consequence of (2.22) we get

$$J_3(J^{\pm} \mid m\rangle) = (m \pm 1)J^{\pm} \mid m\rangle, \tag{2.23}$$

which implies that in any multiplet there exists j such that $J^+ \mid j\rangle = 0$, which is called the highest weight.

4. systematic application of the lowering operator starting from the state of the highest weight (or of rising operator starting from the state with lowest weight) allows us to construct all states of any multiplet.

5. acting by J^- on the state with highest weight yields $J^- \mid j\rangle = N_j \mid j-1\rangle$ where the normalization factor N_j can be determined as follows:

$$N_j^* N_j \underbrace{\langle j-1 \mid j-1\rangle}_{1} = \langle j \mid \underbrace{J^+ J^-}_{J^- J^+ + J_3} \mid j\rangle = j \underbrace{\langle j \mid j\rangle}_{1}, \tag{2.24}$$

where we used the basic fact that $J^+ \mid j\rangle = 0$. As a result we find, using particular sign convention, $N_j = \sqrt{j}$.

6. In fact we can start with any state $\mid j-k\rangle$ for which we define the normalization factor as follows :

$$J^- \mid j-k\rangle \equiv N_{j-k} \mid j-k-1\rangle \Rightarrow J^+ \mid j-k-1\rangle = N_{j-k} \mid j-k\rangle. \tag{2.25}$$

Repeating the above procedure of application of $J^+ J^-$ to $\mid j-k\rangle$ yields

$$N_{j-k}^2 = \langle j-k \mid \underbrace{J^+ J^-}_{[J^+, J^-] + J^- J^+} \mid j-k\rangle = \underbrace{\langle j-k \mid J_3 \mid j-k\rangle}_{j-k} + \langle j-k \mid J^- \underbrace{J^+ \mid j-k\rangle}_{\underbrace{N_{j-k+1} \mid j-k+1\rangle}_{N_{j-k+1}^2}},$$

$$\tag{2.26}$$

which implies the recurrence relation between the normalization coefficients:

$$N_{j-k}^2 - N_{j-k+1}^2 = j-k. \tag{2.27}$$

Writing the full sequence of such recurrence relations

$$\begin{array}{ccccc}
N_j^2 & - & N_{j+1}^2 & = & j \\
N_{j-1}^2 & - & N_j^2 & = & j-1 \\
& \vdots & & \vdots & \\
N_{j-k}^2 & - & N_{j-k+1}^2 & = & j-k \\
& \vdots & & \vdots & \\
N_{-j}^2 & - & N_{-j+1}^2 & = & -j
\end{array} \tag{2.28}$$

and summing the first $k+1$ lines on both sides of these relations we easily find (taking into account that $N_{j+1} = 0$)

$$N_{j-k}^2 = \sum_{i=0}^{k}(j-i) = (k+1)\left[j - \frac{k}{2}\right] = (k+1)\left[j - k + \frac{k}{2}\right]. \tag{2.29}$$

For $k = 2j$ we get $\Leftrightarrow m = -j \Rightarrow N_{-j} = 0$ which implies that the multiplet with highest weight j has $2j+1$ states and that j has to be of the form $2j = l$ with l integer. Note that by summing all the lines in (2.28) we get trivial identity $0 = 0$.

A general state within the multiplet with highest weight j, characterized by the eigenvalue m of J_3, will be denoted as $\mid j, m\rangle$. In terms of the numbers j, m the normalization factor (2.29) can be rewritten as

$$N_{j,m}^2 = \frac{1}{2}(j-m+1)(j+m). \tag{2.30}$$

22

Fundamental representation of SU(2): Recall that the generators of SU(2) can be written by means of Pauli 2×2 matrices σ_i as $J_i = \frac{1}{2}\sigma_i$, where

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.31}$$

This representation can be used to construct all other multiplets of SU(2). This can be done exploiting the concepts of the **direct sum and product** of multiplets. Out of the direct product of basic SU(2) doublets

$$\left[\underbrace{\mathbf{D}^{(1/2)} \otimes \mathbf{D}^{(1/2)} \cdots \otimes \mathbf{D}^{(1/2)}}_{n \text{ times}} \right]_{i_1,i_2,\cdots,i_n;j_1,j_2\cdots,j_n}, \tag{2.32}$$

the matrices of which are given as

$$D^{(1/2)}_{i_1,j_1} D^{(1/2)}_{i_2,j_2} \cdots D^{(1/2)}_{i_n,j_n}, \tag{2.33}$$

one can construct any multiplet of SU(2). The above direct product of multiplets is *reducible* and obviously symmetric under the permutations of the indices $1, 2, \cdots, n$. Individual irreducible components of this reducible representation are then characterized by particular type of symmetry of states on which they act.

Starting from the symmetric state given as the direct product of n one particle states $\mid 1/2, 1/2\rangle$ and applying the lowering operator J^- we get all states of the multiplet corresponding to the highest weight $j = n/2$. In other words the subspace of fully symmetric states of n basic doublets corresponds to the multiplet with $j = n/2$. Physically, all these states describe the system of n **identical** noninteracting particles, each of them corresponding to the fundamental representation $\mathbf{D}^{(1/2)}$. For instance the direct products of two and three SU(2) doublets decomposes as follows

$$\mathbf{2} \otimes \mathbf{2} = \mathbf{3}_s \oplus \mathbf{1}, \quad \mathbf{2} \otimes \mathbf{2} \otimes \mathbf{2} = \mathbf{4}_s \oplus \mathbf{2}_{m,s} \oplus \mathbf{2}_{m,a} \tag{2.34}$$

where the subscript "s" denotes representation symmetric under the permutations of the product representations whereas "ms" (mixed symmetric) and "ma" (mixed antisymmetric) denote two equivalent doublet representations, which differ in their symmetry properties under permutation of the first two doublets. The adjective "mixed" reflects the fact that as far as other permutations are concerned these representation possess no definite symmetry.

2.4 Application of SU(2) group: the isospin

Let us consider a system of finite number of protons and neutrons, neglecting all effects of electromagnetic and weak interactions (like the small difference between proton and neutron masses etc.). This approximation makes sense as nucleons interact much more intensively via **strong** interactions. Starting from late forties, experiments had shown quite convincingly that under these assumptions strong interactions between protons and neutrons are **charge invariant**. This observation and its generalization to be discussed in the next section form the basis of the quark-parton model.

Remark 2.2 *There is a very close analogy, based on the same mathematics used in their description, between the isospin symmetry of strong interactions and rotational symmetry of a two-particle spin-spin interaction Hamiltonian $H_{int} \propto \vec{s_1}\vec{s_2}$.*

There are various ways how to describe such a system. A particularly suitable one is based on the use of **creation** and **annihilation** operators (P_α^+, P_α for protons and N_α^+, N_α for neutrons in state α), which allow us to build all multiparticle states via their action on the vacuum state of the Hamiltonian, denoted as $\mid 0\rangle$ and excited states, denoted generically $\mid s\rangle$:

$$\begin{array}{llll} P_\alpha^+ \mid s\rangle &= \mid s + \text{proton in state } \alpha\rangle &, \quad P_\alpha \mid s\rangle &= \mid s - \text{proton in state } \alpha\rangle \\ P_\alpha \mid 0\rangle &= 0 &, \quad \langle 0 \mid P_\alpha^+ &= 0. \end{array} \tag{2.35}$$

The commutation relations of creation and annihilation operators are assumed to be the following:

$$\{P_\alpha^+, P_\beta\} \equiv P_\alpha^+ P_\beta + P_\beta P_\alpha^+ = \delta_{\alpha,\beta}, \tag{2.36}$$

$$\{P_\alpha^+, P_\beta^+\} = \{P_\alpha, P_\beta\} = 0, \tag{2.37}$$

$$\left[P_\alpha^+, N_\beta^+\right] = [P_\alpha, N_\beta] = \left[P_\alpha^+, N_\beta\right] = \left[P_\alpha, N_\beta^+\right] = 0. \tag{2.38}$$

The first two of the above three equations enforce, when restricted to the case $\alpha = \beta$, the Pauli exclusion principle, which stipulates that there cannot be two identical fermions in any given state. For $\alpha \neq \beta$ they guarantee the antisymmetry of wave functions of systems of protons or neutrons. The third set of equalities tells us that protons are different from neutrons and so we can interchange the order of application of respective operators, be it creation or annihilation, with impunity.

These operators may be used to describe, for instance, the repulsion between two protons, with interaction Hamiltonian given as

$$H_{int} = \sum_{\alpha,\beta} P_\alpha^+ P_\alpha V_{\alpha\beta}(r) P_\beta^+ P_\beta, \tag{2.39}$$

where the interaction potential V(r) may in principle depend on the coordinate r as well.

Out of $P_\alpha^+, P_\alpha, N_\alpha^+, N_\alpha$ we can construct the operators $J^\pm, J_3$ introduced earlier (let me denote them as $T^\pm, T_3$):

$$T^+ \equiv \frac{1}{\sqrt{2}} \sum_\alpha P_\alpha^+ N_\alpha; \ \ T^- \equiv \frac{1}{\sqrt{2}} \sum_\alpha N_\alpha^+ P_\alpha, \tag{2.40}$$

$$T_3 \equiv \frac{1}{2} \sum_\alpha \left(P_\alpha^+ P_\alpha - N_\alpha^+ N_\alpha\right). \tag{2.41}$$

It is straightforward to show, starting from the anticommutation relations for P_α^+ etc., that

$$\left[T_3, T^\pm\right] = \pm T^\pm; \ \ \left[T^+, T^-\right] = T_3 \tag{2.42}$$

as required for the generators of SU(2) group. The group we have just constructed is called the **isospin** group. The invariance of strong interactions with respect to it implies that the appropriate Hamiltonian H_s commutes with all these generators:

$$\left[H_s, T^\pm\right] = [H_s, T_3] = 0. \tag{2.43}$$

Writing the state vector of the nucleon in a two-component column form
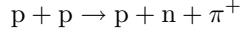
$$| \psi \rangle = \left(\begin{array}{c} \psi_1 \\ \psi_2 \end{array} \right) = \psi_1 | p \rangle + \psi_2 | n \rangle \tag{2.44}$$

as a superposition of pure proton and neutron states $| p \rangle | n \rangle$, the interpretation of the transition operators $T^\pm$ becomes clear: they transform neutron to proton and vice versa. The relation between the third component T_3 of the isospin and the electric charge Q reads

$$Q = T_3 + \frac{B}{2}, \tag{2.45}$$

where B is the baryon number of the system ($B = 1, -1, 0$ for the nucleon, antinucleon and pion respectively).

Remark 2.3 *The power of this formalism as a way to describe charge invariance of strong interactions becomes highly nontrivial when it is extended to cover the interactions of nucleons with pions and other hadrons like, for instance,*

$$p + p \rightarrow p + n + \pi^+$$

To extend the above construction to cover the pions as well, we first introduce the creation and annihilation operators for all their charged states. To avoid double superscripts let me denote them as follows: $a_{\pi^+}, a_{\pi^-}, a_{\pi^0}$ for the annihilation operators and similarly for the creation ones. The commutation relations defining the associated algebra is the same as for the nucleons, but with one very important change: the anticommutator in (2.37) etc. is replaced with the *commutator*! It is nontrivial, but easy to verify that in both cases one gets the same commutation relations for the generators of the SU(2) group. For pions the

generators can be written, bearing in mind their physical interpretation. The expressions for the generators describing a system of both nucleons and pions then read

$$T^+ \equiv \frac{1}{\sqrt{2}} \sum_\alpha P_\alpha^+ N_\alpha + 1 \sum_\alpha \left(a_{\pi^+}^+(\alpha) a_{\pi^0}(\alpha) + a_{\pi^0}^+(\alpha) a_{\pi^-}(\alpha) \right), \tag{2.46}$$

$$T^- \equiv \frac{1}{\sqrt{2}} \sum_\alpha N_\alpha^+ P_\alpha + 1 \sum_\alpha \left(a_{\pi^-}^+(\alpha) a_{\pi^0}(\alpha) + a_{\pi^0}^+(\alpha) a_{\pi^+}(\alpha) \right), \tag{2.47}$$

$$T_3 \equiv \frac{1}{2} \sum_\alpha \left(P_\alpha^+ P_\alpha - N_\alpha^+ N_\alpha \right) + 1 \sum_\alpha \left(a_{\pi^+}^+(\alpha) a_{\pi^+}(\alpha) - a_{\pi^-}^+(\alpha) a_{\pi^-}(\alpha) \right), \tag{2.48}$$

where I have deliberately written "1" in front of the second contributions expressing the fact that pions carry one unit of isospin. The application of the isospin symmetry can be illustrated on the following simple

Example: the scattering of pions on nucleons. From the point of view of the isospin symmetry the system of one pion and one nucleon is described by means of the direct product of a doublet and a triplet of SU(2). The states of definite charge combinations can be written as follows

$$
\begin{aligned}
|\,\pi^+ p\rangle &= |\,1,1\rangle\,|\,\tfrac{1}{2},\tfrac{1}{2}\rangle & = |\,\tfrac{3}{2},\tfrac{3}{2}\rangle, & \langle \pi^+ p \,|\, S \,|\, \pi^+ p\rangle &= a_{3/2}, \\
|\,\pi^+ n\rangle &= |\,1,1\rangle\,|\,\tfrac{1}{2},-\tfrac{1}{2}\rangle & = \sqrt{\tfrac{1}{3}}\,|\,\tfrac{3}{2},\tfrac{1}{2}\rangle + \sqrt{\tfrac{2}{3}}\,|\,\tfrac{1}{2},\tfrac{1}{2}\rangle, & \langle \pi^+ n \,|\, S \,|\, \pi^+ n\rangle &= \tfrac{1}{3}a_{3/2} + \tfrac{2}{3}a_{1/2}, \\
|\,\pi^- p\rangle &= |\,1,-1\rangle\,|\,\tfrac{1}{2},\tfrac{1}{2}\rangle & = \sqrt{\tfrac{1}{3}}\,|\,\tfrac{3}{2},-\tfrac{1}{2}\rangle - \sqrt{\tfrac{2}{3}}\,|\,\tfrac{1}{2},-\tfrac{1}{2}\rangle, & \langle \pi^- p \,|\, S \,|\, \pi^- p\rangle &= \tfrac{1}{3}a_{3/2} + \tfrac{2}{3}a_{1/2}, \\
|\,\pi^- n\rangle &= |\,1,-1\rangle\,|\,\tfrac{1}{2},-\tfrac{1}{2}\rangle & = |\,\tfrac{3}{2},-\tfrac{3}{2}\rangle, & \langle \pi^- n \,|\, S \,|\, \pi^- n\rangle &= a_{3/2}, \\
|\,\pi^0 p\rangle &= |\,1,0\rangle\,|\,\tfrac{1}{2},\tfrac{1}{2}\rangle & = \sqrt{\tfrac{2}{3}}\,|\,\tfrac{3}{2},\tfrac{1}{2}\rangle - \sqrt{\tfrac{1}{3}}\,|\,\tfrac{1}{2},\tfrac{1}{2}\rangle, & \langle \pi^0 p \,|\, S \,|\, \pi^0 p\rangle &= \tfrac{2}{3}a_{3/2} + \tfrac{1}{3}a_{1/2}, \\
|\,\pi^0 n\rangle &= |\,1,0\rangle\,|\,\tfrac{1}{2},-\tfrac{1}{2}\rangle & = \sqrt{\tfrac{2}{3}}\,|\,\tfrac{3}{2},-\tfrac{1}{2}\rangle + \sqrt{\tfrac{1}{3}}\,|\,\tfrac{1}{2},-\tfrac{1}{2}\rangle, & \langle \pi^0 n \,|\, S \,|\, \pi^0 n\rangle &= \tfrac{2}{3}a_{3/2} + \tfrac{1}{3}a_{1/2}.
\end{aligned}
\tag{2.49}
$$

The coefficients in front of the states with definite total isospin, called **Clebsh-Gordan** coefficients, can be obtained directly by application of the above described technique of highest weight.

The S-matrix elements of various channels are thus related as there are only two independent amplitudes, corresponding to full isospin equal to $1/2$ and $3/2$. These can also be used to write down the amplitudes of other processes like

$$\langle \pi^0 p \,|\, S \,|\, \pi^+ n\rangle = \frac{\sqrt{2}}{3} \left(a_{3/2} - a_{1/2} \right) = \langle \pi^- p \,|\, S \,|\, \pi^0 n\rangle. \tag{2.50}$$

A simple manifestation of these relations is the ratio of the cross-sections to produce the Δ resonance in various pion-nucleon channels like, for instance,

$$r \equiv \frac{\sigma(\pi^+ p \to \Delta^{++})}{\sigma(\pi^- p \to \Delta^0)} = \frac{\sigma(\pi^+ p \to \Delta^{++} \to \pi^+ p)}{\sigma(\pi^- p \to \Delta^0 \to \pi^- p) + \sigma(\pi^- p \to \Delta^0 \to \pi^0 n)} = \frac{1}{\frac{1}{9} + \frac{2}{9}} = 3. \tag{2.51}$$

2.5 Weights and roots of compact Lie algebras

In this section I briefly outline the extension of the technique of the highest weight to general compact Lie groups. The basic idea of this extension is to divide the basis vectors of the associated vector space of a given compact Lie algebra $\mathcal{A}$ into two groups

- $H_i, i = 1, \cdots m$ such that

$$[H_i, H_j] = 0, \forall i, j, \tag{2.52}$$

  where m, called the **rank** of the group, is the *maximal* number of mutually commuting elements of $\mathcal{A}$, forming the **Cartan subalgebra**. They can be chosen hermitian and thus interpreted as generators.

- $n - m$ remaining elements E_α (n being the number of generators) such that

$$[H_i, E_\alpha] = \alpha_i E_\alpha, \tag{2.53}$$

  where the m dimensional, *nonzero* vectors α are called the **roots** of the Lie algebra.

We adopt the following normalization of H_i, E_α (λ is defined in (2.16))

$$\begin{aligned} \langle E_\alpha \mid E_\beta \rangle &= \lambda^{-1}\mathrm{Tr}(E_\alpha^+ E_\beta) &= \delta_{\alpha\beta} \\ \langle H_i \mid H_j \rangle &= \lambda^{-1}\mathrm{Tr}(H_i H_j) &= \delta_{ij}. \end{aligned} \tag{2.54}$$

The Cartan subalgebra contains those generators which are analogies of J_3 of SU(2), where there was just one of them, while the E_α are generalizations of the lowering and rising operators $J^\pm$. By definition we can diagonalize all H_i simultaneously, defining **real** vectors $\mu = (\mu_1, \cdots, \mu_m)$ of **weights** of a given multiplet **D**

$$H_i \mid \mu, \mathbf{D} \rangle = \mu_i \mid \mu, \mathbf{D} \rangle, \tag{2.55}$$

the number of which is limited by the dimension of **D**. There are several simple consequences of the above definitions:

1. Roots α are actually weights of the adjoint representation as in this representation

$$H_i \mid E_\alpha \rangle = \mid [H_i, E_\alpha] \rangle = \alpha_i \mid E_\alpha \rangle. \tag{2.56}$$

   Consequently α_i are vectors of *real numbers*.

2. Taking the hermitian conjugate of (2.53) we find

$$[H_i, E_\alpha]^+ = -[H_i, E_\alpha^+] = \alpha_i E_\alpha^+ \quad \Rightarrow \quad E_\alpha^+ = E_{-\alpha}, \tag{2.57}$$

   i.e. E_α are **not** hermitian.

3. Using the commutation relations we easily get

$$H_i(E_{\pm\alpha} \mid \mu, \mathbf{D} \rangle) = (\mu \pm \alpha)_i E_{\pm\alpha} \mid \mu, \mathbf{D} \rangle \Rightarrow E_{\pm\alpha} \mid \mu, \mathbf{D} \rangle = N_{\pm\alpha,\mu} \mid \mu \pm \alpha, \mathbf{D} \rangle, \quad N_{-\alpha,\mu} = N_{\alpha,\mu-\alpha}^* \tag{2.58}$$

   and analogously to the case of the SU(2) algebra, the normalization factors $N_{\alpha,\mu}$ can be chosen *real*.

4. As

$$H_i(E_\alpha \mid E_{-\alpha} \rangle) = E_\alpha \underbrace{H_i \mid E_{-\alpha} \rangle}_{-\alpha_i \mid E_{-\alpha} \rangle} + \underbrace{[H_i, E_\alpha]}_{\alpha_i E_\alpha} \mid E_{-\alpha} \rangle = 0 \tag{2.59}$$

   the state $E_\alpha \mid E_{-\alpha} \rangle$ has zero weight and can therefore be expressed as linear combination of the vectors $\mid H_i \rangle$. Using the normalization (2.16) one moreover finds (exercise 2.2)

$$E_\alpha \mid E_{-\alpha} \rangle = \mid [E_\alpha, E_{-\alpha}] \rangle = \sum_j \alpha_j \mid H_j \rangle \Leftrightarrow [E_\alpha, E_{-\alpha}] = \alpha_j H_j. \tag{2.60}$$

The construction of multiplets of a general compact Lie group follows closely that of the SU(2) group. We shall again introduce the concept of highest weight, but as this concept is in general case somewhat ambiguous, let me start with an arbitrary state $\mid \mu, \mathbf{D} \rangle$ of **D** and apply to it powers of the operators $E_{\pm\alpha}$. As **D** is finite dimensional, there *must* exist nonnegative integers p, q such that

$$E_\alpha^{p+1} \mid \mu, \mathbf{D} \rangle = E_{-\alpha}^{q+1} \mid \mu, \mathbf{D} \rangle = 0. \tag{2.61}$$

Proceeding similarly as in the case of SU(2)

$$\langle \mu, \mathbf{D} \mid \underbrace{[E_\alpha, E_{-\alpha}]}_{\underbrace{\alpha_j H_j}_{\underbrace{\alpha\mu \mid \mu, \mathbf{D} \rangle}_{\alpha\mu}}} \mid \mu, \mathbf{D} \rangle = \langle \mu, \mathbf{D} \mid E_\alpha \underbrace{E_{-\alpha} \mid \mu, \mathbf{D} \rangle}_{\underbrace{N_{-\alpha,\mu} \mid \mu - \alpha, \mathbf{D} \rangle}_{\mid N_{-\alpha,\mu} \mid^2}} - \langle \mu, \mathbf{D} \mid E_{-\alpha} \underbrace{E_\alpha \mid \mu, \mathbf{D} \rangle}_{\underbrace{N_{\alpha,\mu} \mid \mu + \alpha, \mathbf{D} \rangle}_{\mid N_{\alpha,\mu} \mid^2}} \tag{2.62}$$

we get the following set of equations

$$
\begin{array}{rcll}
| N_{\alpha,\mu+\alpha(p-1)} |^2 & - & \overbrace{| N_{\alpha,\mu+\alpha p} |^2}^{0} & = \alpha(\mu+p\alpha) \\
& \vdots & \vdots & \\
| N_{\alpha,\mu} |^2 & - & | N_{\alpha,\mu+\alpha} |^2 & = \alpha(\mu+\alpha) \\
| N_{\alpha,\mu-\alpha} |^2 & - & | N_{\alpha,\mu} |^2 & = \alpha\mu \\
| N_{\alpha,\mu-2\alpha} |^2 & - & | N_{\alpha,\mu-\alpha} |^2 & = \alpha(\mu-\alpha) \\
& \vdots & \vdots & \\
\underbrace{| N_{\alpha,\mu-(q+1)\alpha} |^2}_{0} & - & | N_{\alpha,\mu-q\alpha} |^2 & = \alpha(\mu-q\alpha).
\end{array}
\tag{2.63}
$$

Summing the first $p+1$ of the above equations we get the explicit expression

$$
N^2_{-\alpha,\mu} = N^2_{\alpha,\mu-+)\alpha} = (p+1)\left[\alpha\mu - \frac{p}{2}\alpha^2\right],
\tag{2.64}
$$

which reduces to (2.29) for the SU(2) group when we set $k = p$, $\mu = j - k$ and $\alpha = \pm 1$. Summing all equations in (2.63) we get

$$
0 = (p+q+1)(\alpha\mu) + \alpha^2\left(\frac{p(p+1)}{2} - \frac{q(q+1)}{2}\right) \Rightarrow \frac{2\alpha\mu}{\alpha^2} = q - p,
\tag{2.65}
$$

which determines the relation between the *roots* of the group and *weights* of its representations. Considering in particular the adjoint representation, with weights equal to the roots themselves we get the following condition for any pair of roots:

$$
\frac{2\alpha\beta}{\alpha^2} = q - p = m; \quad \frac{2\beta\alpha}{\beta^2} = q' - p' = m' \Rightarrow \frac{(\alpha\beta)^2}{\alpha^2\beta^2} \equiv \cos^2\vartheta = \frac{mm'}{4},
\tag{2.66}
$$

where m, m' are integers, determining the angle between the roots considered as vectors in m-dimensional space. The integers p, q describe the shifting of the weight $\mu = \beta$ by means of the operator E_α, while in the case of p', q' the roles of α and β are reversed. The above formula implies that only those values of m, m' are allowed for which $mm' = 0, 1, 2, 3, 4$. For SU(2) we have $m = m' = 0, \pm 2$ as the roots corresponding to $J^\pm$ are one-dimensional and equal ± 1.

Composition of roots: Using the Jacobi identity, which implies

$$
[H_i, [E_\alpha, E_\beta]] = -[E_\beta, \underbrace{[H_i, E_\alpha]}_{\alpha_i E_\alpha}] - [E_\alpha, \underbrace{[E_\beta, H_i]}_{-\beta_i E_\beta}] = (\alpha+\beta)_i[E_\alpha, E_\beta],
\tag{2.67}
$$

we conclude that

- $[E_\alpha, E_\beta] \propto E_{\alpha+\beta}$ **iff** the root $\alpha + \beta$ does exist,

- $[E_\alpha, E_\beta] = 0$ if $\alpha + \beta$ is not a root, but $\alpha \neq -\beta$,

- $[E_\alpha, E_{-\alpha}] = \alpha_j H_j$ according to (2.60).

Example: consider the SU(3) group of unitary matrices 3×3 with unit determinant, which will play a crucial role in the following applications. The associated algebra is formed by 8 traceless hermitian matrices 3×3. The eight generators T_a can be chosen in many different ways, the simplest and particularly convenient being that of Gell-Mann, who introduced the following matrices (now called Gell-Mann matrices)

$$
\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},
$$
$$
\lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \lambda_8 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.
\tag{2.68}
$$
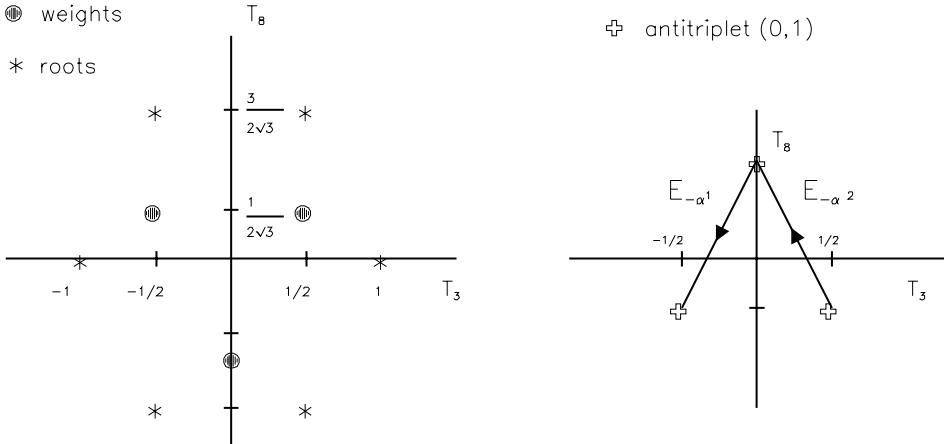
Figure 2.1: Roots and weights of the fundamental representations of SU(3).

In terms of these matrices, the eight generators T_a can be taken as

$$T_a \equiv \frac{1}{2}\lambda_a \Rightarrow \text{Tr}(T_a T_b) = \frac{1}{2}\delta_{ab}, \tag{2.69}$$

where the second relation fixes their normalization. In the case of SU(3) group there are two commuting generators, $H_1 \equiv T_3, H_2 \equiv T_8$ with the following column vectors

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} ; \; e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} ; \; e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{2.70}$$

as their common eigenstates. The eigenvalues (h_1, h_2) of H_1, H_2, corresponding to these states are equal to $(\frac{1}{2}, \frac{1}{2\sqrt{3}})$, $(-\frac{1}{2}, \frac{1}{2\sqrt{3}})$,$(0, -\frac{1}{\sqrt{3}})$. It is straightforward to show that

$$\left[H_1, \frac{1}{\sqrt{2}}(T_1 \pm iT_2) \right] = \pm 1 \left[\frac{1}{\sqrt{2}}(T_1 \pm iT_2) \right], \quad \left[H_2, \frac{1}{\sqrt{2}}(T_1 \pm iT_2) \right] = 0 \tag{2.71}$$

$$\left[H_1, \frac{1}{\sqrt{2}}(T_4 \pm iT_5) \right] = \pm\frac{1}{2} \left[\frac{1}{\sqrt{2}}(T_4 \pm iT_5) \right], \quad \left[H_2, \frac{1}{\sqrt{2}}(T_4 \pm iT_5) \right] = \pm\frac{\sqrt{3}}{2} \left[\frac{1}{\sqrt{2}}(T_4 \pm iT_5) \right], \tag{2.72}$$

$$\left[H_1, \frac{1}{\sqrt{2}}(T_6 \pm iT_7) \right] = \mp\frac{1}{2} \left[\frac{1}{\sqrt{2}}(T_6 \pm iT_7) \right], \quad \left[H_2, \frac{1}{\sqrt{2}}(T_6 \pm iT_7) \right] = \pm\frac{\sqrt{3}}{2} \left[\frac{1}{\sqrt{2}}(T_6 \pm iT_7) \right], \tag{2.73}$$

which implies that the operators E_α are given as

$$E_{\pm 1,0} = \frac{1}{\sqrt{2}}(T_1 \pm iT_2); \;\; E_{\pm 1/2,\pm\sqrt{3}/2} = \frac{1}{\sqrt{2}}(T_4 \pm iT_5); \;\; E_{\mp 1/2,\pm\sqrt{3}/2} = \frac{1}{\sqrt{2}}(T_6 \pm iT_7). \tag{2.74}$$

The basic triplet, represented by their weights, together with the roots is displayed in Fig. 2.1. According to (2.66) we find $mm' = 1$, i.e. the angles between roots of SU(3) algebra are multiples of 60° degrees.

The isospin group SU(2) is a subgroup of SU(3) and the transition operators $E_{\pm 1,0}$ can be identified with the rising and lowering operators $J^\pm$ of the previous subsection. Similarly the other two pairs of transition operators $E_{\pm 1/2,\pm\sqrt{3}/2}$ and $E_{\mp 1/2,\pm\sqrt{3}/2}$ correspond to the other two SU(2) subgroups of SU(3), usually called U-spin and V-spin respectively. The isospin subgroup is singled out in the particular choice (2.68) of the SU(3) generators by selecting as one of the diagonal matrices the matrix λ_3, corresponding to the projection of the isospin.

2.6 Simple roots of simple Lie algebras

It is clear that one doesn't need all the transition operators E_α in order to sweep through a whole given multiplet. Recall that in the case of SU(2) we had two such operators, $J^\pm$, but actually only one of them

was really necessary, the other one working essentially in opposite direction. The same occurs for general compact Lie groups. First we define some **ordering** of operators, which will allow us to tell what "lowering" and "rising" means and thus distinguish between "positive" and "negative" roots. As the eigenvalues of operators from Cartan subalgebra form m–dimensional vectors $\mu = (\mu_1, \mu_2, \cdots, \mu_m)$ we define

Definition 2.2 *The m-dimensional vector $\mu = (\mu_1, \mu_2, \cdots, \mu_m)$ is called **positive** if its first nonzero element is positive. Similarly for negative vectors.*

Corollary 2.1 *The ordering of weights and roots is then defined via the relation*

$$\mu^1 > \mu^2 \Leftrightarrow \mu^1 - \mu^2 > 0,$$

where the superscript labels different vectors.

It is obvious that for any **finite dimensional** representation **D** (recall that for compact Lie groups all **irreducible** representations *are* finite dimensional) there *exists* a weight which may be called **highest** in the above defined sense. Similarly, all the roots can be divided into positive and negative ones the former called "rising" and the latter "lowering". An important concept is that of a **simple** root:

Definition 2.3 *A positive root which cannot be expressed as a sum of two other positive roots is called a **simple** root.*

Corollary 2.2 *For any pair α, β of simple roots the difference $\alpha - \beta$ **is not** a root. Consequently*

$$E_{-\alpha} \mid E_\beta \rangle = \mid \underbrace{[E_{-\alpha}, E_\beta]}_{0} \rangle = 0 \Rightarrow q = 0 \;\; \Rightarrow \;\; \cos \vartheta = -\frac{\sqrt{pp'}}{2} < 0 \qquad (2.75)$$

implying $\vartheta \in (\pi/2, \pi)$. For SU(3), which has three positive (and three negative) roots, two, $\alpha^1 = (1/2, \sqrt{3}/2)$ and $\alpha^2 = (1/2, -\sqrt{3}/2)$, are simple, spanning the angle 120°.

Simple roots have several properties which will be important in further considerations.

Proposition 2.5 *Simple roots are linearly independent.*

Proposition 2.6 *Each positive root Φ can be written as a sum*

$$\Phi = \sum_\alpha k_\alpha \alpha$$

*of simple roots α with **nonnegative** coefficients k_α.*

Proposition 2.7 *The number k of simple roots of a simple Lie algebra is equal to its rank m.*

Proofs: The first two of the above claims are obvious and their proofs therefore left to the reader. The proof of the third one goes as follows. Clearly k cannot be bigger than m as roots are m-dimensional vectors. Assume that $k < m$. In a suitable basis all simple roots will then have the first component equal to zero. This, however, means that the first component of every root Φ vanishes and we have

$$[H_1, E_\Phi] = \Phi_1 E_\Phi = 0.$$

Consequently the generator H_1 commutes with *all* elements of the algebra and thus forms an invariant subalgebra by itself. This, however, is impossible in a simple algebra.

The general classification of compact Lie groups is based on systematic exploitation of the formula (2.75) which is conveniently expressed in the form of **Dynkin** diagrams. These diagrams describe the number and mutual orientation of all simple roots and are discussed in detail in [6].

Fundamental weights and fundamental representations: concepts which generalize the basic doublet representation $\mathbf{D}^{(1/2)}$ of SU(2) and its highest weight 1/2. Consider the highest weight μ of a representation **D** and form (2.66) for all **simple** roots:

$$\frac{2\alpha^i \mu}{(\alpha^i)^2} = q^i - \underbrace{p^i}_{0} = q^i \quad i = 1, \cdots, m, \qquad (2.76)$$

where the set $q^i, i = 1, \cdots, m$ of nonnegative integers *fully* characterizes the highest weight of a representation **D** and thereby also the whole representation. We now define a special class of highest weights μ^j by the condition

$$\frac{2\alpha^i \mu^j}{(\alpha^i)^2} = \delta_{ij}; \quad i, j = 1, \cdots, m \tag{2.77}$$

A general highest weight μ can be expressed in terms of the weights μ^j as follows:

$$\mu = \sum_{i=1}^m q^i \mu^i; \quad \mu^i = (\mu_1^i, \mu_2^i, \cdots, \mu_m^i). \tag{2.78}$$

We then define

Definition 2.4 *The highest weights μ^j are called the* fundamental weights *and the corresponding multiplets* $\mathbf{D}^{(i)}$ fundamental representations.

In the following the multiplets will be denoted in either of the three equivalent ways:

- By means of its highest weight μ as $\mathbf{D}^{(\mu)}$,

- using the vector $q \equiv (q_1, q_2, \cdots, q_m)$,

- or by its dimensionality, like **8**.

The sum (2.78) corresponds to the fact that any irreducible representation **D**, with highest weight μ, can be obtained as a multiple direct product of fundamental representations $\mathbf{D}^{(i)}$ of the form

$$\mathbf{D} = \underbrace{\mathbf{D}^{(1)} \otimes \cdots \otimes \mathbf{D}^{(1)}}_{q^1 \ times} \otimes \underbrace{\mathbf{D}^{(2)} \otimes \cdots \otimes \mathbf{D}^{(2)}}_{q^2 \ times} \cdots \underbrace{\mathbf{D}^{(m)} \otimes \cdots \otimes \mathbf{D}^{(m)}}_{q^m \ times}. \tag{2.79}$$

Example: in the case of SU(3) we have *two* fundamental weights, corresponding to the fact that there are two simple roots α^1, α^2:

$$\begin{array}{llll}
\alpha^1 = (\frac{1}{2}, \frac{\sqrt{3}}{2}) & \Rightarrow & \mu^1 = (\frac{1}{2}, \frac{1}{2\sqrt{3}}), \\
\alpha^2 = (\frac{1}{2}, -\frac{\sqrt{3}}{2}) & \Rightarrow & \mu^2 = (\frac{1}{2}, -\frac{1}{2\sqrt{3}}).
\end{array} \tag{2.80}$$

The first of them corresponds to the basic (defining) representation as given in (2.69), which acts on the Hilbert space spanned by the triplet of states (2.70). The second one corresponds to another fundamental representation, which from the point of view of group theory is equally "fundamental" as the first one. Having its application to quark model in mind, let me call the first multiplet **quark** triplet and denote it as $\mathbf{3} \equiv (1,0)$. It will turn out that the second triplet $(0,1)$ can be identified with **antiquarks** and I shall therefore call it **antitriplet** and denote $\overline{\mathbf{3}} \equiv (0,1)$. Both of these triplets are displayed in Fig. 2.2b. It is illustrative to carry out the explicit construction of this second fundamental representation of SU(3). Starting from the state of highest weight, i.e the point with coordinates $(h_1, h_2) = \mu^2 = (1/2, -1/2\sqrt{3})$ and taking into account that $q^1 = 0$ for this weight, only the application of $E_{-\alpha^2}$ leads to nonvanishing result, i.e. the point $(0, 1/\sqrt{3})$. Going on, further application of $E_{-\alpha^2}$ leads to zero as $q^2 = 1$ and we are thus forced to apply $E_{-\alpha^1}$ which brings us to the leftmost point $(-1/2, -1/2\sqrt{3})$, where the procedure finally stops. One could attempt to apply E_{α^1} in the second stage instead of $E_{-\alpha^1}$, but this leads to zero due the fact that

$$E_{\alpha^1} E_{-\alpha^2} \mid \mu^2 \rangle = E_{-\alpha^2} \underbrace{E_{\alpha^1} \mid \mu^2 \rangle}_{0} + \underbrace{[E_{\alpha^1}, E_{-\alpha^2}]}_{0} \mid \mu^2 \rangle = 0 \tag{2.81}$$

where the first zero is obvious and the second one uses the basic property of simple roots as given in **Corollary 2.2**.

This example suggests a general strategy for constructing any multiplet from its highest weight μ:

- Apply to the state $\mid \mu, \mathbf{D} \rangle$ all possible combinations of lowering operators corresponding to **simple** roots α^i:

$$\prod_{i=1}^k E_{-\alpha^i} \mid \mu, \mathbf{D} \rangle.$$

One might object that all roots should be used for this purpose but actually the simple ones are sufficient.

a) $\mu^r = \mu - \dfrac{2(\alpha\mu)\alpha}{\alpha^2}$ b)   ⊕ triplet (1,0)
                                                          ✢ antitriplet (0,1)

reflection along the root α
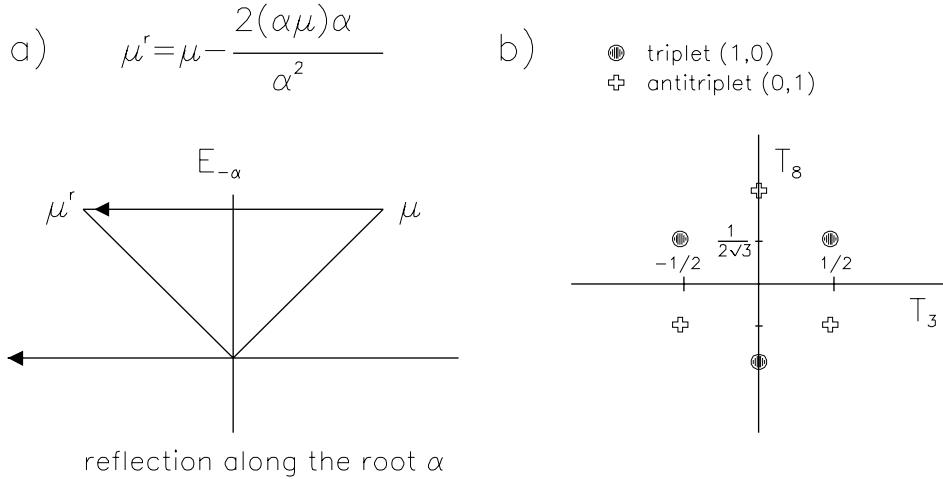
Figure 2.2: a) Weyl symmetry of weights, and b) the fundamental representations of SU(3).

- At each step check whether the suggested application of $E_{-\alpha^i}$ on the reached state $\mid \nu \rangle$ is "legal", i.e. whether the roots α^i and the weight ν satisfy the condition

$$\frac{2\alpha^i \nu}{(\alpha^i)^2} = q^i - p^i.$$

- For weights, which can be reached via different paths one has to determine whether they correspond to different states or not.

To solve this last problem the following lemma can be useful

Proposition 2.8 *Let $\mid A\rangle, \mid B\rangle \in \mathcal{H}$ be two states from a Hilbert space $\mathcal{H}$. They are linearly dependent* **iff**

$$\frac{\langle A \mid B\rangle\langle B \mid A\rangle}{\langle A \mid A\rangle\langle B \mid B\rangle} = 1. \tag{2.82}$$

Weyl group of symmetries of multiplets: another useful tool for the construction of multiplets is the so called *Weyl group* of symmetries. This group exploits the existence in a compact Lie algebra of the SU(2) subalgebras spanned by the generators

$$\frac{\alpha_i H_i}{\alpha^2}, \quad \frac{E_{\pm\alpha}}{\sqrt{\alpha^2}}, \tag{2.83}$$

associated with any root α (not necessarily simple). Taking one of these subgroups and applying powers of $E_{\pm\alpha}$ successively to any weight μ leads in final effect to the "reflected" state, characterized by the weight

$$\mu^r = \mu - \frac{2(\alpha\mu)}{\sqrt{\alpha^2}} \frac{\alpha}{\sqrt{\alpha^2}}, \tag{2.84}$$

which reduces to trivial symmetry operation $-k = k - 2k$ for SU(2) group. The reflection operation (2.84) is illustrated in Fig. 2.2a. The existence of the second fundamental representation of SU(3) $\overline{\mathbf{3}}$ and the relation of its weights to those of the defining triplet $\mathbf{3}$

$$
\begin{array}{rcl}
\mu^2 & = & -(\mu^1 - \alpha^1 - \alpha^2) \\
(\mu^2 - \alpha^2) & = & -(\mu^1 - \alpha^1) \\
\mu^2 - \alpha^2 - \alpha^1 & = & -\mu^1
\end{array}
\tag{2.85}
$$

is an example of a more general relation, concerning the so called **complex representations**.

31
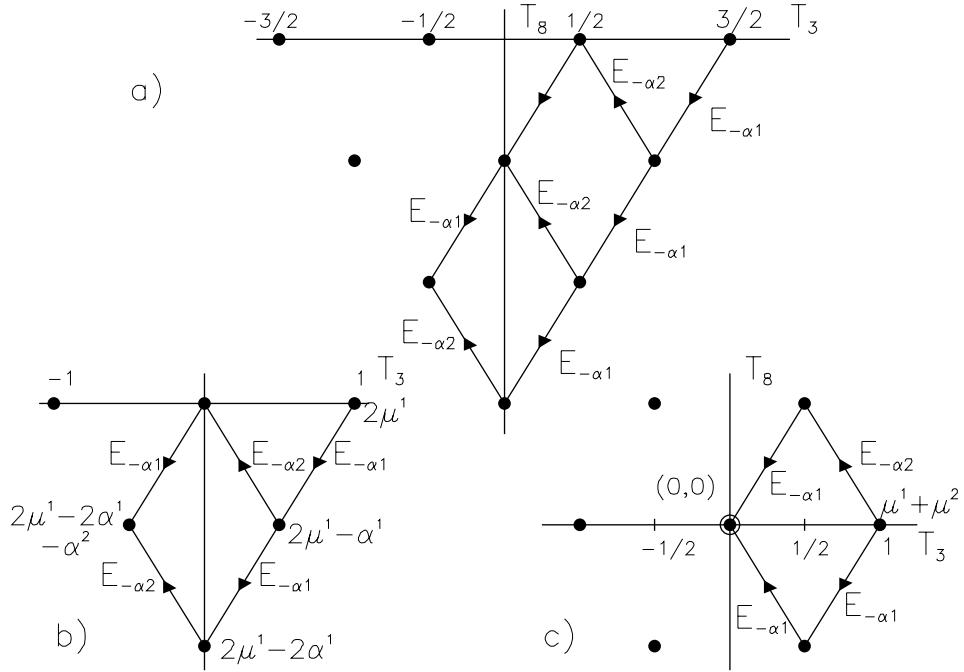
Figure 2.3: Further SU(3) multiplets; a) decuplet; b) sextet; c) octet.

Complex representations: if the matrices T_a of a representation $\mathbf{D}$ satisfy (2.9) so do also the matrices

$$- (T_a)^*,$$

which form the so called *complex conjugate* representation $\overline{\mathbf{D}}$. Clearly this relation is *reflexive* as we have

$$- \left(- (T_a)^*\right)^* = T_a.$$

Definition 2.5 *The representation of a Lie algebra given by matrices T_a is said to be* **real** *if it is equivalent to the complex conjugate representation of matrices $-T_a^*$.*

Corollary 2.3 *As the elements of Cartan subalgebra are hermitian, their eigenvalues are the same for* $\mathbf{D}$ *and* $\overline{\mathbf{D}}$ *and the weights related simply as* $\mu^{\mathbf{D}} = -\mu^{\overline{\mathbf{D}}}$.

Examples:

- SU(2): there is just one fundamental representation, which is *real*, as the Pauli matrices σ^i are equivalent to the negative of their complex conjugates. Nevertheless the relation between the fundamental doublet and its (equivalent) complex conjugate doublet is not without some subtlety if we want to identify definite antiquarks with its elements (Exercise 2.4).

- SU(3): The two fundamental representations are **not** equivalent, the fact that has important consequences in the quark model.

2.7   Examples of further important multiplets of SU(3)

There are several other multiplets of SU(3) which had played important role in the formulation and development of unitary symmetry and the quark model. I will discuss three of them in some detail and briefly mention several other. In all cases the construction of the multiplets is greatly facilitated by taking into account the Weyl reflection symmetry.

2.7.1 Sextet

The sextet $\mathbf{6} \equiv (2,0)$, with the highest weight $2\mu^1 = (1, 1/\sqrt{3})$, is the next simplest multiplet. All of its 6 states depicted in Fig. 2.3b are *unique*. This fact is not trivial, as there are two states, namely those with weights $(-1/2, -1/2\sqrt{3})$ and $(-1, 1/\sqrt{3})$, which can be reached via two different paths, as indicated in the figure. It is, however, easy to show, using the commutation relations between the lowering operators $E_{-\alpha^1}$ and $E_{-\alpha^2}$ that

$$E_{-\alpha^2} E_{-\alpha^1} E_{-\alpha^1} \mid 2\mu^1 \rangle = 2 E_{-\alpha^1} E_{-\alpha^2} E_{-\alpha^1} \mid 2\mu^1 \rangle, \tag{2.86}$$

i.e. that the resulting states *are the same*. The same conclusion can be reached by means of (2.82). As the highest weight of this sextet is *symmetric* under the permutation of the two u-quarks which make it up, the whole multiplet consists of symmetrical combinations of the u, d, s quarks. There are 6 such symmetric combinations, the remaining three being *antisymmetric*. It is illustrative to follow the action of the lowering operators on such quark combinations as it allows us to obtain the same result in a physically transparent way:

$$
\begin{array}{lll}
\mid uu \rangle & & \\
\downarrow E_{-\alpha^1} & & \\
N_{-\alpha^1,\mu^1} \mid us + su \rangle & \xrightarrow{E_{-\alpha^2}} & N_{-\alpha^1,\mu^1} N_{-\alpha^2,\mu^1-\alpha^1} \mid ud + du \rangle \\
E_{-\alpha^1} & & \downarrow E_{-\alpha^1} \\
\downarrow & & (N_{-\alpha^1,\mu^1})^2 N_{-\alpha^2,\mu^1-\alpha^1} \mid sd + ds \rangle \\
2(N_{-\alpha^1,\mu^1})^2 \mid ss \rangle & \xrightarrow{E_{-\alpha^2}} & 2(N_{-\alpha^1,\mu^1})^2 N_{-\alpha^2,\mu^1-\alpha^1} \mid sd + ds \rangle.
\end{array} \tag{2.87}
$$

The same can be done for the three antisymmetric combinations. Starting with $\mid us - su \rangle$ we find that

$$E_{-\alpha^1} E_{-\alpha^2} \mid us - su \rangle = E_{-\alpha^1} N_{-\alpha^2,\mu^1-\alpha^1} \mid ud - du \rangle = N_{-\alpha^2,\mu^1-\alpha^1} N_{-\alpha^1,\mu^1} \mid sd - ds \rangle \tag{2.88}$$

with further application of either of E_α yielding zero. The above explicit construction shows that

$$\mathbf{3} \otimes \mathbf{3} = \mathbf{6} \oplus \overline{\mathbf{3}}; \quad \overline{\mathbf{3}} \otimes \overline{\mathbf{3}} = \overline{\mathbf{6}} \oplus \mathbf{3}. \tag{2.89}$$

Looking at the weights corresponding to these combinations we see that they coincide with those of the antitriplet of antiquarks. This brings us to important conclusion that the *antisymmetric* diquark combinations behave, from the point of view of SU(3) transformations (but not as far as the flavour quantum numbers are concern), as *antiquarks*.

2.7.2 Octet

The octet, denoted as $\mathbf{8} \equiv (1,1)$ and displayed in Fig. 2.3c, had played the very central role in the discovery of SU(3) symmetry of hadrons and the subsequent formulation of the quark model. Also in this case there are states which can be reached from this μ in more then one way. Out of them only the one in the center of the octet is *not unique*. One can proceed exactly as in the previous case to show that the states with the weight $(0,0)$ arrived at by two different paths from that with highest weight $\mu = \mu^1 + \mu^2$ are different. The same can be proven also by means of the lemma (2.82) by showing that for the states

$$\mid A \rangle \equiv E_{-\alpha^1} E_{-\alpha^2} \mid \mu \rangle; \quad \mid B \rangle \equiv E_{-\alpha^2} E_{-\alpha^1} \mid \mu \rangle$$

one finds that

$$\frac{\langle A \mid A \rangle \langle B \mid B \rangle}{\langle A \mid B \rangle \langle B \mid A \rangle} = 4 \quad \Rightarrow \mid A \rangle \neq \mid B \rangle. \tag{2.90}$$

In the following we shall frequently need the fact that octet $\mathbf{8} = (1,1)$ appears in the decomposition of the direct product of the fundamental triplet and antitriplet as well as of three triplets

$$\mathbf{3} \otimes \overline{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1}, \tag{2.91}$$

$$\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{10}_\mathrm{s} \oplus \mathbf{8}_\mathrm{ms} \oplus \mathbf{8}_\mathrm{ma} \oplus \mathbf{1}_\mathrm{a}, \tag{2.92}$$

where the subscripts "ms" (mixed symmetric) and "ma" (mixed antisymmetric) denote similarly as in the case of SU(2) in (2.34), two equivalent octet representations, which differ in their symmetry properties under permutation of the three triplets. Analogously for the product of antitriplets $\overline{\mathbf{3}}$.

33

2.7.3 Decuplet

The last of multiplets which played important role in the formulation of the quark model is the **decuplet** $\mathbf{10} \equiv (3,0)$ with the highest weight $\mu = 3\mu^1$, shown in Fig. 2.3a, which corresponds to fully symmetric combinations of three quarks in the decomposition (2.92). One can again ask about the uniqueness of its states and the answer is the same as for the sextet: *all are unique*.

2.7.4 Other products and multiplets

Other multiplets that had played role in the discovery of unitary symmetry of hadrons and formulation of the quark model are those obtained in the reduction of the following direct products of two multiplets

$$\mathbf{3} \otimes \mathbf{6} = \mathbf{10} \oplus \mathbf{8}, \quad \mathbf{3} \otimes \mathbf{8} = \mathbf{15} \oplus \overline{\mathbf{6}} \oplus \mathbf{3}, \quad \overline{\mathbf{3}} \otimes \mathbf{6} = \mathbf{15} \oplus \mathbf{3}, \tag{2.93}$$

and three triplets and/or antitriplets

$$\mathbf{3} \otimes \mathbf{3} \otimes \overline{\mathbf{3}} = \mathbf{3} \otimes (\mathbf{8} \oplus \mathbf{1}) = (\mathbf{3} \otimes \mathbf{8}) \oplus \mathbf{3} = \mathbf{15} \oplus \overline{\mathbf{6}} \oplus \mathbf{3} \oplus \mathbf{3} \tag{2.94}$$
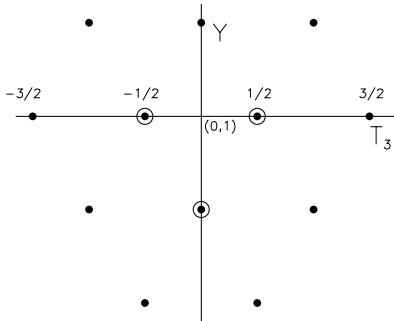
Figure 2.4: The 15-plet. Three inner three weights correspond to two different states.

The above, as well as any other, direct products of the SU(3) multiplets can be reduced using the method of the highest weight. The resulting decompositions may be quite complicated, but there is a method based on the so called **Young tableaux**, which allows fast determination of the results. Young tableau is a diagram associated with a given multiplet $\mathbf{D} = (p,q)$ and describing its symmetry properties under the permutations of the fundamental SU(3) triplets and antitriplets. There is, however, no place to describe here this method in more detail and we refer the interested reader to [6, 24] for comprehensive discussion. The 15-plet, shown in Figure 2.4 and resulting from the decomposition (2.94) of the direct product of two triplets and one antitriplet, was used in the Sakata model, a predecessor of the so called *Eightfold way* (see the next Section for discussion of both schemes). Part of this multiplet has weights that coincide with those of the decuplet. The inner three weights correspond to two states.

One of the most important characteristics of any SU(3) multiplet $\mathbf{D} = (p,q)$ is its dimension $D(p,q)$. This can also be read off the corresponding Young tableau with the result

$$D(p,q) = \frac{1}{2}(p+1)(q+1)(p+q+2). \tag{2.95}$$

2.8 Exercises

1. Evaluate Clebsh-Gordan coefficients of the direct product of SU(2) multiplets $\mathbf{1} \otimes \mathbf{1/2}$.

2. Prove (2.60).

3. Find the matrices λ_3, which correspond to the third projection of the U and V-spins. How do the matrices α_8 look like in these two cases?

4. Argue why only the simple roots are needed to reach from the state with the highest weight all the states of a given multiplet.

5. Prove the lemma (2.82).

6. Show that the fundamental representation of SU(2) is real.

7. Prove (2.90).

8. Show that if we associate the basic doublet of SU(2) with the pair (u, d), the pair of antiquarks which transforms according to identical representation is $(\overline{d}, -\overline{u})$ rather than $(\overline{d}, \overline{u})$.

9. Prove that the elements of the Cartan subalgebra can be chosen as hermitian operators and the roots as real vectors.

10. Show that the three operators defined in (2.83) do, indeed, satisfy the SU(2) commutation relations (2.22).

11. Prove the uniqueness of all states in the sextet $\mathbf{6}$=(2,0) representation of SU(3).

12. Show that the two states in the center of the octet of SU(3) are different.

13. Derive (2.92).

Chapter 3

The road to quarks

In this Chapter the milestones in the development of the current picture of the structure of matter ans forces acting in the microworld are briefly recollected, starting shortly after the discrovery of the atomic nucleus and ending almost exactly 50 years later with the formulation of the Quark Model. There are many interesting, even fascinating circumstances of these developments that are little known and which merit recollection because they throw light on the nontrivial way really fundamental breakthroughs in the quest for understanding the structure of matter have been achieved. Moreover, it is fair to do justice, in this and the following Chapter, to a few theorists, notably Yuval Ne'eman, George Zweig and Yoichiro Nambu, who had made fundamental contributions to the formulation of the quark model, but who for various reasons are often not properly appreciated. For more historical details about the development of the Standard Model in general and the quark model in particular several excellent sources [19, 25–28, 31] can be recommended. Systematic and meticulously documented exposition of the historical development of modern quantum theory is given in the extraordinary book *Inward bound* by Abraham Pais [18], wherefrom many of the quotations found in this text were taken. As most of the original papers are difficult to access I will instead of them cite his book with reference to the page where the appropriate quotation can be found, accompanied with the citation of the original source. Genesis of the theory of strong interactions is also a subject of two essays by Jagdish Mehra, renowned historian of physics, collected in the recent book [40], which gives perhaps the best non-technical description of this subject.

One aspect of the development of the quark model worth emphasizing is the interplay between experiment and theory. Whereas presently experimental input is eagerly awaited to guide the theory towards the physics beyond the Standard Model, in most of the last century it was the other way around and the main preoccupation of theorists was the quest for explanation of new and often puzzling phenomena. I will start by recalling how the very concept of strong interaction (called nuclear force then) did emerge during the two decades after the discovery of the atomic nucleus.

3.1 What are the nuclei made of, what forces hold them together?

Early in 1914, three years after the discovery of atomic nucleus and at about the same time Chadwick proved the continuous character of the nuclear β-decay, it became generally accepted that the nucleus is the seat of all radioactive processes and all nuclei are specified by the numbers A and Z. Moreover, all data supported the conjecture that the interaction between the α-particle and nuclei is purely electromagnetic.

3.1.1 Electrons within nuclei

In attempts to understand the mechanism behind the radioactive processes within the framework of "old" Quantum Mechanics, i.e. essentially the Bohr Model of atom, the first model of the nucleus, suggested in 1913 by van den Broek, assumed that both α-particles and electrons are constituents of nuclei. His suggestion was taken up by Rutherford himself, who considered the α-particle to be composed of two *"four positive electrons (H-particles)* [1] *and two negative"* ([18], p. 230). In general a nucleus X with mass number A and

---

[1] The names used at that time for the proton.

charge Z was assumed to be composed of A H-particles and $A - Z$ electrons, symbolically

$$X = A\text{H} + (A - Z)\text{e}. \tag{3.1}$$

The situation is nicely summarized by Pais ([18], p. 231)

> *Thus Rutherford, though always cautious and averse to speculation, blithely assumed that electrons are nuclear constituents. Actually he would not conceive of this as assumption. Was it not self-evident? Did one not see electrons come out of certain nuclei, in β-processes? To Rutherford, as to all physicists at that time, it was equally sensible to speak of electrons as building blocks of nuclei as it was to speak of a house built of bricks or necklace made of pearls.*
>
> *In actual fact, the H-particle-electron picture of the nucleus is another example of simplicity as a necessary evil. It was a model as inevitable as it was wrong. It is not true that electrons are building blocks of nuclei .... Then how can one understand that electrons do emerge from nuclei in β-decay? Almost exactly 20 years after Rutherford spoke in the Royal Society, Fermi found the answer to that question, using the tools of quantum field theory.*

As for the forces binding H-particles and electrons inside the nuclei, Rutherford was in 1914 quite explicit (see [18], p. 231):

> *"The nucleus, though of minute dimension, is in itself a very complex system consisting of positively and negatively charged bodies bound closely together by intense electrical forces."*

Again, we should not be surprised by his claim as no other forces, beside the gravitational and electromagnetic, were known at that time. The fact that electron cannot be localized at distances smaller than its Compton wavelength $\lambda_e \doteq 400$ fermi, was not yet realized. This model of atomic nuclei has been generally accepted for two decades, though there were signals it had serious problems. One of them concerned the quantitative understanding of nuclear binding energies, as described in [18], pp. 232-234. What is, however, really remarkable is that already in 1919 Rutherford himself observed the first clear evidence of the new force actually responsible for nuclear binding.

3.1.2 First glimpse of nuclear force

This observation resulted from a series of experiments performed by Rutherford in 1916-1919 in his Manchester laboratory in which he had scattered α-particles from various nuclei, heavy as well as light, including the hydrogen. The latter experiments required different experimental arrangement as well as different theoretical framework than those on heavy nuclei because the target nuclei recoil could not be neglected. As in such collisions most of the energy of incoming α-particle is converted into the recoil energy of the target hydrogen, these collisions are less effective for probing the structure of target particles. On the other hand as the electric charge of the proton is much smaller than that of the gold nuclei, the α-particle can get much closer to its center and the large angle scattering therefore probes much smaller distances. Working out the kinematics we find that the minimal distances (corresponding to head-on collision) accessible in α-H and α-Au scattering are, for the same primary energy, related as follows

$$r_{\text{min}}^{\alpha\text{p}} = r_{\text{min}}^{\alpha\text{Au}} \frac{1}{Z_{\text{Au}}} \left(\frac{m_\alpha + m_{\text{p}}}{m_{\text{p}}} \right) = \frac{2\alpha}{E} \left(\frac{m_\alpha + m_p}{m_p} \right). \tag{3.2}$$

Experimentally, the α-particles scatter predominantly at very small angles and the scattering with large momentum transfer is much easier to investigate by measuring the energy of recoiling protons in a hydrogen vessel. As the scattering of incoming α-particle corresponding to (almost) head-on collisions results in fast target protons recoiling in very forward direction, Rutherford measured the number $N(\vartheta)$ of protons recoiling at angles from zero to ϑ_p and compared it with the prediction of Darwin

$$N_p(\vartheta) = C \frac{\alpha^2}{E_{\text{kin}}^2} \tan^2 \vartheta_p, \tag{3.3}$$

where C is a constant depending on the composition and geometry of the target. This formula diverges for $\vartheta_p = \pi/2$, which corresponds to the soft collision when the incoming α-particle bounces off the target proton at very large impact parameters. Rutherford summed up his observations as follows ([18], p. 238)

For α-particles of range less than 4 cm of air, the distribution and absorption of H-atoms are in fair agreement with theory. .... For α-particles with range 7 cm (corresponding to the full 5 MeV energy, my note) the number of fast H-atoms produced is 30 times greater than the theoretical number.

The range of α-particles in air was in those times the convenient measure of their energy, which Rutherford varied by placing the gold and aluminum foil of known stopping power in front of the hydrogen target. The energy of α-particles at which the results started to deviate from theoretical predictions, based on the assumption that the force acting between the α-particles and proton is described by the Coulomb potential, was estimated by Rutherford to be about 5 MeV, which corresponds to the minimal distance $r_{\min}^{\alpha H} \doteq 3.5$ fm. This led Rutherford to conclude ([18], p. 239)

We have every reason to believe that the α-particle has a complex structure consisting probably of four hydrogen nuclei and two negative electrons. ... it appears significant that $r_{\min}$ is about the same as the accepted value of the diameter of negative electron.... It is of course possible to suppose that the actual law of force .. does not follow the inverse square for very small distances... however, it seems simpler to suppose that the rapid alteration of the force close to the nucleus is due rather to the deformation of its structure and of its constituent parts.

The reference to the classical radius of electron $r_{\mathrm{cl}} \equiv \alpha/m_e \doteq 2.8$ fm was natural as the idea that electron was a sphere of that radius was still widely accepted despite its obvious problems. As we know now Rutherford had observed first clear manifestation of the new type of force, unrelated to the electromagnetism, but failed to break with the theoretical framework prevalent at that time to make the right conclusion. Rutherford's observed disagreement between theory and data on the scattering on light nuclei were soon confirmed by Chadwick and Beiler using an improved technique. These authors were probably the first ones to contemplate the need for completely new kind of force, unrelated to the electromagnetism. Their observation ([18], p. 239)

No system of four hydrogen nuclei and two electrons united by inverse square law force could give a field of force of such intensity over so long extent. We must conclude that either the α-particle is not made up of four H-nuclei and two electrons or the law of force is not the inverse square in the immediate neighborhood of an electric charge. It is simpler to choose the latter alternative particularly as other experimental and theoretical considerations point to this direction..... The present experiments do not seem to throw any light on the nature of the forces at the seat of an electric charge, but merely show that the forces are of great intensity.. It is our task to find some field of force which will reproduce these effects.

is considered by Pais to mark the birth of strong interactions [2]. I share this assessment.

3.1.3 Birth of the "new" quantum mechanics

In the middle of the next decade the "new" quantum mechanics was formulated by de Broglie, Heisenberg, Schrödinger and others. I use the adjective "new" to distinguish this theory from the "old" quantum theory of atoms developed by Bohr and Sommerfeld, but it is important to keep in mind that the fundamental break with the basic conceptual framework of the classical mechanics had been done already in this "old" quantum theory. Soon afterwards, Dirac proposed his relativistically invariant equation for the electron thereby laying foundations for the quantum field theory.

It is also worth noting that still within the framework of the "old" quantum theory of Bohr and a couple of months before the papers of Heisenberg Pauli invented in all but name the concept of the spin of electron. Analysing the spectra of anomalous Zeeman effect he introduced a new degree of freedom of the electron, which he called "the two-valuedness, not describable classically".

3.1.4 Pauli and neutrino hypothesis

Although neutrinos do not participate in strong interactions, they play such a vital role in the nature that a brief recollection of the circumstances in which it had been invented is called for. The story starts in

---

[2]I will henceforth use the term "strong interaction" instead of "nuclear force"

1914 when, after 17 years of experimental investigation, Chadwick showed convincingly that the spectrum of electrons or positrons in the nuclear β-decay is continuous. This experimental fact, which seemed to violate the sacred principle of energy and monentum conservation, lead Heisenberg and Bohr, among others, to question its validity in the individual processes in the microworld. For Pauli, however, abandoning energy and monentum conservation was so outrageous, that he proposed, as alternative explanation, the existence of the so far unobserved particle, which should accompany electrons or positrons in the mentioned β-decays. In the letter of December 4, 1930 to the participant of the conference on radioactivity in Tübingen he wrote

> *Dear Radioactive Ladies and Gentlemen,*
>
> *As the bearer of these lines, to whom I graciously ask you to listen, will explain to you in more detail, how because of the "wrong" statistics of the Nitrogen and Lithium 6 nuclei and the continuous beta spectrum, I have hit upon a desperate remedy to save the "exchange theorem" of statistics and the law of conservation of energy. Namely, the possibility that there could exist in the nuclei electrically neutral particles, that I wish to call neutrons, which have spin 1/2 and obey the exclusion principle and which further differ from light quanta in that they do not travel with the velocity of light. The mass of the neutrons should be of the same order of magnitude as the electron mass and in any event not larger than 0.01 proton masses. The continuous beta spectrum would then become understandable by the assumption that in beta decay a neutron is emitted in addition to the electron such that the sum of the energies of the neutron and the electron is constant...*
>
> *I agree that my remedy could seem incredible because one should have seen these neutrons much earlier if they really exist. But only the one who dare can win and the difficult situation, due to the continuous structure of the beta spectrum, is lighted by a remark of my honoured predecessor, Mr Debye, who told me recently in Bruxelles: "Oh, It's well better not to think about this at all, like new taxes". From now on, every solution to the issue must be discussed. Thus, dear radioactive people, look and judge.*
>
> *Unfortunately, I cannot appear in Tubingen personally since I am indispensable here in Zurich because of a ball on the night of 6/7 December. With my best regards to you and also Mr. Back.*
>
> *Your humble servant,*
> *W. Pauli*

Note that, interestingly, Pauli was more wrong than right in his description of the role of his "neutron". First, the particle emitted in nuclear β-decay plays no role in explaining the statistics of the Nitrogen and Lithim 6 nuclei. The particle that provides such explanation, and which is now called "neutron", was discovered by James Chadwick two years after Pauli's proposal. In 1934, at a seminar on his recent theory of beta-decay, Fermi was asked whether the neutral particle emitted in the nuclear beta-decay was the same as Chadwick's neutron. Apparently, Fermi clarified that he was talking about a different particle which he referred to as neutrino ("little neutral one"). Second, Pauli's particle, i.e. electron antineutrino in current parlance, does not "exist in the nuclei" in any sense but is created in the decay. Finally, also his prediction that its mass "should be of the same order of magnitude as the electron mass" turned out to be wrong, at least by a factor of 10^5! Nevertheless, Pauli was right in the one, but crucial point, namely that his particle has spin 1/2 and is emitted together with electron in the β-decay.

Pauli thought his proposal of the "neutron" was too speculative, and did not publish it in a scientific journal until 1934, by which time Fermi had already developped a theory of beta decay incorporating the neutrino. In fact the first published mention of the netrino appeared on 17th June 1931 in the New York Times which carried the following news:

> *A new inhabitant at the heart of the atom was introduced to the world of physics today when Dr. W. Pauli of the Institute of Technology in Zurich, Switzerland, postulated the existence of particles or entities which he christened "neutrons".*

Pauli's hypothesis got a general acceptance during the Salvay coneference in October 1993, where Perrin put forward the conjecture that neutrino may be massless. However, it took two more decades before the existence of (electron) neutrino has been confirmed by Reines and Cowan who observed its collision with proton resulting in neutron and positron.

3.1.5 1932-34: three years that changed all

Despite the formulation of a new theoretical framework for the description of atoms, for about a decade there was no significant progress in the quest for understanding the structure of the nucleus and forces binding it together. Then, during the two years 1932-34, two fundamental experimental discoveries, accompanied by three theoretical breakthroughs had dramatically changed almost the whole picture of the nucleus. The two, almost simultaneous, but unrelated, experimental discoveries occurred in 1932 and concerned the observation of two new particles of fundamental importance: the positron and the neutron.
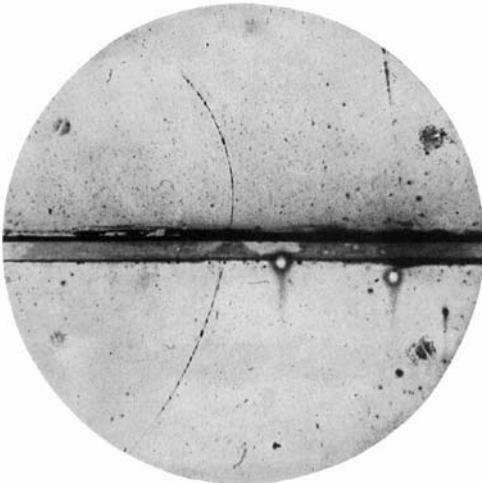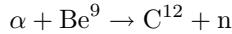

Fig. 1. A 63 million volt positron ($H\rho = 2.1 \times 10^6$ gauss-cm) passing through a 6 mm lead plate and emerging as a 23 million volt positron ($H\rho = 7.5 \times 10^4$ gauss-cm). The length of this latter path is at least ten times greater than the possible length of a proton path of this curvature.

Figure 3.1: The picture taken by Anderson in August 1932 showing the first evidence for positron. Taken from [43].

In August 1932 Carl Anderson had observed, using cloud chamber placed in a magnetic field, the first example of *"a positively charged particle comparable in mass and magnitude of charge with an electron"*. Anderson has apparently not known Dirac theory of electrons and had therefore not interpreted his observation as discovery of antielectron, but his observation was clearly the first evidence for what is now called positron. The crucial ingredient of Anderson's apparatus was the lead plate put across the cloud chamber which slowed down charged particles and in combination with the curvature allowed him to determine the direction of the flight and therefrom also the sign of the electric charge. His observation was rapidly confirmed using a novel mode of operation of cloud chambers, called "counter controlled cloud chamber", the precursor of modern counter detectors. These chambers, later improved by adding more lead plates, had become - together with nuclear emulsions - the main detection technique until the invention of bubble chambers two decades later. Soon afterwards, Anderson observed also the process of the creation of electron-positron pair from some neutral particles, presumably photons, as they traversed lead plate in his chamber. For the first time process in which particles changed identities had been observed confirming the basic ideas of the quantum field theory.

The discovery of neutron is an extraordinarily interesting story in its own right, but for space reasons I will mention just its most important aspects. In all probability it was Joliot-Curies who first observed it in an experiment described in their paper (see [18], p. 399) published on January 28, 1932 under the title *The emission of high energy photons from hydrogenous substances irradiated with penetrating alpha rays*, but they draw a wrong conclusion from their observation. Immediately after reading their paper, Chadwick repeated their experiment by bombarding the berylium target with α-particles. Analyzing the recoil velocities of charged particles produced in collisions of the neutral particle "n" emerging from the reaction

$$\alpha + \text{Be}^9 \rightarrow \text{C}^{12} + \text{n}$$

in the hydrogen or nitrogen medium, he came to the conclusion that the neutral particle "n" must have the mass close to that of the proton and cannot therefore be the energetic photon, as conjectured by Curies. The neutron was born. The circumstances surrounding its discovery provide another illustration of the point that drawing correct conclusion from an unexpected observation is as important as that observation itself. These two experimental discoveries provided the necessary input to theoretical considerations that had in the end lead to the formulation of theories of strong and weak interactions acting between the constituents of nuclei.

The first step towards the theory of nuclear forces was done shortly after the discovery of the neutron by Heisenberg in his theory of nuclear structure based on the assumption that nuclei are composed of protons and neutrons. The electrons, however, did not disappear from his model, but were assigned important role in the mechanism of forces acting between protons and neutrons. He built his theory on the analogy with molecular forces. He imagined that the electron provides a mechanism of new force between the proton and neutron by being "exchanged" between them, similarly as electron in the H_2^+ ion. The idea of exchange mechanism of fundamental forces is due to this Heisenberg assumption. On the other hand, the forces between two neutrons were assumed to be much smaller (so no charge invariance, yet!) than those between

proton and neutron, again in analogy with forces acting within the neutral H_2 molecule. Note, however, that Heisenberg force was not of electromagnetic origin, merely the mechanism was modeled on it. No new force beside the Coulomb electrostatic forces was postulated to act between two protons. The main importance of Heisenberg theory, beside the fact that it assumed the proton-neutron structure of nuclei, was that it was formulated within the nonrelativistic quantum mechanics. The shortcomings, namely the lack of charge invariance and the impossible task assigned to electrons to mediate so short distance force were soon to become obvious, but the first step was taken.

The main obstacle to further progress was in fact the very idea, which seemed so natural at that time, that the neutron is a sort of bound state of the proton and electron. In the words of Chadwick *"The neutron may be pictured as a small dipole, or perhaps better, as a proton embedded in an electron".* [3] This picture of the neutron was shared by most physicists, including Heisenberg, but not all. With the benefit of hindsight we know this picture could not be working as it implied that the neutron had integer spin, but at that time neutron spin was not yet known. D. Ivanenko was perhaps the first to suggest that it should be considered on the same footing as the proton (see [18], p. 411):

> *"We do not consider the neutron as built up of an electron and a proton, but as an elementary particle. Given this fact we are obliged to treat neutrons as possessing spin 1/2 and obeying Fermi-Dirac statistics."*

The step that had dramatically changed the basic theoretical framework for the description of nuclear phenomena was taken by Fermi at the end of 1933, when he put forward his new theory of the β-decay of nuclei and neutron. I will not discuss his theory of what we today call weak interactions, but mention it for two reasons: it was the first theory formulated within the language of relativistic quantum field theory and it was also the first successful theory of intranuclear force beside the electromagnetic one. The first point was particularly important as it banished forever the classical, but wrong idea that if something disintegrates, the decay products must have existed inside the decaying particle. In quantum field theory particles are created and destroyed as a consequence of very notion of quantum field. Applied to β-decay it meant that there is no reason to assume electrons are constituents of nuclei, which in turn dealt a heavy blow to Heisenberg theory of nuclear forces. The decisive step towards the formulation of strong interaction was made by Yukawa in late 1934 when he found the relation between the range of a force and the mass of conjectured mediating particle. It was known for more than two decades that this force acted at short distances only, of the order of a few fermi, but Heisenberg and other theorists still thought it could be mediated by exchange on electrons, which has a Compton wavelength 400 fm. In fact this step was crucial not only for the theory of strong interactions, but for the quantum field theory in general. In Yukawa's own words (quoted in [40], p. 1196) recalling his considerations one night in October 1934:

> *The nuclear force is effective at extremely short distances, of the order of 2 fm. That much I knew already. My new insight was the realization that this distance and the mass of the new particle that I was seeking are inversely related to each other. Why had I not noticed that before? The next morning I tackled the problem of the mass of the new particles and found it to be about two hundred times that of electron. It also had to have the charge of plus or minus that of electron.*

In the paper that soon followed Yukawa gave his idea mathematical form and wrote for the first time the formula for the potential corresponding (in static approximation) to the exchange of particle with the mass μ, known henceforth as Yukawa potential:

$$U(r) \equiv \pm \frac{g}{r} \exp(-\mu r), \tag{3.4}$$

where μ, the mass of exchanged particle and is just the inverse of the range of the force. Yukawa made a decisive breakthrough but his theory had still several shortcomings:

- His new force acted between protons and neutrons only, those between two protons and two neutrons were not considered. This implied that his original formulation of strong interactions was not charge invariant.

---

[3]In fact the existence of a particle that would be this sort of bound state of proton and electron had been predicted by Rutherford in his *Bakerian lecture* of 1920 and was in line with his idea of the structure of nucleus. Chadwick just followed his teacher's line of reasoning.

- Although the conjectured particles mediating his new force were charged, their interaction with electromagnetic field was not considered.

- The spin dependence of his new force was not discussed and thus no conjecture was made about the spin of Yukawa particle itself.

3.1.6   1935-1938: final touches

The above shortcomings of the Yukawa theory were removed during the following four years as a result of theoretical analyses of binding energies of light nuclei, like H^3 and He^4, and first experimental data on the scattering of protons and neutrons. The first of them showed convincingly that pp and nn forces are equal and the second suggested that they, moreover, equal to those acting between proton and neutron in the p-n symmetric state (which is the spin singlet state). Charge invariance of strong interactions had become established.

The formalism incorporating this property of strong interaction was developed by Cassen and Condon [41], who emphasized the analogy between spin-dependent forces described by means of Pauli matrices, and the forces acting between the proton and neutron. This analogy was based on the observation that proton and neutron can be understood as two members of a doublet, in the same way as two spin states of particles with spin 1/2. Using Pauli matrices to describe the transformation of this doublet under the rotation in the "isospin space" they showed that the available data implied that the form of the interaction potential must be of the form

$$W_{12}(r) = a + b\vec{\tau}^{(1)}\vec{\tau}^{(2)}. \tag{3.5}$$

The charge symmetry of strong interaction thus means these forces are invariant under the rotation in the space of internal degree of freedom of the nucleon, later called by Wigner "isotopic spin", and now isospin for short. The assumption of isospin invariance of strong interactions had one important implication: the extension of the Pauli exclusion principle, which now included isospin among the variables describing the state of the nucleon.

Written in terms of the potential (3.5) acting between two nucleons the strong interactions looked isospin invariant, but there was still a problem to extend this property to Yukawa theory, which described proton-neutron forces only and thus had only two charged mediating particles. The theory of nuclear force that respects the isospin symmetry was completed in 1938 by Nicholas Kemmer. To incorporate the charge independence of forces between proton and neutrons in the Yukawa theory, Kemmer introduced a neutral companion of the charged Yukawa mesons, extending thereby the concept of isospin to particles mediating the strong force. His step amounted to adding to the interaction hamiltonian of the Yukawa theory, written generically as (ϕ denoting complex scalar field)

$$H_{\text{int}} = g\left(\overline{p}n\phi^* + \overline{n}p\phi\right) = \frac{g}{\sqrt{2}}\overline{\psi}\left(\tau_+\phi^* + \tau_-\phi\right)\psi = g\overline{\psi}\left(\tau_1\phi_1 + \tau_2\phi_2\right)\psi, \tag{3.6}$$

where τ_i, τ_+, τ_- are standard isospin matrices and

$$\phi_1 \equiv \frac{\phi^* + \phi}{2}, \ \phi_2 \equiv \frac{\phi - \phi^*}{2i}, \quad \psi \equiv \begin{pmatrix} \psi_p \\ \psi_n \end{pmatrix}. \tag{3.7}$$

the third term involving the neutral scalar field ϕ)3

$$H_{\text{int}} = g\overline{\psi}\left(\tau_1\phi_1 + \tau_2\phi_2 + \tau_3\phi_3\right)\psi = g\overline{\psi}\vec{\tau}\psi\vec{\phi}. \tag{3.8}$$

Contrary to 3.6 the expression 3.8 In (3.8) ϕ_3 stands for a neutral meson field of the same mass as the charged fields ϕ, ϕ^*. The form (3.8) is manifestly invariant under the simultaneous rotation of the nucleon doublet and meson triplet in the isospin space. It took some time to elucidate the space-time properties of this new theory [4], but its main structure, which has become the model for further development of the theory of strong interaction, was ready.

---

[4]In particular to show, as did Rarita and Schwinger, that the correct coupling of nucleons and mesons is pseudoscalar.

3.2   1938-1947: decade of uncertainty

Since about 1934 it had been known that cosmic rays contained two components: the soft, mostly electrons and photons, producing showers, and one, which penetrates matter much more easily than as expected for electrons on the basis of the theoretical calculations of Bethe and Heitler. Physicists thus faced two options: either there were new particles in cosmic rays or the theory of electron bremstrahlung breaks down at high energies. In 1937 C. Anderson and Seth Neddermeyer had made the decisive step [29] towards the discovery of the *mesotron* (as was the name of what we now call muon) by showing that the distribution of energy losses clearly prefers the former option. Based on this observation, they had concluded that

> *"..there exist particles of unit charge but with mass (which may not have a unique value) larger than that of a normal free electron and much smaller than that of a proton".*

Their conclusion was immediately confirmed by Street and Stevenson, and accepted by physics community. The words in brackets were pertinent as in 1937 it was by no means evident that penetrating component of cosmic rays is composed of just one type of particle. Note the difference from the discovery of the positron, which was essentially a one man, one picture show.

In fact it took a decade to determine the mass of the mesotron with sufficient precision and thereby show that there is just one such particle with the mass around 100 MeV. It took about the same time to show convincingly that the mesotron interacted only very weakly with nuclear matter and could not therefore be identified with the particle predicted by Yukawa to mediate strong interactions. The three body nature of mesotron decay and the value of its spin were not established until 1949.

The real Yukawa particles, charged pions $\pi^{\pm}$ in current parlance, had been found in 1947 by Powell and collaborators by analyzing the tracks left by cosmic rays in a specially manufactured nuclear emulsion. The neutral pion, π^0, was found only in 1950 in the collisions of photons with protons at Brookhaven. Although further development had shown that Yukawa theory failed to describe strong interactions of hadrons quantitatively, by introducing the concept of isospin it paved the way to unitary symmetry, quark model and eventually Quantum Chromodynamics.

3.3   1947-1955: strange discovery and its consequences

The sequence of experimental findings and theoretical ideas leading to the concept of strangeness as another internal quantum number is described in great detail and with much insight by one of the participants in [31]. Many interesting details concerning the role and contributions of Murray Gell-Mann, one of the principal persons of the drama, can be found in [27] as well as in his personal reminiscences [33]. The circumstances surrounding the discovery of strange particles and the development of theoretical ideas for their explanation were quite complicated, but also very interesting, even fascinating. I will not attempt to describe all the twists and turns of the voyage culminating with the discovery of the Ω^- particle on January 31, 1964. I will merely mention the key experimental findings and outline the related theoretical developments. As we have seen on two previous occasions (Curies, Rutherford), observing a completely new phenomenon is one thing and its proper interpretation quite another. This was also the case with the discovery of strange particles. It is very likely that the first strange particle was observed by Leprince-Ringuet in Summer 1943, four years before the real discovery, in his cloud chamber exposed to cosmic rays in Pyreneen. On one of his pictures he recorded an event in which a positively charged particle coming from above knocks out an electron of the gas in the cloud chamber. Kinematic analysis of this event, assuming elastic scattering of the unknown particle on electron (i.e. no unobserved neutral particles) and measuring the momenta of the incoming particle and knocked-off electron from curvature in the magnetic field, yielded for the mass of the

	mode	%
K^+	$\mu^+\nu_\mu$	63.51
	$\pi^+\pi^0$	21.16
	$\pi^+\pi^+\pi^-$	5.89
	$\pi^+\pi^0\pi^0$	1.73
	$\pi^0\mu^+\nu_\mu$	3.18
	$\pi^0 e^+\nu_e$	4.82
K^0_S	$\pi^+\pi^-$	68.61
	$\pi^0\pi^0$	31.39
Λ	$\pi^- p$	63.9
	$\pi^0 n$	35.8
Σ^+	$\pi^0 p$	51.57
	$\pi^+ n$	48.31
Σ^0	$\Lambda\gamma$	100
Σ^-	$\pi^- n$	99.85
Ξ^-	$\Lambda\pi^-$	99.89
Ξ^0	$\Lambda\pi^0$	99.54

Table 3.1: Main decay channels of some of strange mesons and baryons.

incoming particle the value of about 500 ± 6 MeV, amazingly close to the mass of K^+ meson (494 MeV). Leprince-Ringuet published his result, but as it was just one isolated event, no one paid much attention to it. It was not so surprising after all, as there was nothing qualitatively new or unexpected on the mentioned
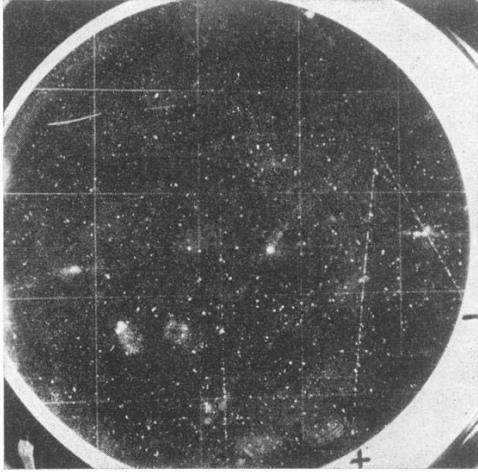
FIG. 2. V^0-disintegration in which the positive particle is probably a π-meson.
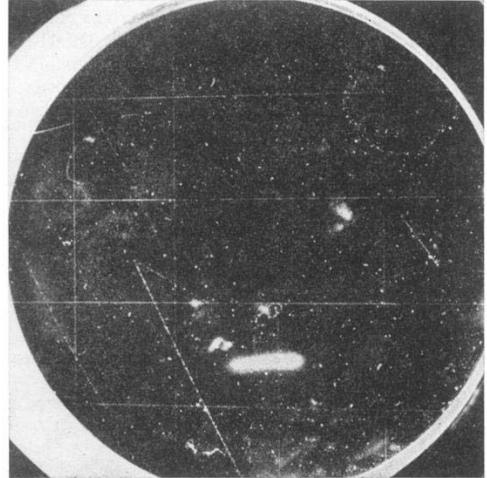
FIG. 1. V^0-disintegration resulting in a heavily ionizing proton.

Figure 3.2: Two events showing the presumable decay of K^0 (left) and Λ (right) observed in a cloud chamber placed in magnetic field. Taken from [30].

event, just the mass of the incoming particle did not coincide with that of any known particle. In view of the accuracy of momentum and angle measurements the Leprince-Ringuet event was interesting but not more.

The discovery of strange particles started in earnest in December 1947 when Butler and Rochester reported two events recorded in their magnetic lead plate cloud chamber exposed to cosmic rays. One of their pictures showed the first "V^0-particle", i.e. neutral particle produced in the lead plate by cosmic rays and decaying into a pair of oppositely charged particles. The second showed "kink" on the charged particle coming from above that was interpreted as the decay of positively charged particle. Kinematical analysis of both decays under the assumption that the decay particles were pions gave for the masses of unknown particles values in the region 440 ± 100 MeV and 540 ± 100 MeV. What they saw were almost certainly the decays of K^0 and K^+ mesons. It took more than two years to confirm their observation by another group using lead plate cloud chamber [5], but before that the Bristol group lead by Cecil Powell and employing nuclear emulsion, found an event that, after careful analysis, was interpreted as a decay of a heavy charged particle into three charged pions. Determination of its mass via the measurement of ionization and multiple scattering gave the value close to 500 MeV. This event was very likely the decay of K^+ into three charged pions. During 1950-1952 it became clear that there are actually two V^0-particles: one with mass of around 500 MeV and decaying into the pair of charged pions, the other with mass around 1100 MeV and decaying into the π^-p pair. Two of the early pictures showing decays of these particles are reproduced in Fig. 3.2.

Soon after their first observation, it became clear that these new particles behaved "strangely" [6]: they were produced in collisions of cosmic rays with nuclei at rates comparable to those of pions, but decayed many orders of magnitude slower than as expected for decay mediated by strong interaction. The pion also decays slowly but as pions are the lightest strongly interacting particles, the long lifetime was understood as the decay went via weak interactions. But why could not V_1^0, meson with the mass of around 500 MeV, decay into two charged pions with lifetime of the order 10^{-23}s as expected for decays mediated by strong interactions? The same question concerned the decay of baryonic V_2^0 with mass around 1100 MeV: what prevents this baryon to decay fast into the π^-p pair?

The situation with charged new particles was much more complicated and it took 5 years, and essential contribution from experiments at the new generation of accelerators, to clarify it. Paradoxically the first baryon that was established with reasonable certainty was the cascade hyperon Ξ^- (then called V_2^-), which was observed in 1952 by the Manchester group in a magnetic cloud chamber exposed to cosmic rays at Pic du Midi. Though very rare, the particle was firmly established on the basis of just a few events, because its practically unique decay mode $\Xi^- \to \Lambda \pi^-$ has a clear experimental signature: kink on the negative charged

---

[5]This group included Carl Anderson, who thus participated in the discovery of three fundamental particles: positron, muon and strange particles.

[6]Hence the name given to them by Gell-Mann in 1955.

track accompanied by the nearby decay of $\Lambda \to \pi^- \mathrm{p}$. In Summer 1953 the situation thus looked as follows:

- there was reasonably well established decay modes of

  - two neutral V^0-particles, one V_1^0 decaying into $\pi^+\pi^-$ pair, the other, V_2^0 into $\pi^- \mathrm{p}$,
  - negatively charged baryon V_2^- (Ξ^- today) decaying into $\pi^- V_2^0$,

- first signals of positively charged baryon V_2^+, decaying into the proton and some neutral particles,

- several events showing decays of positively charged mesons, sometimes lumped together as "V_1^+", into various combinations of pions and muons,

- no signs of a negative strange meson.
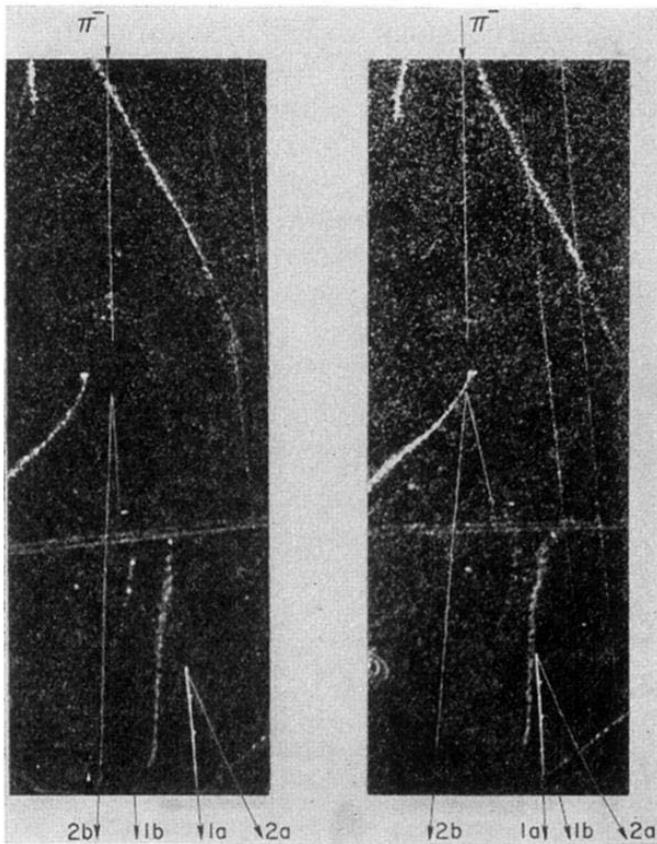
FIG. 2. Case D. Photograph of a 1.5-Bev π^- producing two neutral V particles in a collision with a proton. Tracks 1a and 2a, believed to be proton and π^-, respectively, are the decay products of a Λ^0. A ϑ^0 is probably seen to decay into π^+ (1b) and π^- (2b). Because of the rather "foggy" quality of this picture tracks 1b, 2a, and 2b have been retouched for better reproduction.

Figure 3.3: Pair production of neutral strange particles.

The V_1^+ events mentioned above seemed unrelated as some of these decays looked like two body decays, some were clearly of multibody nature. The precision of mass determination was not sufficient enough to determine whether these were decays modes of just one particle or whether more particles were involved. We now know that in all cases the decaying particle was K^+ meson, but at that time the possibility that a single particle decays into so many different channels (see Table 3.1) was not yet common. In this situation, on August 21, 1953 Gell-Mann submitted a short paper of just a half page [32] containing the suggestion [7]. His solution amounted to a particular assignment of isospin to the new "strange" particles that prevented them from decaying into the observed decay modes via strong interactions, but allowed them to proceed via weak interactions, which explained their long lifetime. Gell-Man assumed that the observed mesons V_1^0 and V_1^+ form an isospin doublet and the observed baryons V_2^0 and $V_2^\pm$ an isospin triplet. To appreciate his foresight and intuition, one has to realize that the evidence for V_2^+ was rather shaky and that for mesonic V_1^- simply absent. Note that this assignment went against the example set by nucleons and pions, which form isopsin doublet and triplet respectively. Assigning strange mesons V_1^+ and V_1^0 to isospin doublet was quite courageous as it implied that the antiparticle to neutral meson V_1^0 was not identical with V_1^0 as is the case of π^0 meson! In another important paper published two years later [37] Gell-Mann and Pais provided the theoretical framework for the treatment of neutral mesons that are not eigenstates under the transformation particle→antiparticle. The above assignment was designed with the sole purpose of preventing the observed decays like (in present day notation) $K_S^0 \to \pi^+\pi^-$ and $\Lambda \to \pi^-\mathrm{p}$ to proceed via strong interaction. With the benefit of hindsight we know that Gell-Mann's success was due in part to his genius but that he also benefited from the lack of precise experimental information which paradoxically helped him. The point is that the strange baryons he assigned to an isospsin triplet were in fact Ξ^-, Λ and Σ^+, each of them belonging, as we now know, actually to a different

---

[7]Shortly after the same proposal was made by Nakano and Nishijima in [34]

isomultiplet and only the last one being really a member of the supposed isospin triplet! It took another year and the arrival of first data from Cosmotron, the newly commissioned 3 GeV proton synchrotron at Brookhaven, to measure the masses with sufficient precision to see that V_2^0 (now Λ with 1115 MeV) and V_2^+ (now Σ^+ with 1189 MeV) cannot belong to the same isomultiplet and to find the true negative companion of Σ^+ and Σ^-.

The first experiments with pion beams at Brookhaven in 1953 had also demonstrated important feature of strange particle production called associated production. This feature, suspected for some time but not obvious in cosmic ray events, followed directly from the conservation of strangeness which required that strange particles are produced by strong interaction always in pairs of opposite strangeness. Example of one of the first events recorded at Brookhaven [36] showing the collision π^-p$\to K^0\Lambda$ is reproduced in Fig. 3.3. Another of pictures presented in [36] showed the associated production of charged strange particles, most likely the event π^-p$\to K^+\Sigma^-$, which turned out to provide the crucial test of Gell-Mann's scheme. Note that this decay conserves third component of the isospin ($-1 + 1/2 = 1/2 - 1$), and can thus proceed via strong interactions, whereas the opposite charge combination $K^-\Sigma^+$ does not. Though the first experiment at Brookhaven [36] produced just one such event it was a clear indication that his scheme is the right one. This observation has also definitely killed the earlier scheme of Pais, in which strange particles could be produced in strong interactions from nonstrange ones in even numbers, but without restrictions concerning strangeness (or equivalently, isospin), as the scheme proposed by Pais allowed both processes, π^-p$\to K^+\Sigma^-$ as well as π^-p$\to K^-\Sigma^+$, to go via strong interaction.

Although their isospin assignment obviously worked, there was still the question: is there any deeper reason for it? The first step in answering this question was taken by Nishijima in the paper [35] submitted in July 1954. In this paper he observed that his and Gell-Mann's peculiar isospin assignment can be reformulated in terms of a new quantum number, called by him "ν-charge" (for which Gell-Mann later coined the name "strangeness") which is conserved in strong interactions, but violated by weak force. In modern notation he wrote down the following relation between T_3, the third component of isospin, the **hypercharge** $Y \equiv (B + S)$ (B stands for baryon number) and electric charge Q

$$Q = T_3 + \frac{B + S}{2}. \tag{3.9}$$

This formula is presently usually called "Gell-Mann-Nishijima relation", but this is in my view inappropriate as Nishijima was clearly the first to write it down, a year before this relation appeared in Gell-Mann's contribution to 1955 Conference in Pisa. At the time Nishijima published his paper [35] containing the relation (3.9), Gell-Mann and Pais, in a contribution to 1954 Conference at Glasgow (delivered by Pais), summarized their understanding of Gell-Mann's model in the following words [38]

> In the present model,... each particle is labelled by three quantum numbers; they are I, I_Z and Q. In the usual and more restricted picture of charge invariance there are only two quantum numbers since Q and I_Z are not then independent. We have not been able to find, in terms of an enlarged group of symmetry operations, any interpretation of the variation of $Q - I_Z$ from multiplet to multiplet.

Gell-Mann has undoubtedly made many vital contributions to modern particle theory, but on three key moments his role has been overstated and he is credited with discoveries in which other people played more important role. One likely reason for that is his penchant for coining right names for new notions like strangeness, quark, Eightfold way, and Quantum Chromodynamics. This my viewpoint, not widely shared, is based on extensive study of original sources, rather than quoting the conventional "wisdom". I will return to this aspect of Gell-Mann's personality at two more occasions: later in this Section and the Section on QCD.

In SU(3) symmetry the conservation of isospin is equivalent to the conservation of strangeness, so this relation by itself does not carry any new restriction on possible strong decays. Nevertheless, the introduction of strangeness was crucial because it opened the way to unitary symmetry and finally the quark model. Within the quark model the peculiar isospin assignments of Gell-Mann and Nakano and Nishijima are simply a consequence of the fact that the strange quark is an isospin singlet. By the time of the annual Conference on high energy physics in 1955 in Pisa, the properties of the observed strange particles were established and the concept of strangeness became the cornerstone of further developments. At this conference Gell-Mann predicted the existence of what we now call Σ^0 and Ξ^0 and, most remarkably, in the Appendix to the written

contribution also that of the Ω^- baryon, assigned to an isosinglet. The first was found in 1956, second in 1959 and the last in 1964. Ω^- was to play a crucial role in the confirmation of the Eightfold Way that brought Gell-Mann the Nobel prize for physics in 1969.

3.4   1956-1960: paving the eightfold way

The second half of fifties, after the Conference at Pisa, where the problem of production and decay of strange particles had, in essence, been solved, was marked by several important experimental discoveries made with accelerators at Brookhaven and Berkley, where new 6 GeV Bevatron entered service in 1955. Preeminent among them was the discovery, in 1955, of the antiproton, first antibaryon ever seen, followed by antineutron a year later. As mentioned above, Σ^0 and Ξ^0 hyperons, predicted by Gell-Mann in 1955, were found in 1956 and 1959. All these discoveries used a novel detection technique invented by Glaser in 1952 had further developed by Alvarez: the bubble chamber. It dominated as detection technique for more than two decades and its innumerable contributions culminated by the discovery of neutral currents in 1973.

On the theoretical side attempts were undertaken to understand the growing number of particles that had been found since the discovery of pions. To understand the atmosphere in which theorists of that time framed their considerations, it should be remembered that the second part of fifties was the time of general decline of interest in quantum field theory in general and in application to strong interactions in particular. The reasons were two-fold. The language and methods of Quantum field theory fell from favor as a result of the failure of the program of Landau and his collaborators to give renormalization of QED good mathematical as well as physical meaning. I will return to this point in section 8 on renormalization. As far as the strong interactions were concerned, field theory seemed unable to provide the framework for successful theory based on standard perturbative methods applied to Yukawa theory.

However on October 1st, 1954, just about the time quantum field theory fell out of favour as far as the strong interactions were concerned, two young american theorists, Robert Mills and Chen Ning Yang, published the paper called *Conservation of Isotopic Spin and Isotopic Gauge Invarianc* [46], which became one of the milestones on the way to our current understanding of the fundamental forces acting in the microworld. In their work Yang and Mills generalized the concept of gauge invariance of Hermann Weyl, who used it to derive Quantum Electrodynamics (QED), to strong interactions between nucleons. The crucial difference between these two theories rests in the fact that in QED the gauge invariance concerns the phase transformation of an elektron field, which are described by abelian group $U(1)$, whereas in strong interaction the transformations of the nucleon dublet in (3.8) correspond to nonabelian group $SU(2)$. This difference has tremendous consequences, which took decades to fully appreciate. The paper [46] was left unnoticed for more than five years and by Yang's own account [47], it took him twenty years to understand the mathematical consequences thereof. Most recent books on Quantum Field Theory present nonabelian gauge theories in the framework of modern mathematical tools like fiber bundles and differential geometry, but as Yang himself has frequently emphasized the actual discovery was done in a very down to earth manner, essentially by trial and error. The interview with Yang in [47] is extraordinarily revealing as it concerns a crucial aspect of present theoretical particle physics, namely the relation between physics and mathematics.

Instead of quantum field theory most of the theorists followed Heisenberg in his suggestion to make the S-matrix the primary theoretical concept, without relying on standard field theoretic methods to calculate it. The approach that took up this idea, usually called analytic S-matrix theory [42], concentrated on the exploitation of the rather subtle properties of the scattering matrix (hence the name) that were previously derived from quantum field theory, but which in this novel approach were simply postulated. The most ambitious approach within this rather widely defined program was the so called *bootstrap* model, based on the idea of "nuclear democracy" invented and pursued by G. Chew. [8] it It assumed that combining the analytical properties of the S-matrix, which were related to particle spectrum, with the requirements following from conservation laws, unitarity and self-consistency will in the end lead to unique determination of the S-matrix itself. The analytic S-matrix was considered as the best theoretical framework for formulation of the theory of strong interactions until the discovery of asymptotic freedom in 1973. This, almost overnight, changed the situation: quantum field theory was resurrected and the S-matrix returned to where it belongs, i.e. at the end of calculations based on quantum field theory.

---

[8]The idea that all hadrons are equally elementary strongly influenced Gell-Mann and his interpretation of quarks as mathematical entities rather than real constituents of hadrons. If all elementary objects are to be on equal footing then quarks cannot be observed.

Name	Model	Isotopic Spin	Strangeness	Ordinary Spin
$\mathfrak{N}$		1/2	0	1/2
$\overline{\mathfrak{N}}$		1/2	0	1/2
Λ		0	−1	1/2?
$\overline{\Lambda}$		0	1	1/2?
π	$\mathfrak{N}+\overline{\mathfrak{N}}$	1	0	0
$\theta(\tau)$	$\mathfrak{N}+\overline{\Lambda}$	1/2	1	0?
$\overline{\theta}(\overline{\tau})$	$\overline{\mathfrak{N}}+\Lambda$	1/2	−1	0?
Σ	$\mathfrak{N}+\overline{\mathfrak{N}}+\Lambda$	1	−1	1/2?
Ξ	$\overline{\mathfrak{N}}+\Lambda+\Lambda$	1/2	−2	1/2?

Here $\mathfrak{N}$ and $\overline{\mathfrak{N}}$ denote nucleon and antinucleon respectively, whereas Λ and $\overline{\Lambda}$ denote Λ^0 and anti-Λ^0 respectively[3].

Figure 3.4: Composition of mesons and baryons according to Sakata.

Fermi-Yang model

In 1948 Fermi and Yang were the first [49] to consider pions not as elementary particles, but rather as bound states of nucleon-antinucleon pairs($N\overline{N}$). Their paper starts with the words

> In recent years several new particles have been discovered which are currently assumed to be "elementary", that is, essentially, structureless. The probability that all such particles should really be elementary becomes less and less as their number increases.

In essence they postulated the pions to be bound states of nucleon-antinucleon pairs of appropriate charge combination. To produce pions with mass about 140 MeV from nucleons with 940 MeV, they had to assume a very strong attractive force in $N\overline{N}$ states, but no such force in other ones.

Their model was in principle, though not in practice, in conflict with the Yukawa theory of strong interactions, as in the latter pions play a fundamental role of force carriers. According to Fermi and Yang, the virtual $N\overline{N}$ states created in the vicinity of baryons *"have a tendency to pair formation of nucleons antinucleons, which will be predominantly formed in the bound states, i.e. as π-mesons.* Their model was precursor of all subsequent models of hadron structure, but has never become more than a qualitative framework of thinking about this question.

Two more remarks concerning this model are in order. First, from present day perspective, it may seem rather surprising that Fermi and Yang saw a problem in the possibility that "all such particles" were elementary, when we realize that "all" meant at that time just the nucleons, $\pi^{\pm}$ mesons and just a few first signals of strange particles.

Second, the subject of the paper shows the extent of Fermi's genius, probably the last physicist who made fundamental contributions to both theory and experiments.

Sakata model

In 1956 the idea of Fermi and Yang was taken up by Sakata [48], who extended it to strange particles by adding the Λ-baryon as third fundamental baryon and $\overline{\Lambda}$ as third antibaryon. In his own words [48]:

> In our model the new particles are considered to be composed of four kinds of fundamental particles in the true sense, that is nucleon, antinucleon, Λ and anti-Λ.

So, for instance, kaons were regarded as bound states $N\overline{\Lambda}$ and antikaons as $\overline{N}\Lambda$. Strange baryons were assumed to be formed from three fundamental particles and antiparticles as follows: $\Sigma = N\overline{N}\Lambda$, $\Xi = \overline{N}\Lambda\Lambda$. This model was an important step forward, as it reduced the problem of strange particles to that of the Λ hyperon. On the other hand, it singled out three particles, proton, neutron and the Λ, and their antiparticles, as fundamental building blocks, although experimentally they did not appear to be more fundamental than other known baryons. Moreover, the model had never been elaborated into a real calculational scheme.

Ne'eman model

In February 1961, before any of the vector mesons were found, Yval Ne'eman submitted a paper [50] in which the theory of strong interactions was derived from the principle of gauge invariance, clearly and without any reservations. In my view this paper represents a real milestone as it starts from the idea of Yang-Mills [46] to use the requirement of local gauge invariance as the basic principle upon which to build the theory of strong interactions. Within this framework the vector mesons play a fundamental role of gauge bosons which mediate strong interactions between all other strongly interacting particles, like pseudoscalar mesons and baryons.

Based on the experimental evidence then available, Ne'eman assigned the observed mesons to an octet (he called the then yet undiscovered η meson $\pi^{0,}$) and wrote, following the recipe of Yang-Mills the interaction term, which contained just one coupling constant and which described, beside the coupling of vector mesons to baryons and pseudoscalar mesons, also the self-coupling of vector mesons. Once Ne'eman chose the symmetry group and representations occupied by baryons and pseudoscalar mesons, he got a unique theory, which is in one important aspect a direct predecessor of Quantum Chromodynamics. Let me emphasize that the crucial point of his construction is not the choice of representations occupied by baryons and pseudoscalar mesons but the choice of the underlying symmetry group SU(3) and the use of gauge invariance to generate the force acting between them.

The central concept of gauge theories is that of the corresponding charge. In QED this charge is the electric charge and the gauge boson is the photon, which itself is electrically neutral, does not change the electric charge when being absorbed or emitted and cannot therefore interact locally with other photons. In Ne'eman's theory, the role of the charge is played by the flavor quantum number. Gauge bosons of his theory, the vector mesons, carry flavor and do therefore change the flavor or baryons and pseudoscalar mesons. Moreover, as they carry flavor, they do couple locally to each other as well. In QCD the role of gauge bosons, in Ne'emans's theory played by eight vector mesons K^*, ρ, ω and ϕ, is taken over by eight gluons, which carry the charge called *color*, hence Quantum Chromodynamics. With this in mind it seems appropriate to call Ne'eman's theory *Quantum Flavordynamics.*

Gell-Mann models

In January 1961 Gell-Mann wrote a preprint called *The Eightfold way: a theory of strong interaction* [58], which made an had enormous influence on further development of the unitary symmetry. As the paper of Ne'eman it is based on the idea of Yang and Mills and builds the SU(3) nonabelian gauge theory. Vector mesons, none of which was at that time observed expperimentally, are introduced as gauge bosons and various phenomenologically interesting processes and decays are calculated. It goes to much greater detail than the paper of Ne'eman, but for reasons that are difficult to understand and have to do with Gell-Mann's complicated personality, this masterpice has never been published! According to his biographer George Johnson [59] the content of the preprint [58] was submitted to the Physical Review in March 1961 but then withdrawn and rewritten. The revised version, which was eventually published as [60], was submitted at the end of September, after the discovery of the first vector mesons, K^*, ρ and ω, contains very little from the original preprint [58] and does not even mention the idea of Yang-Mills gauge theories.

In [60] Gell-Mann modified the Sakata model in one important aspect. He took up its basic assumption that hadrons are made out of triplets of some basic fields, but dropped the other assumption, i.e. that the baryons proton, neutron and Λ are elementary, whereas all other hadrons, mesons as well as baryons are their bound states. In this modification all hadrons, including proton, neutron and Λ are composed of some "hidden" objects (see [60], p.1079) and are thus treated on the same footing. The nontrivial consequence of this approach, following from the decomposition

$$\mathbf{3} \otimes \overline{\mathbf{3}} \otimes \mathbf{3} = \mathbf{15} \oplus \overline{\mathbf{6}} \oplus \mathbf{3} \oplus \mathbf{3} \tag{3.10}$$

is that physical states of baryons may occupy any of the representations appearing on the l.h.s. of (3.10). For instance, the nucleon doublet and the triplet of Σ may be part of the same anti-sextet $\overline{\mathbf{6}}$, together with some so far unknown baryon isosinglet, but without the nucleon and Ξ doublets. Alternatively, Σ triplet may occupy, together with the Ξ doublet but without the nucleon one, the pentuplet $\mathbf{15}$. The answer which of these options is realized in nature could not be answered by purely theoretical arguments, but had to be left to experiments. The physical states of mesons occupy, as in original Sakata model, the octet.

The vestiges of the original preprint [58] can be found in a somewhat hidden form in Section VIII of [60]:

In Section VIII we propose, as an alternative to the symmetrical Sakata model, another scheme with the same group, which we call "eightfold way". Here the baryons, as well as mesons, can form octets and singlets, and the baryons N, Λ, Σ and Ξ are supposed to constitute an approximately degenerate octet.

This suggestions relies on the relation (2.92)

$$\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{10} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1}. \tag{3.11}$$

Note that that this scheme can be easily distinguished from the one based on (3.10) as is predicts quite different multiplet structure. Most importantly, the baryons N, Λ, Σ and Ξ form just one multiplet.

In this paper the assignment of vector mesons to octet followed from the experimental fact but had no deeper theoretical reason as in [58]. The "effective coupling" of octets of pseudoscalar mesons and baryons, based on the relation

$$\mathbf{8} \otimes \mathbf{8} = \mathbf{27} \oplus \mathbf{10} \oplus \overline{\mathbf{10}} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1} \tag{3.12}$$

was assumed to be of the following generic form, which is just a straightforward generalization of the old Yukawa coupling of pions and nucleons

$$g_1(\overline{B}\gamma_5 B\pi + \overline{B}\gamma_5 \pi B), \tag{3.13}$$

where B and π stand for baryon and pseudoscalar meson octets. For vector meson-baryon coupling Gell-Mann assumed form that represents direct generalization of electromagnetic coupling of electrons to photons

$$g_2(\overline{B}\gamma_\mu B W^\mu + \overline{B}\gamma_\mu W^\mu B), \tag{3.14}$$

where W_μ denotes the octet of vector mesons. No self-interaction of vector mesons or their coupling to pseudoscalar ones was assumed.

The paper [60] is written in a style, which has become typical of all Gell-Mann's later contributions: the author makes interesting suggestion which differs from other approaches existing at that time, but presents it merely as "an alternative" to other possibilities, without spelling out clearly his preference. This stands in marked contrast to the paper of Ne'eman, as well as the original Gell-Mann preprint [58] where the authors clearly and unequivocally lay out their model. In the case of the eightfold way this trait of many of Gell-Mann's papers is expressed by the following statement ([60], p. 1081):

If the meson spectrum is consistent with broken unitary symmetry, we should examine the baryons and see whether the various baryon states fit into the representations $\mathbf{3}, \mathbf{6}$ and $\mathbf{15}$ (or the representations $\mathbf{1}, \mathbf{8}, \mathbf{10}, \mathbf{10}^*$ and $\mathbf{27}$ that arise in an alternative form of unitary symmetry).

The idea of *Eightfold way* forms actually only a tiny part of the rather long paper [60], devoted mostly to the discussion of equal time commutation relations of various currents, dispersion relations and other heavy artillery of the apparatus of S-matrix theory. The paper ends with a sentence that has become indispensable part of almost all future papers of Gell-Mann and which shows how deeply rooted he was in this framework and how little relevance he ascribed to any attempts to consider hadrons as composite particles.

Nowhere does our work conflict with the program of the Chew et al. of dynamical calculation of the S-matrix from strong interactions using dispersion relations. If something like Sakata model is correct, then most of the mesons are dynamical bound states or resonances …Those particles for which there are fundamental fields (like n, p, Λ in a specific field-theoretic model) would presumably occur as CDD poles or resonances in the dispersion relations.

If there are no fundamental fields and no CDD poles, all baryons and mesons being bound or resonant states of one another, models like Sakata will fail; the symmetry properties we have abstracted can still be correct, however.

3.5 1961-1964: from resonances to quarks

The first half of sixties was marked by a proliferation of experimental discoveries, some of them spurious, of the so called *resonances* [9] accompanied by a surge of related theoretical research.

---

[9]In the wider sense, the concept of resonance is a generalization of the conventional bound states to unstable particles. It is well-defined using the language of of analytic S-matrix, where it corresponds to the single pole in complex plane of appropriate
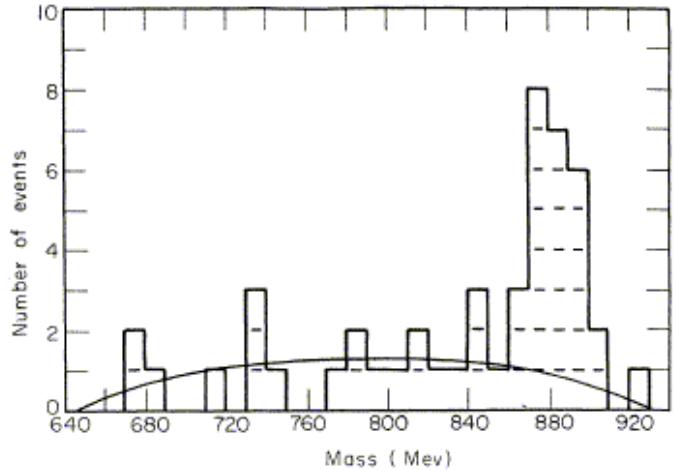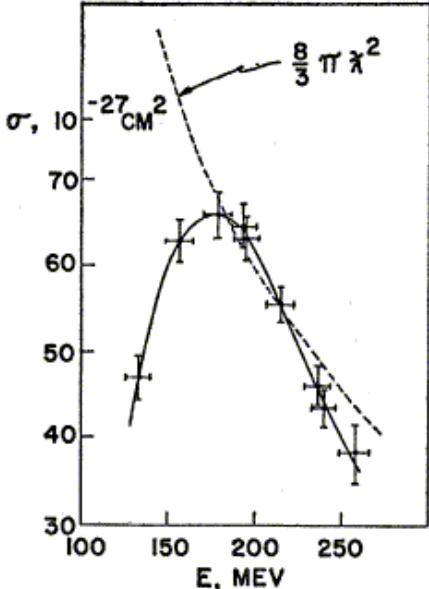
FIG. 2. Mass spectrum of the $\overline{K}^0$-π^- system. The solid line represents the phase-space curve normalized to background events.

Figure 3.5: The plots showing the first baryon and meson resonances. At left the energy dependence of the total cross section $\sigma(\pi^- p)$ as observed in [44] exhibits a peak corresponding to the Δ^0 resonance. The original text to the figure showing the first evidence [45] for the K^{*-} meson resonance is reproduced.

The first indication of such resonances was found by Fermi and collaborators back in 1952 using the π^- beams from old Chicago proton synchrotron. Due to limited energy only the onset of typical resonance behavior could, however, be observed in the $\pi^+ p$ channel. In the left part of Fig. 3.5 this resonance, now called Δ^0, is clearly visible as peak in the energy dependence of the cross section $\sigma(\pi^- p \to \pi^- p)$ measured in 1954 by a different group at Carnegie Institute of Technology using the π^- beam supplied by the 450 MeV proton synchrotron [44]. This plot is the first ever picture of a resonance. In 1955 S. Lindenbaum and C. Yuan showed, using the π^+ and π^- beams of energies up to 750 MeV provided by the the Brookhaven Cosmotron, that the resonance observed by Fermi exists in all four possible πN channels and thus has isospin 3/2. No information on its spin was then available. As discussed below, this particle later played crucial role in the formulation of the quark model and introduction of the concept of colour.

The right plot in Fig. 3.5 shows the distribution of the invariant mass of $\overline{K}^0 \pi^-$ pair produced in $\pi^- p$ collisions at Berkeley Bevatron in February 1961, which contains evidence for the a state around 880 MeV [45], now known as K^{*-}. This particle, later shown to have spin 1, was the first vector meson discovered. Soon after it followed ρ [51], other charge states of the Δ, ω [52], η [54], and ϕ [53]. The last one, ϕ, had a puzzling decay pattern, which, as we shall see below, provided a crucial input for the formulation of the quark model. Interestingly, three months before the discovery of the first meson resonances, K^*, the second baryon one, $\Sigma^{*\pm}$, was seen in $K^- p$ collisions at Berkeley [55]. The plot showing this resonance is reproduced in Fig. 3.6. It should be stressed that the pace of experiment progress was not without false traces leading to "discoveries" that had not withstood the time. An example of such a "discovery", announced by a group of renowned experimentalists [56] in a paper under the title *Evidence for $\pi^+\pi^-$ resonance at 395 and 520 MeV effective mass*, is shown in Fig. 3.7. The fact that the, at first glance rather impressive, evidence for both of these resonances turned out to be spurious, illustrates the point made in the footnote above, namely that not every bump in the distribution of some invariant mass, here in the invariant mass of $\pi^+\pi^-$ pairs, must be a genuine resonance.

Next decisive moment on the path to the formulation of the quark model occurred at the International Conference on High Energy Physics in Geneva in the Summer of 1962, where the discoveries of two more

kinematical variable. In contrast to genuine bound states, which lie on the positive real semi-axis, resonances are located on the second sheet of the cut complex plane. Their presence there is revealed by characteristic peaks in the dependence of cross sections on the corresponding energy variable, but it should be emphasized that this signature is not unambiguous and more sophisticated methods must be used to prove that an observed bump represents a genuine resonance. Ignoring this lesson had lead to a number of spurious discoveries that were not confirmed. In the narrower sense the term "resonance" denotes hadrons that can decay by strong interaction and have therefore very short lifetime, of the order 10^{-23}s. This lifetime is so short that resonances do not leave any measurable tracks and their existence can be ascertained only by observing the above mentioned peaks with the width of around 100 MeV.
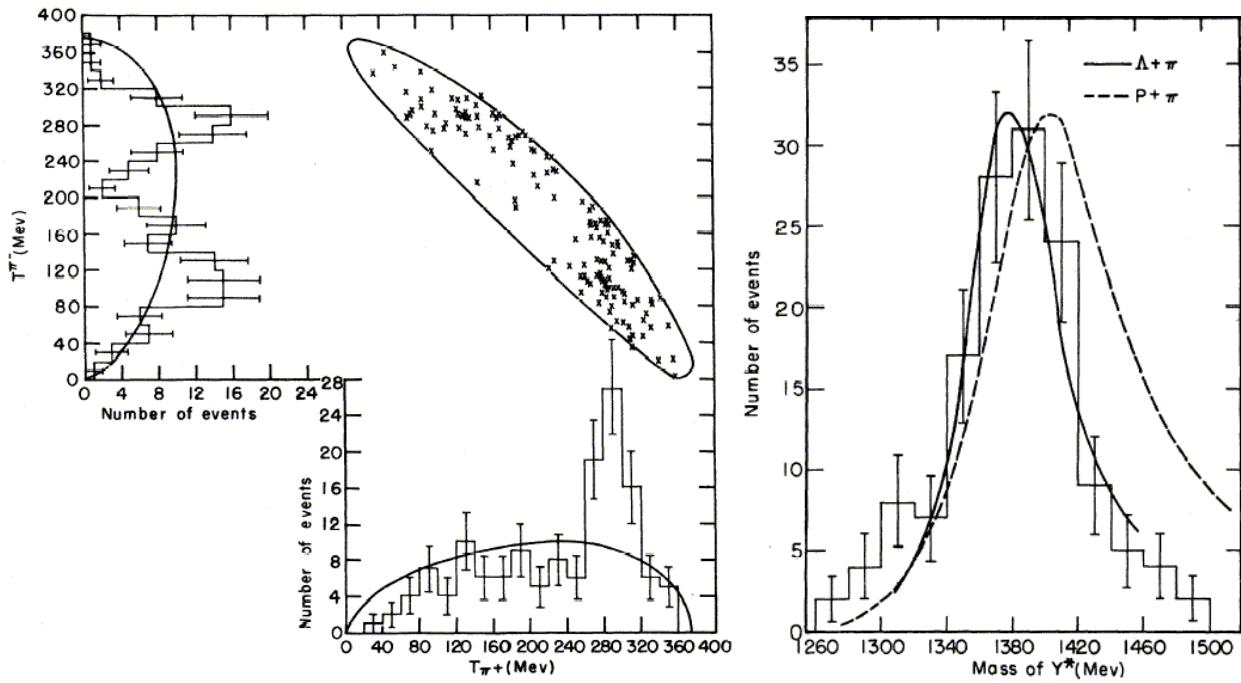
Figure 3.6: The Dalitz plot demonstrating the existence of $\Sigma^{*\pm}$, then called Y^*, in the channel $K^-\text{p} \to \Lambda\pi^+\pi^-$ (T_{π^+} and T_{π^-} denote the energies of produced pions).
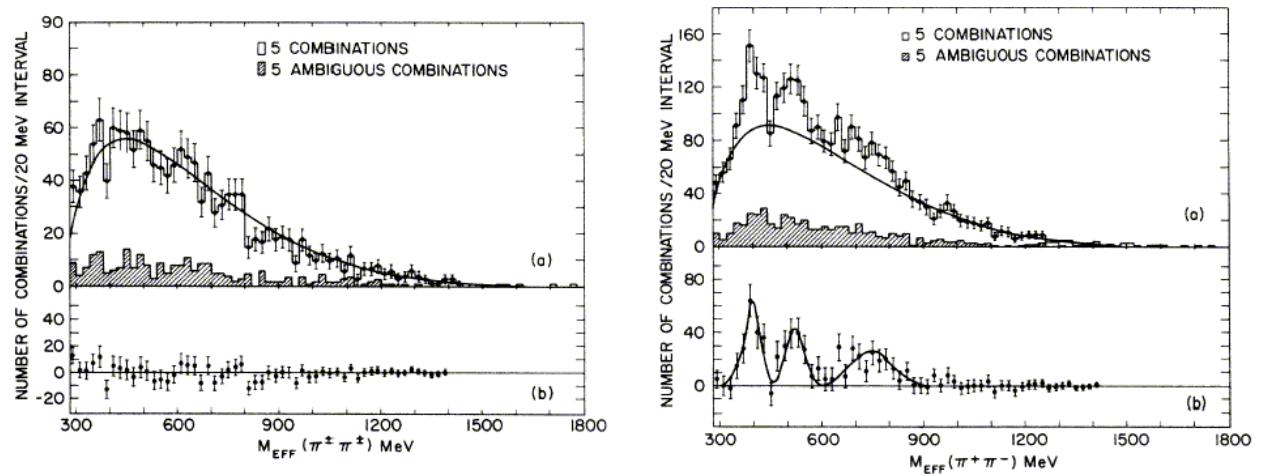

Figure 3.7: Example of a discovery that was not confirmed.
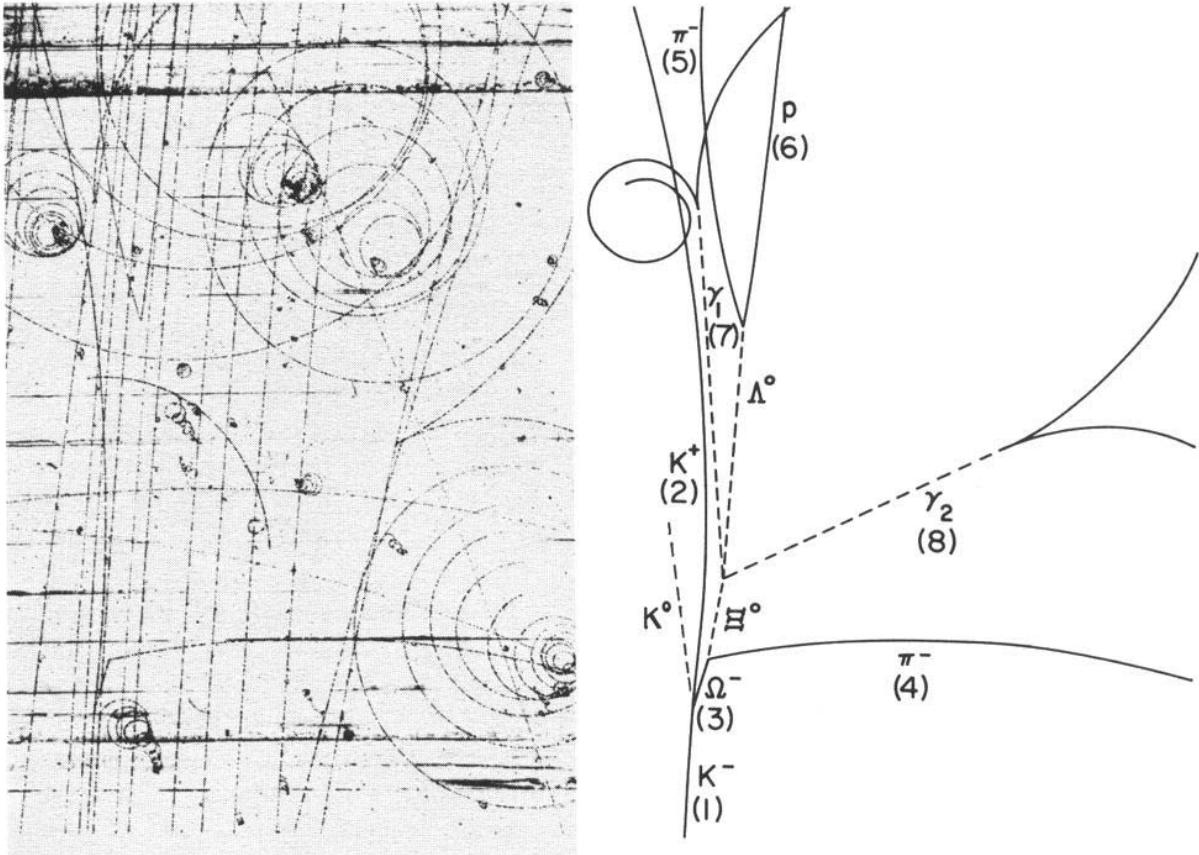
FIG. 2. Photograph and line diagram of event showing decay of Ω^-.

Figure 3.8: The first Ω^- observed in bubble chamber experiment at Brookhaven on 31.1.1964 [66].

baryon resonances, $\Xi^*(1530)$ and Ξ^{*0}, were announced. Although this Conference was the occasion of an official announcement of their discovery, the rumors about them went around for some weeks and reached many leading theorists of that time, including Ne'eman and Gell-Mann. From the testimony of contemporaries it seems clear that both of them interpreted these discoveries, which fit very well into their model, as vindicating its correctness. Both of them also used the measurement of their masses as input for predicting the existence and mass of (in current notation) Ω^-, the last baryon resonance missing in the baryon decuplet. After the presentation of experimental results, Gell-Mann strode to the blackboard and predicted the existence of Ω^- hyperon with mass of about 1670 MeV. The circumstances surrounding his exposition are quite interesting [10] and show that luck plays role even in science. Gell-Mann's prediction was immediately addressed in several experiments in which Ω^- was looked for in one of the most likely channel

$$K^-\mathrm{p} \to K^+ K^0 \Omega^-,\tag{3.15}$$

with the expected decay modes $\Omega^- \to \Lambda K^-$, $\Omega^- \to \Xi^0 \pi^-$ and $\Omega^- \to \Xi^- \pi^0$. On January 31st 1964 Nicholas Samios found the event, reproduced in the left part of Fig. 3.8, which was rather unambiguously interpreted in a way depicted in the right part of that figure. The last member of baryon decuplet was found and the *Eightfold way* triumphed. For his prediction Gell-Mann was awarded the Nobel prize for physics in 1969.

---

[10]According to one source [26], after the experimental presentation of $\Xi^*(1530)$ and Ξ^{*0} discoveries, both Gell-Mann and Ne'eman raised hand asking for word, but as Gell-Mann was sitting in the front row he got the word. According to [27] only Ne'eman raised but the chairman called, "Gell-Mann", probably because the latter was better known. In any case I think Ne'eman deserves the same credit for the prediction of Ω^- as Gell-Mann.

Chapter 4

Quark Model

In this Chapter the elements of the *constituent* quark model [1] are introduced and some of its applications discussed. Beside the idea of quarks as fundamental building blocks of matter, the application of the quark model to the spectrum of hadrons had lead to introduction of another fundamental concept of present theory of strong interactions: the *color*. This quantum number, which plays crucial role in the phenomenon of **quark confinement** [2] has become the cornerstone of Quantum Chromodynamics, the theory of strong forces between colored objects, to be discussed in Chapter 7.

4.1 January 1964: birth of the quark model

In January 1964, even before the discovery of Ω^- and the confirmation of the *Eightfold way*, two theorist published papers that heralded the birth of the quark model. Both the personalities and positions from which they approached the problem of dynamical origins of the Eightfold way could hardly be more different. On one side Murray Gell-Mann [61], an established and widely respected theorist with many previous works to his credit, on the other George Zweig [62], a young postdoc who was just spending a year at CERN. Each of them have approached the problem from quite different directions as illustrated by comparing the introductory paragraph of [61]

> *If we assume that the strong interactions of baryons and mesons are correctly described in terms of the broken "eightfold way", we are tempted to look for some fundamental explanation of the situation. A highly promised approach is surely dynamical "bootstrap" model for all strongly interacting particles within which one may try to derive isotopic spin and strangeness conservation and broken eightfold symmetry from self-consistency alone. Of course, with only strong interactions the orientation of the asymmetry in the unitary space cannot be specified; one hopes that in some way selection of specific components of the F-spin by electromagnetism and the weak interactions determines the choice of the isotopic spin and hypercharge directions.*

with the abstract to [62]

> *Both mesons and baryons are constructed from a set of three fundamental particles, called aces. The aces break up into isospin doublet and singlet. Each ace carries baryon number 1/3 and is fractionally charged. SU_3 (but not the Eightfold way) is adopted as a higher symmetry for the strong interactions. The breaking of this symmetry is assumed to be universal, being due to the mass difference among the aces. Extensive space-time and group theoretic structure is then predicted for both mesons and baryons, in agreement with existing experimental information. Quantitative speculations are presented concerning resonances that have not as yet been definitely classified into representations of SU_3. A weak interaction theory based on right and left handed aces is used to predict rates for $|\Delta S| = 1$ baryon leptonic decays. An experimental search for the aces is suggested.*

---

[1] Also called *additive quark model*. These adjectives will be omitted in the following.
[2] I.e. the fact that quarks do not exist in nature as isolated free objects like, for instance, electron or proton

Figure 4.1: Symbols used by Zweig for his aces and wave functions of baryons.

Zweig's quarks

Inside [62] one finds detailed and extensive elaboration of the points mentioned in the above abstract. Except for effects related to color it contains practically all quantitative predictions that go now under the name of *additive* or *constituent* quark model, some of which will be discussed in more detail in Section 4.2. Replacing Zweig's names for quarks (aces) and isospin doublets (deuces) and triplets (trays) with the terminology introduced in part by Gell-Mann we, for instance, easily recognize in Fig. 3 of [62], reproduced in Fig. 4.1, the expression for the anti-symmetric part of the proton wave function (4.12).

The personal description of his journey to quarks, full of interesting details concerning the circumstances in which he worked and the people who surrounded him, can be found in the talk on the history of the quark model given by Zweig in [64]. I know of no better source of unbiased information on this topic. At one point Zweig explains what was at the origin of his approach to quarks: it was the discovery of the ϕ meson with its puzzling decay pattern. The fact that ϕ decayed predominantly (in 83 %) into $K\overline{K}$ pairs rather than to $\rho\pi$ pair (12.9 %) or $\pi^+\pi^-\pi^0$ (2.7 %), contrary to expectations based on the phase space arguments, was for Zweig, and for all theorists, "totally unintelligible". To understand this puzzling phenomenon Zweig developed the phenomenological rule now carrying his name, which allowed him to understand the decay pattern of ϕ as a consequence of the fact that ϕ meson can be considered a bound state of the strange quark and its antiquark. Separating this $s\overline{s}$ pair leads to creation of $u\overline{u}$ and $d\overline{d}$ pairs which recombine with the "constituent" s and $\overline{s}$ quarks into kaons, hence the dominance of $K\overline{K}$ final states. More on this rule below.

Gell-Mann's quarks

In [61] Gell-Mann assumes that

A formal mathematical model based on field theory can be built up for the quarks exactly as for p, n, Λ in the old Sakata model, for example with all strong interactions ascribed to a neutral vector meson field interacting symmetrically with the three particles.

Within such a framework he then writes down the expression for the electromagnetic and weak currents, where the fractionally charged quarks appear explicitly

$$J_\alpha^{\text{elmag}} = \frac{2}{3}\overline{u}\gamma_\alpha u - \frac{1}{3}\overline{d}\gamma_\alpha d - \frac{1}{3}\overline{s}\gamma_\alpha s, \quad J_\alpha^{\text{weak}} = \overline{u}\gamma_\alpha(1-\gamma_5)(d\cos\theta_C + s\sin\theta_C), \tag{4.1}$$

where θ_C is the so called Cabbibo angle, to be introduced later, and uses them in the sense of the above quotation. For Gell-Mann quarks have always remained [3] basically a mathematical concept, devoid of any physical reality, which served him merely in order to derive various commutation relations between hadronic currents that could then be "abstracted" and used in dispersion relations. In this paper the term "quark" was used for the first to denote the members of the fundamental SU(3) triplet and was practically immediately generally accepted.

In my view, we owe our current understanding of quarks, and gluons, within QCD to Zweig rather than to Gell-Mann. It was Zweig's insistence on the some sort of physical reality of quarks what lies at the heart of Feynman's parton model and the whole framework of current perturbative QCD. In our current picture of the structure of hadrons, quarks (in the sense of constituent quarks) are not just mathematical concept, but a sort of quasi-particle, appropriate for simple description of their static properties. This is particularly true for hadrons made of heavy quarks, which, as outlined in Section 4.7, can be reasonably well described by nonrelativistic quantum mechanics.

Unfortunately, it took almost a decade, the personal influence of people like Feynman and Bjorken, and finally the invention of QCD for this viewpoint to take hold. The non-conformity of Zweig's ideas, combined with his age and the crushing inertia of journal editors prevented his work from being properly published. It has remained only as a preprint, difficult to access even in CERN Library, which enters the collection of papers on quark model [62]. What is, however, really sad is that the non-conformity of his idea had caused Zweig problems with getting appointments even years after he formulated his "ace model", which, as no one now dares to deny, is basically correct.

4.2 Quarks with flavor and spin: the SU(6) symmetry

In 1964, after the discovery of the Ω^- hyperon and the formulation of the quark model, the world of elementary particles looked as follows. All observed hadrons could be arranged (identifying $H_2 = T_8 = (\sqrt{3}/2)Y$)into multiplets of the SU(3) group shown in Fig. 4.2. The grouping of hadrons of the same spin, parity and baryon number in SU(3) octets and decuplets was based essentially on their masses and isospin symmetry. Moreover, all evidence indicated that the strangeness was conserved by strong interactions exactly. The following features of the above arrangement are worth noting:

- Mass differences within the isospin multiplets are much smaller than those between the different SU(2) multiplets within one unitary multiplet, indicating that the full SU(3) symmetry is broken much more strongly than the subgroup of isospin symmetry.

- The mean masses of isospin multiplets are *increasing* functions of the absolute value of the *strangeness*.

- The mass pattern is especially simple in the baryon decuplet, where the four isospin multiplets are spaced nearly equidistantly, the mass separation being roughly 150 MeV.

- Mass relations are much more complicated in the octets, in particular in the center (where, as we know, the state is **not** uniquely defined by its weight). In particular η, η' cause problems to accommodate their masses within the unitary multiplets [9].

The fact that strong interactions conserve third component of isospin and strangeness exactly implies that the strong interaction Hamiltonian H_s commutes with T_3 and T_8:

$$\left[H_s, T^\pm\right] = [H_s, T_3] = [H_s, T_8] = 0. \tag{4.2}$$

---

[3]Despite his statements to the contrary after the invention of the parton model and QCD [63].

baryon octet with spin 1/2   strangeness   mass [MeV]

n (ddu)   Y p (uud)   0   p   939.6
                                               n   938.3

Σ⁻ (dds) Σ⁰ (uds) Σ⁺ (uus)   −1   Σ⁻ 1197
−1   Λ   1 T₃   Σ⁰ 1192
                                               Σ⁺ 1189
                                               Λ   1115

Ξ⁻(dss)   Ξ⁰ (ssu)   −2   Ξ⁰ 1320
          −1   Ξ⁻ 1314

octet of pseudoscalar mesons

$\overline{K^0}$ (d$\bar{s}$) Y K⁺ (u$\bar{s}$)   1 $\overline{K^0}$ 498
                                               K⁺ 494

π⁻ (d$\bar{u}$) π⁰ (q$\bar{q}$) π⁺ (u$\bar{d}$)   0   π⁺⁻ 140
−1   η   1 T₃   π⁰ 135
                                               η   548

K⁻ ($\bar{u}$s) K⁰ ($\bar{d}$s)   −1   K⁰ 498
          −1   K⁻ 494

octet of vector mesons

$\overline{K^{*0}}$(d$\bar{s}$) Y K^{*+}(u$\bar{s}$)   1 $\overline{K^*}$ 890

ρ⁻ (d$\bar{u}$) ρ⁰ (q$\bar{q}$) ρ⁺ (u$\bar{d}$)   0   ρ   770
−1   ω   1 T₃   ω   783

K^{*-}($\bar{u}$s) K^{*0}($\bar{d}$s)   −1   K* 890
          −1

decuplet of baryons with spin 3/2

Δ⁻ (ddd) Δ⁰(ddu) Y Δ⁺ (uud) Δ⁺⁺(uuu)   0   Δ   1236

Σ^{*-} (dds) Σ^{*0} (uds) Σ^{*+} (uus)   −1   Σ* 1385
−1   1 T₃

Ξ^{*-}(ssd) Ξ^{*0} (ssu)   −2   Ξ* 1530
          −1

Ω (sss)   −3   Ω   1670

Figure 4.2: Basic SU(3) multiplets of baryons and mesons.

In the middle of sixties the question whether quarks exist in nature as individual free particles, i.e. in the same sense as, for instance, leptons, was open and serious effort were undertaken to find them. The fact that quarks should have fractional electric charges $\pm\frac{2}{3}, \pm\frac{1}{3}$, as well as fractional baryon numbers $\pm\frac{1}{3}$, made them look exotic, but there was on the other hand no obvious theoretical reason why these "exotics" should not exist in nature. The experimental search for free quarks went on until the late 1970, when it became increasingly clear that they do not exist in this way but are forever bound inside hadrons. The mechanism of this "quark confinement" will be discussed below in Section 4.6.

Quark composition of four basic unitary multiplets of hadrons is indicated in Fig. 4.2. In addition, there are many other, fully or partially, filled SU(3) multiplets. For meson octets, the states in the center are linear combinations of the three $q\overline{q}$ pairs: $u\overline{u}$, $d\overline{d}$, $s\overline{s}$. It is clear that the information on quark composition itself *doesn't* uniquely specify the corresponding hadron and that we need to know more about its quantum numbers and/or wave function to distinguish, for instance, ω from ρ^0 or proton from Δ^+.

From the relations (2.92) one sees why *three* quarks, and antiquarks, are needed to form the experimentally observed pattern of meson octets and baryon octet and decuplet. The natural question why the quarks don't form other possible combinations, like diquarks (qq pairs), 4q configurations etc., took about a decade to answer qualitatively and another decade to do so more quantitatively within the QCD. I shall address this question, together with the crucial feature of quark confinement, in the last section of this chapter.

A natural extension of the SU(3) flavor symmetry takes into account the fact that quarks are fermions with spin 1/2 and thus exist in 6 different states. The idea of extending the SU(3) flavor symmetry to SU(6) flavor-spin symmetry was originally suggested on the basis of the relations between masses in the baryon octet and decuplet. Fig. 4.2 shows that typical differences between isospin multiplets within both the baryon octet and decuplet ($\approx 150 - 200$ MeV) are about the same as the difference between the average masses of these SU(3) multiplets. Consequently assembling all the 56 baryonic states of different flavor-spin combinations ($4 \times 10 + 2 \times 8 = 56$) into one higher multiplet seemed justified. This idea got support from the observation that the *fully symmetric* multiplet $\mathbf{56}$=(3,0) of SU(6) has just the right number of states 56, and decomposes with respect to the unitary and spin subgroups SU(3) and SU(2) as follows:

$$\mathbf{56} = (\mathbf{10}, \mathbf{4}) \oplus (\mathbf{8}, \mathbf{2}), \tag{4.3}$$

where the first symbol in the brackets refers to SU(3) flavor and the second to SU(2) spin subgroups. The full decomposition of the direct product of three quark sextets reads [4]

$$\mathbf{6} \otimes \mathbf{6} \otimes \mathbf{6} = \mathbf{56}_s \oplus \mathbf{70}_{\mathrm{ms}} \oplus \mathbf{70}_{\mathrm{ma}} \oplus \mathbf{20}_{as}, \tag{4.4}$$

where, as in the case of the product (2.92) of triplets of SU(3) group, the subscripts "ms" and "ma" denote representations with particular symmetry under the permutation of first two sextets. Note the presence in the decomposition (4.4) of both the fully symmetric $\mathbf{56}_s$ and antisymmetric $\mathbf{20}_{as}$ multiplets. Analogously to (4.3) we find the following decompositions with respect to the product subgroup SU(3)⊗SU(2) of SU(6)

$$\mathbf{70} = (\mathbf{10}, \mathbf{2}) \oplus (\mathbf{8}, \mathbf{4}) \oplus (\mathbf{8}, \mathbf{2}) \oplus (\mathbf{1}, \mathbf{2}), \tag{4.5}$$

$$\mathbf{20} = (\mathbf{8}, \mathbf{2}) \oplus (\mathbf{1}, \mathbf{4}), \tag{4.6}$$

For mesons we get

$$\mathbf{6} \otimes \overline{\mathbf{6}} = \mathbf{35} \oplus \mathbf{1}, \tag{4.7}$$

where the decomposition with respect to SU(3)⊗SU(2) subroup reads

$$\mathbf{35} = (\mathbf{8}, \mathbf{1}) \oplus (\mathbf{8}, \mathbf{3}) \oplus (\mathbf{3}, \mathbf{1}). \tag{4.8}$$

Mathematically the hypothesis of SU(6) symmetry implies that the wave functions of all baryons populating the $\mathbf{56}$-plet, i.e. all those in Fig. 4.2, must be *fully symmetric* under any permutation of the constituent quarks. One way of constructing such wave functions is to combine the known properties of SU(3) (flavor) and SU(2) (spin) subgroups. It is straightforward to show that the following expression for the states of the baryon **octet**

$$\Phi(\mathbf{8}, \mathbf{2}) = \frac{1}{\sqrt{2}} \left(\Phi^{(8)}_{m,s} \Psi^{(2)}_{m,s} + \Phi^{(8)}_{m,a} \Psi^{(2)}_{m,a} \right) \tag{4.9}$$

---

[4]In the simper case of isospin dublet of (u, d) with spin we have $\mathbf{4} \otimes \mathbf{4} \otimes \mathbf{4} = \mathbf{20}_s \oplus \mathbf{20}_{\mathrm{ms}} \oplus \mathbf{20}_{\mathrm{ma}} \oplus \mathbf{4}_{as}$.

is indeed *fully* symmetric with respect to any permutation of the quarks in the direct product (4.4). For the multiplet (**10**,**4**) the fully symmetric wave functions are simply products of corresponding SU(3) (flavor) decuplet and SU(2) (spin) quartet wave functions, which each *separately* are symmetric. In Table (4.1) the wave functions of all baryons in the octet with spin 1/2 are listed.

In fact most of the expressions for the wave functions in Table 4.1 can be obtained exploiting the isospin and other two SU(2) subgroups of SU(3) and taking into account the following decomposition (2.34). To construct the fully symmetric quartet $\mathbf{4_s}$ as well as the remaining doublets of definite symmetry from the three fundamental doublets is a simple exercise in the Clebsh-Gordan coefficients (exercise 3.1). Note a difference between the proton wave function and that of the Δ^+ resonance although both these states have the same quark content *uud*. In the following we use the notation in which the third projection of the spin of a fermion (baryon or quark) is labeled by vertical arrow ($\uparrow$ for $\frac{1}{2}$ and $\downarrow$ for $-\frac{1}{2}$) to distinguish them from the third projection of the isospin.

	$\Phi^{(8)}_{(m,s)}$	$\Phi^{(8)}_{(m,a)}$
p	$\frac{1}{\sqrt{6}}[(ud+du)u-2uud]$	$\frac{1}{\sqrt{2}}[(ud-du)u]$
n	$\frac{1}{\sqrt{6}}[(ud+du)d-2ddu]$	$\frac{1}{\sqrt{2}}[(ud-du)d]$
Σ^+	$\frac{1}{\sqrt{6}}[(us+su)u-2uus]$	$\frac{1}{\sqrt{2}}[(us-su)u]$
Ξ^0	$\frac{1}{\sqrt{6}}[(us+su)u-2ssu]$	$\frac{1}{\sqrt{2}}[(us-su)s]$
Σ^-	$\frac{1}{\sqrt{6}}[(ds+sd)d-2dds]$	$\frac{1}{\sqrt{2}}[(ds-sd)d]$
Ξ^-	$\frac{1}{\sqrt{6}}[(ds+sd)s-2ssd]$	$\frac{1}{\sqrt{2}}[(ds-sd)s]$
Σ^0	$\frac{1}{2\sqrt{3}}[s(ud+du)+(dsu+usd)-2(du+ud)s]$	$\frac{1}{2}[(dsu+usd)-s(du+ud)]$
Λ	$\frac{1}{2}[(dsu-usd)+s(du-ud)]$	$\frac{1}{2\sqrt{3}}[s(du-ud)+(usd-dsu)-2(du-ud)s]$

Table 4.1: The flavor part of wave functions of the baryon octet.

The first six entries in the above table are simple generalizations of the nucleon (p,n) wave functions, using three distinct SU(2) subgroups of SU(3), based on the pairs of (u,d), (u,s) and (d,s) quarks respectively. The two states in the middle of the baryon octet require subtler arguments. They are both composed of the same combination (u,d,s) of quarks, but differ by SU(3) and SU(2) flavor quantum numbers:

- The wave function of Σ^0 can be obtained from Σ^+ by application of the lowering operator $E_{-(\alpha^1+\alpha^2)}$, corresponding to the isospin subgroup.

- the wave function of Λ is determined by the requirement that it has isospin 0, and must therefore be a linear combination of the following terms

$$\frac{(ud-du)s}{\sqrt{2}}; \quad \frac{s(du-ud)}{\sqrt{2}}; \quad \frac{dsu-usd}{\sqrt{2}}, \tag{4.10}$$

combined with the condition that it is *orthogonal* to the combination describing Σ^0 as well as to the SU(3) singlet

$$\frac{1}{\sqrt{6}}(uds-dus+dsu-sdu+sud-usd) \tag{4.11}$$

The explicit expression for the wave function of the proton with spin projection $\frac{1}{2}$, displaying both flavor and spin structure, reads

$$
\begin{aligned}
\mid p,\uparrow\rangle \;=\;& \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{6}}\mid(ud+du)u-2uud\rangle\frac{1}{\sqrt{6}}\mid(\uparrow\downarrow+\downarrow\uparrow)\uparrow-2\uparrow\uparrow\downarrow\rangle\right.\\
&\left.+\frac{1}{\sqrt{2}}\mid(ud-du)u\rangle\frac{1}{\sqrt{2}}\mid(\uparrow\downarrow-\downarrow\uparrow)\uparrow\rangle\right).
\end{aligned}
\tag{4.12}
$$

Note that the decompositions (4.3) and (4.6) imply that the baryon octet with spin 1/2 can be placed not only in fully symmetric 56-plet, but also into the fully antisymmetric 20-plet. It is straightforward to construct, for instance, fully antisymmetric wave function of the proton:

$$| p, \uparrow \rangle = \frac{1}{\sqrt{6}} \left(| uud \rangle | \uparrow \downarrow \uparrow \rangle - | uud \rangle | \downarrow \uparrow \uparrow \rangle - | udu \rangle | \uparrow \uparrow \downarrow \rangle + | udu \rangle | \downarrow \uparrow \uparrow \rangle + | duu \rangle | \uparrow \uparrow \downarrow \rangle - | duu \rangle | \uparrow \downarrow \uparrow \rangle \right) \quad (4.13)$$

and by interchanging $p \leftrightarrow n$ similarly for the neutron.

Any of these wave functions can now be used to calculate one of the basic static properties of baryons, their magnetic moments. Table 4.2 contains the present experimental situation for baryons. In the nonrelativistic quantum mechanics the operator of the magnetic moment of a *pointlike* fermion with spin 1/2, electric charge (in units of positron charge) Z and mass m is defined as

$$\hat{\vec{\mu}} \equiv \frac{Ze}{2m} \vec{\sigma} = \frac{Ze}{m} \vec{s} \quad (4.14)$$

where e is the positron charge in absolute units and $\vec{\sigma} \equiv (\sigma_1, \sigma_2, \sigma_3)$ is the vector of Pauli matrices. The magnetic moment of a given fermion with spin $\frac{1}{2}$ is then defined as the expectation value of the, say, third component of (4.14) in the state with the projection of the spin pointing in this third direction. For pointlike fermions this expectation value is given simply as $Ze/2m$, while for particles with internal structure the situation is more complicated, as is clear from an example of the neutron, which has nonvanishing, negative magnetic moment, despite the fact that it has zero total electric charge.

Assuming SU(6) symmetry of the baryon octet [5] and treating its members as pointlike fermions with spin 1/2 we get for the baryon B of the mass m_B [6]

$$\mu_B = |\vec{\mu}_B| \equiv \mu_N \frac{m_p}{m_B} \langle B, \uparrow | Q \sigma_3 | B, \uparrow \rangle \quad (4.15)$$

where $Q = T_3 + Y/2$, the operator of electric charge, is an element of flavor SU(3) subgroup of SU(6). Consequently the operator $Q \sigma_3$ is an element of SU(6) algebra and thus its matrix elements in the baryon octet states are *related*. As we know how the direct product $Q \sigma_3$ acts on any state from the direct product of the corresponding flavor and spin Hilbert spaces, it is straightforward to evaluate its matrix elements in the wave functions of Table 4.1:

$$\langle p, \uparrow | Q \sigma_3 | p, \uparrow \rangle = \frac{4Q_u - Q_d}{3} = 1, \quad (4.16)$$

$$\langle n, \uparrow | Q \sigma_3 | n, \uparrow \rangle = \frac{4Q_d - Q_u}{3} = -\frac{2}{3}, \quad (4.17)$$

$$\langle \Lambda, \uparrow | Q \sigma_3 | \Lambda, \uparrow \rangle = Q_s = -\frac{1}{3}, \quad (4.18)$$

where Q_u, Q_d, Q_s are electric charges of u, d and s quarks respectively. The first two numbers are in excellent agreement with experiment in the sense that they reproduce to remarkable accuracy the *ratio* of the measured magnetic moments of the proton and the neutron, while for the Λ hyperon the agreement is much worse. This is to be expected due to significant mass difference between the nucleon doublet and Λ. Note that if, on the other hand, baryon wave functions were given by the *ma* or *ms* parts of (4.9), the r.h.s. of (4.16-4.18) would equal Q_u, Q_d and Q_s, respectively. Assuming isospin symmetry, this would, in turn imply $\mu_n/\mu_p = -1/2$, in clear disagreement with experiment. So magnetic moments of baryons provide an independent strong argument for assigning the baryons into a fully symmetric **56**-plet of the SU(6) group.

In constituent quark model most of the SU(3) and SU(6) symmetry breaking is attributed to differences in quark masses. In this model not the observable baryons, but the constituent quarks u, d, s behave like a pointlike Dirac fermions, with masses m_u, m_d, m_s and magnetic moment operator $e_q \sigma_3 / 2m_q$. Isospin symmetry implies $m_u \doteq m_d \equiv m$, but as the SU(3) symmetry is violated in masses of hadrons at the level of a few hundreds of MeV, we expect similar difference between m_s and m. Under this assumptions magnetic moment of a baryon B (polarized in "third" direction) is then given simply as a sum of the magnetic moments of its constituent quarks:

$$\mu_B \equiv \sum_{u,d,s} \mu_q = \sum_{q=u,d,s} \langle B, \uparrow | \mu_3^q | B, \uparrow \rangle. \quad (4.19)$$

---

[5]This implies neglecting mass differences, poor approximation outside isospin multiplets.

[6]Magnetic moments of baryons are usually expressed in units of the so called **nuclear magneton**, $\mu_N \equiv e/2m_p$, where m_p is the proton mass.

Using the explicit expressions for baryon wave functions given in Table 3.1, the above formula allows us to express baryon magnetic moments in terms of those of u, d and s quarks:

$$\begin{array}{lll}
\mu_p = (4\mu_u - \mu_d)/3, & \mu_n = (4\mu_d - \mu_u)/3, & \mu_\Lambda = \mu_s, \\
\mu_{\Sigma^+} = (4\mu_u - \mu_s)/3, & \mu_{\Sigma^-} = (4\mu_d - \mu_s)/3, & \mu_{\Sigma^0} = (2\mu_u + 2\mu_d - \mu_s)/3, \\
\mu_{\Xi^0} = (4\mu_s - \mu_u)/3, & \mu_{\Xi^-} = (4\mu_s - \mu_d)/3, & \mu_{\Omega^-} = 3\mu_s.
\end{array} \tag{4.20}$$

Solving the first three equations for μ_u, μ_d, μ_s we get (in nuclear magnetons)

$$\mu_u = +1.852; \quad \mu_d = -0.972; \quad \mu_s = -0.613, \tag{4.21}$$

corresponding to $m_u = 338$ MeV, $m_d = 332$ MeV and $m_s = 510$ MeV. Quark masses obtained in this way are called **constituent** masses. In the next Chapter we shall encounter the so called **current** quark masses, which are substantially different from those in (4.21). Having determined magnetic moments of u, d and s quarks from those of proton, neutron and Λ, we can predict magnetic moments of five other baryons for which measurements are available. The results, displayed in Table 4.2, show a good, though not perfect, agreement with the data.

Note that assigning the baryon octet to the fully antisymmetric 20-plet would lead to gross disagreement with data. For instance, the wave function of the proton given in (4.13) would imply

$$\mu_p = \mu_d < 0, \quad \mu_n = \mu_u > 0, \quad \Rightarrow \frac{\mu_p}{\mu_n} = \frac{\mu_d}{\mu_u} = -\frac{1}{2}, \tag{4.22}$$

i.e. predicting thus both wrong signs and wrong magnitudes. Measurement of the magnetic moments of the baryon octet thus provides strong support for the assignment of baryons to the fully symmetric 56-plet of SU(6).

The mass difference $\Delta \equiv m_u - m_s \doteq 180$ MeV is consistent with mass differences between isotopic mutliplets containing different number of strange quarks. In the additive quark model the proton therefore looks like a nucleus, with most of its mass concentrated in the masses of its weakly bound **constituent quarks**, each having the mass of about 330 MeV.

4.3 Spin structure of the baryons

Given the baryon in a state with definite spin projection, say $| B, \uparrow \rangle$, it is natural to introduce the probabilities $P_q^B(\uparrow\uparrow)$, $P_q^B(\uparrow\downarrow)$ of finding in this state a quark q with the spin *parallel* or *antiparallel* to the spin of the baryon B. Using these probabilities we can then define for each quark q the quantities

$$\Delta^B(q) \equiv P_q^B(\uparrow\uparrow) - P_q^B(\uparrow\downarrow), \tag{4.23}$$

which have a clear physical interpretation as the fraction of the spin of the baryon B carried by the constituent quark q. The normalization of the probabilities P_q^B is such that

$$P_q^B(\uparrow\uparrow) + P_q^B(\uparrow\downarrow) = N^B(q), \tag{4.24}$$

where $N^B(q)$ is the number of constituent quarks of the flavor q in the baryon B. The fact that the spins of all quarks q add up to the spin of the baryon B moreover implies

$$\sum_q \Delta^B(q) = 1. \tag{4.25}$$

p	n	Λ	Σ^+	Σ^-	Ξ^0	Ξ^-	Ω
2.793	-1.913	-0.613±0.004	2.458±0.010	-1.160±0.025	-1.25±0.014	-0.651±0.003	-2.02±0.05
input	input	input	2.674	-1.092	-1.435	-0.494	-1.839

Table 4.2: Current status of experimental determination (first row) of and theoretical predictions (second row) for baryon magnetic moments in units of μ_N. Magnetic moments of the proton and neutron are known with large accuracy ($2.10^{-9}\%$ and $2.10^{-8}\%$ respectively).

The evaluation of the probabilities P_q^B is a simple exercise using the explicit expressions of the baryon wave functions constructed above. For instance for the experimentally most interesting case of the proton we find

$$P_u^p(\uparrow\uparrow) = \frac{5}{3}, \; P_u^p(\uparrow\downarrow) = \frac{1}{3}, \; P_d^p(\uparrow\uparrow) = \frac{1}{3}, \; P_d^p(\uparrow\downarrow) = \frac{2}{3} \Rightarrow \Delta^p(u) = \frac{4}{3}, \; \Delta^p(d) = -\frac{1}{3}, \qquad (4.26)$$

i.e. the up and down quarks carry the entire spin of the proton.

The spin structure of the baryons, and in particular the nucleon, as revealed in the deep inelastic scattering, which will be discussed in the next Chapter on the **parton model**, is currently one of the most interesting open problems in particle physics. Although the constituent quarks of the naive additive quark model are not sufficient to account for the rather surprising results, we shall see that they nevertheless play an important role in the understanding of this phenomenon.

4.4 The Zweig rule

As already mentioned above, the Zweig rule is a phenomenological rule which explains remarkable regularities in cross sections and decay rates in terms of the constituent quarks. The basic idea of this rule is depicted in Fig. 4.3 which shows the so called "quark flow" diagrams for several decay channels of vector mesons ρ^0, ϕ as well as for the process

$$\pi^+ + p \rightarrow \pi^+ + \pi^0 + p. \qquad (4.27)$$

The solid lines in Fig. 4.3 describe the "flow" of quarks with definite flavour from the initial to final states or back. All the processes can be divided into two classes

- **Zweig allowed** ones: those for which the *planar* [7] flow diagrams are **connected**, i.e. the diagram cannot be separated into two parts without cutting some quark lines, like in Fig. 4.3a,b,d.

- **Zweig forbidden** ones: those for which the diagrams are **disconnected**, like in Fig. 4.3c.

The Zweig rule then says that the Zweig forbidden processes are *suppressed* with respect to the Zweig allowed ones. The degree of this suppression is not unique and depends on many other aspects of the process in question. The best illustration of this rule is the decay of the Φ meson into its various channels. On the basis of the Zweig rule and Fig. 4.3 we conclude that from the two channels

$$\phi \rightarrow K^+ K^-; \quad \phi \rightarrow \rho^+ \pi^- \qquad (4.28)$$

the second should be suppressed with respect to the first one. And indeed, despite that fact that the $K^+ K^-$ channel is barely open, with only about 30 MeV left for kinetic energy of the decay products, it still dominates the $\rho\pi$ channel (with about 130 MeV left) by a factor of 40! The application of the Zweig rule to the decay of the J/ψ meson played a crucial role in the discovery of the **charmed** quark (see the next Section). As we shall see later, the "returning" lines in the quark diagrams like that in Fig. 4.3c can be interpreted as those in which some of the quarks *annihilate* to gluons.

4.5 Problems and puzzles of the Quark Model

The original quark model as proposed by Zweig and Gell-Mann was remarkably successful in explaining the observed hadrons in terms of basic constituents and in particular in calculating baryon magnetic moments, but it had also several serious problems:

- The baryons were assigned to the *fully symmetric* **56**-plet of SU(6), despite the fact that quarks are **fermions** and their wave functions should therefore be *antisymmetric* with respect to the interchange of *all* their characteristics (space-time as well as internal).

- Quarks had not been observed in nature as free particles. Neither had there been any indications for the existence of other "exotic" states, like those of the **sextet**, which would be populated by symmetric diquark combinations.

---

[7] The term "planar" means that the diagrams are drawn in the plane and we furthermore require that there are *no crossings* of quark lines. The reason for that provision is that by allowing crossings we could make any disconnected diagram connected simply by crossing some quark lines there and back, as shown in Fig. 4.3e.
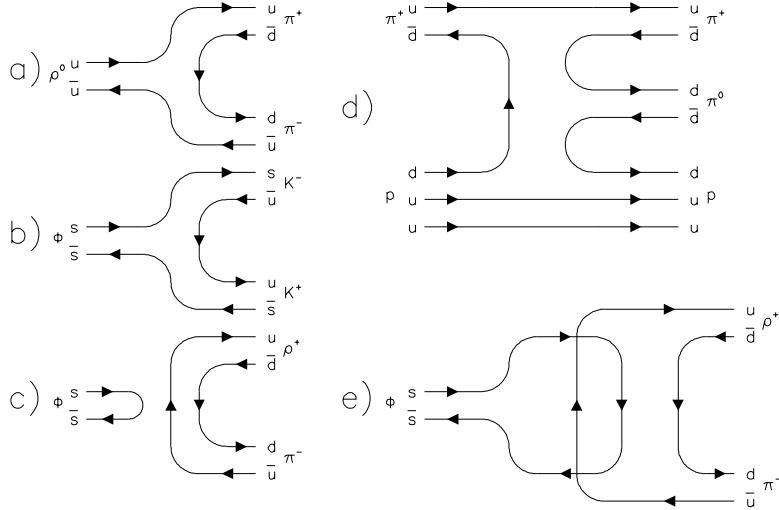
Figure 4.3: Examples of the Zweig rule: the decay of ϕ meson and pion-nucleon scattering.

- There are also no signs of the existence of particles, which would correspond to states like $2q2\overline{q}$, or $4q\overline{q}$, formed by combinations of "allowed" states $q\overline{q}$ and $3q$.

There were several suggestions how to solve the first, most pressing problem - that of quark statistics.

1. If quarks do not exist as free particles, the problem can simply be ignored. This was the attitude of Gell-Mann.

2. The spatial parts of the ground state wave functions could be antisymmetric. Although in principle possible, no realistic model of forces leading to such wave functions has been constructed.

3. In another attempt at the conventional explanation Sakita [67] suggested to assign the baryon octet to the fully antisymmetric 20-plet of SU(6), as in (4.13). This solves the statistics problem for the octet of baryons with spin 1/2, but not for the decuplet of baryons with spin 3/2. As shown in (4.22) it, moreover, leads to completely wrong values of baryon magnetic moments.

4. Very unconventional solution of the statistics problem had been proposed by O. Greenberg [68] who assumed that quarks are not fermions, but the so called *parafermions* of rank three. This solved the statistics problem because there can be at most three such parafermions in each state.

It became soon evident that the idea of quark parastatistics is similar, though not quite equivalent, to assuming that each quark flavour exists in **three** different **color** states [8] and the observed hadrons correspond to **white**, i.e. **colorless**, combinations. In the language of group theory quarks transform like the fundamental triplet of a *new* SU(3) group, called **color** SU(3) and denoted for distinction $\mathrm{SU}_c(3)$. The observed hadrons are postulated to be **color singlets**. The wave functions of all baryons are thus of the form

$$| \mathrm{baryon}^{\alpha\beta\gamma}\rangle = \underbrace{\epsilon^{ijk} \,| \, q_i^{\alpha}\rangle \, | \, q_j^{\beta}\rangle \, | \, q_k^{\gamma}\rangle}_{\text{totally antisymmetric}}, \qquad (4.29)$$

where $i, j, k = 1, 2, 3$ are the *color* indices while α, β, γ specify the *flavour*. As an example let us write down explicitly the wave function for the Δ^{++} baryon:

$$\begin{aligned} | \Delta^{++}\rangle \;=\;& | \, u_1\rangle \, | \, u_2\rangle \, | \, u_3\rangle - | \, u_2\rangle \, | \, u_1\rangle \, | \, u_3\rangle + | \, u_2\rangle \, | \, u_3\rangle \, | \, u_1\rangle - \\ & | \, u_3\rangle \, | \, u_2\rangle \, | \, u_1\rangle + | \, u_3\rangle \, | \, u_1\rangle \, | \, u_2\rangle - | \, u_1\rangle \, | \, u_3\rangle \, | \, u_2\rangle. \end{aligned} \qquad (4.30)$$

---

[8]Parastatistics corresponds to SO(3) group, rather than SU(3) as does color. The invention of concept of parastatistic and its use in solving puzzles of the quark model, as well as the whole way leading to the concept of color as dynamical quantity is described in an excellent reminiscence [69]. It was Abraham Pais, who first used the terms "green", "red" and "yellow" to denote the three different states of a given quark flavor [70].

According to (2.92) to make color singlets for baryons, we need at least as many colors as there are quarks from which they are composed. No diquark is therefore observable, but in principle the color singlet hypothesis doesn't by itself rule out the existence of states like $4q\overline{q}$ mentioned above. For mesons there is no such limitation, as for any number of colors the direct product of the fundamental representation and its complex conjugate contains the singlet **1**. Although the hypothesis of hadrons as color singlets proved extremely fruitful, it had not truly solved the puzzles of the quark model, but merely recast them into another question: **why are there only color singlet states in the nature?**

The crucial step towards answering this question was taken by Yochiro Nambu shortly after the appearance of Greenberg's paper [68]. In the contributions [71, 72] at two conferences in early 1965, Nambu put forward the basic framework in which quark confinement follows from dynamics of what he called "super-strong interactions" between the "fundamental objects" [9], i.e. quarks. The really fundamental ingredient of his model, which was phrased in terms of nonrelativistic approximation of the corresponding field theory, was the assumption that the "superstrong force" acting between quarks is (see [72], p. 138)

> .... mediated by an octet of gauge fields $G_\mu, \mu = 1, ..., 8$, coupled to the infinitesimal $SU(3)''$ generators (currents) λ''_μ of the triplets, with the strength g. (The $SU(3)''$ denotes the color $SU(3)$ in current notation, remark by J.Ch.)

By introducing 8 gauge fields, Nambu had fully exploited the nonabelian nature of the underlying color $SU(3)$ symmetry of quarks! From the point of view of solving the problem of quark statistics this, however, was not necessary, as one could do with just one gauge field. Note that the above statement represents the explicit formulation of the essence the present day QCD! For Nambu his superstrong force was responsible for quark confinement and the exchange 8 gauge fields (i.e. gluons in QCD) was motivated by the fact that *"For a system containing altogether N particles, the exchange of such fields between a pair then results in an interaction energy"* that allowed him to explain why only color singlet states may exist in

	u	d	s
Q_1	1	0	0
Q_2	1	0	0
Q_3	0	-1	-1

Table 4.3: Charge assignment in Hahn-Nambu model.

nature. In the next Section I will elaborate on Nambu's model using the modern notation and terminology as employed in [76]. Let me stress again that by assuming that the force acting between quarks is mediated by 8 gauge bosons of color $SU(3)$ symmetry, which themselves carry the color, Nambu had clearly laid down the foundations of QCD 8 years before its definite formulation as quantum field theory. Unfortunately, more than with the introduction of nonabelian nature of strong force, Nambu's name is connected with the so called Hahn-Nambu model of colored quarks with integral electric charges. In this model [75] quarks with a given flavor but different colors were assigned different integer electric charges in such a way that their color-averaged values were equal to fractional electric charges of the Gell-Mann-Zweig quark model. The Hahn-Nambu model differed from the latter in several aspects (for instance, some of the gauge fields carried electric charge) but was identical to it as far strong interactions were concerned. Consequently, the fact that it has since been ruled out by experiments on deep inelastic scattering of leptons on nucleons and other processes to be described in the next Chapter, does nor change the fact that Nambu's model contains the very fundamental ingredient of QCD!

4.6   Color to the rescue: Quasinuclear colored model of hadrons

The basic idea of a mechanism of color confinement has been introduced first by Nambu in the framework of nonrelativistic quantum mechanics [72]. In his model the interaction potential follows from the basic assumption that the interaction is mediated by the exchange of the octet of colored gauge bosons and has, as we shall see, the following properties:

- Quarks as individual "particles" are *infinitely* heavy and thus not observable.

- The forces acting between quarks are *attractive* in color *singlet* channels, resulting in bound systems of *finite* mass.

- In all other channels the forces are repelling and the systems thus infinitely heavy and unobservable.

- The force $F_{q\overline{q}}$ between a $q\overline{q}$ pair in a meson is *twice* bigger than the force F_{qq} acting between each of three pairs of quarks inside any baryon.

---

[9]It seems that Nambu was not aware of Greenberg's paper. Neither did he use the words "quark" or "ace" for members of his fundamental triplets. I will, however, use the former term.

In this model the binding energy thus *cancels* the infinite masses of constituent quarks in color singlets, as we want, but leaves the other states too heavy to be observable [10].

The interquark potential of Nambu is a straightforward extension of a typical spin-spin interaction, where the forces between any system of fermions and antifermioms are described by two-body potentials $V_{ij}(r)$, acting between all pairs i, j of quarks, of the form

$$V_{ij} = v(\vec{r}_{ij})\vec{s}_i\vec{s}_j, \tag{4.31}$$

where $\vec{s}_i$ are the usual spin matrices of SU(2) and the functions $v(\vec{r}_{ij})$ contain the dependence on space coordinates as well as all other quantum numbers except the spin. The full potential energy of a system of n quarks and antiquarks is then given as

$$V(n) = \sum_{i<j}\langle n \mid V_{ij} \mid n \rangle = \frac{v}{2}\langle n \mid \left[\left(\sum_{i=1}^{n}\vec{s}_i\right)^2 - \sum_{i}^{n}\vec{s}_i^2\right] \mid n \rangle = \frac{v}{2}\left[s(s+1) - \frac{3}{4}n\right], \tag{4.32}$$

where $\mid n \rangle$ denotes the state vector of the system, $v \equiv \langle \mid v_{ij}(r_{ij}) \mid \rangle$ is the mean value of the interquark potential taken in the space coordinates, s is total spin of the system and $3/4 = (1/2)(1/2 + 1)$ is just the square of the spin of each individual quark. Analogously, introducing the *color* interaction in the form (λ_i are the familiar Gell-Mann matrices)

$$\frac{1}{8}\sum_{i\neq j}^{n} v(\vec{r}_{ij})\vec{\lambda}_i\vec{\lambda}_j \tag{4.33}$$

we get for the full potential energy of such a system

$$V(n) = \frac{v}{2}(C - nc), \tag{4.34}$$

where C and c are the eigenvalues of the **Casimir operator** [11]

$$\mathbf{C} \equiv \sum_{a=1}^{8} T^a T^a \tag{4.35}$$

	SU(3)	**C**	**C-nc**	**C/c**
q	triplet	4/3	0	1
$\overline{q}$	antitriplet	4/3	0	1
qq	antitriplet	4/3	-4/3	1
qq	sextet	10/3	2/3	5/2
$q\overline{q}$	singlet	0	-8/3	0
$q\overline{q}$	octet	3	1/3	9/4

Table 4.4: Values of the quadratic Casimir operator and related quantities.

in a given multiplet (p, q):

$$C(p,q) \equiv \langle n \mid \mathbf{C} \mid n \rangle = \langle T_3^2 \rangle + 2\langle T_3 \rangle + \frac{3}{4}\langle Y \rangle^2 = \frac{1}{3}(p^2 + pq + q^2) + (p+q). \tag{4.36}$$

Table 4.4 lists the values of $C(p,q)$ for several important multiplets of SU(3). Expressing the total mass of such a system as the sum over the masses M_q of the constituent quarks (for simplicity assumed all the same) and potential energy (4.34)

$$M(n) = nM_q + V(n) = n\left(M_q - \frac{cv}{2}\right) + \frac{Cv}{2} \tag{4.37}$$

the infinite quark masses M_q are seen to cancel by potential energy $V(n)$ provided $v = 2M_q/c$ and

$$M(n) = C\frac{v}{2} = \frac{C}{c}M_q. \tag{4.38}$$

Taking into account the dependence of $v(r_{ij})$ on its arguments but still assuming that it is the same for all pairs i, j leads to the same results, except for the fact that v in (4.38) and other equations is now the mean value of $v(r)$ in the wave function of the corresponding color multiplet. The resulting formula (4.38) has all the properties we need:

- In color singlet states $C = 0$ implying vanishing total mass [12] of the system, which makes it *observable.*

---

[10]This model doesn't in fact rely on actual infiniteness of free quark mass M_q. What is essential is the fact that M_q must be *large* compared to currently accessible energies so that free quarks cannot be observed even if the confinement were not exact.

[11]Casimir operators are quadratic, cubic or higher order forms constructed from the generators of a given Lie algebra, which *commute* with all its elements. Themselves, they are *not* elements of this algebra, but we can work out their commutators with any element of the Lie algebra using the commutators of generators. Because of their property, Casimir operators are **invariants** of any multiplet. The full set of these invariants can be used as another way of specifying the multiplets. In the case of SU(3) there are *two* Casimirs, (4.35) and a cubic one, which, however, is much more complicated and has no simple use in the quark model.

[12]The potential doesn't have to cancel the whole mass M_q. What is important is that the remaining uncanceled part is *finite.*

- In all nonsinglet channels, on the other hand, $C \neq 0$ and thus the system has a mass proportional to M_q. This makes it as unobservable as the quarks themselves.

- The forces between $q\overline{q}$ and qq pairs in mesons and baryons are in the relation: $F_{q\overline{q}} = 2F_{qq}$.

This simple model demonstrates the mechanism of color confinement in nonrelativistic quantum mechanics. It is much more difficult to show that it naturally follows from interaction of colored quarks in the relativistic quantum field theory as well. There are indications, based on numerical calculations in lattice gauge theory, that Quantum Chromodynamics, does, indeed, have this property.

4.7 The arrival of charm

Shortly after Zweig and Gell-Mann formulated their quark model, in which three quark flavors u, d and s were the basic entities, theorists started to speculate about the possible existence of the fourth, named *charm* by Bjorken and Glashow [77].

The original argument in favor of the fourth quark was based on the quest for symmetry between quarks and leptons. Recall, that in 1962 the group led by Lederman, Schwartz and Steinberger found compelling evidence for the existence of the muon neutrino ν_μ. This suggested that there might be the fourth quark, that would complete the second generation of the fundamental fermions. Further, and actually much more urgent, reasons for its existence came in the early seventies from attempts to solve two serious problems in the theory of weak interactions (see [19] for details):

- Strong suppression of the so called *flavor changing neutral currents* (FCNC), i.e. processes like $K^+ \to \pi^+e^+e^-$ (Br=2.7 10^{-7}) with respect to usual charge current process $K^+ \to \pi^0e^+\nu_e$ (Br=0.048).

- The problem with the so called axial anomalies.

Both of the above problems are naturally resolved by assuming the existence of the fourth quark with electric charge 2/3 and $I_3 = 1/2$ which forms an isospin doublet with the strange quark. The mass m_c of the charm quark must not be too large if its contributions to the FCNC processes should solve the first problem, which gave the upper bound on m_c of about 2 GeV. From the theoretical point of view charm was thus badly needed. In fact is was so badly needed that Sheldon Glashow (1979 Nobel Laureate) ended his talk at the 1974 Conference on Experimental Meson Spectroscopy, just three months before its discovery, with the following challenge to the present experimentalists:

> *There are just three possibilities:*
>
> *1. Charm is not found and I eat my hat.*
>
> *2. Charm is found by hadron spectroscopers, and we celebrate.*
>
> *3. Charm is found by outlanders, and you eat your hats.*

As we shall see, his hat was spared. Turning now to the experimental side, it is worthwhile to recall that the first glance of the charm quark was very probably seen already in 1970 in an experiment carried out at the Brookhaven 30 GeV proton synchrotron [78, 79]. In this experiment Lederman [13] and collaborators studied the distribution of the invariant mass of $\mu^+\mu^-$ pairs produced in collisions of protons with uranium

$$p + U \to \mu^+\mu^- + \text{anything}, \tag{4.39}$$

Only the $\mu^+\mu^-$ pairs were observed as all hadrons were absorbed in the heavy target. This so called "beam dump" arrangement had its advantages, primarily the simplicity of the detector, but lead to poor mass resolution, of about $0.1m_{\mu\mu}$. The results of their measurement at the primary energy of 29.5 GeV is reproduced in the upper right plot of Fig. 4.4 and compared to the dashed curve representing smooth background assumed to be proportional to $1/m_{\mu\mu}^5$. Clear shoulder in the distribution of $m_{\mu\mu}$ is seen in the region $3-5$ GeV, but poor mass resolution prevented mode detailed study. Similar shoulder was observed for other energies between 22 and 30 GeV, suggesting the effect is real. Interestingly, and perhaps surprisingly in view of the urgent need for charm from theoretical side, the authors conclude in the first letter [78] that their

---

[13]Who was to receive, with Schwartz and Steinberger, the 1988 Nobel prize for Physics for the above mentioned discovery of the muon neutrino.
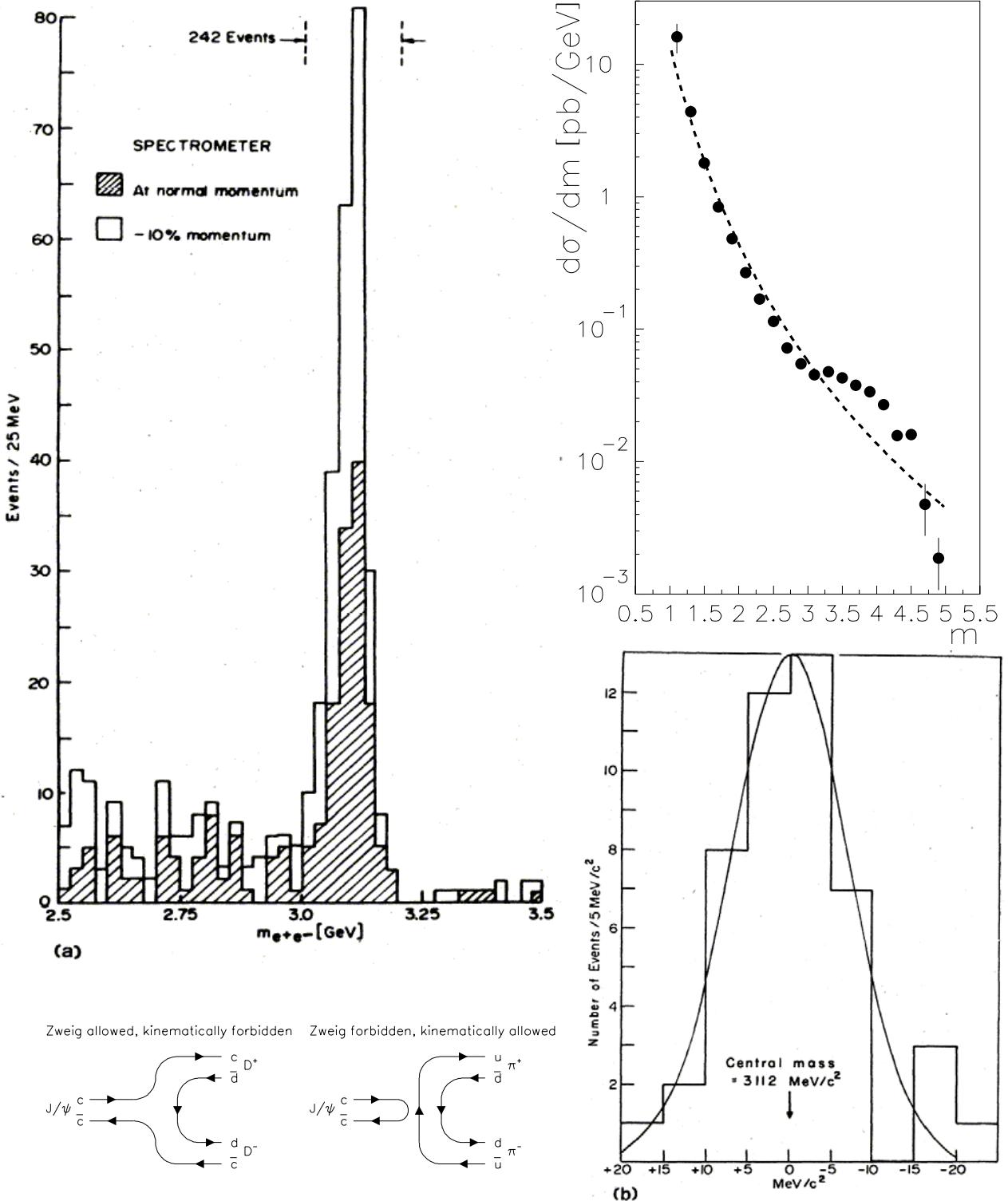
Figure 4.4: The distribution of the invariant mass $m_{e^+e^-}$ showing the peak corresponding to J/ψ particle (upper left) [73]; analogous distribution as measured in ref. [78] in 1970 (upper right), detailed shape of the J/ψ peak (lower right) and the Zweig rule applied to decay of J/Ψ (lower left).

data "exhibit no resonant structure" and in their 1973 comprehensive paper [79] that the distinct excess of the observed cross section over the "reference curve" may be due to the production of a "resolution-broadened resonance" but it may also be interpreted as "merely a departure from the overly simplistic and arbitrarily normalized $1/m^5$ dependence." In fact Lederman and collaborators were searching for the so called "heavy photons" which were expected in some models and which were to decay into lepton pairs. The J/ψ is, of course, just such a "heavy photon" but experimentalists expected a broad resonance.

The search for "heavy photons" was pursued also by the group at Brookaven lead by Samuel Ting. Ting, who had previously done experiments at DESY, had an extensive experience with identification of electrons, which was crucial for his aim. At Brookhaven he basically repeated the investigation done in [78, 79] but with a superior detector and using Berylium as target. His detector, a magnetic double arm spectrometer with Cherenkov counters for reliable identification of electrons and positrons, had mass resolution of about 5 MeV and was ready in late 1974. At about the same time first results from the new SPEAR electron-positron collider at SLAC, commissioned in 1973, appeared, showing the rise of the total cross section for e^+e^- annihilation to hadrons (scaled by that of e^+e^- annihilation to $\mu^+\mu^-$ pairs) in the region up to $\sqrt{s} = 6$ GeV. Burton Richter, who lead the SLAC-LBL team, summarized the situation at the 1974 International Conference on High Energy Physics, held in London in July, in a figure reproduced in Fig. 4.5. The rise was significant, but, as we know now, the $c\bar{c}$ resonances fell through because of their very narrow width, which requires precise tuning of the primary energy to hit upon it.

No such fine tuning is needed if the narrow resonances are looked for via the e^+e^- or $\mu^+\mu^-$ pairs produced in hadron-hadron collisions. Knowing the results of [78,79] Ting could concentrate on the interesting region of m_{ee} between 3 and 5 GeV and got fast the first results [73] shown in Fig. 4.4. The spectrum showed a clear evidence for a narrow resonance at 3.1 GeV, which Ting called J, and which is now know as J/ψ. Its observed width was compatible with experimental resolution, which implied that its true width had to be much smaller.

On the other hand, once you know where the narrow resonance is the electron-positron annihilation into hadrons is much better place where to study its mass and other properties. Shortly after the Summer conference in 1974, the group lead by Richter found by a careful scan of the whole accessible region a peak in total cross section $\sigma(e^+e^- \to \text{hadrons})$ at 3.097 GeV, which they christened ψ [74]. Zooming in on the nearby region, the SLAC-LBL group found soon also its first recurrence, called ψ', at 3.68 GeV (see upper plots in Fig. 4.6) and later the whole spectrum of states.

Figure 4.5: Early data on $\sigma(e^+e^- \to had)$.

The interpretation of J/ψ (and ψ') as $c\bar{c}$ resonances was based on the remarkable fact that this heavy vector meson decays into hadrons via *strong* interactions about 1000 times slower than similar hadronic resonances like ρ etc. The unusually small width of only 70 keV (now this number stands at 87 keV) was explained by the Zweig rule as a consequence of the fact that the D–mesons (not yet observed at the time of J/ψ discovery) are too heavy for J/ψ to decay via Zweig *allowed* channel into their pairs, and thus all decays have to go via Zweig *forbidden* ones, as illustrated in the lower left part of Fig. 4.4. Although the charm hypothesis was from the very beginning high on the list of possible explanations of the observed J/ψ and ψ', its confirmation had to wait until the early 1976, when the mesons carrying *open* charm were found at SLAC in reanalysis of older data (see [74] for details). These mesons, denoted now as D, D^*, D_s, D_s^*, are bound states of c (or $\bar{c}$) with light quarks u, d and s or their antiquarks. From the point of view of this interpretation it was crucial that the charmed mesons $D^\pm$ were found (see lower right plot in Fig. 4.6) to decay as $D^\pm \to K^-\pi^+\pi^+$ but not in a mode expected for conventional strange resonances, i.e. $D^\pm \to K^+\pi^+\pi^-$.

Interestingly the first charmed baryon was found even before the charmed mesons in neutrino-proton collisions at Brookhaven in early 1975 [80]. One particularly clear photograph from hydrogen bubble chamber, reminiscent of the discovery of Ω^-, in which three positive and two negative particles were accompanied by

Figure 4.6: Energy dependence of the total cross section of e^+e^- annihilation to hadrons with peaks corresponding to J/ψ and ψ' (left) and distributions of invariant masses of $K^\pm K^0$ and $K^\pm \pi^\mp \pi^\mp$.

the unambiguous Λ, was convincingly interpreted as the process

$$\nu_\mu p \to \mu^- \Lambda \pi^+ \pi^+ \pi^+ \pi^- \tag{4.40}$$

which violates the sacred $\Delta S = \Delta Q$ rule of standard weak interactions. On the other hand such final states are expected if the primary process $\nu_\mu + d \to \mu^- + c$ or $\nu_\mu + s \to \mu^- + c$ creates the charm quark c which then hadronizes into charm mesons D or baryons Λ_c. The point is that as these hadrons contain the charm quark c, their weak decay proceeds as $c \to s + W^+$ and must thus contain strange mesons K^- or $\overline{K}^0$ or baryons Λ in the final state. With the discovery of the fourth quark the SU(3) flavor symmetry was extended into
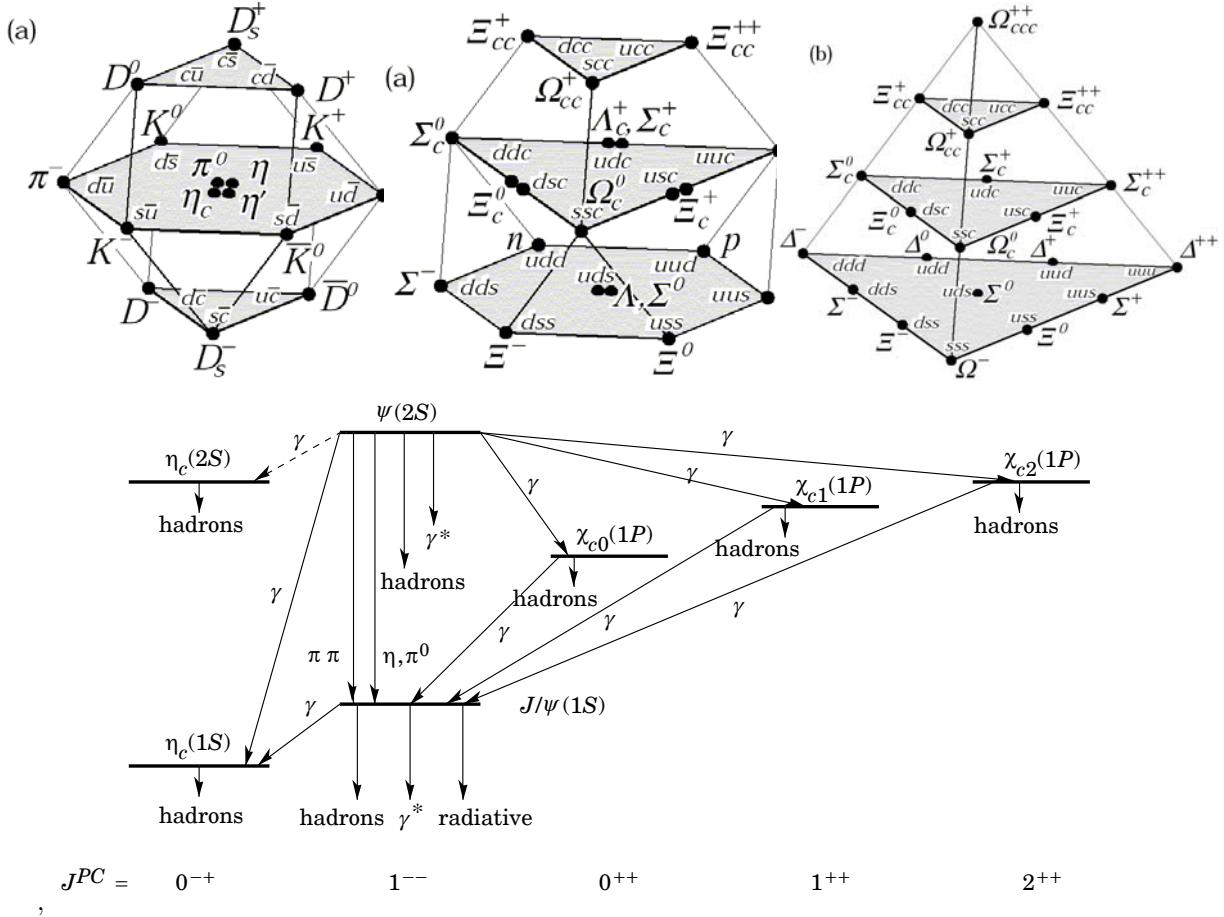
Figure 4.7: The basic SU(4) multiplets of mesons and baryons in the quark model (upper part) and the spectrum of the lowest lying $c\bar{c}$ bound states (lower part).

SU(4) one. There are many similarities between SU(3) and SU(4) groups, but there are several important differences as well:

- As SU(4) algebra has rank 3, the Cartan subalgebra has *three* elements and there are thus *three* mutually commuting generators. Their eigenvalues may again be used to label the states of a given multiplet. Besides T_3 and Y, the generator, denoted C, corresponding to the conserved charm quantum number is usually selected for this purpose.

- There are three simple roots and consequently *three* fundamental representations. One of them is the **defining** representation, **4**, formed by the set of 4×4 matrices satisfying the appropriate commutation relations and describing the transformations of the basic quark quartet. Its complex conjugate representation, $\overline{\mathbf{4}}$, transforms the four antiquarks. Contrary to the case of SU(3) there is, however, *another* fundamental representation, which has no analogy in SU(3) group. This third fundamental quartet appears, for instance, in direct product of three quartets (see below) and is labeled $\mathbf{4}_{as}$ as it

corresponds to fully antisymmetric combination of *three* quarks with *four* flavors. In SU(3) such an antisymmetric combination forms a *singlet*. The relation (3.9) generalizes to

$$Q = T_3 + \frac{Y + C}{2}. \tag{4.41}$$

- The most important direct products, describing systems of three quarks or q$\overline{\text{q}}$ pair read

$$\mathbf{4} \otimes \mathbf{4} \otimes \mathbf{4} = \mathbf{20}_s \oplus \mathbf{20}_{m,s} \oplus \mathbf{20}_{m,a} \oplus \mathbf{4}_{as}, \tag{4.42}$$

$$\mathbf{4} \otimes \mathbf{\overline{4}} = \mathbf{15} \oplus \mathbf{1}. \tag{4.43}$$

- The SU(4) multiplets can be graphically represented in three dimensional space by polyhedrons. According to (4.43) mesons form 15-plets, while baryons fill two kinds of 20-plets, a fully symmetric 20-plet $\mathbf{20}_s$ which contains the SU(3) baryon octet with spin 1/2 and another 20-plet, $\mathbf{20}_{m,s}$ with mixed symmetry, containing the SU(3) decuplet and corresponding to spin 3/2, see Fig. 4.7.

The large mass splittings between charmed and noncharmed mesons and baryons indicate that except for the part of SU(4) symmetry responsible for the conservation of the charm quantum number in strong interactions, the rest of this symmetry is broken even more strongly than the SU(3) symmetry.

The discovery of charmed quark has brought compelling evidence for the reality of quarks. The investigation of the properties of $c\overline{c}$ bound states, summarized in Fig 4.7, has shown that these states can be described to a good approximation by means of nonrelativistic quantum mechanics using the interquark potential $V(r)$ of the form [15]

$$V(r) \equiv -\frac{4}{3}\frac{\alpha_s}{r} + \kappa r, \tag{4.44}$$

where α_s is the strong interaction coupling, analogous to α in Quantum Electrodynamics (see Chapter 7 for definition). The first part of (4.44), dominating at short distances, is motivated by perturbative QCD, whereas the second describes in a phenomenological manner the confinement.

It is quite clear that the heavy quarks behave in the way envisioned by Zweig. The discovery of charm with its rich spectrum of states that can easily be understood within nonrelativistic quantum mechanics has thus definitely buried the approach advocated by Gell-Mann. Without treating quarks in a way reminiscent of nucleons inside nuclei, we would never be able to arrive at the predictions that agree so much with the experimental data!

4.8   To be or not to be?

For a few months after the discovery of J/ψ, but still before the confirmation of its nature by observation of charmed mesons, there seemed to be just two complete generations of fundamental fermions. However, already during the next year, 1975, evidence of new physics began to emerge from analysis of unlike sign dimuon events observed in e$^+$e$^-$ annihilations at SLAC by a group led by Martin Perl. After about a year of detailed scrutiny of their properties and estimates of other possible sources, these events were interpreted as coming from the production of pairs of heavy leptons $\tau^\pm$, with mass of about 1.8 GeV, followed by their decays satisfying the universality of weak interactions:

$$\text{e}^+\text{e}^- \to \tau^+\tau^-, \quad \tau^- \to \nu_\tau \mu^- \overline{\nu}_\mu \text{ or } \tau^- \to \nu_\tau \text{e}^- \overline{\nu}_e, \quad \tau^+ \to \overline{\nu}_\tau \mu^+ \nu_\mu \text{ or } \tau^+ \to \nu_\tau \text{e}^+ \nu_e, \tag{4.45}$$

where ν_τ is the conjectured neutrino associated with the charged $\tau^\pm$. Experimentally the τ-lepton had been identified by the presence in the final state of combinations $\mu^+\text{e}^-$ or $\mu^-\text{e}^+$. The existence and identity of this third neutrino has finally been established only two years ago. The discovery of the τ-lepton opened the gates to the third generation and led to searches for its conjectured quark members.

The story of this search is short, but as in the case of charm, quite illuminating though it took different twists and turns. The stage of the drama was the 400 GeV proton synchrotron at Fermilab and the main protagonist again Leon Lederman. His team looked again for dilepton pairs produced in proton-Berylium collisions using double arm spectrometer with magnetic field for momentum measurements. The resulting dilepton mass resolution was 2%, much worse than in Ting's experiment at Brookhaven. Within the period of 18 months between January 1976 and September 1977 the group published 4 papers [81–84] with quite
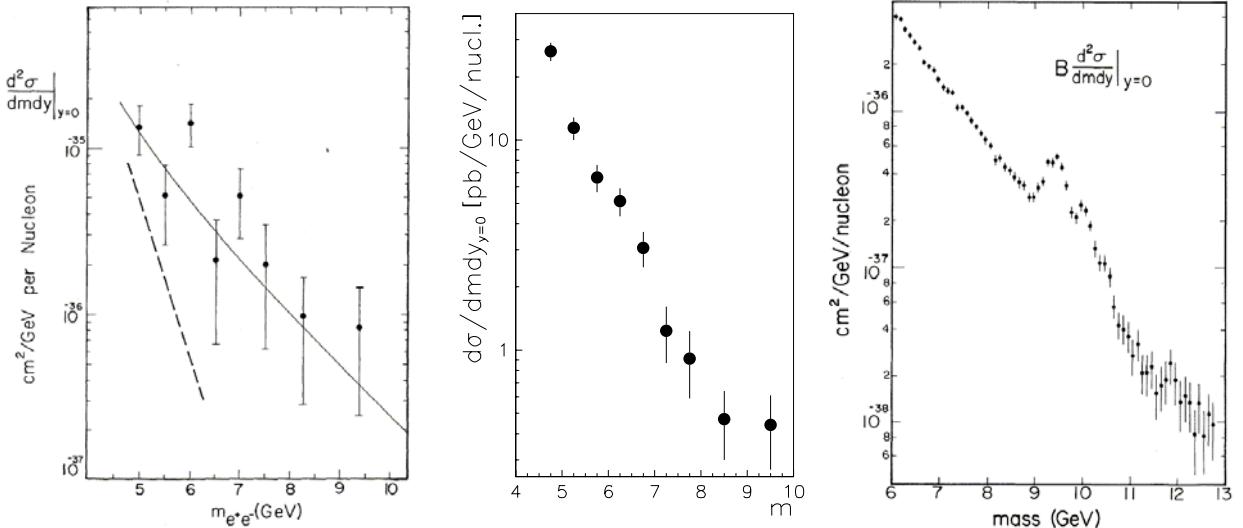
Figure 4.8: Twists and turns of the discovery of the bottom quark. Evidence of "a new resonance at 6 GeV" [81] observed in the distribution of invariant mass of e^+e^- pairs (left), but shortly afterwards ruled out by the data [82] on $\mu^+\mu^-$ pairs (middle), and replaced by the observation of the dimuon resonances at 9.44 and 10.06 GeV in [84] (right).

conflicting claims. The first of them [81], reporting the results based on the analysis of e^+e^- pairs, claims to have found cluster of events with invariant mass m_{ee} in the region between 5.8 and 6.2 GeV suggesting "that the data contain a new resonance at 6 GeV." The relevant plot, reproduced in the left part of Fig. 4.8, shows, indeed, an enhancement of the spectrum $d\sigma/dm_{ee}$ at $m_{ee} \doteq 6$ GeV, where the data are above the solid line, representing theoretical prediction based on QCD calculation of the continuum of dilepton pairs, but the accuracy of the measurement was clearly too poor to warrant such a definite claim. Indeed, repeating their investigation using $\mu^+\mu^-$, the group found (see middle plot in Fig. 4.8) no evidence for the 6 GeV resonance seen in [81]. Improving further their detector the same group had finally reported [83,84]
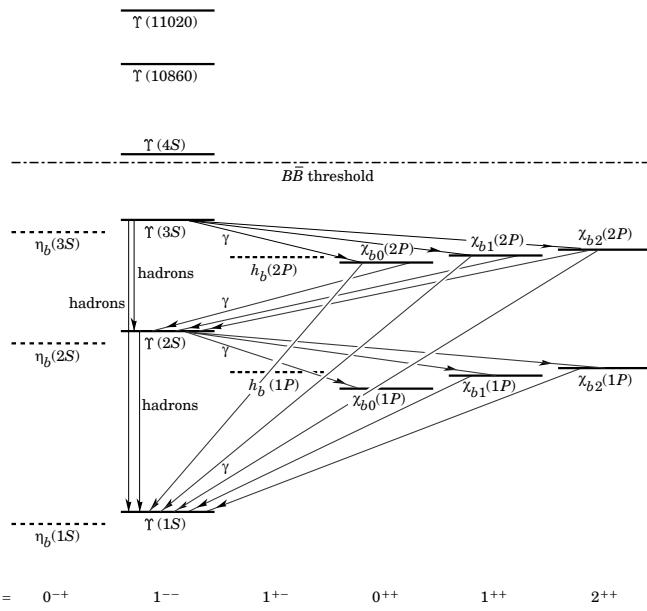
Figure 4.9: The spectrum of the lowest lying $b\bar{b}$ bound states.

the clear and compelling evidence for two new heavy resonances at $m_{\mu\mu} = 9.41$ GeV and $m_{\mu\mu} = 10.06$ GeV, with observed widths of about 1 GeV, compatible with the mass resolution of their detector. These states, denoted Υ and Υ', are actually very narrow: their widths, 52 and 44 keV respectively, are comparable to

72

those of J/ψ and ψ'. Contrary to the charm family, there are, however, three narrow $b\bar{b}$ states, the third one, Υ'', at 10.44 GeV. Since then a lot of experimental attention has been paid to the measurement of the spectrum of $b\bar{b}$ bound states, the so called *bottomonia*. The results, summed up in Fig. 4.9, show a spectrum of states similar to that of charmonium, which turns out to be even better described within the framework of nonrelativistic quantum mechanics based on the potential (4.44) than the $c\bar{c}$ bound states.

4.9   Lone at the top

Since the discovery of the fifth quark in 1977 the search went on to find the sixth quark, called *top*, which would complete the third generation. Many theorists tried to estimate its mass on the grounds of theoretical considerations taking into account the known pattern of quark masses, but all failed to forecast its huge value.

However, precise LEP measurements of the properties of the Z and W bosons combined with the theoretical predictions of the Standard Model (which depend on m_t through radiative corrections) had finally lead to the prediction of the top quark mass in the range $170 - 180$ GeV. This allowed experimentalists at Fermilab to narrow their search for the top quark until they found it among complicated multijet final states originating from the basic subprocess, depicted in Fig. 4.10

$$q + \overline{q} \to t + \overline{t} \to b + \overline{b} + W^+ + W^- \tag{4.46}$$

with the intermediate vector bosons $W^\pm$ decaying subsequently into two jets or pairs of leptons. The search
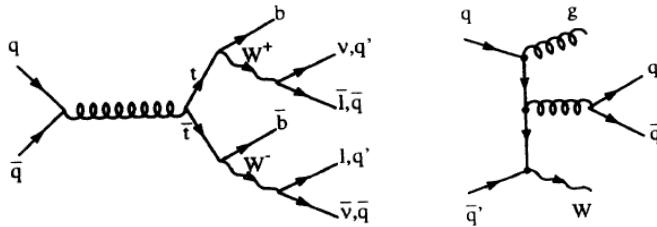
Figure 4.10: The basic mechanism of the top quark production at TEVATRON energies (left) and the main background process (right).

concentrated on the channel, called "lepton and jets", in which one of the W's decays leptonically and the other W into two jets. This process had in principle lead to four jets, two of them coming from the b-quark, and one charged lepton (with the accompanying neutrino remaining undetected).

This channel, which accounts for roughly 35% of all final states has manageable background coming predominantly from Standard Model process of associated production of W and jets, described by diagrams like that in the right part of Fig. 4.10. Statistically significant excess of events with 3 and 4 jets and a charged lepton over the Standard model prediction, shown in Fig. 4.11, had been observed. Performing kinematic fits of the 7 events with four jets to the hypothesis of four jets (two of which had to be identified as b-quark jets by the presence of secondary vertex signalling the decay of a B-meson), one charged lepton and one massless missing particle (the associated neutrino) allowed the CDF Collaboration to determine the mass of the top quark in each event. The resulting distribution, shown in the right part of Fig. 4.11, could best be fitted assuming the top quark mass to be $m_t = 174 \pm 10(stat) \pm 12(syst)$ GeV, remarkably close to the predicted value mentioned above. The discovery of the top quark was thus a triumph of the Standard Model.

Compared to the lighter quarks, the top quark has one property that singles it out: due to its huge mass, its lifetime,

$$\tau_t \doteq \tau_\mu \left(\frac{m_\mu}{m_t} \right)^5 \doteq 2\ 10^{-22} s \tag{4.47}$$

is so short, in fact comparable to that of strongly decaying resonances, that it decays via weak interaction before it can hadronize! This implies there are no $t\bar{t}$ bound states, analogous to the rich spectrum of $c\bar{c}$ or $b\bar{b}$ states, or bound states of the top quark with the other quarks, similar to charmed (or bottom) mesons and baryons. Note that for charm or bottom quarks the approximate formula (4.47) yields, taking into account

73
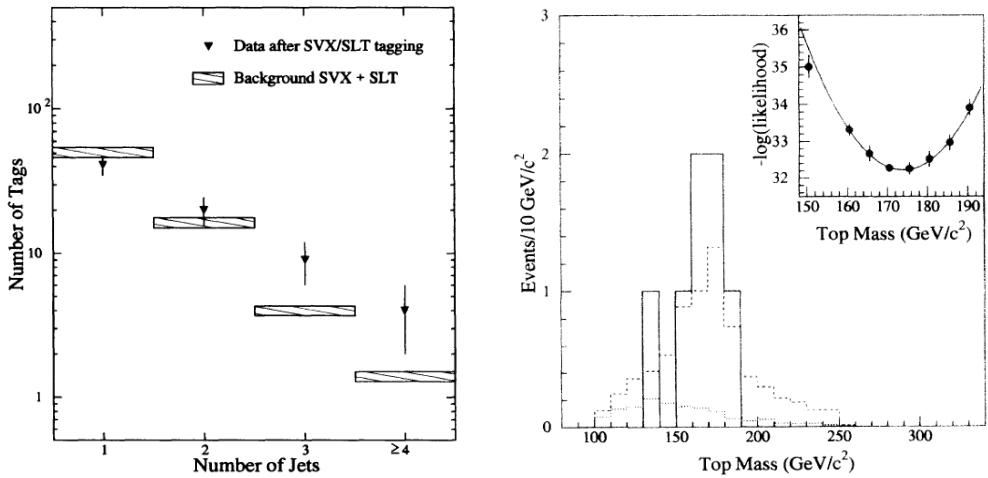
Figure 4.11: The first evidence for the top quark as observed by the CDF Collaboration at Fermilab [85]. Top mass distribution for the data (solid histogram), the W+jets background (dots) and the sum of background plus Monte Carlo $t\bar{t}$ for $M_{top} = 175$ GeV (dashed). The inset shows the logarithmic likelihood fit used to determine the top mass.

the (small!) CKM matrix elements describing the transitions $b \to u + W^-$ and $b \to c + W^-$, lifetimes in the range of picoseconds.

4.10   Exercises

1. Derive (4.12).

2. Construct wave function of Σ^0 from that of Σ^+.

3. Prove (4.164.18).

4. Evaluate magnetic moments of all baryons in Table 3.2, assuming $m_u = m_d \equiv m$, but $m_s \neq m$.

5. Derive (4.26).

6. Show that the Casimir operator (4.35) commutes with all generators of SU(3).

7. Derive (4.36).

Chapter 5

Elements of the parton model

5.1 Kinematics

Let me first recall the basic notions and kinematical relations concerning the scattering of charged as well as neutral leptons (l, ν_e, ν_μ) on nucleons. For definiteness, all relations will be written for lepton-proton scattering

$$l(k) + \text{proton}(P) \rightarrow l'(k') + X; \quad l, l' = e, \mu, \nu_e \nu_\mu, \tag{5.1}$$

where X denotes *any* final state allowed by conservation laws and the letters in brackets stand for four–momenta of the corresponding particles. A special case of (5.1) is the elastic lepton-nucleon scattering, when X=proton and $l = l'$. There are two types of lepton-nucleon scattering processes, called

- **neutral current** processes, when $l' = l$. These processes can be mediated by the exchange of either the photon or the (neutral) intermediate vector boson Z, but in the kinematical range we shall be interested in, the latter contribution can be safely neglected.

- **charged current** processes, when the electric charges of the initial and final leptons l, l' differ by one unit. These processes are mediated by the exchange of charged intermediate vector bosons $W^\pm$.

Kinematical considerations discussed below are common to both types of processes. The selection of variables describing (5.1) is a matter of convention, but the following set is commonly used (see Fig. 5.1) [1]

$$S \equiv (k + P)^2 = M_p^2 + 2kP = M_p(2E_{lab} + M_p), \tag{5.2}$$

$$Q^2 \equiv -q^2 \equiv -(k - k')^2 = 2kk' = 4EE' \sin^2(\vartheta/2), \tag{5.3}$$

$$y \equiv \frac{qP}{kP} = \frac{E_{lab} - E'_{lab}}{E_{lab}} = \frac{\nu}{E_{lab}} = \left(\frac{s - M_p^2}{s}\right) \frac{1 - \cos \vartheta^*}{2}, \tag{5.4}$$

$$x \equiv \frac{Q^2}{2Pq} = \frac{Q^2}{2M_p \nu} = \frac{Q^2}{Q^2 + W^2 - M_p^2}, \tag{5.5}$$

$$W^2 \equiv (q + P)^2 = \frac{Q^2(1 - x)}{x} + M_p^2. \tag{5.6}$$

The variable W has a clear interpretation as the invariant mass of the hadronic system X, produced by the absorption of the exchanged photon by the target nucleon. Notice that only the final state electron momentum k' is needed (together with initial momenta k, P) to evaluate x, y and Q^2 and that in (5.3) the last equality holds in *any* reference system. There are two reference systems in which the process (5.1) is usually analyzed: the laboratory system, where the target proton is at rest and the center-of-mass system (CMS), where the colliding lepton and proton have equal but opposite momenta. Notice, that all five quantities defined above are relativistic *invariants*. In the laboratory frame they can be expressed by means of the scattering angle ϑ_{lab} and the energies E_{lab}, E'_{lab} of the primary and scattered lepton. For instance

$$E'_{lab} = \frac{E_{lab} - (W^2 - M_p^2)/2M_p}{1 + E_{lab}\left(1 - \cos \vartheta_{lab}\right)/M_p}. \tag{5.7}$$

---

[1]In the following m will denote generic lepton masses and M_p the proton mass. In most cases the former will be neglected, but the latter kept.
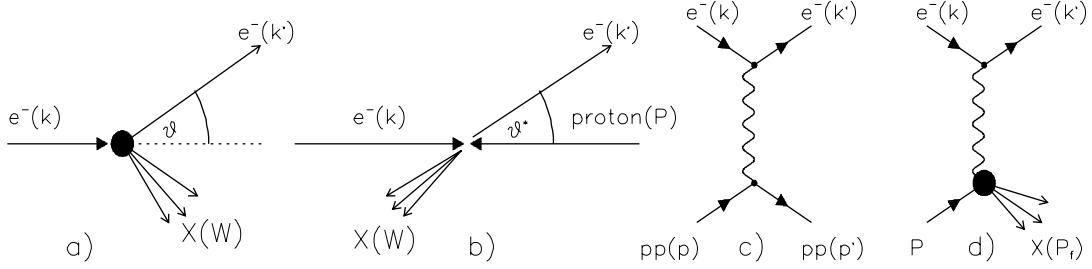
Figure 5.1: The DIS scattering in laboratory (a) and CMS (b) frames and the lowest order Feynman diagram describing the scattering on a pointlike proton, denoted "pp" (c).

The variable S specifies the initial state of the colliding electron and proton, while x, y, q^2, ϑ and E'_{lab}, describe, for given S and the azimuthal angle ϕ, the final state of an electron in (5.1). As, however, only two of them are independent, one can thus choose any of the pairs (x, y), (x, q^2), (y, q^2), or $(E'_{lab}, \vartheta_{lab})$ to uniquely specify the state of the scattered electron provided the colliding particles are *unpolarized* [2]. Unless mentioned otherwise, this will always be assumed. In modern notation the pairs (x, y) or (x, Q^2) are most often used to write down the cross–section of (5.1). Note that these variables are related, for given total CMS energy squared S, by a simple formula

$$Q^2 = xy(2kP) = xy\left(S - M_p^2\right). \tag{5.8}$$

In elastic scattering of unpolarized electrons and protons the final state is described, beside the azimuthal angle, by one independent variable only, for which one can take any of the following Q^2, y, ϑ or E'_{lab}. In the following the scattering angle ϑ_{lab} or $y \doteq (1 - \cos \vartheta^*)/2$ will usually be used for this purpose [3]. For elastic scattering we find that x and W^2 are fixed by kinematics, $x = 1$ and $W^2 = M_p^2$, and the relation (5.7) implies that ϑ and E'_{lab} are related

$$E'_{lab} = \frac{E_{lab}}{1 + E_{lab}\left(1 - \cos \vartheta_{lab}\right)/M_p}. \tag{5.9}$$

The term "deep inelastic scattering" means that both invariants Q^2 and Pq are large with respect to M_p

$$Q^2 = -q^2 \gg M_p^2; \quad \nu \gg M_p. \tag{5.10}$$

The so called **Bjorken limit** corresponds to the idealized case when $Q^2 \to \infty, Pq \to \infty$ but the ratio $x \equiv Q^2/2Pq$ remains finite.

The above relations can be used, with the trivial replacement $M_p \to M_{nucl}$, for the description of the scattering of leptons on nuclei as well.

5.2 Dynamics

The basic quantity describing the general scattering process

$$1 + 2 \to 3 + 4 + \cdots n \tag{5.11}$$

is the differential cross–section

$$d\sigma = \frac{(2\pi)^4}{\mid \vec{v}_1 - \vec{v}_2 \mid} \frac{1}{2E_1} \frac{1}{2E_2} \left|M_{if}\right|^2 \prod_{i=3}^{n} \frac{d\vec{p}_i}{(2\pi)^3 2E_i} \delta^4(p_1 + p_2 - \sum_{i=3}^{n} p_i) \, SF, \tag{5.12}$$

---

[2]In the case of the scattering of unpolarized particles the trivial dependence on ϕ is integrated over.

[3]Quantities in the CMS will be labeled by an asterisk and the symbol $\doteq$ will be employed to denote approximate equality, neglecting the mass of the proton or target parton. Recall that the mass of an electron is always neglected.

where E_i is the energy of particle i with velocity $\vec{v_i}$ (in units of c), M_{if} denotes the invariant amplitude of the process (5.11), corresponding to the following normalization of incoming and outgoing one–particle states $\mid E, \vec{p} \rangle$ of energy E and momentum $\vec{p}$

$$\langle E', \vec{p}\,' \mid E, \vec{p} \rangle = (2\pi)^3 2E\delta^3(\vec{p} - \vec{p}\,') \tag{5.13}$$

and the statistical factor SF

$$SF \equiv \prod_i \frac{1}{n_i!} \tag{5.14}$$

takes into account that there are n_i identical particles of type i in the final state. The expression (5.12) assumes that the particles 1, 2 collide head on, otherwise more general expression would replace the product $\mid \vec{v_1} - \vec{v_2} \mid 2E_1 2E_2$, which is invariant with respect to Lorentz boosts along the direction of these colliding particles. The above normalization, applicable to massive as well as massless particles, will be used for both bosons and fermions.

All the analyses described bellow were carried out using the Quantum Electrodynamics, formulated in 1950 by Feynman, Schwinger, Dyson and Tomonaga. Feynman rules for incoming and outgoing particles corresponding to the above normalization as well as for interaction vertices, are given in the Appendix.

Elastic scattering of electrons on pointlike proton

As in the case of the Rutherford experiment, we have to know the cross section (5.12) of elastic lepton-proton scattering for the case of pointlike proton with momentum p and mass M. [4] In the lowest order of perturbative quantum electrodynamics (QED) this process is described by the one photon exchange diagram of Fig. 5.1c. The square of the corresponding matrix element M_{if}, summed over the spins s_3, s_4 of the final and averaged over the spins s_1, s_2 of the initial fermions, reads

$$\left| \overline{M_{if}} \right|^2 = \frac{1}{4} \sum_{spins} \left| [\overline{u}(k', s_4)\gamma_\mu u(k, s_2)] \, e^2 \frac{-ig^{\mu\nu}}{q^2} [\overline{u}(p', s_3)\gamma_\nu u(p, s_1)] \right|^2 . \tag{5.15}$$

The quantities $u, \overline{u}$ are the usual Dirac bispinors, satisfying the free particle Dirac equations

$$(\not{p} - m)u(p) = 0; \quad \overline{u}(p)(\not{p} - m) = 0 \tag{5.16}$$

and are normalized as

$$\overline{u}(p)u(p) = -\overline{v}(p)v(p) = 2m; \quad u^+(p)u(p) = v^+(p)v(p) = 2E \tag{5.17}$$

and the photon propagator is taken in the Feynman gauge. Some of the relations concerning the evaluation of traces of Dirac matrices are reviewed in the Appendix. Using the standard technique, (5.15) can be written as the contraction of two tensors

$$\left| \overline{M_{if}} \right|^2 = \frac{e^4}{Q^4} L^{(1)\mu\nu} L^{(2)}_{\mu\nu}, \tag{5.18}$$

where

$$L^{(1)}_{\mu\nu} \equiv \frac{1}{2}\text{Tr}\left[(\not{k}' + m)\gamma^\mu(\not{k} + m)\gamma^\nu\right] = 2\left[k^\mu k'^\nu + k^\nu k'^\mu - g^{\mu\nu}(kk' - m^2)\right] \tag{5.19}$$

is associated with the upper, electron vertex of Fig. 5.1c. Similarly, $L^{(2)}_{\mu\nu}$ is associated with the lower, proton, vertex and is given by (5.19) with the substitutions $k \to p, k' \to p', m \to M$. Carrying the contraction as indicated in (5.18) we get

$$
\begin{aligned}
L^{(1)\mu\nu} L^{(2)}_{\mu\nu} &= 8[\underbrace{(k'p')}_{kp}(kp) + (k'p)\underbrace{(kp')}_{k'p} - M^2\underbrace{(kk')}_{-q^2/2} - \underbrace{m^2(pp') + 2m^2 M^2}_{\text{neglected}}] \\
&= 8(kp)^2 \left[1 + \left(\frac{k'p}{kp}\right)^2 + \frac{M^2 q^2}{2(kp)^2}\right] \\
&= 8(kp)^2 \left[1 + (1-y)^2 - \frac{M^2 y}{kp}\right] \xrightarrow[kp\to\infty]{} 8(kp)^2 \left[1 + (1-y)^2\right], \tag{5.20}
\end{aligned}
$$

---

[4] In view of further considerations I use in this paragraph the symbols p, M to denote momentum and mass of the pointlike proton, reserving P, M_p for those of the *real* one.

where $(kp) = (s - M^2)/2$ specifies the primary energy of the colliding particles and $(k'p)/(kp) = 1 - y$ describes the final state. (5.20) takes, as indicated, particularly simple form for $s \to \infty$. All the quantities in (5.20) are relativistic invariants, but it is also useful to express it in terms of the scattering angle ϑ in a particular reference frame. In the CMS of the colliding particles we have

$$y \equiv \frac{qp}{kp} = \left(\frac{s - M^2}{s} \right) \frac{1 - \cos \vartheta^*}{2} \doteq \frac{1 - \cos \vartheta^*}{2} \Rightarrow \mathrm{d}y = - \left(\frac{s - M^2}{2s} \right) \mathrm{d} \cos \vartheta^* \doteq - \frac{\mathrm{d} \cos \vartheta^*}{2}, \qquad (5.21)$$

the integral over the phase space elements gives

$$\int \int \frac{\mathrm{d}^3 \vec{p}\,'}{2E_{p'}} \frac{\mathrm{d}^3 \vec{k}\,'}{2E_{k'}} \delta^4(k + p - k' - p') = \int \left(\frac{s - M^2}{8s} \right) \mathrm{d}\Omega \doteq \frac{1}{8} \int \mathrm{d}\Omega \qquad (5.22)$$

and the flux factor reduces to

$$\frac{1}{2E_k} \frac{1}{2E_p} \frac{1}{\mid v_1 - v_2 \mid} = \frac{1}{2(s - M^2)}, \qquad (5.23)$$

where we have retained the nonzero mass M of the target "pointlike proton". Substituting the above expressions into (5.12), integrating over ϕ and replacing $\mathrm{d} \cos \vartheta^*$ with $\mathrm{d}y$ using (5.21) we finally get

$$\frac{\mathrm{d}\sigma}{\mathrm{d}y} = \frac{2\pi\alpha^2(2kp)}{Q^4} \left[1 + (1 - y)^2 - \frac{M^2 y}{kp} \right] \Rightarrow \frac{\mathrm{d}\sigma}{\mathrm{d}Q^2} = \frac{2\pi\alpha^2}{Q^4} \left[1 + (1 - y)^2 - \frac{M^2 y}{kp} \right], \qquad (5.24)$$

where we have used the fact that $Q^2 = y2kp$. Carrying out the same reduction in the laboratory frame yields

$$L^{\mu\nu}_{(1)} L^{(2)}_{\mu\nu} = 16EE'M^2 \left[\cos^2 \frac{\vartheta_{lab}}{2} - \frac{q^2}{2M^2} \sin^2 \frac{\vartheta_{lab}}{2} \right] \Rightarrow \qquad (5.25)$$

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega_{lab}} = \frac{\alpha^2 \cos^2(\vartheta_{lab}/2)}{4E^2 \sin^4(\vartheta_{lab}/2)} \frac{E'}{E} \left[1 + \frac{Q^2}{2M^2} \tan^2 \frac{\vartheta_{lab}}{2} \right] \xrightarrow[M \to \infty]{} \frac{\alpha^2 \cos^2(\vartheta_{lab}/2)}{4E^2 \sin^4(\vartheta_{lab}/2)} = \sigma_{\mathrm{Mott}} \qquad (5.26)$$

where σ_{Mott} gives the differential cross–section corresponding to the scattering of a spinless particle off an infinitely heavy proton and the factor $E'/E = 1 - Q^2/2ME < 1$ takes into account the recoil. Apart from the presence of $\cos^2(\vartheta/2)$, which is characteristic of spin $1/2$ fermion, the Mott cross section represents relativistic generalization of the Rutherford formula (1.17) for elastic scattering in the Coulomb potential.

The same calculation can be repeated in the case that the pointlike target particle is not a fermion with spin $1/2$ but a boson with spin 0. Feynman rules lead to the following simple expression for the vertex boson-γ-boson in the notation as in Fig. 5.1: $-ie(p + p')^\mu$. The tensor $L^{(2)}_{\mu\nu}$ is then trivial and we get

$$L^{(1)\mu\nu} L^{(2)}_{\mu\nu} = 8(kp)^2 \left[2(1 - y) - \frac{M^2 y}{kp} \right]. \qquad (5.27)$$

Both of the above results can be written in a single compact form

$$\frac{\mathrm{d}\sigma}{\mathrm{d}y} = \frac{4\pi\alpha^2(2kp)}{Q^4} \left[(1 - y) + \varepsilon \frac{y^2}{2} - \frac{M^2 y}{2kp} \right], \qquad (5.28)$$

or equivalently

$$\frac{\mathrm{d}\sigma}{\mathrm{d}Q^2} = \frac{4\pi\alpha^2}{Q^4} \left[(1 - y) + \varepsilon \frac{y^2}{2} - \frac{M^2 y}{2kp} \right] \underset{s \to \infty}{\Longrightarrow} \frac{4\pi\alpha^2}{Q^4}, \qquad (5.29)$$

where

- $\varepsilon = 1$ for pointlike fermion with spin $1/2$,

- $\varepsilon = 0$ for pointlike boson with spin 0.

To see a difference between the scattering on fermions and bosons requires therefore large y, i.e. large angle scattering.

Elastic scattering of electrons on real proton

For the real proton the situation is more complicated because the strong interactions modify the pointlike coupling $\overline{u}(p')[ie\gamma^\mu]u(p)$. In general this simple coupling has to be replaced with a structure, compatible with gauge and Lorentz invariance and parity conservation [5]

$$\overline{u}(p')\left[F_1(Q^2)\gamma^\mu + \kappa\frac{F_2(Q^2)}{2M_p}i\sigma^{\mu\nu}q_\nu\right]u(p); \;\; \sigma^{\mu\nu} \equiv \frac{i}{2}\left[\gamma^\mu\gamma^\nu - \gamma^\nu\gamma^\mu,\right] \tag{5.30}$$

where $F_1(Q^2), F_2(Q^2)$ are the **elastic electromagnetic formfactors** of the proton and κ is a number, conveniently introduced in order to allow us to set $F_2(0) = 1$. No such parameter is needed in front of F_1 as gauge invariance implies $F_1(0) = 1$. Moreover, it is an interesting exercise in the nonrelativistic reduction of the Dirac equation to show that κ gives just the **anomalous magnetic moment** of the proton (i.e. 1.79).

The evaluation of the differential cross section using (5.30) is straightforward but tedious. The result

$$\frac{d\sigma}{d\Omega_{lab}} = \sigma_{\text{Mott}}\frac{E'}{E}\left\{F_1^2(Q^2) + \frac{Q^2}{4M_p^2}\left[2\tan^2\frac{\vartheta_{lab}}{2}\left(F_1(Q^2) + \kappa F_2(Q^2)\right)^2 + \kappa^2 F_2^2(Q^2)\right]\right\}, \tag{5.31}$$

to be compared to (5.26), was derived first by Rosenbluth in 1950. The pointlike "Dirac proton" corresponding to $F_1 = 1, \kappa = 0$, includes the interaction of pointlike magnetic moment equal to $\mu_p = 1$. Setting $F_1(Q^2) = F_2(Q^2) = 1, \kappa = 1.793$ defines the pointlike proton with anomalous magnetic moment. For the real proton the formfactors F_i are steeply falling functions of Q^2. Instead of F_1, F_2 it is convenient to introduce the **electric and magnetic** formfactors

$$G_E(Q^2) \equiv F_1(Q^2) - \frac{Q^2}{4M_p^2}\kappa F_2(Q^2) \Rightarrow G_E(0) = 1, \tag{5.32}$$

$$G_M(Q^2) \equiv F_1(Q^2) + \kappa F_2(Q^2) \;\; \Rightarrow \;\; G_M(0) = 1 + \kappa = \mu_p. \tag{5.33}$$

In Fig. 5.5a the ratio G_M/μ_p is plotted as a function of Q^2 and fitted to the dipole formula, which implies that at large Q^2, $G_M \propto 1/Q^4$! Similarly for $G_E(Q^2)$. In terms of these new formfactors ($\tau \equiv Q^2/4M_p^2$)

$$\frac{d\sigma}{d\Omega_{lab}} = \sigma_{\text{Mott}}\frac{E'}{E}\left[\frac{G_E^2 + \tau G_M^2}{1 + \tau} + 2\tau G_M^2\tan^2(\vartheta_{lab}/2)\right]. \tag{5.34}$$

Deep inelastic scattering of electrons on nucleons

We now come to the process which played decisive role in the formulation of the parton model. In inelastic scattering of *unpolarized* electrons on *unpolarized* protons there are *two* independent variables which uniquely specify (besides the trivial dependence on azimuthal angle) final state of the scattered electron. One of the convenient choices is the pair of variables x, y. The basic quantity of interest is then the double differential cross–section $d\sigma/dydx$. For electromagnetic interactions its general form reads, written in three equivalent ways, which differ by the choice of two independent variables on which the cross section depends

$$\frac{d\sigma}{dE'd\Omega_{lab}} = \sigma_{\text{Mott}}\frac{1}{M}\left[W_2(x, Q^2) + 2W_1(x, Q^2)\tan^2\frac{\vartheta}{2}\right]. \tag{5.35}$$

$$\frac{d\sigma}{dxdy} = \frac{4\pi\alpha^2(2kP)}{Q^4}\left[\left(1 - y - \frac{M_p^2 xy}{S}\right)F_2(x, Q^2) + \frac{1}{2}y^2 2x F_1(x, Q^2)\right], \tag{5.36}$$

$$\frac{d\sigma}{dxdQ^2} = \frac{4\pi\alpha^2}{Q^4}\left[\left(1 - y - \frac{M_p^2 xy}{S}\right)\frac{F_2(x, Q^2)}{x} + \frac{1}{2}y^2 2F_1(x, Q^2)\right] \underset{S\to\infty}{\longrightarrow} \frac{4\pi\alpha^2}{Q^4}\frac{F_2(x, Q^2)}{x}, \tag{5.37}$$

where $S \equiv (k + P)^2$ is the square of the total CMS energy. Noting that $Q^2 = 2(kP)xy$ the limit in the last expression follows from the fact that for fixed x and Q^2 the limit $s \to \infty$ implies $y \to 0$. The unknown functions $F_i(x, Q^2)$, or equivalently, $W_i(x, Q^2)$, are called **inelastic electromagnetic formfactors** or **structure functions** of the proton. They are usually written as functions of x, Q^2, but this is merely a

---

[5]See next subsection for the derivation in the case of a more complicated inelastic scattering.

matter of convenience as any one of the three variables x, y, Q^2 can, for a given S, be expressed as a function of the remaining two. The form of (5.36) follows from four fundamental properties of electromagnetic interactions: Lorentz invariance, unitarity, gauge invariance and parity conservation.

Its derivation proceeds in several steps. First, the tensor $L_{\mu\nu}^{(2)}$ associated in the case of elastic scattering on a pointlike proton with the lower vertex in Fig. 5.1d, will be replaced with the general Lorentz tensor that can be formed out of the momenta [6] P, q and basic tensors of rank two, i.e. symmetric metric tensor $g^{\mu\nu}$ and totally antisymmetric Levi–Civita pseudotensor $\epsilon_{\mu\nu\alpha\beta}$:

$$W_{\mu\nu}(P,q) = -W_1 g_{\mu\nu} + W_2 \frac{P_\mu P_\nu}{M_p^2} + iW_3 \epsilon_{\mu\nu\alpha\beta} P^\alpha q^\beta + W_4 q_\mu q_\nu + W_5 \left(P_\mu q_\nu + P_\nu q_\mu\right) + iW_6 \left(P_\mu q_\nu - P_\nu q_\mu\right). \quad (5.38)$$

In the above decomposition $W_i(P,q)$ are unknown functions of P and q, which depend on the internal structure of the proton . The presence of the imaginary unit in front of W_3 and W_6 is merely a matter of convention. With this choice of W_i's the hermiticity of the tensor $W_{\mu\nu}$, i.e. $W_{\mu\nu} = W_{\nu\mu}^*$, implies that all $W_i(P,q)$ are *real functions*. Not all of them are, however, independent. Parity conservation requires $W_3(P,q) = 0$ and gauge invariance, expressed as the condition

$$q^\mu W_{\mu\nu}(P,q) = \underbrace{\left[-W_1 + W_4 q^2 + W_5 (Pq)\right]}_{=0} q_\nu + \underbrace{\left[W_2 \frac{Pq}{M_p^2} + W_5 q^2\right]}_{=0} P_\nu + \underbrace{iW_6 \left[(Pq)q_\nu - q^2 P_\nu\right]}_{\text{imaginary} \Rightarrow W_6 = 0} = 0 \quad (5.39)$$

leads to three relations: $W_6 = 0$ and

$$W_5(P,q) \quad = \quad -W_2(P,q) \frac{Pq}{q^2 M_p^2}, \quad (5.40)$$

$$W_4(P,q) \quad = \quad W_1(P,q) \frac{1}{q^2} + W_2(P,q) \frac{(Pq)^2}{q^4 M_p^2}. \quad (5.41)$$

Out of the original six functions W_i only two are thus independent. In terms of W_1, W_2 eq. (5.38) reads

$$W_{\mu\nu}(P,q) = -W_1(P,q) \left(g_{\mu\nu} - \frac{q_\mu q_\nu}{q^2}\right) + \frac{W_2(P,q)}{M_p^2} \left(P_\mu - \frac{Pq}{q^2} q_\mu\right) \left(P_\nu - \frac{Pq}{q^2} q_\nu\right). \quad (5.42)$$

Dropping the structures proportional to q^μ or q^ν, which vanish after contraction with the leptonic tensor $L^{\mu\nu}(k,k')$, due to its transversality (i.e. $q^\mu L_{\mu\nu} = 0$), we get

$$W_{\mu\nu}(P,q) = -W_1(P,q) g_{\mu\nu} + W_2(P,q) \frac{P_\mu P_\nu}{M_p^2}. \quad (5.43)$$

Performing the contraction $L_{\mu\nu} W^{\mu\nu}$ neglecting the electron mass m but keeping the proton mass M_p

$$L^{\mu\nu}(k,k') W_{\mu\nu}(P,q) = 2 \left(k^\mu k'^\nu + k^\nu k'^\mu - g^{\mu\nu}(kk')\right) \left(-W_1 g_{\mu\nu} + W_2 \frac{P_\mu P_\nu}{M^2}\right) \quad (5.44)$$

we find, after some straightforward algebra and using basic kinematics,

$$L^{\mu\nu} W_{\mu\nu} = \frac{4(kP)}{y} \left[\frac{\nu}{M_p} W_2 \left(1 - y - \frac{M_p^2 x^2 y^2}{Q^2}\right) + 2xW_1 \frac{y^2}{2}\right]. \quad (5.45)$$

Instead of W_1, W_2 it is common to introduce another pair of functions

$$F_1 \equiv W_1, \quad F_2 \equiv \frac{\nu}{M} W_2, \quad (5.46)$$

in term of which (5.45) reads

$$L^{\mu\nu} W_{\mu\nu} = \frac{4(kP)}{y} \left[F_2(x,Q^2) \left(1 - y - \frac{M_p^2 x^2 y^2}{Q^2}\right) + 2xF_1(x,Q^2) \frac{y^2}{2}\right] \quad (5.47)$$

---

[6] For scattering of unpolarized leptons on unpolarized nucleons there are no other four-momenta available. In the case of polarized particles the general decomposition of $W_{\mu\nu}$ includes also structures containing their polarization vectors.

Inserting (5.47) into the general expression (5.12) for the corresponding cross–section and working out the differential $\mathrm{d}^3 k'$ in terms of $\mathrm{d}x\mathrm{d}y$ we finally arrive at (5.36).

Recalling that in elastic scattering $x = 1$ by kinematics, we find the following relations between elastic and inelastic formfactors:

$$F_2^{\text{inel}}(x = 1, Q^2) = (F_1^{\text{el}}(Q^2))^2 + \frac{\kappa^2 Q^2}{4M_p^2} \left(F_2^{\text{el}}(Q^2)\right)^2, \tag{5.48}$$

$$2F_1^{\text{inel}}(x = 1, Q^2) = \left(F_1^{\text{el}}(Q^2) + \kappa F_2^{\text{el}}(Q^2)\right)^2, \tag{5.49}$$

where the superscripts (not used elsewhere) "inel" and "el" were employed in the above relations merely to emphasize the difference between elastic and inelastic formfactors.

5.3 *Cross–sections of virtual photons

The functions W_1, W_2 introduced in (5.38), can be given intuitively clear interpretation in terms of the total cross–sections $\sigma(\gamma^* p)$ of the "collision" between the virtual photon and the proton in the lower vertex of Fig. 5.1d. In this "collision" $W = \sqrt{Q^2(1 - x)/x}$ plays the role of the total center of mass energy and Q^2 that of the magnitude of the mass squared of the incoming photon. While real photons exist only in two spin states (spin parallel or antiparallel to its momentum), corresponding to two polarization four-vectors ϵ_μ, no such restriction holds for virtual photons, which exist altogether in *four independent* spin states, described by polarization vectors $\epsilon_\mu^{(i)}$, $i = 1, 2, 3, 4$.

As will be discussed in detail in Section 6.1 any treatment of photons, either real or virtual, requires the selection of a particular **gauge**, which leads to additional condition on the polarization four-vectors ϵ_μ. In the so called Lorentz gauge this condition implies $\epsilon(q)q = 0$. There are three independent polarization vectors, which satisfy this *gauge fixing condition*. Assuming $\vec{q}$ to point in the "third" direction and denoting $q = (q_0, q_1, q_2, q_3)$ they can be chosen as the following real vectors:

$$\epsilon_{\mathrm{T}}^{(1)} = (0, 1, 0, 0); \quad \epsilon_{\mathrm{T}}^{(2)} = (0, 0, 1, 0); \quad \epsilon_{\mathrm{L}} = \frac{1}{\sqrt{Q^2}}(q_3, 0, 0, q_0). \tag{5.50}$$

The first two describe *transverse* and the third *longitudinal* polarizations of the virtual photon with momentum q. They satisfy the following normalization conditions $(\epsilon_{\mathrm{T}}^{(1,2)})^2 = -1, (\epsilon_{\mathrm{L}})^2 = 1$ and appear in the decomposition of the metric tensor $g_{\mu\nu}$

$$-g_{\mu\nu} = -\frac{q_\mu q_\nu}{Q^2} + \epsilon_{\mathrm{T},\mu}^{(1)} \epsilon_{\mathrm{T},\nu}^{(1)} + \epsilon_{\mathrm{T},\mu}^{(2)} \epsilon_{\mathrm{T},\nu}^{(2)} - \epsilon_{\mathrm{L},\mu} \epsilon_{\mathrm{L},\nu}. \tag{5.51}$$

Note the difference in the normalization of transverse and longitudinal photons!

Leaving for the moment the overall normalization factor C, which contains flux and other factors undetermined, the total cross section of a collision of the virtual photon of momentum q and polarization vector $\epsilon_\mu = (\epsilon_0, \vec{\epsilon})$ with the proton at rest (i.e. with $P = (M, 0, 0, 0)$) is given as

$$\sigma(\gamma^* p; S) = C\epsilon^\mu W_{\mu\nu}(P, q)\epsilon^\nu = C\left(-W_1 \epsilon^2 + \epsilon_0^2 W_2\right), \tag{5.52}$$

where $W_{\mu\nu}$ is the same hadronic tensor as introduced above. For the three polarizations in (5.50) we introduce *longitudinal* and *transverse* cross–sections as follows

$$\sigma_{\mathrm{T}}(\gamma^* \mathrm{p}) \equiv \frac{1}{2}\left(\sigma_{\mathrm{T}}^{(1)} + \sigma_{\mathrm{T}}^{(2)}\right) = CW_1, \tag{5.53}$$

$$\sigma_{\mathrm{L}}(\gamma^* \mathrm{p}) = C\left(-W_1 + \frac{q_3^2}{Q^2}W_2\right) = C\left(-W_1 + \left(1 + \frac{Q^2}{4M_p^2 x^2}\right)W_2\right), \tag{5.54}$$

$$\sigma(\gamma^* \mathrm{p}) \equiv (\sigma_{\mathrm{T}} + \sigma_{\mathrm{L}}) = C\left(1 + \frac{Q^2}{4M_p^2 x^2}\right)W_2, \tag{5.55}$$

wherefrom

$$CW_1 = \sigma_{\mathrm{T}} \qquad\qquad \Rightarrow 2xF_1 = \frac{2x}{C}\sigma_{\mathrm{T}} \equiv F_{\mathrm{T}} \tag{5.56}$$

$$CW_2 = (\sigma_{\mathrm{T}} + \sigma_{\mathrm{L}})\frac{4M_p^2 x^2}{Q^2 + 4M_p^2 x^2} \Rightarrow \quad F_2 = \frac{2x}{C}\left[\frac{Q^2}{Q^2 + M_p^2 x^2}\right](\sigma_{\mathrm{T}} + \sigma_{\mathrm{L}}) \doteq F_{\mathrm{T}} + F_{\mathrm{L}}, \tag{5.57}$$

81

where the last equality holds for $M \to 0$. Defining the ratio of the longitudinal and transverse cross–sections

$$R(x, Q^2) \equiv \frac{\sigma_{\mathrm{L}}(x, Q^2)}{\sigma_{\mathrm{T}}(x, Q^2)} \tag{5.58}$$

and taking into account that

$$\frac{W_2}{W_1} = \left(1 + \frac{\sigma_{\mathrm{L}}}{\sigma_{\mathrm{T}}}\right)\left(\frac{Q^2}{Q^2 + \nu^2}\right) = (1 + R)\left(\frac{Q^2}{Q^2 + \nu^2}\right) \tag{5.59}$$

we get

$$\frac{2xF_1}{F_2} = \frac{1 + 4M_p^2 x^2/Q^2}{1 + R}. \tag{5.60}$$

In addition to F_1, F_2 it is useful to introduce the so called **longitudinal structure function**

$$F_{\mathrm{L}}(x, Q^2) \equiv F_2(x, Q^2)\left(1 + \frac{4M_p^2 x^2}{Q^2}\right) - 2xF_1(x, Q^2), \tag{5.61}$$

in terms of which the ratio (5.58) reads

$$R(x, Q^2) = \frac{F_L(x, Q^2)}{2xF_1(x, Q^2)}. \tag{5.62}$$

Inverting the relations (5.56)-(5.57) we get $\sigma_{\mathrm{T}}, \sigma_{\mathrm{L}}$ in terms of F_1, F_2:

$$\sigma_{\mathrm{T}}(x, Q^2) = CF_1(x, Q^2), \tag{5.63}$$

$$\sigma_{\mathrm{L}}(x, Q^2) = \frac{C}{2x}\left[F_2(x, Q^2)\left(1 + \frac{4M_p^2 x^2}{Q^2}\right) - 2xF_1(x, Q^2)\right] = C\frac{F_{\mathrm{L}}(x, Q^2)}{2x}. \tag{5.64}$$

Similarly inverting (5.60) we get F_1 in terms of F_2 and R

$$2xF_1(x, Q^2) = F_2(x, Q^2)\left[\frac{1 + 4M_p^2 x^2/Q^2}{1 + R(x, Q^2)}\right] \tag{5.65}$$

and consequently the brackets of (5.36) can be written in terms of F_2 and R as follows

$$F_2(x, Q^2)\left[1 - y + \frac{y^2}{2}\left(\frac{1 + 4M_p^2 x^2/Q^2}{1 + R(x, Q^2)}\right) - \frac{M_p^2 x^2 y^2}{Q^2}\right]. \tag{5.66}$$

To measure both structure functions F_1, F_2 as functions of x, Q^2 is experimentally very difficult and in fact impossible in a single experiment at one primary energy. According to (5.66) separating $F_1(x, Q^2)$ from $F_2(x, Q^2)$ at fixed values of x, Q^2 is possible only if y can be varied. Taking into account that $Q^2 = Sxy$, this, however, requires that the measurement is performed at two primary energies at least. Moreover, for this separation the double differential cross–section is usually written as

$$\frac{\mathrm{d}\sigma}{\mathrm{d}x\mathrm{d}Q^2} = \Gamma\left(\sigma_{\mathrm{T}}(x, Q^2) + \varepsilon(y)\sigma_{\mathrm{L}}(x, Q^2)\right), \ \varepsilon(y) \equiv \frac{2(1 - y)}{1 + (1 - y)^2}, \ \Gamma \equiv \frac{4\pi\alpha^2}{CQ^4}\left(1 + (1 - y)^2\right). \tag{5.67}$$

Note that the function $\varepsilon(y)$ can be interpreted as the ratio of flux factors of longitudinal and transverse virtual photons. In terms of $F_2(x, Q^2)$ the total $\gamma^* p$ cross–section reads

$$\sigma(\gamma^* p; W, Q^2) = \frac{CF_2(x, Q^2)}{2x}\frac{Q^2 + 4M_p^2 x^2}{Q^2} = \frac{CF_2(W, Q^2)}{2x}\left[\frac{(W^2 - M_p^2) + Q^2\left(Q^2 + 2W^2 + 2M_p^2\right)}{(W^2 - M_p^2) + Q^2\left(Q^2 + 2W^2 - 2M_p^2\right)}\right]. \tag{5.68}$$

In the second equality above I have deliberately written F_2 as as a function of Q^2 and W instead of Q^2 and x to emphasize the fact that W plays the role of the total CMS primary energy in the $\gamma^* p$ collision. Note

that in the limit $Q^2 \to 0$ and W fixed, where (5.68) goes over to the cross–section of a *real* photon, x actually *vanishes*!

In the last step the normalization factor C will be determined. This involves some ambiguity related with the definition of the flux factor of virtual photons. Recall that for real photons with polarization vector ϵ eq. (5.12) leads to a unique formula for the total cross–section

$$\sigma(\gamma p; W\epsilon) = \frac{4\pi^2\alpha}{E_\gamma M_p}\epsilon^\mu\left(-W_1(W,0)g_{\mu\nu} + \frac{P_\mu P_\nu}{M^2}W_2(W,0)\right)\epsilon^\nu, \tag{5.69}$$

wherefrom for real photon $C = 4\pi^2\alpha/E_\gamma M_p$. Using the same formula for virtual photons and recalling that $E_{\gamma^*} = (W^2 - M^2 + Q^2)/(2M)$ the direct generalization would be

$$C = \frac{8\pi^2\alpha}{W^2 - M_p^2 + Q^2}. \tag{5.70}$$

However, for virtual photons the standard definition of C is (5.70), but *without* Q^2 in its denominator, i.e.

$$C = \frac{8\pi^2\alpha}{W^2 - M_p^2} = \frac{8\pi^2\alpha x}{Q^2(1 - x)} \tag{5.71}$$

so that F_2 and $\sigma(\gamma^*p)$ are then related as follows

$$F_2(x, Q^2) = \sigma(\gamma^*p; W, Q^2)\left[\frac{Q^2}{Q^2 + 4M_p^2x^2}\right]\frac{Q^2(1 - x)}{4\pi^2\alpha}, \tag{5.72}$$

$$\sigma(\gamma^*p; W, Q^2) = \frac{F_2(x, Q^2)}{Q^2}\left[\frac{Q^2 + 4M^2x^2}{Q^2}\right]\frac{4\pi^2\alpha}{1 - x}. \tag{5.73}$$

Keeping Q^2 in denominator of (5.70) would mean replacing $(1 - x)$ in (5.72) by unity. Note that as the virtual photon goes on shell, $Q^2 \to 0$, with W kept fixed, $x \to 0$ and the square bracket in (5.72)-(5.73) approaches unity

$$\lim_{Q^2\to 0}\frac{Q^2 + 4M_p^2x^2}{Q^2} = \lim_{Q^2\to 0}\frac{\left(W^2 - M_p^2\right) + Q^2\left(Q^2 + 2W^2 + 2M_p^2\right)}{\left(W^2 - M_p^2\right) + Q^2\left(Q^2 + 2W^2 - 2M_p^2\right)} = 1. \tag{5.74}$$

Consequently $F_2(W, Q^2)$ must behave as $\mathcal{O}(Q^2)$ at small Q^2!

5.4   Electrons as tools for investigating nuclear structure

The general idea of scattering experiments was outlined in previous Section 1. Rutherford used α-particles for his studies, but already in 1914 Franck and Hertz used as first electrons for this purpose. Electrons were soon recognized as the best probes of hadrons because their electromagnetic interactions are well-understood. Since the commissioning Mark III Linac at Stanford University in 1953 they had been systematically used by a group led by Robert Hofstadter for the investigation of the structure of nuclei and later also of nucleons. These investigations earned Hofstadter the Nobel prize for physics in 1961.

In first experiments the structure of various nuclei was probed by means of the elastic scattering of electrons with energies up to 150 MeV. The results confirmed what was expected, namely that they have a *finite* size, but also showed that the boundary of nuclei is not sharp. In one of the first papers on this subject [86] several nuclei were investigated, among them gold, as in the classic Rutherford experiment. The apparatus used for these investigations was just a scaled up version of the arrangement used by Geiger and Marsden. The one used for the scattering on protons is shown in Fig. 5.2. Because of the heavy target, the differential cross section for elastic scattering on gold nuclei is given to a good approximation by the formula for scattering on an infinitely heavy charge, or equivalently, a fixed external source. Making use of the earlier results of Mott, Feschbach derived a simple formula [90] for the scattering of relativistic electrons on such heavy nuclei provided they behaved as a pointlike source of the Coulomb field. His result coincides to first order in α with the Mott cross section σ_{Mott}, but take into account higher orders of Born series. These are negligible for the proton or light nuclei, but become important for medium to heavy ones. For electron
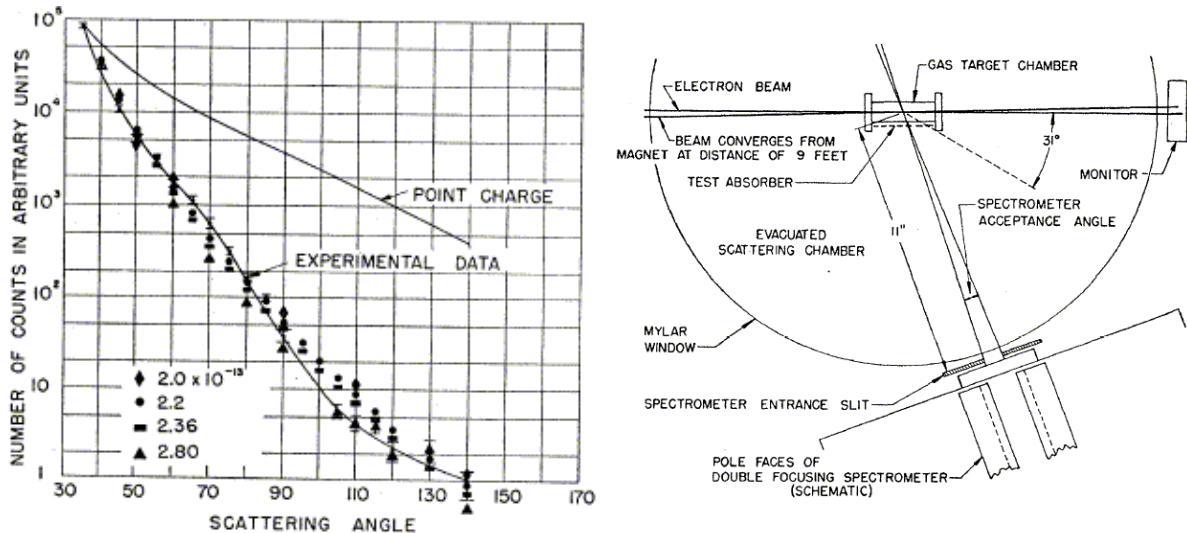
Figure 5.2: Left: The angular distribution of electrons scattered from a gold foil at 125 MeV as measured in [86]. The point charge calculation of Feschbach is indicated. Theoretical points based on the Born approximation for exponential charge distribution are shown for different values of a, demonstrating the sensitivity of angular distribution to change of the nuclei radius. All curves are normalized at 35°. Right: Schematic layout of the detector used by Hofstadter in the early electron-proton scattering experiment [87].

energies (125 and 150 MeV) used in these experiments the momentum transfer $Q = \sqrt{Q^2}$ was restricted to rather small values $Q \leq 0.1$ GeV, corresponding to distances of about 2 fm. Recall in this context that in 1911 Marsden and Geiger managed to restrict the size of gold nuclei to less than about 30 fm. The results obtained at SLAC and presented in Fig. 5.2 show large deviation from Feschbach prediction assuming pointlike gold nuclei, but are in very good agreement with the theory in which the pointlike cross section is multiplied by formfactors obtained in Born approximation for the "exponential" distribution $\rho(r) = \rho_0 \exp(-r/a)$ with $a = 2.3$ fm. This parameter provides some measure of the nuclei size, but the standard definition of the "radius" uses the so called root mean square (rms) radius defined as $r_{ch} \equiv \sqrt{\langle | r^2 | \rangle}$. For the exponential charge distribution $r_{ch} = \sqrt{12}a = 7.95$ fm.

5.5 Probing the structure of proton by elastic ep scattering

In 1955 Hofstadter and his group turned attention to the investigation of the structure of individual nucleons. Using the detector shown in Fig. 5.2 and beams of electrons with energies up to 200 MeV, they performed a series of experiments [87, 88] the results of which earned Hofstadter the Nobel prize for physics in 1961.

In essence, their experiments have shown that proton does not behave as a pointlike source of the Coulomb field, but rather as a particle with *finite* radius of the order of 1 fm. This conclusion was based on the observation (see left part of Fig. 5.3) that the differential cross section for scattering of relativistic electrons on protons differs from those corresponding to the pointlike proton ("Dirac curve") as well as those including the anomalous part of proton magnetic moment as given by the formula (5.31) with $F_1 = 1, F_2 = 1.793$ ("anomalous curve"). The fact that the latter curve is above the data had been interpreted as evidence that the proton "is not a point" [87]. Fitting the experimental distribution in the left plot of Fig. 5.3 to the Mott cross section multiplied by the formfactor (assumed the same for charge and magnetic moment) lead to the determination of root-mean-square radius of the proton $r_{ch} = r_{mag} = 0.7 \pm 0.24$ fm. Note that the available energy severely limited the momentum transfer squared to $Q^2 \leq 0.1$ GeV^2. In 1955 the electron energy was increased to 550 MeV, which allowed Hofstadter to extend the measurements into the region of momentum transfer $Q^2 \leq 0.5$ GeV^2 and substantially improve their precision. The main results concerned the measurement of the Q^2-dependence of proton formfactors, exemplified by the plot in Fig. 5.3, which confirmed the earlier measurement [87]. These investigations were completed in early sixties using 1000 MeV electron beam from the upgraded Mark III Linac [89]. Among other things the measurements showed that

FIG. 8. The proton form factor for various energies and momentum transfers as measured in early experiments using the 36-inch spectrometer facility at HEPL. The value of F^2 was calculated from the original Rosenbluth formula which defined form factors $F_1(Q^2)$ and $F_2(Q^2)$. F_1 corresponds to the form factor for a Dirac (spin-$\frac{1}{2}$) proton, and F_2 to the form factor for the anomalous magnetic moment. In the analysis of the data it was assumed that $F_1 = F_2$. At higher values of Q^2 it became evident that $F_1 \neq F_2$, but rather that $G_E = G_M/\mu_p$ for the proton, and the use of the G's then became universal.

Figure 5.3: Left: Comparison of data obtained in [87] with theoretical calculations described in the text. The best fit corresponds to exponential charge distribution with rms radius equal to 0.74 fm. Right: Q^2 dependence of proton formfactors as measured [88]. and plotted in units of inverse square fm, which means that the point 12.5 fm^{-2} corresponds to $Q^2 = 0.5$ GeV^2.

within the accuracy of the measurement the charge and magnetic formfactors coincide, as assumed in the earlier work [87, 88].

5.6 What does the deep inelastic scattering of electrons tell us about the structure of nucleons?

The successful operation of 1 GeV Linac at Stanford and fundamental importance of the results obtained there on the structure of nucleons had led to the decision to built a new electron linear accelerator with energies up to 20 GeV, using the infrastructure available at the newly created Stanford Linear Accelerator Center (SLAC). The primary aim of the new machine was the extension of the program of elastic scattering experiments and the investigation of quasi-elastic scattering, i.e. the electroproduction of resonances, like Δ, etc. Just for completeness the group also wanted to look at the inelastic continuum, which was inaccessible in previous experiments of Hofstadter. However, the planning of the machine was not influenced by any considerations that possible point-like substructure of the nucleon might be observed in electron scattering. The two miles long machine was commissioned in late 1966 and in 1967 the group of experimentalists from SLAC and MIT (Massachusetts Institute of Technology) started a series of experiments aiming primarily at the investigation of elastic electron proton scattering in much wider range of momentum transfer Q^2, but making possible also the first serious study of deep inelastic electron-proton scattering.

The schematic layout of the detector used in SLAC experiments, shown in Fig. 5.4, indicates the vast increase in size and complexity of the apparatus and the whole experimental arrangement. For the first time intensive on-line computing was used. On the other hand, the basic set-up of the measurements is basically the same as those used by Rutherford and Hofstadter. In particular the rotating spectrometer "arms", which consisted of various track chambers and magnets for momentum measurement, are closely reminiscent of the rotating microscope in Rutherford experiment (see Fig. 1.3). Only the dimensions and weight scaled up by approximately two orders of magnitude.

The first results on the elastic as well as inelastic electron-proton scattering were obtained in early 1968 and were reported at the IV International Conference on High Energy Physics in Vienna in July of that year by SLAC Director W. Panofsky and later published in [91, 92]. Some of these results are shown in Fig. 5.5.

85

Figure 5.4: Schematic layout of the spectrometer used in experiments with deep inelastic scattering of electrons on protons at SLAC.

Figure 5.5: The Q^2 dependence of the new SLAC experiments on the structure of proton a): elastic formfactor $G_M \equiv F_1 + \kappa F_2$, fitted by the dipole formula $(1/(1 + Q^2/0.71))^2$, and b): inelastic formfactor at fixed W.

Those on the elastic scattering confirmed the rapid decrease of the elastic formfactors [7] observed at lower Q^2 in earlier experiments, extending the accessible range up to $Q^2 \simeq 30$ GeV^2. In the part concerning the inelastic electron-proton scattering his talk contains also the first mention of the possibility "... *that these data might give evidence on the behavior of point-like charged structures in the nucleon.*"

However, this was not the prevailing point of view of theorists at that time. In fact there were few clear ideas about the expected Q^2-dependence of inelastic formfactors, i.e. the structure functions $F_1(x, Q^2)$ and $F_2(x, Q^2)$. Although the quark model had been formulated three years earlier and had some success in explaining the static properties of nucleons, almost nobody thought it might be relevant for the dynamics of high energy electron-proton scattering. To illustrate the prevalent mood of that time, let me quote from Gell-Mann's introductory talk at the XIII International Conference on High Energy Physics, held in Berkeley in Summer 1967, just at the time the first measurements of inelastic electron-proton collisions made by the MIT-SLAC group were under way:

> *We consider three hypothetical and probably fictitious spin 1/2 quarks, ... We make the known bound and resonant states of the mesons formally out of $q\bar{q}$ and the known bound and resonant baryon states out of qqq. .....*
>
> *Now what is going on? What are these quarks? It is possible that real quarks exist, but if so they have high threshold for copious production, many GeV; if this threshold comes from their rest mass, they must be very heavy and it is hard to see how deeply bound states of such heavy real quarks could look like $q\bar{q}$, say, rather than a terrible mixture of $q\bar{q}$, $qq\bar{q}\bar{q}$, and so on. Even if there are light real quarks and the threshold comes from very high barrier, the idea that mesons and baryons are made primarily of quarks is difficult to believe, since we know that, in the sense of dispersion theory, they are mostly, if not entirely, made up out of one another. The probability that a meson consists of a real quark pair rather than two mesons or a baryon and antibaryon must be quite small. Thus it seems to me that whether or not real quarks exist, the q or $\bar{q}$ we have been talking about are mathematical entities that arise when we construct representations of current algebra, which we shall discuss later on.....*
>
> *If the mesons and baryons are made of mathematical quarks, then the quark model may perfectly well be compatible with bootstrap hypothesis, that hadrons are made up out of one another.*

To understand the circumstances in which the SLAC experiments took place let me recall that in late sixties the only theoretical framework for the analysis of inelastic electron-proton scattering was based

---

[7]This behavior has a simple explanation, discussed already in Section 1. Imagine the target proton has some internal structure, i.e. its electric charge is distributed over some constituents. If the beam electron hits some of them and the collision is hard enough, the proton will tend to disintegrate, decreasing thereby the probability that the struck constituent and the proton remnant will recombine back to the proton state. The "harder" the collision, the less probable this recombination will be and hence the steeply falling elastic formfactor.

on the analytic properties of the scattering matrix (so called "analytic S-matrix" approach). Within this framework, and using rather heavy artillery of the so called "current algebra", it was possible to derive several relations involving measurable quantities. For instance, Bjorken derived the following restriction on the integrals over the structure functions $F_2^p(x, Q^2)$ and $F_2^n(x, Q^2)$

$$\frac{1}{M} \int \mathrm{d}\nu \left[W_2^p(\nu, Q^2) + W_2^n(\nu, Q^2) \right] = \int \frac{\mathrm{d}x}{x} \left[F_2^p(x, Q^2) + F_2^n(x, Q^2) \right] \geq \frac{1}{2}. \tag{5.75}$$

Callan and Gross derived the sum rule for similar combination, but without the $1/x$ factor

$$\frac{Q^2}{2M^2} \int \frac{\mathrm{d}\nu}{\nu} \left[W_2^p(\nu, Q^2) + W_2^n(\nu, Q^2) \right] = \int \mathrm{d}x \left[F_2^p(x, Q^2) + F_2^n(x, Q^2) \right] \leq \frac{1}{2}, \tag{5.76}$$

which should hold in the limit $Q^2 \to \infty$. In both cases no assumption concerning the structure of the nucleon was made. Many theorists who, like Bjorken, made vital contributions to theoretical analysis of deep inelastic electron-nucleon scattering, actually expected the data on deep inelastic electron-proton scattering to rule out the quark model for ever! It is certainly revealing that just at the time the first experiment at SLAC started data taking, James Bjorken closed his talk at the International Symposium on Electron and Photon Interactions that took place directly in Stanford in September 1967 with the words:

> But the indication is that there does not yet seem to be any large cross sections which this model of point-like constituents suggests. Additional data is necessary and very welcome in order to completely destroy the picture of elementary constituents.

which were echoed in the Kurt Gottfried's comment following Bjorken's talk:

> "I think Professor Bjorken and I constructed sum rules in hope of destroying the quark model."

The basic open question concerned the behavior of structure functions $F_1(x, Q^2)$ and $F_2(x, Q^2)$ as functions of Q^2 for W^2 outside the resonance region, i.e. for $W^2 = Q^2(1-x)/x \gtrsim 2$ GeV^2, in the limit $Q^2 \to \infty$. Will they vanish as rapidly as the elastic formfactors or could they eventually approach nonzero functions of x? Shortly before the first data from SLAC appeared Bjorken had suggested [94] the latter possibility, hence usually called "Bjorken scaling", but his derivation was based on several questionable assumptions within the framework of current algebra and few other theorists, perhaps including Bjorken himself, took it seriously. With the benefit of hindsight we know that his argumentation was, indeed, flawed, as within Quantum Chromodynamics the structure functions $F_i(x, Q^2)$ vanish for any fixed $x > 0$ when $Q^2 \to \infty$. However, and this is the crucial point, they do so very slowly, much more slowly than the elastic formfactors of nucleons! Although Bjorken's suggestion proved to be wrong if taken literally, it prepared the ground for the concept of approximate scaling, which was observed at SLAC and which became the starting point for Feynman's parton model to be discussed below.

In fact, had the theorists in late sixties taken the idea that nucleons are in some sense composite objects, the slow Q^2-dependence of structure functions describing the deep inelastic electron-proton scattering could be expected on quite general grounds. Note that for large Q^2 the steep dependence of elastic formfactors $G_E(Q^2) \simeq G_M(Q^2) \propto 1/Q^4$ (implying $1/Q^8$ dependence of elastic cross sections!) can be easily understood as follows. Even for very large Q^2 it is in principle possible that the momentum transferred to the nucleon will dissipate among its constituents in such a way that the nucleon will not break up but will recoil as whole but this process will be very rare. Hence the Q^2-dependence of G_M and G_E. In most cases the nucleon experiencing hard collision with an electron will break up leading to final states with typically many particles. The probability of any specific final state will again depend on the mechanism of recombination of the constituents from the broken nucleon into this final state and will thus also be strongly Q^2 suppressed. However, if we sum over all hadronic final states by measuring the inclusive inelastic electron-proton cross section (5.35), (5.37) or (5.36), there is no reason to expect any strong Q^2 dependence!

Despite this low expectations, the data on deep inelastic scattering, some of which are reproduced in Fig. 5.6, turned out to be both surprising and of far-reaching consequences. In recognition of their importance for the development of parton model and eventually the formulation of Quantum Chromodynamics, three of their authors, R. Taylor, H. Kendall and J. Friedman were awarded the 1990 Nobel Prize for Physics [8]. The main features of these results are the following (the notation used in Fig. 5.6 is related to that used in this text as follows: $\nu W_2 = F_2, \omega = 1/x, W^2 = -Q^2 + M_p^2 + 2M\nu = Q^2(1-x)/x + M_p^2$):

$$W^2 = 2m_p\nu - Q^2 + M_p^2$$

$$W^2 = Q^2(\omega - 1)$$
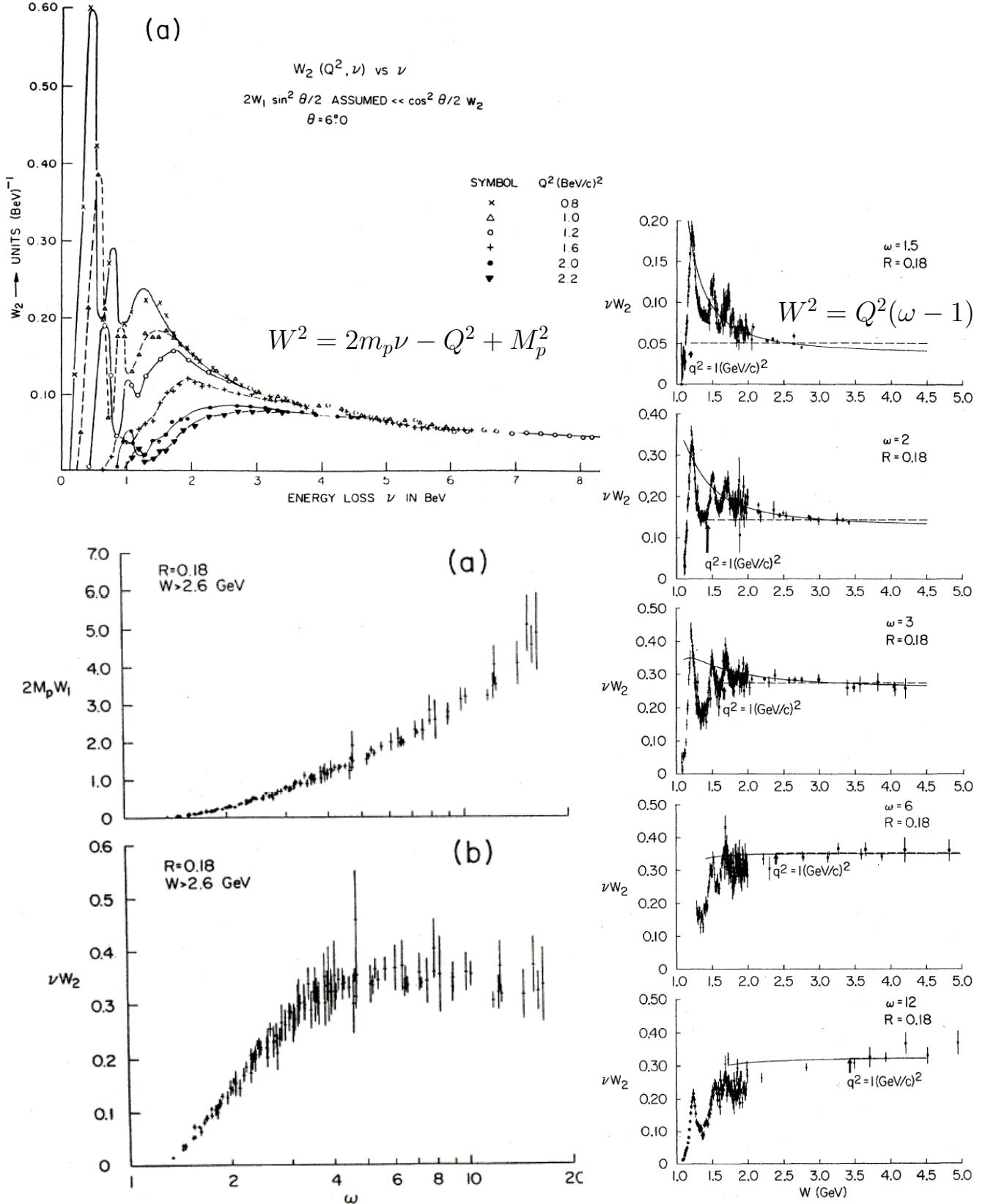
Figure 5.6: SLAC results, taken from [91–93].

- Varying the primary energy of the electron beam allowed the MIT-SLAC group to determine, for a fixed pair x, Q^2, separately the two structure functions $F_1^p(x, Q^2)$ and $F_2^p(x, Q^2)$. The measured value of the ratio (5.62) $R = 0.18$ was then used, neglecting its possible x and Q^2 dependence, in the following considerations.

- In the resonance region, i.e. for $W^2 \simeq 1 - 4$ GeV, the structure function $F_2(W^2, Q^2)$ rapidly falls off with increasing Q^2, similarly as the elastic formfactor $F_2(Q^2)$. Three resonances are clearly visible in the upper left plot of Fig. 5.6, the first corresponding to $\Delta(1236)$. Their positions as functions of ν move with increasing Q^2 to the right reflecting the fact that $\nu = (W^2 + Q^2)/2M_p$.

- On the other hand, in the continuum region, i.e. roughly for $W \gtrsim 2$ GeV, $F_2^p(x, Q^2)$ varies with Q^2 only very slowly. This is shown in the right part of Figs. 5.5 and 5.6.

- In this region the data are moreover consistent with the Bjorken scaling hypothesis. This is suggested by the right plot of Fig. 5.6 which shows the approximate constancy of $F_2(\omega, W^2)$ considered as a function of $W^2 = Q^2(\omega - 1) + M_p^2$ for fixed $\omega = 1/x$, and in the lower left plot as the approximate Q^2-independence of $F_2(x = 1/\omega, Q^2)$ considered as a function of $\omega = 1/x$.

- Integrating over the accessible kinematical region of both x and Q^2 gave the following values for two important sum rules mentioned above.

$$I_1^p = \int_{0.05}^1 \frac{\mathrm{d}x}{x} F_2^p(x, Q^2) = 0.78 \pm 0.04, \quad I_2^p = \int_{0.05}^1 \mathrm{d}x F_2^p(x, Q^2) = 0.172 \pm 0.009. \tag{5.77}$$

5.7 Emergence of the parton model

The first MIT-SLAC results on deep inelastic electron-proton scattering, presented at the Vienna Conference in Summer 1968, had naturally attracted the attention of Feynman, Bjorken and other theorists working in this field. Nevertheless, their analysis and interpretation had been almost entirely the work of Feynman. During several visits to SLAC in Autumn 1968 he had developed the basic ideas of the **parton model**, using the intuitive language of parton distribution functions (PDF), which remain the basic theoretical concept also in the Quantum Chromodynamics. Interestingly, there are no written records of Feynman's visits to SLAC and his original explanation of SLAC data, but there is no doubt that it was he who formulated and worked out the parton model during these visits. The Nobel Laureates [8], as well as Bjorken [98] and other theorists who had taken up Feynman's ideas and further developed them, explicitly acknowledge his authorship. Feynman himself gave a comprehensive formulation of his model in the book [4].

The first published paper on the parton model and his application to the analysis of deep inelastic scattering of leptons on nucleons was written by Bjorken and Paschos in April 1969 [98], after the preliminary SLAC results were presented at the Vienna Conference but five months before they were published in two letters [91, 92]. The basic idea of Feynman's parton model is to represent the inelastic electron-nucleon scattering as quasi-free scattering from point-like constituents within the proton when viewed from a frame in which the proton has infinite momentum and provided the collision is deeply inelastic, i.e. $Q^2 \gg M_p^2$. The electron-proton center of mass system is at high energies a good approximation of such a frame. Note that the choice of a preferred reference frame (or, better, family of frames) is at the heart of the parton model, despite the fact that any frame is in principle as good as this one and we use manifestly relativistic invariant variables like x, y or Q^2. For instance, parton model cannot be easily formulated in proton rest frame. In an infinite momentum frame the transverse momenta p_T of proton constituents induced by the internal "Fermi motion", characterized by $p_T \lesssim M_p$, as well as the masses of nucleons and partons can be neglected with respect to the longitudinal one and we can write the parton four-momentum as $p = \eta P$. In the deep inelastic electron-proton collisions these transverse components can be safely neglected also in the evaluation of the corresponding cross section, which can be written as an **incoherent sum** of cross sections from elastic scattering on individual charged partons. Note that momentum conservation at the lower vertex in Fig. 5.1c,

$$(p')^2 = (p + q)^2 = p^2 + q^2 + 2pq \Rightarrow 2pq = -q^2 = Q^2 \tag{5.78}$$

implies that the variable x in (5.5) coincides within the parton model with fraction η introduced above.

The above description of the scattering of electrons from individual partons is all we need for the calculation of differential cross sections of inclusive electron-proton scattering, which depend on the kinematical
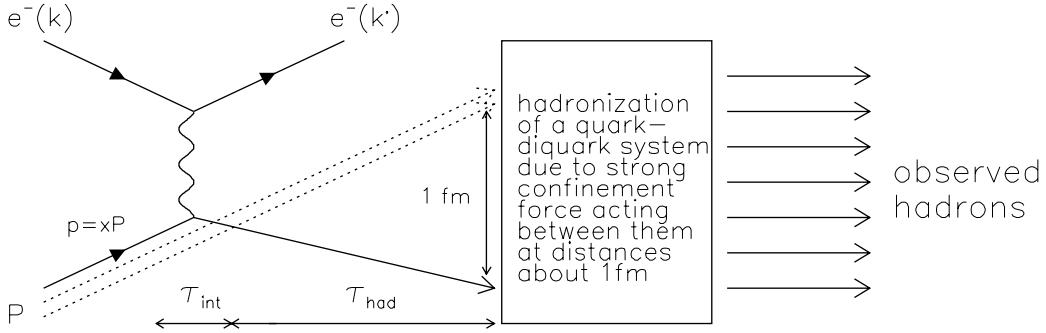
Figure 5.7: Deep inelastic scattering on a real proton in the parton model. The incoming electron scatters on a single parton, while the remaining diquark flies unperturbed on, until the forces of confinement, acting between them at distances larger than about 1 fm, cause their conversion into observable hadrons. For this picture to be consistent the time separation τ_{had} between these two stages must be significantly larger than the characteristic time τ_{int} of the electron–parton scattering.

variables describing the scattered electron only. To make the model selfconsistent Feynman had also added the basic mechanism of **hadronization** i.e. the conversion of final state partons into observable hadrons, to be described in Section 5.11. In his picture of deep inelastic scattering this process occurs *much later* than the scattering of electrons from partons and *does not* therefore influence it. The overall picture of the collision is sketched in Fig. 5.7.

Consider now the differential cross section $\mathrm{d}\sigma/\mathrm{d}x\mathrm{d}Q^2$ as given in (5.37) at large energies. i.e.

$$\frac{\mathrm{d}\sigma}{\mathrm{d}x\mathrm{d}Q^2} = \frac{4\pi\alpha^2}{Q^4}\frac{F_2(x,Q^2)}{x} \underset{kp\to\infty}{\longrightarrow} \frac{4\pi\alpha^2}{Q^4}\frac{F_2(x,Q^2)}{x}, \tag{5.79}$$

where only one structure function appears. Noting that in this limit the differential cross section of elastic scattering of electrons from a pointlike fermion or boson of electric charge e_p, given in (5.29) for unit electric charge, equals

$$\frac{\mathrm{d}\sigma}{\mathrm{d}Q^2} = \frac{4\pi\alpha^2 e_p^2}{Q^4}\left[1 - y + \varepsilon\frac{y^2}{2} - \frac{M_p y}{2kp}\right] \underset{kp\to\infty}{\longrightarrow} \frac{4\pi\alpha^2 e_p^2}{Q^4} \tag{5.80}$$

and is thus the same for both types of partons, Feynman wrote down $F_2(x,Q^2)/x$ in terms of the functions $d_i(x)$ describing the probability distribution to find inside the proton parton species i with fraction x of proton momentum and electric charge e_i (in units of the positron charge) as

$$\frac{F_2(x)}{x} = \sum_i e_i^2 d_i(x) \Rightarrow F_2(x) = x\sum_i e_i^2 d_i(x) \tag{5.81}$$

Approximate scaling, i.e. independence of the measured $F_2(x,Q^2)$ of Q^2 is clearly indispensable for this expression to make sense. Various models for the parton distribution functions had been proposed starting from the assumption that proton can be imagined as a superposition of states with different, but fixed, number of partons. In the simplest of them one assumes that in the N-parton configuration, which appears in the proton with the probability $P(N)$, all partons carry the same momentum fraction $x_N = 1/N$ and thus

$$F_2(x) = \sum_N P(N)\frac{\delta(x - 1/N)}{N}\langle\sum_{i=1}^N e_i^2\rangle \Rightarrow \int_0^1 \mathrm{d}x F_2(x) = \sum_N P(N)\frac{\langle\sum_{i=1}^N e_i^2\rangle}{N}, \tag{5.82}$$

where $\langle\sum_{i=1}^N e_i^2\rangle/N$ is the mean square electric charge in the N-parton configuration. In this model the integral I_2^{p} defined in (5.77) can thus the interpreted as mean square charge per parton in the proton. In fact this interpretation holds in the whole class of models of distribution functions $d_i(x)$ in which the N-parton distribution function $f(x_1, x_2, ..., x_N)$ is a symmetric function of its arguments [98].

In the above considerations, holding strictly speaking for infinite collision energy $\sqrt{S}$, only $F_2(x,Q^2)$ structure function appears. At finite energies also $F_1(x,Q^2)$ does contribute and one may attempt to extract
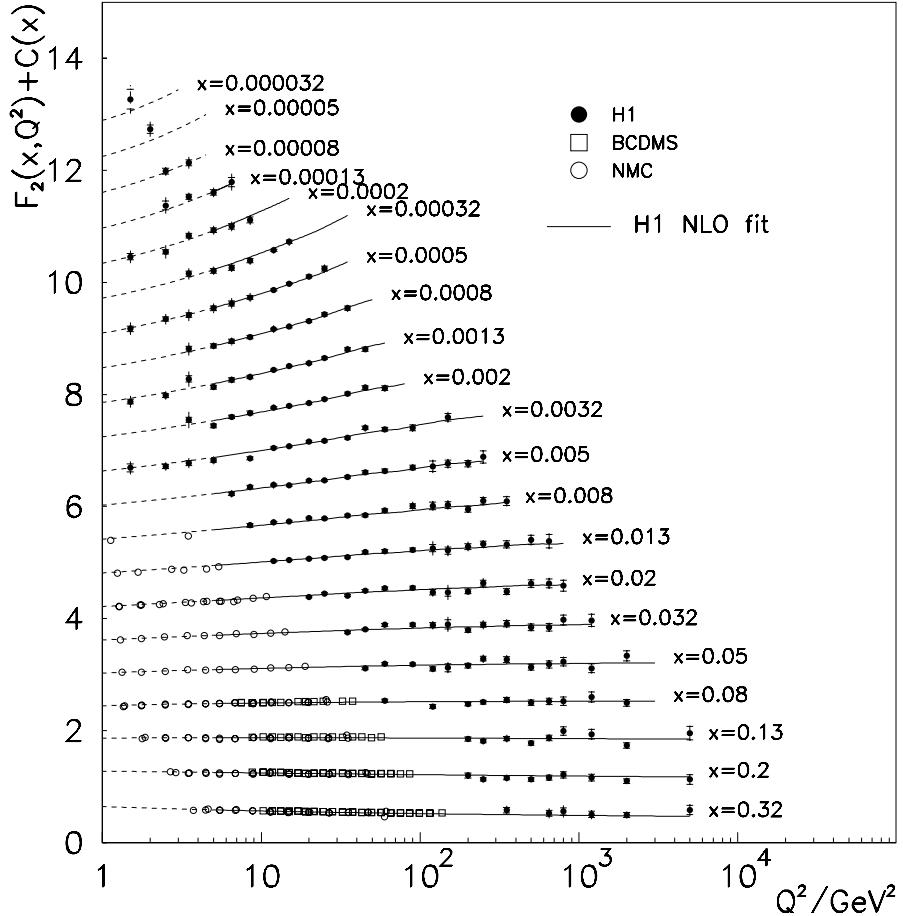
Figure 5.8: The Q^2 dependence of the proton structure function $F_2^{ep}(x, Q^2)$ for several fixed values of x. Old SLAC data are complemented by the recent results from the H1 Collaboration at DESY.

both of them from experiment, using the relation (5.67), which requires performing the measurements for fixed x, Q^2 at two energies. This has, indeed, been done at SLAC with the results $R = 0.18$, which implies that to a good approximation the measured structure functions satisfy the so called **Callan-Gross relation** $F_2(x) = 2xF_1(x)$. This in turn means that charged partons are fermions with spin 1/2! Note that for scalar partons (5.28) implies $F_1(x) = 0$, which is clearly ruled out by data. The same holds for vector, i.e. spin 1, gluons, which provides one of the experimental facts for ruling out the Hahn-Nambu model [75] of integral charge colored quarks.

Comparing formula (5.28) for the scattering on pointlike fermion with spin 1/2 with the general form (5.36) for the scattering on the proton we see that the last, mass terms, coincide provided we identify $M = xM_p$, i.e. the rest mass of interacting quarks should depend on its momentum! This conceptually unacceptable conclusion suggests that the formulation of the parton model as sketched above is strictly speaking satisfactory only for the *massless* protons and partons, or in the so called **infinite momentum frame**, where all particles are moving with infinite momentum. In practice, however, the parton model is phenomenologically relevant any time Q^2, which measures the "hardness" of the collision between the electron and proton, is much larger than the nucleon and quark masses, i.e. when $Q^2 \gg m_{\mathrm{p}}, m_q$.

The spin 1/2 nature of charged partons, suggested by data, immediately raised the question of their relation to the **constituent quarks** of the additive quark model, discussed in Chapter 4. Although there is, indeed, a close relation between them, these two concepts are nevertheless *not* identical. Nevertheless, it has become generally accepted practice to call the original parton model of Feynman in which the basic charges constituents of nucleons are identified as far as their quantum numbers are concerned with quarks,

the *Quark Parton Model* and denoted QPM [8].

The fact that the QPM assumes that electrons scatter on *free* quarks, despite the fact that no quarks have so far been observed in the nature and thus one would expect them to be *tightly* bound inside the proton (and other hadrons), may seem difficult to understand, and has originally been the source of suspicion of many physicists. The reason why the QPM nevertheless works so well will be explained in the Section 7.

The measured structure functions F_1, F_2 are in fact not exactly Q^2-independent, as assumed in the QPM, but depend, as indicated already by the original SLAC measurements, also weakly on Q^2. An example of the one of the most recent measurements of Q^2 dependence of the nucleon structure functions $F_2(x, Q^2)$ in electromagnetic interactions is shown in Fig. 5.8. This weak Q^2 dependence will be ignored in this chapter. We shall return to it again in Chapter 10, which deals primarily with the description of these *scaling violations*.

The fact that experimental value of the integral $I_2^{\mathrm{p}} = 0.17 \pm 0.009$, representing in the parton model the mean square charge per parton in the proton, was about half the value $1/3 = (4/9 + 4/9 + 1/9)/3$ expected in the simplest form of the parton model had for some time been a cause for concern. For instance, Bjorken and Paschos [98] found that *"the observed cross section is uncomfortably small"* even if the three quarks were supplemented by the sea of quark-antiquark pairs. The data clearly indicated that there must be also electrically neutral partons in the proton but it took another year, until the invention of QCD in 1973, to realize it.

5.8 Parton distribution functions and their basic properties

In the rest of this text the distribution functions $d_i(x)$ of current quarks inside the proton will be denoted as $q(x)$, $q = u, d, s, c$ and as $\overline{q}(x)$ for the corresponding antiquarks. Note that isospin symmetry allows us to use these distribution functions for the description of scattering on neutron as well. The analyses of experimental data as well as theoretical considerations suggest that the concept of parton distribution function does not have to be introduced for bottom and top quarks, which are too heavy for that purpose. For proton and neutron the quark distribution functions have been measured in a number of experiments with beam electrons, muons and (anti)neutrinos, using various techniques. It took a lot of work to extract from vast amount of experimental data on structure functions F_1, F_2 distribution functions of individual current quarks, primarily because the former are always combinations of the latter, for instance

$$F_2^{\mathrm{ep}}(x) = x \left(\frac{4}{9}[u(x) + \overline{u}(x) + c(x) + \overline{c}(x)] + \frac{1}{9}[d(x) + \overline{d}(x) + s(x) + \overline{s}(x)] \right), \qquad (5.83)$$

$$F_2^{\mathrm{en}}(x) = x \left(\frac{4}{9}[d(x) + \overline{d}(x) + c(x) + \overline{c}(x)] + \frac{1}{9}[u(x) + \overline{u}(x) + s(x) + \overline{s}(x)] \right). \qquad (5.84)$$

A particularly useful combination is the average of (5.83), (5.84)

$$F_2^{\mathrm{eN}} \equiv \frac{1}{2}\left(F_2^{\mathrm{ep}} + F_2^{\mathrm{en}}\right) = \frac{5}{18}x \underbrace{\left[u(x) + d(x) + \overline{u}(x) + \overline{d}(x) + s(x) + \overline{s}(x) + c(x) + \overline{c}(x)\right]}_{\Sigma(x)} - \frac{1}{6}xk(x), \qquad (5.85)$$

where $k(x) \equiv s(x) + \overline{s}(x) - c(x) - \overline{c}(x)$ is numerically negligible with respect to $\Sigma(x)$. The above combination (5.85) corresponds to the **isoscalar** target and is measured in experiments using nuclear targets.

In going from (5.83) to (5.84) we exploited the *isospin* symmetry of quark distribution functions (u-quark in the proton is the same as d-quark in the neutron and vice versa) and furthermore assumed that the $s, c, \overline{s}, \overline{c}$ distributions are the same in protons and neutrons. Fig. 5.9 displays quark and antiquark distribution functions (multiplied by x) at $Q^2 = 100$ GeV^2 as extracted by three recent *global analyses*, and one older parameterization (denoted HMRSB) obtained by the MRS group about a decade ago. In global analyses, which will be described in more detail in Section 10, one attempts to describe a wide range of physical process with the same set of parton distribution functions. The marked improvement in the determination of PDF from experimental data achieved over the last decade has been the consequence

---

[8]The charged, spin 1/2 partons, introduced by Feynman, are usually called **current** quarks (of definite flavor) to emphasize this difference. Compared to the constituent quarks they are light (about 10 MeV for u, d quarks compared to roughly 300 MeV for the constituent ones) and there is no fixed number of them inside the proton (see discussion below).

of several facts: improvement in the precision and scope of experimental data, progress in perturbative calculations and closer collaboration between experimentalists and theorists in the extraction procedure. There are two important features of quark distribution functions (in Fig. 5.9 shown for $Q^2 = 100$ GeV^2) which merit further comments:

- For $x \to 0$ both quark and antiquark distribution functions behave roughly as $1/x$ causing the integrals

$$\int_0^1 q(x)\mathrm{d}x \tag{5.86}$$

  to *diverge*. This implies that the number of charged partons in the proton is infinite. This at first sight unwelcome conclusion is, however, easy to understand and presents no serious problem for the QPM. To see why, consider the so called **valence** distribution functions, defined as the differences between quark and corresponding antiquark distributions

$$q_{val} \equiv q(x) - \overline{q}(x); \quad q = u, d, s, c \tag{5.87}$$

  It is these valence distribution functions which provide a bridge between quark-partons of the QPM and the quarks of the old additive quark model. The data on u_{val} and d_{val} indicate that these functions are *integrable* and the values

$$\int_0^1 u_{val}(x)\mathrm{d}x \doteq 2; \quad \int_0^1 d_{val}(x)\mathrm{d}x \doteq 1 \tag{5.88}$$

  moreover consistent with the prediction of the additive quark model. One can test these relations by measuring the difference of structure functions $F_2^{ep} - F_2^{en}$ and evaluating the integral

$$GS \equiv \int_0^1 \frac{F_2^{\mathrm{ep}}(x) - F_2^{\mathrm{en}}(x)}{x}\mathrm{d}x = \frac{1}{3}\int_0^1 (u_v(x) - d_v(x))\mathrm{d}x - \frac{2}{3}\int_0^1 (\overline{d}(x) - \overline{u}(x))\mathrm{d}x, \tag{5.89}$$

  which should be equal to 1/3, if $\overline{u}(x) = \overline{d}(x)$ is assumed, as seems natural in view of the isospin symmetry. Nevertheless a recent measurement by the NMC Collaboration at CERN yields GS=0.235±0.026, which is clearly inconsistent with the expected 1/3. The simplest and currently favored interpretation of this discrepancy is the *isospin violation* in the sea quark distributions.

- The integral

$$\int_0^1 xq(x)\mathrm{d}x; \quad q = u, d, s \tag{5.90}$$

  represents the fraction of the proton momentum carried by the quark q and consequently the quantity

$$\int_0^1 x\Sigma(x)\mathrm{d}x \doteq \frac{18}{5}\int_0^1 F_2^{\mathrm{eN}}(x)\mathrm{d}x, \tag{5.91}$$

  gives the fraction of the proton momentum carried by all u, d, s and c quarks and antiquarks together. The experimental value, roughly 0.5, signals that there must be other, electrically neutral, partons in the proton, carrying the remaining half of its momentum. Soon after the experimental measurement of (5.91) these partons were identified with the **gluons**, to be discussed in the section on QCD.

Our present understanding of the origin of antiquarks and gluons inside the proton is indicated in Fig. 5.10. Observed with small resolution power like, for instance, in the small momentum transfer scattering, the proton behaves as a system of three massive **constituent** quarks, bound together by a static potential, which approximates the result of complicated multigluon exchanges mediating the force between quarks. This potential cannot yet be reliably calculated within QCD from the first principles but the results obtained so far using the techniques of lattice gauge theory are encouraging. With higher resolution power, provided by the hard scattering processes, the picture changes and the static potential gives rise to gluons which in turn convert to pairs of quarks and antiquarks, which again radiate gluons and so on and so forth. All the created quarks and gluons are **virtual** and thus have to recombine after some time, but if their virtuality
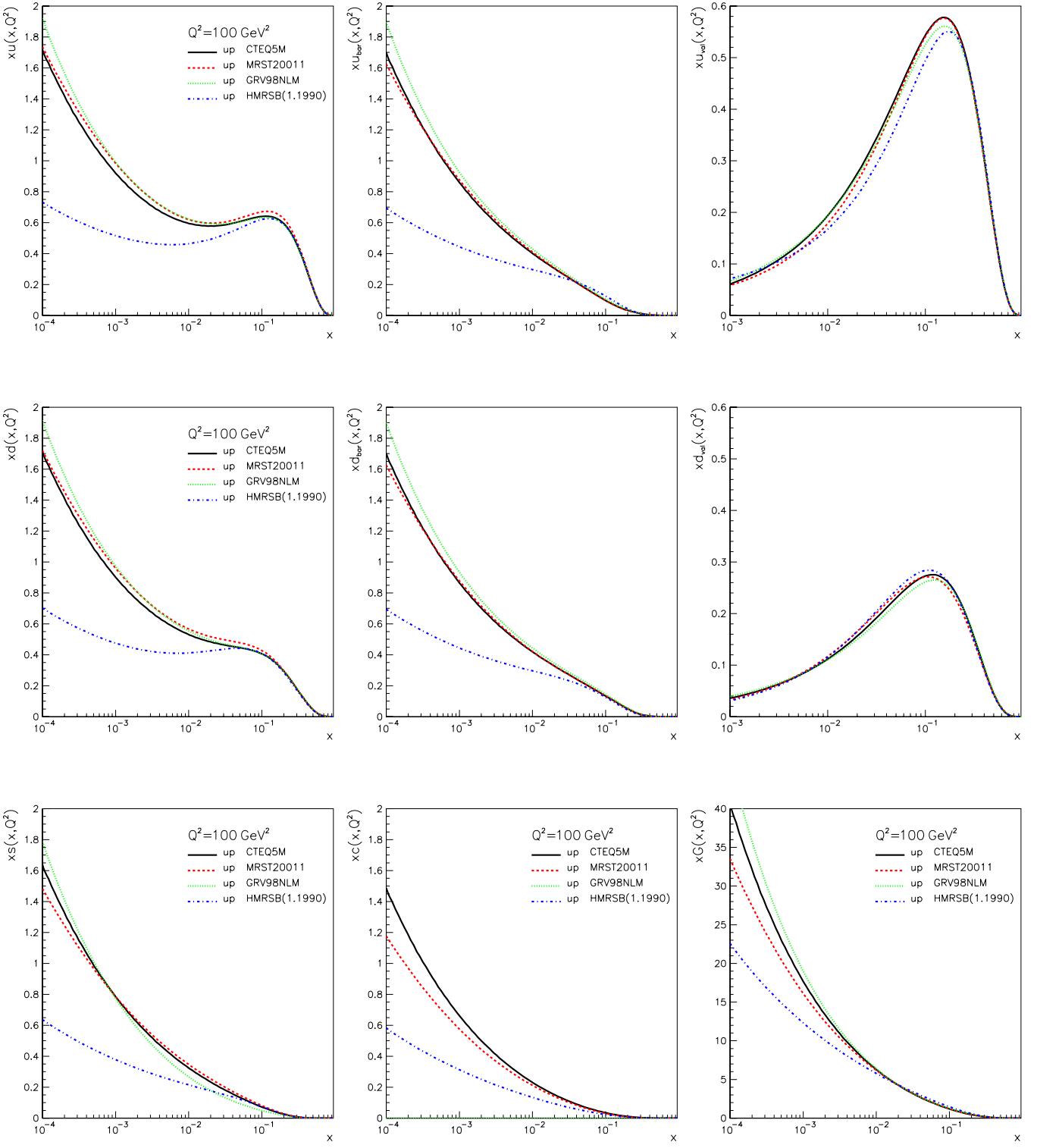
Figure 5.9: Parton distribution functions in the proton for $Q^2 = 100$ GeV^2 as extracted by four different parameterizations, taken from the CERN PDFLIB library of distribution functions. The first three are the recent ones obtained by CTEQ, MRS and GRV groups, the last one (dash-dotted curves) corresponds to a MRS parametrization more than a decade old.
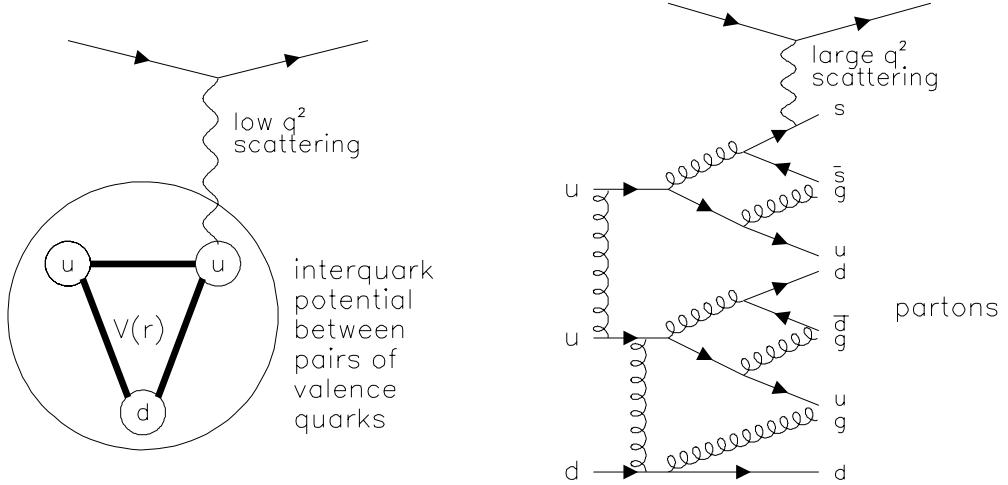
Figure 5.10: Field theoretic "explanation" of the generation of partons from valence quarks interacting via the exchange of gluons, carriers of strong force.

[9] is small compared to $\sqrt{Q^2}$, where Q^2 is some measure of the "hardness" of the collision, they will "live" long enough for the probe electron to scatter on them (nearly) as on the real ones.

This intuitive explanation will be corroborated by more quantitative arguments and calculations in Section 6. In this picture distribution functions of quarks u and d (in the proton) can be divided into the **valence** parts, defined in (5.87), and the **sea** parts, which are by definition identical to the antiquark distribution functions

$$
\begin{array}{llll}
\overline{u}(x) & = & \overline{u}_{sea}(x) & = u_{sea}(x), \\
\overline{d}(x) & = & \overline{d}_{sea}(x) & = d_{sea}(x).
\end{array}
\tag{5.92}
$$

Note that strange and charmed quark distribution functions of the proton have only sea components. Isospin symmetry is usually assumed to hold for the sea distributions. On the other hand, the strange sea in the proton is expected to be suppressed with respect to u_{sea}, due to bigger current mass of the s quark.

5.9   Parton model in neutrino interactions

With the advent of intense beams of ν_μ and $\overline{\nu}_\mu$ at CERN and Fermilab in the early seventies the deep inelastic scattering of neutrinos on nucleons in the **charged current** (CC)

$$
\nu_\mu + N \rightarrow \mu^- + \text{anything}
\tag{5.93}
$$

as well as **neutral current** (NC) channels

$$
\nu_\mu + N \rightarrow \nu_\mu + \text{anything}
\tag{5.94}
$$

has opened new ways of testing the parton model. Because weak interactions *do distinguish* quarks from antiquarks, these processes contributed decisively to the extraction of distribution functions of individual quark flavors.

In the Standard Model the above processes are mediated by the exchange of intermediate vector bosons $W^\pm$ and Z. The relevant lowest order diagrams are shown in Fig. 5.11. In the following I discuss in some detail the charge current processes (5.93) only. In most of the existing applications the exchange of vector bosons $W^\pm$ can be replaced by the **effective four–fermion vertex** of the original Fermi theory [17] [10] which takes the form of product

$$
\frac{G_F}{\sqrt{2}} J_\mu^{(l)} \left(J^{(h)\mu} \right)^+ + \text{h.c.,}
\tag{5.95}
$$

---

[9]Measured by the "off–mass shellness" $k^2 - m^2$, where k and m are the momentum and mass of the virtual particle.

[10]This is a good approximation so long as the momentum transfer at the leptonic vertex in Fig. 5.11 is much smaller than its mass, i.e. $Q^2 \ll M_W^2$.
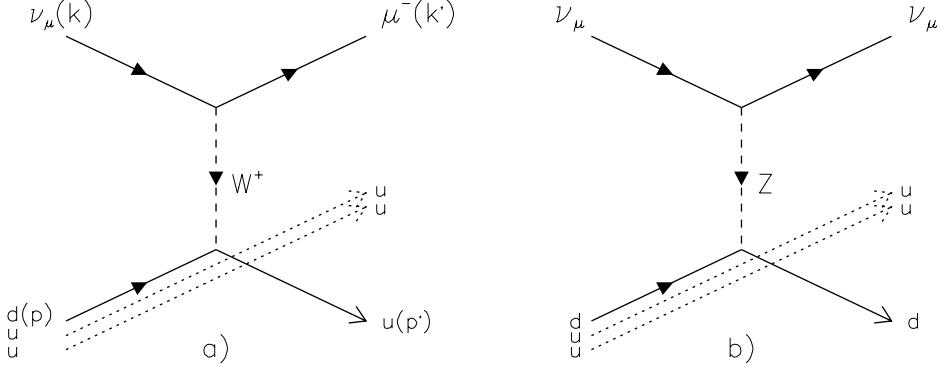
Figure 5.11: Lowest order Feynman diagrams describing the weak interaction of neutrinos with the exchange of W^+ (a) and Z (b) intermediate vector bosons.

of the leptonic and quark currents $J_\mu^{(l)}$, $J_\mu^{(h)}$

$$J_\mu^{(l)} = \sum_{l,l'} \overline{u}_{l'}\gamma_\mu(1-\gamma_5)u_l + \sum_{l,l'} \overline{v}_l\gamma_\mu(1-\gamma_5)v_{l'}, \tag{5.96}$$

$$J_\mu^{(h)} = \sum_{q,q'} \overline{u}_{q'}\gamma_\mu(1-\gamma_5)V_{qq'}u_q + \sum_{q,q'} \overline{v}_q\gamma_\mu(1-\gamma_5)V_{q'q}v_{q'}, \tag{5.97}$$

where the sums run over the pairs of leptons l, l' and quarks q, q' for which $\Delta_l \equiv e_l - e_{l'} = \Delta_q \equiv e_q - e_{q'} = +1$. The *unitary* matrix $V_{qq'}$, called, for two fermion generations, Cabbibo matrix, describes the **mixing** of quarks (antiquarks) with electric charge $-1/3$ $(1/3)$ [17]. The unitarity of this matrix implies that for *fully inclusive* quantities, i.e., so long as we *sum* over all final states in (5.93), $V_{q,\overline{q}}$ can be replaced in (5.97) by a unit matrix.

In (5.96) and (5.97) I have explicitly written out the Dirac bispinors u, v which describe the in– and outgoing states of leptons and quarks (u) and their antiparticles (v), participating in the basic parton level scattering process:

$$l(k) + q(p) \rightarrow l'(k') + q'(p'). \tag{5.98}$$

As in the previous sections, we shall be interested in cross-sections of *unpolarized* leptons and nucleons only. Technically the spin averaging is slightly more complicated due to the presence of the γ_5 matrix. Consider first the scattering of a *massless* neutrino on the quark with mass M. The spin averaged matrix element squared corresponding to one particular term of (5.95) is proportional to the contraction of leptonic and quarkonic tensors

$$L_{\mu\nu}^{(1)}L_{(2)}^{\mu\nu} = \frac{1}{2}\text{Tr}\left[\not{k}'\gamma_\mu(1-\gamma_5)\,\not{k}\gamma_\nu(1-\gamma_5)\right]\frac{1}{2}\text{Tr}\left[(\not{p}'+M)\gamma^\mu(1-\gamma_5)(\not{p}+M)\gamma^\nu(1-\gamma_5)\right]. \tag{5.99}$$

Anticommuting the γ_5 matrix in the above trace we find that the first, i.e. leptonic, tensor is given as

$$\text{Tr}\left[\not{k}'\gamma_\mu\,\not{k}\gamma_\nu(1-\gamma_5)\right] = 4\left[k'_\mu k_\nu + k'_\nu k_\mu - g_{\mu\nu}(kk') - \mathrm{i}\epsilon_{\mu\nu\gamma\delta}k'^\gamma k^\delta\right], \tag{5.100}$$

where $\epsilon^{\mu\nu\alpha\beta}$ is *fully antisymmetric* tensor of rank four. Although the quark mass M is explicitly present in the expression for the corresponding quark tensor, it disappears once the trace in the quarkonic tensor is worked out:

$$\frac{1}{2}\text{Tr}\left[(\not{p}'+M)\gamma^\mu(1-\gamma_5)(\not{p}+M)\gamma^\nu(1-\gamma_5)\right] = \text{Tr}\left[\not{p}'\gamma^\mu\,\not{p}\gamma^\nu(1-\gamma_5)\right] + \frac{M^2}{2}\underbrace{\text{Tr}\left[\gamma^\mu\gamma^\nu + \gamma^\mu\gamma_5\gamma^\nu\gamma_5\right]}_{0} \tag{5.101}$$

and we get, with obvious substitutions $k \rightarrow p$, $k' \rightarrow p'$, exactly one half the result for the leptonic tensor. Contracting the lepton and parton tensors using the identity

$$\epsilon_{\mu\nu\gamma\delta}\epsilon^{\mu\nu}{}_{\alpha\beta} = 2(g_{\alpha\delta}g_{\beta\gamma} - g_{\alpha\gamma}g_{\beta\delta}) \tag{5.102}$$

and employing standard Mandelstam variables for the parton level subprocess

$$s \equiv (k+p)^2, \quad t \equiv (k-k')^2 = (p-p')^2, \quad u \equiv (k-p')^2 = (k'-p)^2 \tag{5.103}$$

we get

$$L^{(1)}_{\mu\nu} L^{\mu\nu}_{(2)} = 32 \left[(kp)^2 + (k'p)^2 + (kp)^2 - (k'p)^2 \right] = 64(kp)^2 = 16 \left(s - M^2 \right)^2, \tag{5.104}$$

where the last two terms came from the contraction (5.102) and are therefore due to the presence of the γ_5 matrix. Notice that only this part can be different for particles and antiparticles. Moreover as this term comes from the contraction of two *antisymmetric* tensors, it is immediately clear what will change when we consider other three combinations. Since the switch to antiparticles, for both neutrinos and quarks, implies the substitutions $k \Leftrightarrow k'$, $p \Leftrightarrow p'$, the antisymmetry of the term proportional to $\varepsilon_{\mu\nu\alpha\beta}$ means that it should be multiplied by a factor of -1 for each of these substitutions . This leads to two different situations:

- Scattering of antineutrinos on antiquarks: As both of the mentioned substitutions are applied *simultaneously* we get exactly the same result as for the neutrino-quark channel.

- Scattering of antineutrino on quark or neutrino on antiquark: here only one factor of -1 multiplies the term $(kp)^2 - (k'p)^2$ in (5.104) and we get

$$L^{(1)}_{\mu\nu} L^{\mu\nu}_{(2)} = 64(k'p)^2 = 16 \left(u - M^2 \right)^2 = 64(kp)^2 (1-y)^2. \tag{5.105}$$

The relation between (5.104) and (5.105) is a particularly simple example of the so called **crossing symmetry**. This symmetry relates the scattering amplitudes corresponding to the crossed channels of the basic $2 \to 2$ process (5.98) and holds under very general conditions even beyond the perturbation theory. The y dependence of the cross-sections $d\sigma/dy$ of ν_μ, $\overline{\nu}_\mu$ interactions can therefore be summarized as follows

$$\frac{d\sigma(\nu + q)}{dy} = \frac{d\sigma(\overline{\nu} + \overline{q})}{dy} = \frac{G_F^2 s}{\pi}, \tag{5.106}$$

$$\frac{d\sigma(\nu + \overline{q})}{dy} = \frac{d\sigma(\overline{\nu} + q)}{dy} = \frac{G_F^2 s}{\pi}(1-y)^2. \tag{5.107}$$

Introducing all numerical factors we end up with the following expressions

$$\frac{d\sigma^{\nu p}}{dxdy} = \frac{G_F^2 Sx}{\pi} \left[(s(x) + d(x)) + (\overline{u}(x) + \overline{c}(x))(1-y)^2 \right], \tag{5.108}$$

$$\frac{d\sigma^{\overline{\nu} p}}{dxdy} = \frac{G_F^2 Sx}{\pi} \left[(\overline{s}(x) + \overline{d}(x)) + (u(x) + c(x))(1-y)^2 \right]. \tag{5.109}$$

As parity is violated in weak interactions, there are three independent structure functions in the general expression for the double differential cross-section $d\sigma/dxdy$ on a proton

$$\begin{aligned}
\frac{d\sigma^{(\nu/\overline{\nu})p}}{dxdy} &= \frac{G_F^2 S}{2\pi} \left[F_2^{(\nu/\overline{\nu})p}(x, Q^2)(1-y) + 2x\frac{y^2}{2} F_1^{(\nu/\overline{\nu})p}(x, Q^2) \pm y \left(1 - \frac{y}{2} \right) x F_3^{(\nu/\overline{\nu})p}(x, Q^2) \right], \\
&= \frac{G_F^2 S}{2\pi} \left[F_2^{(\nu/\overline{\nu})p}(x, Q^2) \left(\frac{1 + (1-y)^2}{2} \right) \pm x F_3^{(\nu/\overline{\nu})p}(x, Q^2) \left(\frac{1 - (1-y)^2}{2} \right) \right], \tag{5.110}
\end{aligned}$$

where in the second equality the Callan-Gross relation between F_2 and F_1 used. Comparing the above expression with parton model formulae (5.108) and (5.109) we find

$$\begin{aligned}
F_2^{\nu p} &= 2x[s(x) + d(x) + \overline{u}(x) + \overline{c}(x)], \tag{5.111} \\
F_2^{\overline{\nu} p} &= 2x[u(x) + c(x) + \overline{s}(x) + \overline{d}(x)], \tag{5.112} \\
x F_3^{\nu p} &= 2x[s(x) + d(x) - \overline{u}(x) - \overline{c}(x)], \tag{5.113} \\
x F_3^{\overline{\nu} p} &= 2x[u(x) + c(x) - \overline{s}(x) - \overline{d}(x)]. \tag{5.114}
\end{aligned}$$

Particularly important are the combinations

$$\frac{1}{2} \left(F_2^{\overline{\nu} p} - F_2^{\nu p} \right) = x[(u - \overline{u}) - (d - \overline{d})], \tag{5.115}$$

$$\frac{1}{2}\left(F_2^{\nu p} + F_2^{\overline{\nu} p}\right) = x(u + d + s + c + \overline{u} + \overline{d} + \overline{s} + \overline{c}) = x\Sigma(x) \doteq \frac{1}{2}\left(F_2^{\nu p} + F_2^{\nu n}\right) \equiv F_2^{\nu N}, \qquad (5.116)$$

$$\frac{1}{2}\left(F_3^{\nu p} + F_3^{\overline{\nu} p}\right) = (u + d + s + c - \overline{u} - \overline{d} - \overline{s} - \overline{c}) = \Delta(x) \doteq \frac{1}{2}\left(F_3^{\nu p} + F_3^{\nu n}\right) \equiv F_3^{\nu N}, \qquad (5.117)$$

where the third equalities in the preceding two expressions are only approximate, neglecting small differences in sea quark distributions. For isoscalar targets F_2, F_3 can be experimentally determined from the following combinations of the measured cross-sections:

$$F_2^{\nu N} \sim \frac{\mathrm{d}\sigma^{\nu N}}{\mathrm{d}x\mathrm{d}y} + \frac{\mathrm{d}\sigma^{\overline{\nu} N}}{\mathrm{d}x\mathrm{d}y}, \qquad (5.118)$$

$$F_3^{\nu N} \sim \frac{\mathrm{d}\sigma^{\nu N}}{\mathrm{d}x\mathrm{d}y} - \frac{\mathrm{d}\sigma^{\overline{\nu} N}}{\mathrm{d}x\mathrm{d}y}. \qquad (5.119)$$

As already mentioned, neutrinos and antineutrinos *distinguish* quarks from antiquarks and thus allow us to separate determination of quark and antiquark distribution functions. This is done by suitable combinations of different beams ($\nu_\mu, \overline{\nu}_\mu$) and targets (n,p), or by selecting specific final states (for instance the D-particles, containing the charm quark). Still it required a lot of data and extensive phenomenological analyses before all the parton distribution functions, including that of gluon, could be determined with reasonable accuracy. The four different parameterizations displayed in Fig. 5.9 were obtained by three groups, which differed in a number of aspects of their analysis: selection of experimental data, theoretical methods employed, treatment of systematic errors, estimate of theoretical ambiguities of perturbation theory. The older HMRSB parameterization, shown merely to indicate the progress achieved in the determination of parton distribution functions in the last decade, is now clearly ruled out by new more precise data. Comparing (5.116) and (5.85) and neglecting in the latter the term (numerically very small) proportional to $k(x)$ we get the following relation between $F_2^{\nu N}$ and F_2^{eN}:

$$F_2^{\nu N} \doteq \frac{18}{5} F_2^{\mathrm{eN}}, \qquad (5.120)$$

Figure 5.12: The comparison of $F_2^{\nu N}(x, Q^2)$ and $(18/5)F_2^{\mathrm{eN}}(x, Q^2)$.

which expresses the **universality** of quark distribution functions in weak and electromagnetic interactions. Although a simple consequence of the QPM, (5.120) is in fact *nontrivial* consistency condition. The numerical factor $\frac{18}{5}$ is a direct reflection of quark charges and thus the experimental confirmation of this relation, illustrated in Fig. 5.12, provides another evidence for their fractional values. It also provides another argument against the Hahn-Nambu model of integral charge colored quarks which predicts $F_2^{\nu N} = 2F_2^{\mathrm{eN}}$. Integrating $F_3^{\nu N}(x, Q^2)$ over the whole kinematical range $x \in (0, 1)$ we get

$$\int_0^1 F_3^{\nu N}(x, Q^2)\mathrm{d}x = \int_0^1 (u_v(x) + d_v(x))\mathrm{d}x, \qquad (5.121)$$

which, according to the QPM, should be equal to 3, corresponding to the fact that there are two valence u quarks and one valence d quark in the proton. This so called **Gross-Llewellyn-Smith sum rule** has been measured experimentally with the result, $2.50 \pm 0.018(stat) \pm 0.078(syst)$, which deviates somewhat from the parton model prediction, but this deviation can be explained in QCD. Together with the sum rule (5.89) measuring the integral in (5.121) allows us to determine separately the integrals over $u_{val}(x)$ and $d_{val}(x)$, which provide a crucial bridge between the additive quark model and the parton model of Feynman [4].

In all the preceding calculations quark masses have been neglected with respect to Q^2. This is a good approximation for u, d and usually also s quark, but becomes questionable for heavy quarks c and b, which can be *produced* from light quarks (present in the nucleons) in weak interactions. As emphasized earlier the consistent incorporation of quark masses in the QPM is not trivial, but there exist a plausible procedure called **slow rescaling**, [99], which is commonly used to account for the effects of c and b quark masses.

99
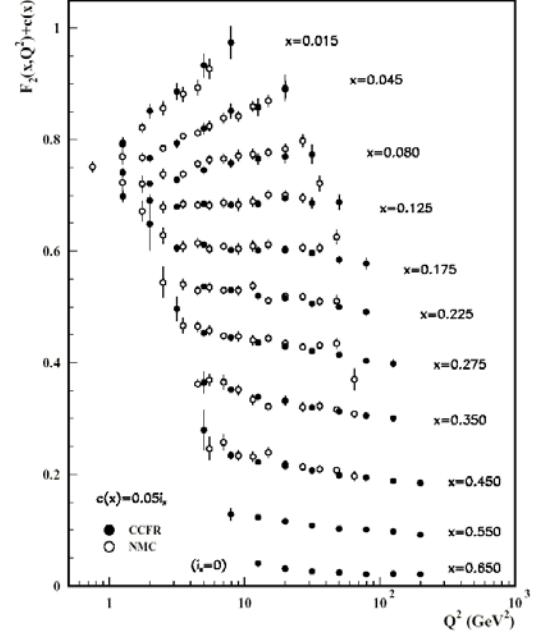
5.10 Polarized nucleon structure functions

Results of the previous section can be used in order to derive in the framework of the parton model simple formulae for the differential cross–sections of the scattering of polarized electrons on polarized protons.

In a completely general analysis of this process, following the derivation in Section 5.6, the fourvector S_μ describing the polarization of the target nucleon allows us to construct additional tensor structures, leading to two additional structure functions. Instead of discussing this topic in complete generality, I shall stay within the framework of the parton model.

To extend the treatment of DIS to polarized electrons and protons, we have to introduce the concept of **polarized** quark distributions functions, $q_i\!\uparrow\!(x)$ $(q_i\!\downarrow\!(x))$, which describe the probability to find a quark of flavor i with a momentum fraction x and spin *parallel* (*antiparallel*) to the spin of the target nucleon. We shall always assume that the proton spin points along its momentum and will neglect all effects of finite proton as well as quark masses. This means that also quark and antiquark spin orientations will point along their momenta. For massless quarks the states with definite spin projections have also a definite handedness, which will allow us to use directly the formulae (5.106)-(5.107) of the previous section. The unpolarized quark distribution functions, introduced earlier, are given as sums

$$q_i(x) = q_i\!\uparrow\!(x) + q_i\!\downarrow\!(x), \;\; \overline{q}_i(x) = \overline{q}_i\!\uparrow\!(x) + \overline{q}_i\!\downarrow\!(x), \tag{5.122}$$

while for polarized quarks we shall also need the other combinations

$$\Delta q_i(x) \equiv q_i\!\uparrow\!(x) - q_i\!\downarrow\!(x) + \overline{q}_i\!\uparrow\!(x) - \overline{q}_i\!\downarrow\!(x). \tag{5.123}$$

The basic quantity of interest is the difference

$$\frac{\mathrm{d}\sigma\!\uparrow\!\downarrow}{\mathrm{d}x\mathrm{d}y} - \frac{\mathrm{d}\sigma\!\uparrow\!\uparrow}{\mathrm{d}x\mathrm{d}y} \tag{5.124}$$

of cross–sections, corresponding to parallel and antiparallel beam and target spin orientations, or better the asymmetry

$$A^{\mathrm{p}} \equiv \frac{\mathrm{d}\sigma\!\uparrow\!\downarrow - \mathrm{d}\sigma\!\uparrow\!\uparrow}{\mathrm{d}\sigma\!\uparrow\!\uparrow + \mathrm{d}\sigma\!\uparrow\!\downarrow}. \tag{5.125}$$

Using (5.106)-(5.107) it is straightforward to show that in the parton model

$$A^{\mathrm{p}}(x) = \frac{(1 - (1-y)^2)\sum_i e_i^2 x \Delta q_i(x)}{(1 + (1-y)^2)\sum_i e_i^2 x (q_i(x) + \overline{q}_i(x))}. \tag{5.126}$$

The so called **polarized structure function** of the proton $g_1^{\mathrm{p}}(x)$ is then given in terms of the asymmetry A^{p} and unpolarized structure function F_2 as

$$g_1^{\mathrm{p}}(x) \equiv \frac{1}{2}\sum_i e_i^2 \Delta q_i(x) = \frac{A^{\mathrm{p}}(x) F_2(x)}{2x D(y)}, \;\; D(y) \equiv \frac{1 - (1-y)^2}{1 + (1-y)^2}. \tag{5.127}$$

Similarly for the neutron. The integrals

$$\Gamma_1^{\mathrm{p}} \;\; \equiv \;\; \int_0^1 \mathrm{d}x g_1^{\mathrm{p}}(x) = \frac{1}{2}\left(\frac{4}{9}\Delta(u) + \frac{1}{9}\Delta(d) + \frac{1}{9}\Delta(s)\right), \tag{5.128}$$

$$\Gamma_1^{\mathrm{n}} \;\; \equiv \;\; \int_0^1 \mathrm{d}x g_1^{\mathrm{n}}(x) = \frac{1}{2}\left(\frac{4}{9}\Delta(d) + \frac{1}{9}\Delta(u) + \frac{1}{9}\Delta(s)\right), \tag{5.129}$$

where $\Delta(q)$ stands for integrals over $\Delta q(x)$ and corresponds to $\Delta^p(q)$ defined in (4.26), have been measured in several experiments. The results of the SMC experiment at CERN [100], where polarized 190 GeV muons collided with stationary target made of polarized protons or neutrons, are shown in Fig. 5.13 and imply

$$\Gamma_1^{\mathrm{p}} \;\; = \;\; +0.136 \pm 0.011, \tag{5.130}$$

$$\Gamma_1^{\mathrm{n}} \;\; = \;\; -0.063 \pm 0.024. \tag{5.131}$$
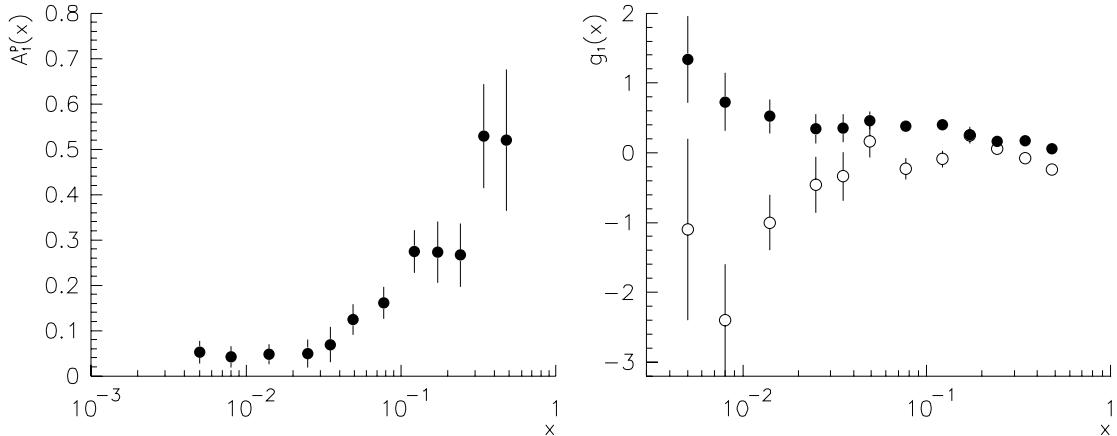
Figure 5.13: The spin asymmetry $A^{\mathrm{p}}(x)$ and the corresponding polarized structure functions $g_1(x)$ for the proton (solid points) and neutron (open points) as measured by the SMC Collaborartion at CERN.

Note that the increase of the asymmetry $A^{\mathrm{p}}(x)$ at large x is primarily due to the vanishing of $F_2(x)$ as $x \to 1$. Summing (5.130) and (5.131) we get

$$\Gamma_1^{\mathrm{p}} + \Gamma_1^{\mathrm{n}} = \frac{5}{18}\left(\Delta(u) + \Delta(d) + \frac{2}{5}\Delta(s)\right) = 0.073 \pm 0.03 \tag{5.132}$$

and neglecting the strange content of the proton yields

$$\Delta(u) + \Delta(d) = 0.26 \pm 0.09, \tag{5.133}$$

which gives the fraction of the spin of the proton carried by the u and d quarks and antiquarks. In a more rigorous treatment (see [101] for details and further references), where strange quarks are not neglected and additional theoretical constraints are imposed on polarized quark distribution functions, the u, d and s quark contributions to the proton spin can be determined separately, we get

$$\Delta(u) = +0.82 \pm 0.04, \tag{5.134}$$
$$\Delta(d) = -0.44 \pm 0.04, \tag{5.135}$$
$$\Delta(s) = -0.11 \pm 0.04. \tag{5.136}$$

The total contribution of quarks and antiquarks to the proton spin

$$\sum_q \Delta(q) \equiv \Delta(u) + \Delta(d) + \Delta(s) = 0.27 \pm 0.11 \tag{5.137}$$

is then positive but small and the strange quark contribution is significantly nonzero. The numbers (5.134)-(5.136) should be compared with the predictions (4.26) of the nonrelativistic quark model, where spins of the u and d constituent quarks add up to unity. In the parton model this is not necessary the case and there are other possible carriers of the proton spin, like gluons and orbital momentum as well as other plausible explanations of the above results. Moreover, the experimental extraction of the values quoted above is very complicated and rests, among other things, on certain assumptions concerning the contribution of various nonperturbative effects. Nevertheless the problem of what carries most of the proton spin represents one of the major puzzles of the present particle physics and remains the subject of intensive experimental as well as theoretical research.

5.11   Space–time picture of DIS and hadronization

As already emphasized, the central assumption underlying the QPM is the separation of a collision into two distinct stages:
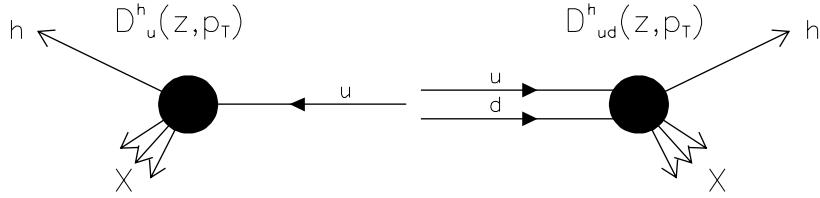
Figure 5.14: Hadronization of quark-diquark system, viewed in hadronic center of mass system. As they fly away, strong force, acting between them at distances larger than about 1 fm causes them to "radiate" hadrons. The big solid points represent the fragmentation functions.

- The "hard" scattering of leptons on *individual* quarks. During this stage the influence of other partons in the nucleon is neglected and the cross–section of the lepton-parton subprocess calculated as if partons were as real as leptons. We thus disregard fact that the outgoing quark and remnant diquark cannot separate to infinity.

- The "hadronization", i.e. that stage of the collision in which the outgoing quark and the proton remnant diquark convert into observable hadrons. This conversion is caused by the force acting between colored quarks as described in Section 3.11, which starts to be very strong when they are about 1 fm apart.

Let me emphasize that the fact that the collision can be divided into such two distinct stages is far from trivial! The first stage is calculable within the simple QPM but to describe the second one is much more complicated. There is no real theory of hadronization, merely more or less sophisticated **models** thereof. The first (and relatively successful) model of hadronization, the so called **independent fragmentation model**, was invented by Feynman and Field in the middle of seventies [102]. Although this model has serious shortcoming in the very assumption of the *independent* fragmentation of colored partons and there are now several more realistic and sophisticated models of hadronization, it is still useful to learn its language. The basic idea of this model is to imagine the hadronization of the outgoing quark-diquark system in Fig. 5.7 as an *independent* fragmentation of a quark and diquark. This assumption is, of course, in principle untenable as the hadronization is due to the *mutual* interaction between them, but it turns out that the resulting distributions of hadrons can in some approximation be *described* using the language of fragmentation functions, introduced below.

In IFM the hadronization box of Fig. 5.7 is approximated as indicated in Fig. 5.14. The basic quantity used in the description of hadronization is the so called **fragmentation function**

$$D_q^h(z, p_T), \tag{5.138}$$

describing the probability that a parton q (quark, antiquark, diquark or gluon) produces ("fragments" into) a hadron h, carrying the fraction z of the original parton energy and the transverse (with respect to the direction of the fragmenting parton) momentum p_T. Very often transverse momentum is integrated over and the resulting fragmentation functions studied as a function of z only. The fragmentation functions are normalized as follows:

$$\sum_h \int_0^1 \int_0^\infty D_q^h(z, p_T) \mathrm{d}z \mathrm{d}p_T = \langle N_q \rangle, \tag{5.139}$$

where $\langle N_q \rangle$ is the average multiplicity of hadrons coming from fragmentation of the quark q. The fragmentation functions are used, for instance, in the analysis of **inclusive** production of hadrons in DIS,

$$\mathrm{e}^- + \mathrm{p} \to \mathrm{e}^- + h + \text{anything} \tag{5.140}$$

i.e. the process in which we are interested not only in the properties of the final state electron, but also in those of the accompanying hadrons. Notice, however, that fragmentation functions *don't* describe the *full* configuration of final state hadrons, as do modern event generators, like HERWIG [106] or PYTHIA [107]. The differential cross-section of the process (5.140) is given as

$$\frac{\mathrm{d}\sigma(\mathrm{ep} \to \mathrm{e}' + \mathrm{h} + \text{anything})}{\mathrm{d}x\mathrm{d}y\mathrm{d}z\mathrm{d}p_T} = \frac{4\pi\alpha^2 S}{Q^4} \left(\frac{1 + (1-y)^2}{2} \right) \sum_i e_i^2 x q_i(x) D_{q_i}^h(z, p_T), \tag{5.141}$$

where the sum runs over all quarks inside the protons involved in the hard scattering with electron. Very often the p_T dependence of fragmentation functions is integrated over and one works with fragmentation functions $D_q^h(z)$ depending on the energy fraction z only. As the dependence on x and y in (5.141) *factorizes*, it is convenient to normalize (5.141) to $\mathrm{d}\sigma/\mathrm{d}x\mathrm{d}y$:

$$\frac{1}{\mathrm{d}\sigma/\mathrm{d}x\mathrm{d}y}\frac{\mathrm{d}\sigma(ep \to e' + h + \text{anything})}{\mathrm{d}x\mathrm{d}y\mathrm{d}z} = \frac{\sum_i e_i^2 q_i(x) D_{q_i}^h(z)}{\sum_i e_i^2 q_i(x)}, \tag{5.142}$$

which is then *independent of y*. Moreover, this quantity is also very suitable from experimental point of view as various systematical errors tend to cancel in the ratio (5.142).

In neutrino interactions the x and y dependences *do not* factorize due to different y–dependences of quark and antiquark cross–sections and we thus get slightly more complicated relation

$$\frac{1}{\mathrm{d}\sigma/(\mathrm{d}x\mathrm{d}y)}\frac{\mathrm{d}\sigma(\nu_\mu p \to \mu^- + h + \text{anything})}{\mathrm{d}x\mathrm{d}y\mathrm{d}z} = \frac{d(x)D_u^h(z) + \overline{u}(x)(1-y)^2 D_{\overline{d}}^h(z)}{d(x) + \overline{u}(x)(1-y)^2}. \tag{5.143}$$

where only the light quarks u, d from the first generation and their antiquarks were taken into account. The characteristic shapes of fragmentation functions for light as well as heavy quarks into pions, kaons or D-mesons (mesons containing the charmed quark c) are discussed in more detail in the next section. Here let us merely emphasize the basic feature of these fragmentation functions: the heavier the quark, the bigger the average fraction of its momentum, carried away by the meson, containing, as its valence quark, the fragmenting quark.

Taking into account quark composition of pions we expect the following symmetry relations between different fragmentation functions:

$$D_u^{\pi^+} = D_d^{\pi^-} = D_{\overline{d}}^{\pi^+} = D_{\overline{u}}^{\pi^-}, \tag{5.144}$$

$$D_u^{\pi^-} = D_d^{\pi^+} = D_{\overline{d}}^{\pi^-} = D_{\overline{u}}^{\pi^+}, \tag{5.145}$$

$$D_s^{\pi^+} = D_s^{\pi^-} = D_{\overline{s}}^{\pi^+} = D_{\overline{s}}^{\pi^-}. \tag{5.146}$$

Let us emphasize the following important fact: both the parton distribution and fragmentation functions have been introduced to describe experimental data on DIS processes. Neither of them can be calculated (at least so far) and must be determined from data. However, once they are extracted using data from one process [11], they can be used in any other one and thus allow nontrivial **predictions** to be made. This property of **universality**, discussed more in the next section, provides the basis for the usefulness of the QPM. Several examples of other processes treated in the QPM are discussed in the next chapter.

5.12   Deep inelastic scattering at small x

In the small x region, roughly defined as $x \leq 10^{-2}$, theoretical description via the Altarelli–Parisi evolution equations encounters problems stemming from appearance of terms behaving typically as powers of $\ln(1/x)$. I shall not discuss theoretical ideas relevant for quantitative understanding of this region, but will merely show the latest experimental results from ep collider HERA at DESY.

In Fig. 5.15 the latest HERA data (H1 and ZEUS Collaborations) on $F_2^{\mathrm{ep}}(x, M)$ are shown together with the older fixed target data from muon–proton scattering at CERN (BCDMS and NMC) and at FERMILAB (E665). The rise of $F_2^{\mathrm{ep}}(x, M)$ at small x for all accessible values of Q^2 is clearly visible. Taking into account the expression (5.68) the same data in small x region are replotted in Fig. 5.16 as the energy (i.e. W) dependence of the total cross–sections $\sigma(\gamma^* p; W, Q^2)$ for fixed Q^2. The figure includes also the data on total cross–section of real photon and indicates that the slope of the W–dependence of $\sigma(\gamma^* p; W, Q^2)$ increases with the photon virtuality Q^2. This feature is a source of intensive theoretical debate.

5.13   Exercises

1. Carry out the contraction (5.20) in the laboratory frame and derive (5.25).

---

[11]Or, realistically, rather one set of processes
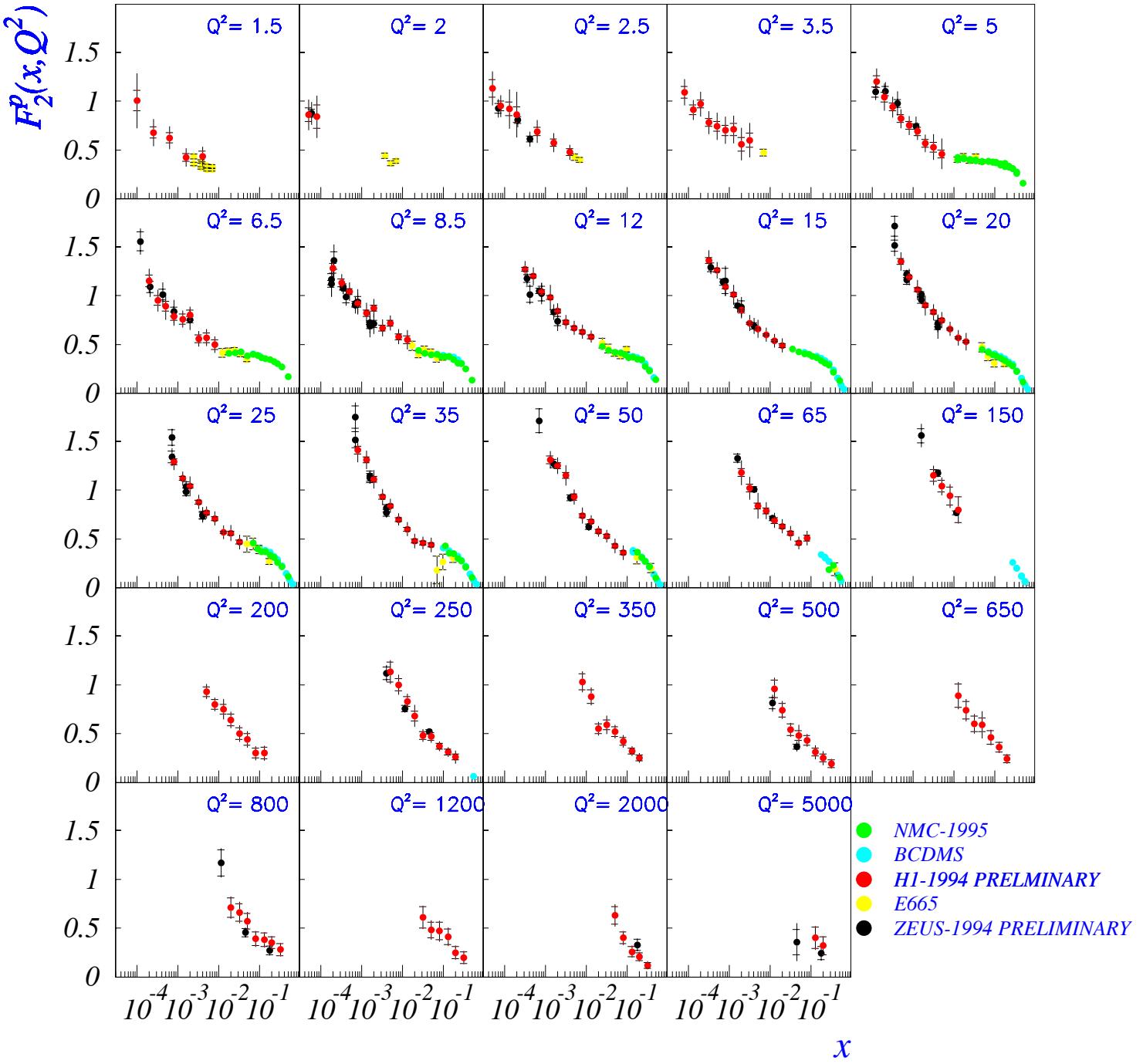
Figure 5.15: Low x data on $F_2^{\mathrm{ep}}(x,Q)$ as measured at HERA (H1 and ZEUS) and in fixed target experiments (BCDMS and E665).

104
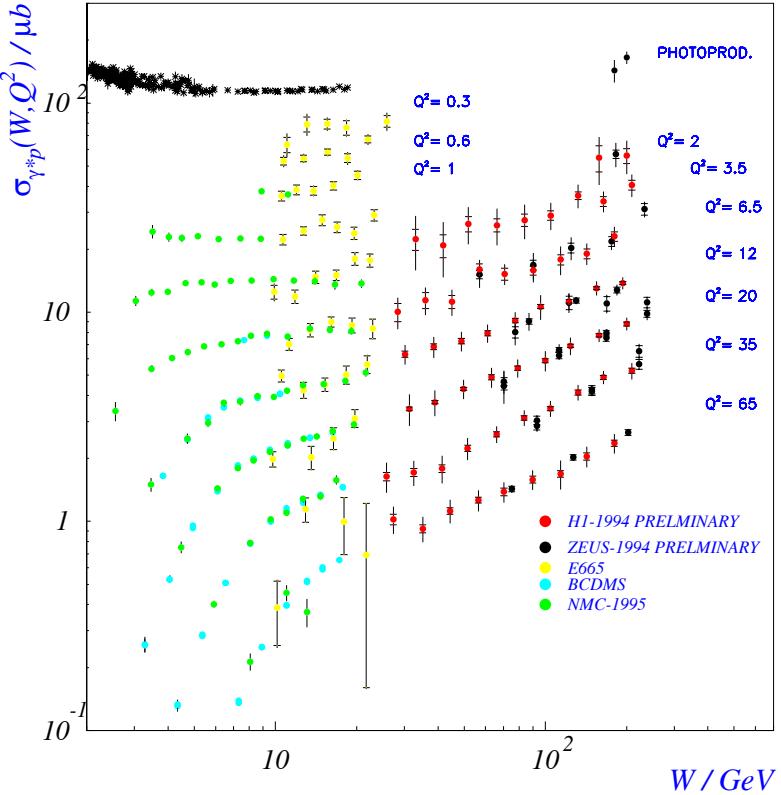
Figure 5.16: Energy dependence of the total cross–section of real and virtual γ^*p collisions.

2. Perform explicitly the integration over δ-function in (5.22).

3. Calculate the scattering of electron on a charged scalar pointlike particle with electromagnetic coupling as defined in the text at the end of section 4.4.

4. Argue why the Mott cross section follows by replacing in the numerator of the Rutherford formula (1.17) $m \to E$ and in its denominator $p \to E$.

5. Prove the so called Gordon decomposition

$$\overline{u}(p')\gamma_\mu u(p) = \overline{u}(p') \left[\frac{(p'+p)_\mu}{2m} + i\frac{\sigma_{\mu\nu}}{2m}(p'-p)^\nu \right] u(p)$$

where $\sigma_{\mu\nu}$ is defined in (5.30) and m is the fermion mass.

6. Derive (5.30) exploiting gauge and Lorentz invariance and parity conservation.

7. Show that due to gauge invariance the value of elastic formfactor F_1 at $q^2 = 0$ is fixed: $F_1(0) = 1$.

8. Show by nonrelativistic reduction of the Dirac equation that the value of κ in (5.30) gives the anomalous magnetic moment of the proton.

9. Derive the general form (5.36) for deep inelastic scattering using gauge and Lorentz invariance and parity conservation.

10. Show that for massless fermions the projection operator $(1 - \gamma_5)/2$ projects out states of particles with *negative* helicity (spin opposite to the momentum of the particle) and states of antiparticles with *positive* helicity (spin in the direction of momentum.

Chapter 6

Parton model in other processes

6.1 Electron-positron annihilations into hadrons at high energies

The first process to which the parton had been applied after its invention in early seventies was the electron-positron annihilation into hadrons at high energies which in the lowest order of QED proceeds via the one photon annihilation into a parton-antiparton pair which subsequently converts into hadrons (hadronizes)

$$e^+e^- \rightarrow q\overline{q} \rightarrow \text{hadrons.} \tag{6.1}$$

This hadronization determines details of the final states in (6.1) but does not change the cross section of this process if we integrate over all possible hadron final states. The evaluation of the parton level cross section proceeds similarly as in the case of lepton-quark scattering. Neglecting the electron mass but keeping that of the quark (M_q), the spin averaged matrix element squared $|\overline{M_{if}}|^2$ can be written in general as

$$|\overline{M_{if}}|^2 = \frac{e^4 e_q^2}{s^2} L_{\mu\nu}^{(1)}(p_1, p_2) L^{(2)\mu\nu}(p_3, p_4) = e^4 e_q^2 \left[a_T(1 + \cos^2 \vartheta) + a_L(1 - \cos^2 \vartheta) \right] \tag{6.2}$$

where e_q denotes the quark charge in units of positron charge. The leptonic tensor $L_{\mu\nu}^{(1)}$ is given as in (5.19)

$$L_{\mu\nu}^{(1)}(p_1, p_2) \equiv \frac{1}{2} \text{Tr} \left[\slashed{p}_1 \gamma_\mu \slashed{p}_2 \gamma_\nu \right] = 2 \left[p_{1\mu} p_{2\nu} + p_{1\nu} p_{2\mu} - g_{\mu\nu}(p_1 p_2) \right], \tag{6.3}$$

whereas its partonic counterpart $L^{(2)\mu\nu}$ depends on its spin. Working out the flux factor and final parton differentials we arrive at the following general expression for the differential cross section

$$\frac{d\sigma}{d\cos\vartheta^*} = \frac{\beta}{32\pi s} |\overline{M_{if}}|^2 = \frac{\pi \alpha^2 e_q^2 \beta}{2s} \left[a_T(1 + \cos^2 \vartheta^*) + a_L(1 - \cos^2 \vartheta^*) \right], \tag{6.4}$$
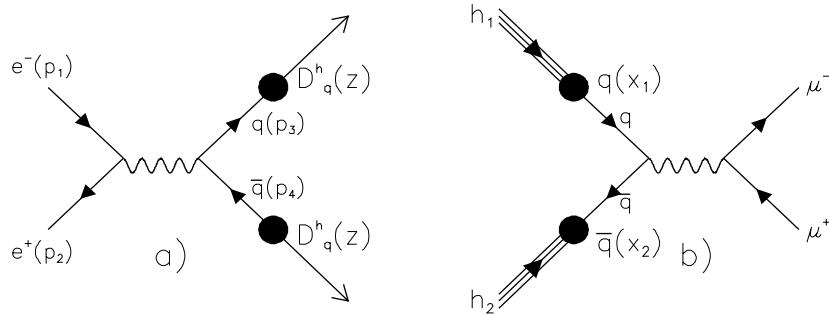

Figure 6.1: a) Lowest order Feynman diagram for the production of a $q\overline{q}$ pair in e^+e^- annihilation and its subsequent fragmentation in IFM; b) inverse process of dilepton pair production in hadronic collisions, via $q\overline{q}$ annihilation into a virtual photon. The big solid blobs stand in (a) for quark fragmentation functions and in (b) for quark distribution functions.

where $\beta = \sqrt{1 - 4M_q^2/s}$. Note that this threshold factor comes exclusively from kinematical considerations. Integrating the above expression over the scattering angle we get

$$\sigma_{tot} = \frac{\pi\alpha^2 e_q^2 \beta}{2s}\left[\frac{8}{3}a_T + \frac{4}{3}a_L\right]. \tag{6.5}$$

Taking into account that

$$\text{partons with } s = 1/2 \quad : \quad L_{\mu\nu}^{(2)}(p_3, p_4) \equiv 2\left[p_{4\mu}p_{3\nu} + p_{4\nu}p_{3\mu} - g_{\mu\nu}(p_4 p_3 + M_q^2)\right], \tag{6.6}$$

$$\text{partons with } s = 0 \quad : \quad L_{\mu\nu}^{(2)}(p_3, p_4) \equiv (p_3 - p_4)_\mu (p_3 - p_4)_\nu, \tag{6.7}$$

we find after straightforward contraction of leptonic and partonic tensors

$$\text{partons with } s = 1/2 \quad : \quad a_T = 1, \quad a_L = 1 - \beta^2, \tag{6.8}$$

$$\text{partons with } s = 0 \quad : \quad a_T = 0, \quad a_L = \beta^2. \tag{6.9}$$

Summing over N_f quarks with spin $1/2$ in the region where $s \gg M_q^2$ we get

$$\sigma_{tot}(s, \text{e}^+\text{e}^- \rightarrow \text{hadrons}) = \sum_{i=1}^{N_f} \sigma_{tot}(s, \text{e}^+\text{e}^- \rightarrow \text{q}_i\overline{\text{q}}_i) = \frac{4\pi\alpha^2}{3s}\sum_{i=1}^{N_f}\beta_i e_i^2 \xrightarrow[s \gg M_i^2]{} \frac{4\pi\alpha^2}{3s}\sum_{i=1}^{N_f} e_i^2. \tag{6.10}$$

It is convenient to normalize (6.10) by $4\pi\alpha^2/3s$, defining thus the dimensionless quantity $R(s)$

$$R(s) \equiv \frac{\sigma_{tot}(s, \text{e}^+\text{e}^- \rightarrow \text{hadrons})}{\sigma_{tot}(s, \text{e}^+\text{e}^- \rightarrow \mu^+\mu^-)} \xrightarrow[s \gg M_q^2]{} \sum_{i=1}^{N_f} e_i^2. \tag{6.11}$$

For scalar quarks $R(s)$ is given by the same expression except for the additional overall factor $1/2$.

The above predictions can be tested in several ways. Measuring the ratio (6.11) is the simplest and clearest test of the basic idea of the QPM. It doesn't need fragmentation functions and is thus *fully* calculable. Comparing (6.11) to data in Fig. 6.2 shows, however, that apart from the resonance structures, which the QPM clearly *cannot* reproduce, there is also a overall normalization factor problem. Although the data outside the resonance region seem to be essentially flat, with the threshold at about 4 GeV, corresponding to the opening of c$\overline{\text{c}}$ channel, they are about a factor 3.5 *higher* than 10/9, the QPM prediction (6.11), taking into account the four active quarks (u,d,s,c). This discrepancy has a simple solution. We have so far ignored the **color** of quarks, both in DIS and in e$^+$e$^-$ annihilations, despite the fact that this degree of freedom triples the number of quark states. Taking it into account in relations (6.10) or (6.11) means simply multiplying their r.h.s. by a factor of 3 and thus bringing them almost to agreement with data. The remaining discrepancy is then explained as higher order correction in QCD. Note that the Hahn-Nambu version of colored quarks predicts for the same quartet of quarks the value $R = 6$, and is thus definitely ruled out by the data.

Another possibility is to look for the manifestation of the specific angular dependence (6.4) of outgoing quarks in the angular distribution of produced **hadrons**. As in the process of quark fragmentation hadrons are produced as **jets** [1] of particles, collimated around the direction of the fragmenting quark, the angle of the latter can be reasonably accurately determined from the direction of these jets. The process of parton fragmentation closely resembles particle showers in cosmic rays. The angular distribution of outgoing jets in the process (6.1) has been measured in a number of experiments and the expected form $(1 + \cos^2 \vartheta^*)$, predicted by the QPM for quarks with spin $1/2$, found in excellent agreement with data. This provides yet another evidence that quarks have spin $1/2$. For scalar quarks the angular distribution would be proportional to $(1 - \cos^2 \vartheta^*)$, which is clearly ruled out by the data.

One can also study the one-particle inclusive production of hadrons h in the process

$$\text{e}^+\text{e}^- \rightarrow h + \text{anything} \tag{6.12}$$

---

[1]The concept of **jets** is introduced and its significance discussed in Chapter 9.
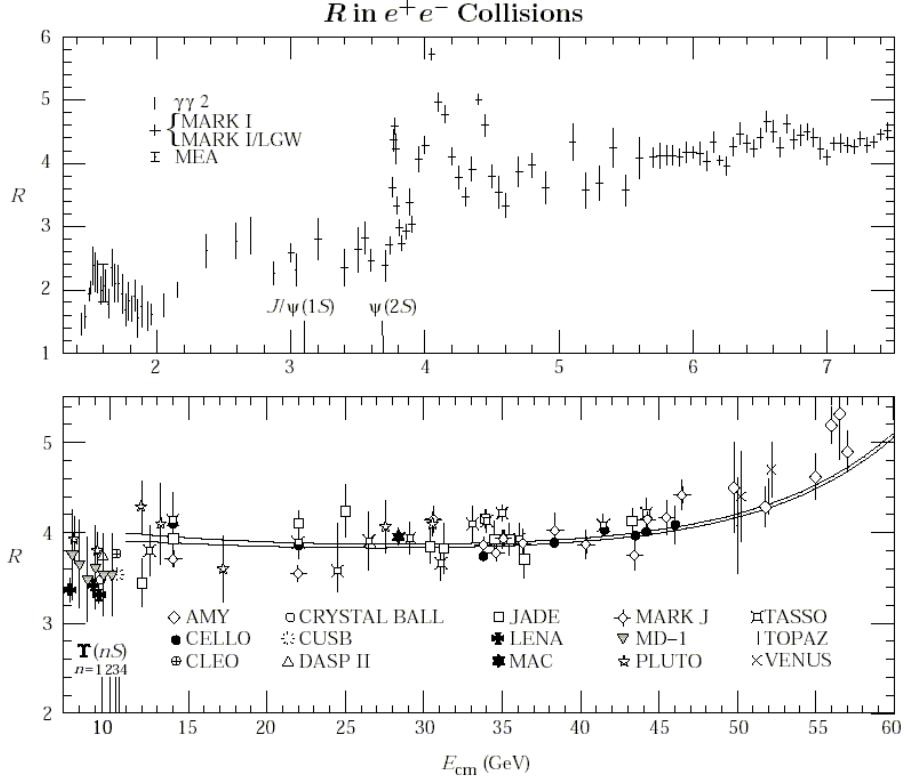
Figure 6.2: A summary of measurements of the ratio $R(s)$ from various experiments at DESY, SLAC, KEK and Frascati. Positions of narrow $c\bar{c}$ states J/ψ and ψ' are indicated, but the resonances themselves are invisible in this plot. The threshold for open charm production lies at 3.77 GeV.

as a function of the energy of final state hadron h, or more conveniently, of the fraction $z \equiv 2E^{(h)}/\sqrt{s}$. This process is described by the diagram in Fig. 6.1a, where the fragmentation functions are represented by big solid blobs. Normalizing $d\sigma/dz$ by $\sigma_{tot}(e^+e^- \to \mu^+\mu-)$, the parton model predicts

$$\frac{1}{\sigma_{tot}(e^+e^- \to \mu^+\mu^-)} \frac{d\sigma}{dz} = \sum_i e_i^2 \left(D_i^h(z) + D_{\bar{i}}^h(z) \right). \tag{6.13}$$

As emphasized above, to make any sense, the fragmentation functions should be *universal*, i.e. independent of the process in which they are used. Comprehensive analyses of the fragmentation functions of light as well as heavy quarks extracted from different processes (mainly e^+e^- annihilations and DIS lepton-nucleon scattering) show convincingly, that this universality holds, indeed, very well. Notice the difference in shapes of fragmentation functions of light and heavy quarks into mesons containing them as valence quarks. While the former peak at very small x, the fragmentation functions of heavy quarks into heavy mesons look very differently and peak at much larger x, see Fig. 6.3 for comparison. The light quark fragmentation functions are conveniently parameterized by analytic formula of the type

$$D_q^h(z) = Az^a(1-z)^b(1+c/z), \tag{6.14}$$

where A, a, b, c are free parameters, depending on the hadron h. The fragmentation function of a heavy quark Q into a hadron containing this heavy quark is usually parameterized by a different form, suggested in [103]

$$D_Q(z) = A\frac{z(1-z)^2}{((1-z)^2 + \varepsilon_Q z)^2}, \tag{6.15}$$

where the free parameter ε_Q, equal to 0.05 for the charm quark and 0.006 for the bottom one, determines the "hardness" of this fragmentation function.
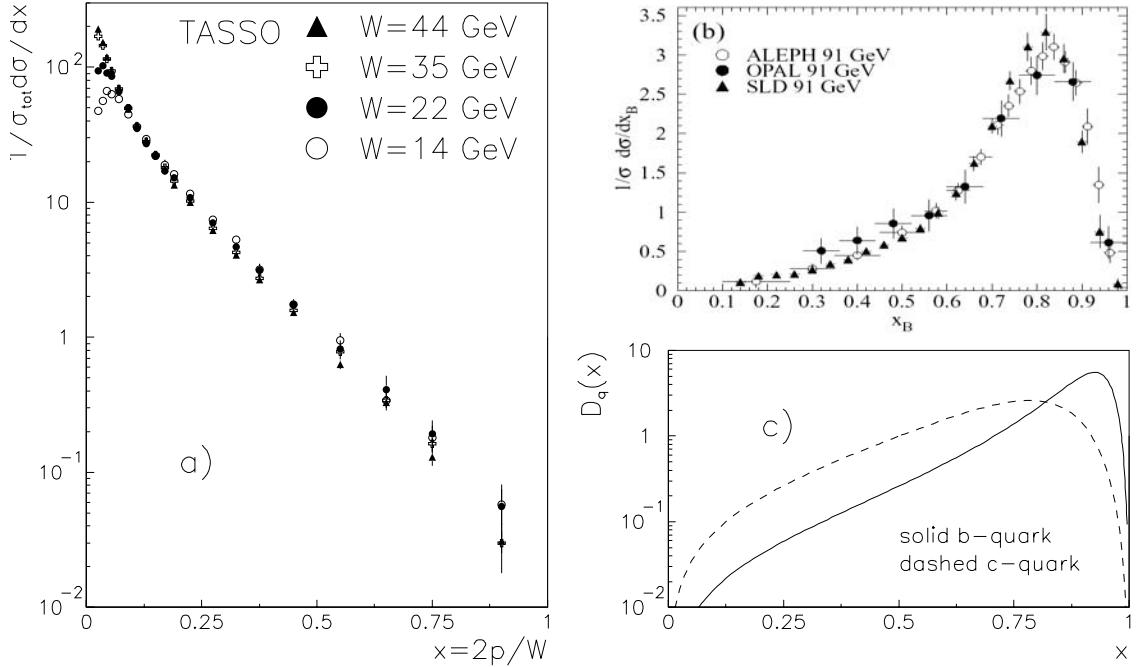
Figure 6.3: The measured shapes of fragmentation functions of light as well as heavy quarks into valence hadrons (i.e. those hadrons containing the fragmenting quark) together with the analytical fits to fragmentation functions of heavy quarks c and b.

The weighted average over all quark fragmentation functions into pions as measured in e^+e^- annihilations by the TASSO Collaboration at DESY for four different primary energies are shown in Fig. 6.3a. The weak dependence of this fragmentation function on the primary energy, suggested by this plot, is of the same origin as the Q^2 dependence of the nucleon structure functions and can also be explained in QCD. The distribution of B-hadrons (i.e. mesons and baryons containing the bottom quark) as measured in e^+e^- annihilations at LEP and SLD for $\sqrt{s} = 91$ GeV is plotted in the upper right part of Fig. 6.3 in order to demonstrate the difference between the fragmentation functions of light and heavy quarks.

6.2 Drell-Yan production of heavy dilepton pairs

The production of dilepton pairs in hadronic collisions

$$h_1 + h_2 \to \mu^+\mu^- + \text{anything} \tag{6.16}$$

proceeds in the parton model via the diagram in Fig. 6.1b. Experimentally the produced $\mu^+\mu^-$ pair can be characterized by its mass m as well as the longitudinal and transverse (with respect to the collision axis) momenta $p_\parallel$ and p_T. Assuming the hadrons h_1, h_2 collide head–on along the z-axis in some reference frame, the produced pair has no net transverse momentum p_T with respect to this axis. Instead of the dilepton mass m and the longitudinal momentum $p_\parallel$ it is convenient to introduce the following two *dimensionless* variables

$$x_F \equiv \frac{2p_\parallel}{\sqrt{S}}; \quad \tau \equiv \frac{4m^2}{S}, \tag{6.17}$$

where S stands for the square of the total CMS energy of colliding hadrons. It is straightforward to express x_F, τ in terms of the fractions $\overline{x}_1, \overline{x}_2$, carried by the annihilating quarks and antiquarks. Neglecting quark masses with respect to $\sqrt{S}$ we have

$$m^2 = x_1 x_2 S; \quad x_F = x_2 - x_1 \tag{6.18}$$
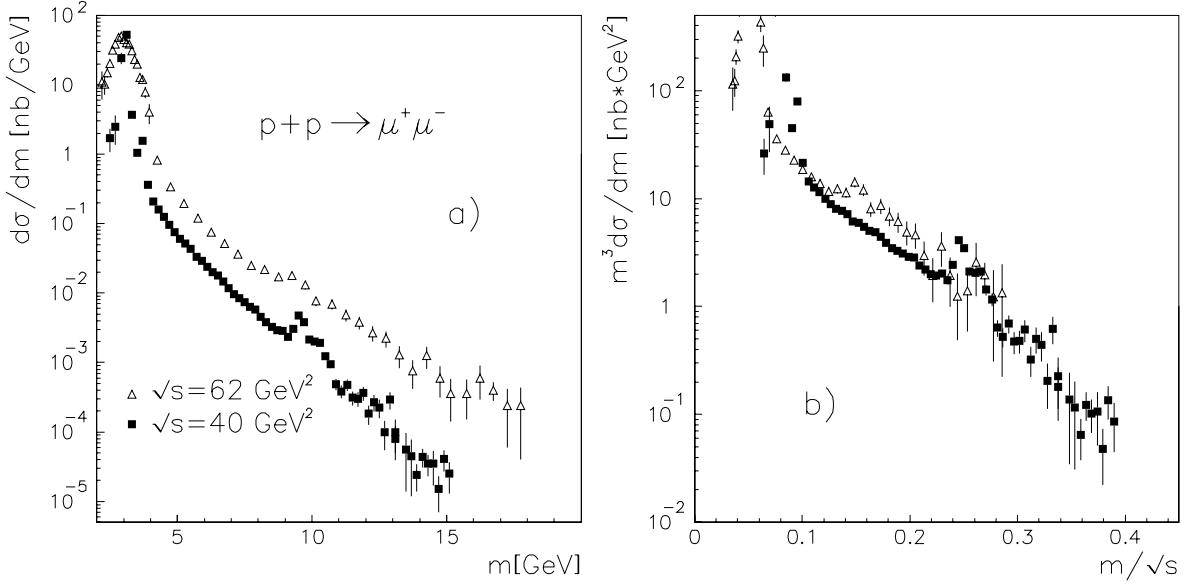
Figure 6.4: DY spectra as measured in pp collisions at CERN, ($\sqrt{s} = 62$ GeV) and Fermilab ($\sqrt{s} = 40$ GeV) displayed as a function of the dilepton mass m (a). In (b) the same spectra, but multiplied by m^3 and expressed as functions of $m/\sqrt{s}$ are compared, lending support to the QPM prediction of scaling (5.20).

where we have chosen the positive z-axis in the direction of the hadron h_2. Solving (6.18) for $\overline{x}_1, \overline{x}_2$ we find

$$\overline{x}_1 = \frac{1}{2}\left(-x_F + \sqrt{x_F^2 + \tau}\right),$$ (6.19)

$$\overline{x}_2 = \frac{1}{2}\left(+x_F + \sqrt{x_F^2 + \tau}\right).$$ (6.20)

The differential cross section describing this process can be written as

$$\frac{\mathrm{d}\sigma}{\mathrm{d}m^2 \mathrm{d}x_F} = \sum_i \int\int \mathrm{d}x_1 \mathrm{d}x_2 q_i(x_1)\overline{q}_i(x_2)\delta(x_1 x_2 S - m^2)\delta(x_2 - x_1 - x_F)\underbrace{\left[\frac{4\pi\alpha^2 e_i^2}{3m^2}\right]}_{\sigma(m^2)},$$ (6.21)

where the quark flavor index i in the above sum is understood to run over *both* quarks and antiquarks and the result (6.10) has been used (with the substitution $s \to m^2$) for $\mathrm{d}\sigma/\mathrm{d}m^2$. Integrating (6.21) over x_2 by means of the second δ-function we are left with the first δ-function in the form

$$\frac{1}{S}\delta(x_1^2 + x_1 x_F - \tau/4) = \frac{1}{S\sqrt{x_F^2 + \tau}}\delta(x_1 - \overline{x}_1) \quad \Rightarrow$$ (6.22)

$$\frac{m^3 \mathrm{d}\sigma}{\mathrm{d}m \mathrm{d}x_F} = \frac{8\pi\alpha^2}{3\sqrt{x_F^2 + \tau}}\left[\sum_i e_i^2 \overline{x}_1 q_i(\overline{x}_1)\overline{x}_2 \overline{q}_i(\overline{x}_2)\right]\frac{1}{3},$$ (6.23)

Several aspects of the DY dilepton production are worth mentioning:

- In Fig. 6.4a typical spectra of DY pairs produced in pp collisions at two different energies and measured in two experiments, are plotted. The structures above the smooth Drell–Yan continuum, described by the QPM model, correspond to the $c\overline{c}$ and $b\overline{b}$ bound states of the charmonium and bottomonium families. Experiments with protons, antiprotons and pions have demonstrated that for masses above roughly 3 GeV the QPM describes this continuum spectrum very well. For smaller masses the DY mechanism is clearly *below* the data.

110

- The quantity on the left hand side of (6.21) was defined in such a way that the r.h.s. of it depends on dimensionless variables x_F, τ only. So *independently* of the particular shape of quark distribution functions the QPM implies a nontrivial prediction by relating dilepton spectra at *different* energies, but the *same* τ. This is an analogue of the scaling phenomena in deep inelastic scattering of leptons on nucleons. Outside the resonance regions, the data plotted in Fig. 6.4b are in reasonable agreement with this basic prediction.

- The dilepton production in hadronic collisions (6.16), besides providing additional checks of basic QPM ideas, has also played a decisive role in the determination of quark distribution functions of **pions and kaons**, which are impossible to measure in conventional DIS processes. Their general features are similar to those of nucleons.

- The careful reader must have noticed that the factor $1/3$ at the end of (6.23) *doesn't* follow from (6.21) but has been added *by hand*. The reason for its presence is related to the color of quarks, which has been ignored in the previous discussion of DIS and thus far also here. The point is that quark distribution functions, as introduced in Section 5.6, correspond to definite flavors, but are *averaged* over the colors. This is fine as long as we don't force a q$\overline{\text{q}}$ pair of definite flavor to annihilate to *colorless* photon. In that case we have to take into account that for each, say, u quark from hadron h_1 with a definite color the antiquark $\overline{u}$ from h_2 will have only $1/3$ chance to have the right *anticolor* needed for the annihilation into the colorless photon. Hence the factor $1/3$.

6.3   Exercises

1. Calculate the angular distribution of scalar partons in e^+e^- annihilations as well as the magnitude of the integrated cross-section.

2. Evaluate the differential cross–section for the production of scalar quarks in e^+e^- annihilations. Compare the results on the angular dependence of the produced quarks as well as on the integrated cross–section with those of spin $1/2$ quarks.

Chapter 7

Basics of Quantum Chromodynamics

This text does not aspire at providing a selfcontained introduction to Quantum Chromodynamics (QCD). Rather it brings a collection of remarks on points I consider important for understanding how perturbative QCD works and in particular how it accommodates the parton model. The reader is assumed to be familiar with basic concepts of quantum field theory, in particular with the language of Feynman diagrams in perturbation theory, as expounded, for instance, in the classical book [2], or modern treatises like [1,9,10,14].

I shall concentrate on general principles of gauge theories, to which QCD, as well as the more familiar **Quantum Electrodynamics** (QED) belong. In most of my considerations I will stay within classical physics, avoiding the formal aspects of quantization of gauge theories, but the basic features of the resulting Feynman rules will be shown to follow from the classical Lagrangian. These rules will then be used to evaluate matrix elements of some important parton level processes.

Throughout this and the following chapters only *perturbative* QCD is discussed and applied. There is a number of nonperturbative approaches that address some of the important questions, like the mechanism of color confinement, hadron masses and wave functions and structure of the QCD vacuum, that remain outside the applicability of perturbative QCD: lattice QCD, QCD sum rule approach [109], potential models [15], $1/N_c$ expansion [110], Schwinger–Dyson and Bethe–Salpeter equations [111], bag models [112], applications of instantons [113], light cone formalism [114]and possibly other.

In order to understand the fundamental importance of nonabelian gauge theories for theoretical foundations of the Standard Model, it is worth recalling the evolution of the idea of quantum field theory from its inception by Dirac in 1928 until its present maturity. The twists and turns of this development are best captured in the words of one of the main protagonists, David Gross [120]:

> *Science progresses in a much more muddled fashion than is often pictured by the history books. This is especially true of theoretical physics, partly because history is written by the victorious. Consequently, historians of science often ignore the many paths that people wandered down, the many false clues they followed, the many misconceptions they had. These alternate points of view are less clearly developed than the final theories, harder to understand and easier to forget, especially as these are viewed years later, when it all really does make sense. Thus reading history one rarely gets the feeling of the true nature of scientific development, in which the element of farce is as great as the element of triumph.*
>
> > *The emergence of QCD is a wonderful example of the evolution from farce to triumph. During a very short period, a transition occurred from experimental discovery and theoretical confusion to theoretical triumph and experimental confirmation.*

7.1 The rise and fall of quantum field theory

Soon after the formulation of Quantum Electrodynamics by Dirac in 1928, theorists like Bethe, Heitler, Oppenheimer, Weisskopf and others have attempted to calculate quantum corrections to lowest order perturbative results for several physical quantities. In modern language, they were attempting to evaluate the effects of loop corrections. In the course of these, from the current point of view rather cumbersome and unsystematic calculations, they had encountered infinities coming from contributions of very small distances.

It took about two decades to develop a systematic procedure how to handle and remove these so called ultraviolet (UV) divergencies. For their decisive contributions towards the formulation of this procedure, called <u>renormalization</u>, Feynman, Schwinger and Tomonaga were awarded the 1965 Nobel Prize for physics.

Although this procedure provided the framework for systematic removal of UV infinities to all orders of perturbation theory, it lacked clear physical interpretation and appeared mathematically rather arbitrary. One of those who had never accepted it was Dirac himself. As late as in 1974, a year already after the discovery of asymptotic freedom and the formulation of QCD, he summed up his attitude towards renormalization with particular clarity and eloquence in [104]. After recalling the way the UV infinities in quantum electrodynamics are removed he says:

> *Hence most physicists are very satisfied with the situation. They say: "Quantum electrodynamics is a good theory, and we do not have to worry about it any more." I must say that I am very dissatisfied with the situation, because this so-called "good theory" does involve neglecting infinities which appear in its equations, neglecting them in an arbitrary way. This is just not sensible mathematics. Sensible mathematics involves neglecting a quantity when it turns out to be small – not neglecting it just because it is infinitely great and you do not want it!*

He then turns to the scenario in which there is a finite "cut-off", or equivalently, the electron has a finite radius, and observes

> *One then has a theory where infinities are gone, a theory that is sensible mathematically. An unfortunate result is that, of course, the relativistic invariance is spoiled. …. I think, however, that that is a lesser evil than departing from standard rules of mathematics and neglecting infinite quantities.*

His conclusion from all that was unambiguous and unequivocal:

> *I disagree with most physicists at the present time on just this point. I cannot tolerate departing from the standard rules of mathematics. Of course, the proper inference from this work is that the basic equations are not right. There must be some drastic change introduced into them so that no infinities occur in the theory at all and so that we can carry out the solution of the equations sensibly, according to ordinary rules and without being bothered by difficulties. This requirement will necessitate some really drastic changes: simple change will not do.*

I think Dirac was right in insisting on the observing the "standard rules" of mathematics but wrong in his final prophesy. As will be briefly reviewed in the next section and discussed in more detail in section 8, some of the nonabelian gauge theories have the property called "asymptotic freedom", which allows us to construct quantum field theories that are free of genuine UV infinities. The fact that even in these theories UV infinities do appear in perturbation theory is due to the fact that in standard perturbation theory physical quantities are expanded in powers of the renormalized coupling parameter, defined at a finite scale.

One aspect of the renormalization procedure that has undergone essential change, thanks to the work of Kenneth Wilson [105] and other theorists, concerns the nature of the coupling parameter itself. For Dirac and his contemporaries the electric charge appearing in the QED lagrangian was considered to have a given value, which might have been screened by quantum effects, but had nevertheless some definite value, and once this value was given the theory was fully defined. In modern view the construction of quantum field theory starts with discretized space-time and involves, as a nontrivial step, the limiting procedure in which the lattice spacing goes to zero [108]. A crucial feature of this limiting procedure concerns the fact that in order to get finite results for physical quantities, the parameters appearing in the original lagrangian must be nontrivial functions of the lattice spacing!

Replacing the lattice spacing with the radius of electron, the construction of lattice field theory follows closely the strategy employed by Landau, Pomeranchuk and their collaborators in the early fifties in their attempts to give a physical content and mathematical sense to the procedure of renormalization in QED. Trying to avoid the standard procedure of subtracting UV infinities they had therefore approached the problem from a different, though related, direction. They put a finite electric charge e_0, usually called "bare" charge, on the sphere of radius r_0 placed in the QED vacuum, calculated how it appears to a test particle at a finite distance $r > r_0$ and investigated what happens when we sent $r_0 \to 0$, as we indeed, must if we want to have Lorentz invariant theory. Note that if we do the same in classical physics, the answer is trivial. As discussed at length in the Introduction, the force at a distance r induced by the bare charge can

be characterized (see eq. (1.11)) by the effective electric charge $e_{\text{eff}}(r)$ which equals $e(r_0) = e_0$ for $r \geq r_0$ and vanishes for $r < r_0$. Consequently, if we shrink the electron radius to zero keeping $e_0 = e(r_0)$ fixed, the effective electron charge $e_{\text{eff}}(r) = e_0$ for <u>all</u> distances r!

In quantum field theory the answer is not that simple as there vacuum is not really "empty" but bustling with activity that is usually described in terms of "virtual pair creation and annihilation". However, the best analogy of what happens when we place the bare electric charge of a finite radius into QED vacuum is with the response of a dielectric medium to analogous action. The force exerted by the inserted bare charge polarizes the elementary electric dipoles, attracting the same sign charges and repelling the opposite ones. The net effect of this action is that the original, bare electric charge is screened, i.e. its magnitude decreases with increasing distance r. If we want to "see" the original bare charge we must come closer to r_0. The problem with this analogy is that the evaluation of the relation between $e_0 = e(r_0)$ (or, more conveniently, the combination $\alpha_0 \equiv e_0^2/4\pi$, called "couplant") and the effective couplant $\alpha(r)$ at the distance $r > r_0$, using standard QED perturbation theory, gives

$$\alpha(r) = \frac{\alpha(r_0)}{1 - \beta_0 \alpha(r_0) \ln(r_0/r)}, \quad \Rightarrow \quad \alpha(r_0) = \frac{\alpha(r)}{1 - \alpha(r)\beta_0 \ln(r/r_0)}. \tag{7.1}$$

where $\beta_0 = 2/3\pi$. The above expression holds for distances larger than Compton wavelength of the electron $r \geq \lambda_e \doteq 400$ fm, whereas below λ_e the effective couplant $\alpha(r)$ becomes essentially constant, approaching the classical value $1/137$. If we now send $r_0 \to 0$ keeping $\alpha_0 = \alpha(r_0)$ fixed, the first equation in (7.1) implies $\alpha(r) \to 0$ for any $r > 0$! In other words, shrinking the electron radius while keeping its charge fixed results in **free, non-interacting theory**! If we want to have finite $\alpha(r)$ at all distances r, the electric charge placed on the sphere must grow with its decreasing radius r_0. In fact the second equation in (7.1) tells us how exactly: it rises and diverges at a <u>finite</u> distance $r_L \simeq \lambda_e \exp(-137/\beta_0)$. We cannot go on shrinking the electron radius further as below r_L the couplant as given in (7.1) turns negative, making the theory meaningless. It was the appearance of this so called **Landau singularity** what had marred the attempts of Landau, Pomeranchuk and their school to formulate QED in a physically motivated and mathematically well-defined way. This, in turn had led to a temporary loss of interest in the lagrangian formalism of the local quantum field theory. An interesting and pedagogical recollection of these early attempts can be found in [119]. Although written back in 1976, mere 3 years after the breakthrough to be discussed in the next Section, it is still worth reading.

Although the failure to formulate the renormalization procedure in a fully consistent manner represented principal problem for quantum field theory, it had not prevented the application of QED to evaluation of quantum corrections to various physical processes. The point is that the distance, r_L, at which we encounter the Landau singularity, is unimaginably small and therefore we can for all practical purposes forget about it. The price we must, however, pay for it, is that the renormalization procedure has the features that had led Dirac to reject it. Most of the physicists had not assumed such radical position and QED was widely used to describe, and with great success, the electromagnetic interactions. It is, however, true that quantum field theory fell into disrepute as the basic theoretical tool for the treatment of strong interaction. Even disregarding the principal problem of the Landau singularity, there were two more reasons for it. For one, perturbation theory seemed inapplicable in the case of large values of the coupling between hadrons within the framework of Yukawa theory or the Eightfold Way. Second, it appeared entirely unsuitable to describe interactions between quarks within the quark model.

The rejection of quantum field theory as the basic theoretical tool for the description of strong interaction was accompanied by the formulation and development of the analytic S-matrix theory, which dispenses with the concept of local fields and works primarily with matrix elements describing transitions between various initial and final states of observable hadrons. This approach had dominated strong interaction theory since middle fifties until the arrival of QCD in 1973. The reader is referred to [119,120] for illuminating description of its origins, achievements and the present role.

7.2 And its remarkable resurrection

As discussed in the previous Section the results of SLAC experiments on deep inelastic electron-proton scattering had led Feynman, Bjorken and a few other theorists to the formulation of parton model. In that model nucleon behaved in hard collisions as a beam of almost noninteracting point-like constituents. By the early 1973 the data had moreover provided compelling evidence for identification of charged partons with

quarks and indirect one for the presence of neutral partons in the nucleon as well. The candidates for these neutral partons were the gluons introduced by Nambu in his model of interquark forces mediated by the exchange of octet of colored vector bosons. Recall also that in Nambu's model the color sector of interquark force provided a crude and semiquantitative explanation of the quark confinement (4.34). By 1973 all the important ingredients of what we now call Quantum chromodynamics were thus available. What remained to be done was to show that this quark-parton model with approximate scaling on one side and the color confinement on the other follows from some field theory.

With the benefit of hindsight, it seems almost trivial to realize that these seemingly contradictory requirements can be accommodated in a field theory that differs from QED in just one "minor" point: the <u>sign</u> of the coefficient β_0 appearing in (7.1). Note that if β_0 is negative the result of the procedure discussed by Landau in QED changes dramatically: in order to keep the corresponding effective couplant (let me denote it $\alpha_s(r)$ in anticipation of the current notation) at a given distance r finite as $r_0 \to 0$, the bare couplant $\alpha_s(r_0)$ does not have to diverge at some finite distance as in QED, but must <u>vanish</u> instead! This has three important consequences:

- the renormalization procedure ala Landau can be pursued to arbitrarily small distances, allowing thus the construction of Lorentz invariant field theory,

- there are in fact not genuine ultraviolet infinities,

- the effective couplant $\alpha_s(r)$ depends on the distance r in opposite manner that $\alpha(r)$ in QED: it decreases at small and grows at large distances, the latter behavior suggesting the quark confinement.

In other words, in theories with $\beta_0 < 0$ the effects of quantum fluctuations in the vacuum lead to **antiscreening** of the appropriate charge! It should be emphasized that the rise in such theories of the effective couplant at large distances must be taken as mere indication, not a real proof of the quark confinement.

Although all the experimental evidence available in late 1972 suggested that only theories with the above antiscreening behaviour of their effective couplant have a chance at being the correct theories of the strong interactions, almost nobody believed that such theory existed at all. The atmosphere of those times is described in great detail and with much insight by David Gross, one of the discoverers of QCD, in his excellent and remarkably candid historical reminiscence [120].

> I decided, quite deliberately, to prove that local field theory could not explain the experimental fact of scaling and thus was not an appropriate framework for description of the strong interactions. Thus, deep inelastic scattering would finally settle the issue as to the validity of quantum field theory.
>
> The plan of the attack was twofold. First, I would prove that "utraviolet stability", the vanishing of the effective coupling at short distances, later called asymptotic freedom, was necessary to explain scaling. Second, I would show that there existed no asymptotically free field theories.

The first part of his plan was relatively simple. To complete his intended destruction of local quantum field theory he set out with his PhD student Frank Wilczek to calculate the coefficient β_0 in nonebalian gauge theories, which represented the only "hole" in the argument Gross was pursuing. I will describe the basics of gauge theories, abelian as well as nonabelian, in the subsection 7.4, so let me just state the result of their calculations: Gross was wrong in his expectation as they found that (in some circumstances) nonabelian gauge theories lead to $\beta_0 < 0$! With an apparent reference to the biblical story of the conversion of the Jew named Saul to the Apostle Paul, he frankly admits [120]:

> For me the discovery of asymptotic freedom was totally unexpected. Like an atheist who has just received a message from a burning bush, I become an immediate true believer. Field theory was not wrong – instead scaling must be explained by an asymptotically free gauge theory of the strong interactions.

This breakthrough observation had changed practically overnight the status of quantum field theory: instead of a rejected child it became the subject of worship of most theorists, prominent among them Gross himself.

The circumstances of the discovery of asymptotic freedom have one trait common to most of the fundamental steps in the development of the Standard model. As in the cases of the discoveries of strangeness (Gell-Mann and Nishijima), Eightfold Way (Gell-Mann and Ne'eman), the Quark model (Gell-Mann and Zweig), also the discovery of asymptotic freedom in nonabelian gauge theories was achieved simultaneously

by two groups. Beside the duo Gross-Wilczek, the same result was obtained simultaneously and independently by another PhD student, David Politzer [122]. Both calculations were submitted on April 24th [121] and May 3rd [122] and published in the same issue of the Physical Review Letters, heralding the dawn of a new age in physics. These papers concerned the property of asymptotic freedom in general nonabelian gauge theory, and though of obvious relevance for strong interactions, had not yet discussed strong interactions of quarks explicitly. This was done in the papers [123, 125], which followed three months later and which contained not only explicit Lagrangian of what we now call QCD, but also detailed analysis of scaling violations using the sophisticated method of operator product expansion. These papers were submitted to Physical Review at the end of July 1973 and the first of them, [123], appearing simultaneously as a FERMILAB preprint [128]. Unquestionably, QCD was formulated and applied for the first time in these two papers and by their authors, with little contribution from other people.

Unfortunately, in most of the history articles and standard textbooks, including those otherwise very respectful, like the excellent web text *Particle Adventure* [126], the path to QCD is distorted and its formulation ascribed to Fritzsch, Gell-Mann and Leutwyler for their paper *Advantages of Color Octet Gluon Picture* [127]. This paper was submitted to Physics Letters B on October 1st, three months after the appearance of the preprint [128], which, as I emphasized above, contains complete formulation of QCD as well as its elaborate application to deep inelastic scattering. Fritzsch, Gell-Mann and Leutwyler refer in their paper to the discovery of asymptotic freedom, but do not quote the preprint [128]. It is hard to believe that this preprint would not be available at CALTECH, where they all worked at that time or otherwise be brought to their attention. In any case even the published version of the paper [123] appeared before that of [127] so there can be no question where the QCD Lagrangian appeared for the first time in published form. Contrary to [123, 125], the paper [127] contains no calculations and, moreover, their own words on page 367 of [127]

> For us the result that the color octet field theory model comes closer to asymptotic scaling than the color singlet is interesting, but not necessarily conclusive, since we conjecture that there may be a modification at high frequencies that produces true asymptotic scaling.

betray that its authors failed to appreciate the very basic consequence of QCD, namely the fact that in asymptotically free theories scaling is not exact, even at asymptotic frequencies, but only an approximate property at arbitrarily large frequencies (i.e. small distances)! To claim that QCD was born in the paper [127] is thus utter nonsense and serious historical falsification.

7.3 Maxwell equations in the covariant form

In classical physics the free electromagnetic field in the vacuum can be described by the electric and magnetic field strengths $\vec{E}(t, \vec{x})$ and $\vec{H}(t, \vec{x})$, which satisfy Maxwell equations

$$\text{rot}\vec{H} = \frac{\partial \vec{E}}{\partial t}, \; \text{div}\vec{E} = 0, \tag{7.2}$$

$$\text{rot}\vec{E} = -\frac{\partial \vec{H}}{\partial t}, \quad \text{div}\vec{H} = 0 \tag{7.3}$$

and appear in the expression for the force with which this field acts on a test particle with electric charge e and velocity v

$$\vec{F} = e\left[\vec{E} + \vec{v} \times \vec{H}\right]. \tag{7.4}$$

The second pair of Maxwell equations (7.3) allows us to express the field strengths $\vec{E}(x, t)$ and $\vec{H}(x, t)$ in terms of the scalar and vector potentials $A_0(x, t)$ and $\vec{A}(x, t)$ as follows

$$\vec{E} = -\text{grad}A_0 - \frac{\partial \vec{A}}{\partial t}, \quad \vec{H} = \text{rot}\vec{A}, \; x^\mu = (t, \vec{x}), \; x_\mu = (t, -\vec{x}) \tag{7.5}$$

because (7.5) automatically implies (7.3). Arranging A_0 and $\vec{A}$ as components of the four-potential

$$A^\mu \equiv (A^0, \vec{A}), \; A^0 = A_0; \quad A_\mu \equiv (A_0, -\vec{A}) \tag{7.6}$$

116

and introducing the antisymmetric tensor $F^{\mu\nu}$, the elements of which are the componets E_i, H_i

$$F^{\mu\nu}(x) \equiv \frac{\partial A^\nu(x)}{\partial x_\mu} - \frac{\partial A^\mu(x)}{\partial x_\nu} \ \Rightarrow F^{0i}(x) = -E_i(x), \ \ F^{ij}(x) = -\epsilon_{ijk}H_k(x), \tag{7.7}$$

allows us to rewrite (7.2) in a manifestly covariant form

$$\partial_\mu F^{\mu\nu}(x) = 0. \tag{7.8}$$

Within the framework of Lagrangian field theory (still at the classical level!) the above equations are just the Euler equations resulting from the Lagrangian

$$\mathcal{L}_{\text{free}}^{\text{Maxwell}} = -\frac{1}{4}F^{\mu\nu}(x)F_{\mu\nu}(x). \tag{7.9}$$

The point worth noting is the following. In classical physics the introduction of the four-potential $A_\mu(x)$ is very convenient as it reduces the original 8 equations (7.2-7.3) for six components E_i and H_i to only four in (7.8). On the other hand, we may in principle dispense with the four-potential A_μ as the expression for the Lorentz force (7.4) contains only the original quantities $\vec{E}$ and $\vec{H}$. As we shall see below, in quantum field theory the four-potential $A_\mu(x)$ is indispensable for proper formulation of interacting electromagnetic field.

7.4 Comments on the basic idea and concepts of gauge theories

In introducing the elements of perturbative QCD I shall draw on the analogy with the structure and principles of **Quantum Electrodynamics** (QED). Although there are several fundamental differences between QCD and QED, most of the QCD formalism is basically an extension of that of QED. In presenting the elements of QCD I shall emphasize the role played by the principle of **local gauge invariance** in the "derivation" of the equations of motion and discuss at length the physical interpretation of the Feynman diagrams. This will help us in the next chapter in understanding the way QPM is incorporated in QCD.

The basic object of QED is the **local fermion** field $\Psi(x)$, depending on space-time coordinates x and describing one type of charged particle with spin 1/2 (lepton or quark). On the classical level it is just the solution of the corresponding Dirac equation. In quantum field they $\Psi(x)$ becomes a **local operator** satisfying certain **anticommutation** relations. Skipping the formalism of quantized fields (for details see, for instance, [1, 2, 10]), I shall merely outline how the requirement of local gauge invariance, combined with Lorentz invariance and some conditions on the type of particles and properties of their interactions we wish to get, determines the basic properties of electromagnetic interactions of charged fermions and photons in QED. The heuristic "derivation" of QED Lagrangian goes as follows:

1. Start with the classical Lagrangian of a noninteracting fermion field $\Psi(x)$ of mass m

$$\mathcal{L}_{\text{free}}^{\text{fermion}} = \overline{\Psi}(x)(\mathrm{i}\ \slashed{\partial} - m)\Psi(x), \tag{7.10}$$

   which generates, via the **Euler–Lagrange** equations of motion, the Dirac equation for a free fermion with spin 1/2

$$(\mathrm{i}\ \slashed{\partial} - m)\,\Psi(x) = 0. \tag{7.11}$$

2. Define the **global gauge transformations** of this field as phase rotations:

$$\Psi' \equiv \exp(\mathrm{i}\alpha)\Psi(x), \tag{7.12}$$

   where α is a *real number*. The set of such transformations forms a group called U(1). Provided α is independent of x, the free fermion Lagrangian (7.10) *is invariant* under this simple transformation.

3. Impose the requirement of **local gauge invariance**, i.e. the invariance with respect to transformations (7.12) but for α *depending on x*. In that case the Lagrangian (7.10) by itself is no longer invariant due to the term

$$-\overline{\Psi}(x)\gamma^\mu\frac{\partial\alpha(x)}{\partial x^\mu}\Psi(x), \tag{7.13}$$

   which comes from the partial derivative in (7.10) and vanishes only for constant α.

4. To recover the local gauge invariance of the fermion Lagrangian one has to introduce another field, which *compensates* the noninvariance of (7.10). This is achieved by introducing the so called **gauge field** $A_\mu(x)$, which is a **vector** field describing a particle with spin 1 and mass M. Its Lagrangian, including the possible mass term, is a simple generalization of (7.9)

$$\mathcal{L}_{\text{free}}^{\text{gauge}} = -\frac{1}{4}F^{\mu\nu}(x)F_{\mu\nu}(x) + \frac{1}{2}M^2 A_\mu(x)A^\mu(x), \tag{7.14}$$

where $F^{\mu\nu}(x)$ is given in (7.7). The corresponding equations of motion (so far on classical level) read

$$\partial_\mu F^{\mu\nu}(x) = -M^2 A^\nu(x). \tag{7.15}$$

To cancel the noninvariant term (7.13) this gauge field is *assumed* to interact with $\Psi(x)$ via the interaction term

$$\mathcal{L}_{\text{int}} = e\overline{\Psi}(x)\gamma^\mu\Psi(x)A_\mu(x) \tag{7.16}$$

and transform *simultaneously* with (7.12) as

$$A'_\mu(x) = A_\mu(x) + \frac{1}{e}\frac{\partial\alpha(x)}{\partial x^\mu}, \tag{7.17}$$

where e, an arbitrary *real number*, determining the strength of the interaction (7.16), will be interpreted as the **electric charge** of the fermion field $\Psi(x)$.

5. There is, however, one casualty of this compensation mechanism, and namely the mass M of the gauge field $A_\mu(x)$. As the mass term $M^2 A_\mu A^\mu$ violates the invariance under the combined fermion and gauge field transformation only *massless* gauge field A_μ is compatible with the principle of local gauge invariance. This step essentially completes the construction, so far on classical level, of the full QED Lagrangian, which reads

$$\mathcal{L}^{\text{QED}} = -\frac{1}{4}F^{\mu\nu}(x)F_{\mu\nu}(x) + \overline{\Psi}(x)(\mathrm{i}\,\slashed{\partial} - m)\Psi(x) + e\overline{\Psi}(x)\gamma^\mu\Psi(x)A_\mu(x), \tag{7.18}$$

which describes the system of interacting Dirac and Maxwell fields. The coupled equations of motion satisfied by these fields read:

$$[\mathrm{i}\gamma_\mu\left(\partial^\mu - \mathrm{i}eA^\mu(x)\right) - m]\Psi(x) = 0, \tag{7.19}$$
$$\partial_\mu F^{\mu\nu}(x) = e\overline{\Psi}(x)\gamma^\nu\Psi(x). \tag{7.20}$$

Note that the Lagrangian (7.18) of interacting fields can be obtained from the sum of free lagrangians (7.10) and (7.14) by the replacement, called **minimal substitution**, of the partial derivative ∂_μ by the **covariant derivative** $D_\mu \equiv \partial_\mu - \mathrm{i}eA_\mu(x)$. The name comes from the fact that $D_\mu\psi(x)$ transforms in the same way (7.12) as fermion field itself

$$D'_\mu\psi'(x) = \exp(\mathrm{i}\alpha(x))D_\mu\psi(x) \;\Rightarrow\; D'_\mu = \exp(\mathrm{i}\alpha(x))D_\mu\exp(-\mathrm{i}\alpha(x)) \tag{7.21}$$

6. Quantization brings in some important conceptual changes, but doesn't modify the essence of the above "derivation". Moreover the form of the interaction term (7.16) suggests that the factor at the basic QED vertex e–γ–e will be proportional to $e\gamma^\mu$.

The above chain of steps does not, of course, constitute a real derivation of QED interaction term from the first principles. We have emphasized the role of the local gauge invariance, but it is true that this principle alone is not sufficient to determine the QED Lagrangian uniquely. For instance we could add to (7.18) any power of the photon kinetic term $F^{\mu\nu}F_{\mu\nu}$, which by itself is locally gauge invariant and thus obtain **selfinteraction of electromagnetic field**! Similarly we could couple the field tensor $F^{\mu\nu}$ directly to the fermion tensor $\overline{\psi}\sigma_{\mu\nu}\psi$, which also represents locally gauge invariant interaction. In this case we would dispense with the gauge potential $A_\mu(x)$ and fermions would interact with photons via their magnetic moments, but this interaction would be of *finite range*. In both of the preceding alternatives the kind of interaction we get at the classical level is significantly different from the *coulomb* forces observed in the nature. The fact that the photon is massless corresponds to our experience that electromagnetic interactions are of

long (actually *infinite*) range, but in classical theory none of the mentioned alternatives can be rejected merely on theoretical grounds. In quantum field theory there is a powerful theoretical principle, which actually rules out both of the mentioned alternatives and which, combined with the requirement of local gauge invariance, determines the QED Lagrangian essentially *uniquely*: **the renormalizability of the theory**. We shall discuss this concept in the next section. On the other hand there is no a priori reason, why only locally gauge invariant field theories should exist in nature. For instance, it is possible to formulate a renormalizable theory of *massive* photons, interacting locally with fermions but such an interaction is again of *finite* range and thus doesn't describe the real world.

The principle of local gauge invariance is probably the most fundamental principle of the current quantum field theory. On the other hand it leads to complications in the description of the very simplest quantity concerning the photons: propagator of the free photon. To see where the problem comes from consider the definition of a free photon propagator $D_{\mu\nu}(x, y)$, which follows from the Maxwell equations for the gauge field $A_\mu(x)$

$$\partial_\mu F^{\mu\nu}(x) = \left(\partial^2 g^{\mu\nu} - \partial^\mu \partial^\nu\right) A_\mu(x) = 0 \tag{7.22}$$

and reads

$$\left(\partial^\mu \partial^\rho - \partial^2 g^{\mu\rho}\right) D_{\rho\nu}(x - y) = \mathrm{i}\delta_{\mu\nu}\delta(x - y). \tag{7.23}$$

The Fourier transform of $D_{\mu\nu}(x - y)$, denoted $D_{\mu\nu}(k)$, satisfies, as a consequence of (7.23), the following matrix equation

$$\left(k^\mu k^\rho - k^2 g^{\mu\rho}\right) D_{\rho\nu}(k) = -\mathrm{i}g^\mu_\nu = -\mathrm{i}\delta_{\mu\nu}. \tag{7.24}$$

However, as the matrix $(k_\mu k_\nu - k^2 g_{\mu\nu})$ is *singular* the equation (7.24) has no solution! [1] The reason why this is so lies precisely in the fact that with each $A^{(0)}_\mu(x)$ satisfying the equation of motion (7.22), all the fields $A_\mu(x)$, gauge equivalent [2] to $A^{(0)}_\mu(x)$, do so as well. One way how to avoid this problem is to select from each class of gauge equivalent fields $A_\mu(x)$ merely those satisfying some additional **gauge fixing condition**,

$$f(A_\mu(x)) = 0. \tag{7.25}$$

and solve the equation (7.24) merely on the subset of functions satisfying (7.25). Alternatively we may give (7.23) a good meaning by adding to (7.18) the so called **gauge fixing term** of the form

$$-\frac{1}{2\alpha_G} \left(f(A_\mu(x))\right)^2, \tag{7.26}$$

where α_G is an arbitrary real number called **gauge parameter**. There are various classes of such gauge fixing terms, here I mention two of them, which are most often used in perturbation theory:

- The class of **covariant** gauges, where

$$f(A_\mu) \equiv \partial_\mu A^\mu(x). \tag{7.27}$$

  In this case the addition of (7.26) modifies the equations of motion (7.22) to

$$\partial_\mu F^{\mu\nu}(x) + \frac{1}{\alpha_G}\partial^\nu \left(\partial_\mu A^\mu\right) = \left(\partial^2 g^{\mu\nu} - (1 - 1/\alpha_G)\partial^\mu \partial^\nu\right) A_\mu(x) = 0 \tag{7.28}$$

  and the definition equation for the propagator in this gauge

$$\left((1 - 1/\alpha_G)k^\mu k^\rho - k^2 g^{\mu\rho}\right) D_{\rho\nu}(k) = -\mathrm{i}g^\mu_\nu \tag{7.29}$$

  has a well–defined solution. Writing $D_{\mu\nu}(k)$ as

$$D_{\mu\nu}(k) \equiv \mathrm{i}\frac{d_{\mu\nu}}{k^2} \tag{7.30}$$

  it is easy to convince oneself that (7.29) is satisfied for

$$d_{\mu\nu} = -g_{\mu\nu} + (1 - \alpha_G)\frac{k_\mu k_\nu}{k^2}. \tag{7.31}$$

---

[1] For *massive* photons k^2 should be replaced in (7.24) by $k^2 - m^2$, which renders it regular and we find without problems $d_{\mu\nu} = -g_{\mu\nu} + k_\mu k_\nu/m^2$.

[2] Gauge fields related by the gauge transformation (7.17) are called **gauge equivalent**.

Clearly *any* nonzero α_G is good for this purpose, but $\alpha_G = 1$, defining the so called **Feynman gauge**, makes the propagator particularly simple for calculations. Absence of the gauge fixing term, corresponding formally to the case $\alpha_G = \infty$, makes (7.31) obviously ill-defined.

- The class of **axial** gauges with the gauge fixing term given as

$$f(A_\mu(x)) \equiv c_\mu A^\mu(x), \tag{7.32}$$

where c_μ is an arbitrary, but *fixed* four–vector with the dimension of mass. In this case the modified equations of motion read

$$\partial_\mu F^{\mu\nu}(x) - \frac{1}{\alpha_G} c^\nu (c_\mu A^\mu) = \left(\partial^2 g^{\mu\nu} - \partial^\mu \partial^\nu - \frac{1}{\alpha_G} c^\mu c^\nu \right) A_\mu(x) = 0 \tag{7.33}$$

and the corresponding equation for the propagator

$$\left(k^\mu k^\rho + \frac{1}{\alpha_G} c^\mu c^\rho - k^2 g^{\mu\rho} \right) D_{\rho\nu}(k) = \mathrm{i} g^\mu_\nu \tag{7.34}$$

has again a well–defined solution with

$$d_{\mu\nu} = -g_{\mu\nu} + \frac{k_\mu k_\nu \left(\alpha_G k^2 - c^2 \right)}{(kc)^2} + \frac{k_\mu c_\nu + k_\nu c_\mu}{kc}. \tag{7.35}$$

For $c^2 = 0$ we talk about *lightlike* gauges. Setting in addition $\alpha_G = 0$ we get

$$d_{\mu\nu} = -g_{\mu\nu} + \frac{k_\mu c_\nu + k_\nu c_\mu}{kc}. \tag{7.36}$$

which is particularly suitable for perturbative calculations. Axial gauges have several important advantages with respect to the covariant ones, but, on the other hand, also some problems, related to the potential zeros of the product (kc) in denominator of (7.35).

The terms added in (7.28) and (7.33) to the original classical equations of motion (7.22) vanish for the fields $A^\mu(x)$ satisfying the gauge fixing condition (7.25). On the other hand not all the solutions $A^\mu(x)$ of the modified equations (7.28) and (7.33) satisfy the condition (7.25). For such solutions the additional terms $1/\alpha_G \partial^\nu (\partial_\mu A^\mu)$ and $-1/\alpha_G c^\nu (c_\mu A^\mu)$ *do not vanish* and consequently the corresponding tensor $F^{\mu\nu}$ *depends* on the choice of the gauge fixing condition and the value of α_G, in violation of the gauge invariance of the theory. In other words, in classical physics the addition of the gauge fixing term (7.26) to the Lagrangian (7.18) is *not equivalent* to selecting only those fields $A_\mu(x)$ satisfying the condition (7.25). It is inconsistent with gauge invariance and thus illegal.

The situation is quite different in quantum field theory. In QED the addition to the Lagrangian (7.18) of the gauge fixing term (7.26) *is consistent* with the gauge invariance of the quantized theory, but the proof that the results for **physical quantities** are independent of

- the choice of the function $f(A_\mu)$ and

- the value of gauge parameter α_G

is by no means trivial and is an integral part of the whole quantization procedure! The simplest way how to prove this invariance is to work within the framework of the path integral formalism, where, roughly speaking, all elements of a class of gauge equivalent fields contribute to the propagators ((7.31) or (7.35)), but with **weights** proportional to the exponential of the gauge fixing term (7.26). So in quantum field theory the addition of gauge fixing term <u>does not</u> mean selecting one of the "gauge copies" like in classical electrodynamics. This holds only in the case of the *Landau gauge*, corrisponding to $\alpha_G = 0$ in (7.26).

The basic theoretical tool of QED has been and remains the perturbation theory, where the physical (as well as some other) quantities are expressed as expansions in powers of the electric charge e^2, appearing in the Lagrangian (7.18). [3] Let us recall that in perturbation theory individual terms of these power expansions are graphically represented by the **Feynman diagrams**, which contain three basic elements:

---

[3]The crucial concept of **renormalized** electric charge is discussed in the Chapter 8.

- External legs, describing the in and outgoing *real* particles,

- Propagators, corresponding to *virtual* intermediate states.

- Interaction vertices, describing the local interaction of quantized fields.

The one–particle states describing *free*, noninteracting particles before and after the collision, are represented by plane waves of definite momenta (see Appendix). In the case of photons the **polarization vector** $\epsilon_\mu(k)$, describing the polarization states of a photon with momentum k, depends also on the choice of gauge we work in. In the Landau gauge of the class of covariant gauges $\epsilon_\mu k^\mu = 0$ but it still means that there are *three* independent solutions. [4] Two of them

$$\epsilon^{(1)} \equiv (0,1,0,0), \quad \epsilon^{(2)} \equiv (0,0,1,0) \tag{7.37}$$

describe the physical, **transverse** photons, while the third one

$$\epsilon^{(3)} \equiv (1,0,0,1) \tag{7.38}$$

corresponds to the *unphysical* **longitudinal** ones. This latter animal is unphysical because the corresponding tensor $F^{\mu\nu}(x)$ vanishes and must therefore *decouple* from all physical quantities. In other covariant gauges *all four* independent polarization states have to be taken into account, but again the two unphysical ones mutually cancel in the physical quantities.

Both the propagators and wave functions representing external particles are independent of a particular form of the interaction term. The latter determines the rule for the interaction vertex of QED, which, as I already emphasized, closely reflects its *classical* form (7.16). Apart from the rules for the powers of the imaginary unit, essentially everything else can be read off this classical expression.

At this place I would like to emphasize the importance and usefulness of the concept of a "virtual" particle, or, more correctly, a "virtual" state. Its (cautious) use is very helpful in getting an intuitive understanding of various complicated calculations in perturbation theory. The point is that although "virtual" particles are just intermediate states in quantum evolution, they behave *nearly as real particles* if observed for sufficiently short time interval. This happens in the hard scattering processes (like those discussed in Chapters 5 and 6) via the interaction with some "testing" particle, (electron, muon, etc.). The measure of the virtuality of a virtual particle of mass m and momentum k is basically its off–mass shellness $m_{\text{virt}}^2 \equiv k^2 - m^2$, but what really determines whether this virtual particle behaves nearly like real is actually the ratio m_{virt}/Q, where "Q" denotes the "hardness" of the collision. [5] So even states far off mass-shell may behave nearly like real particles when observed for short enough time.

In QCD, the chain of steps leading to its Lagrangian is similar but there is a *fundamental* difference due to the **nonabelian** character of the corresponding group of gauge transformations. This, in turn, is due to the fact that **quarks exist in three color states** and thus the basic mathematical quantity describing a quark of a particular flavor (like u, d, s, etc.) is the matrix in the color space

$$\Psi(x) \equiv \begin{pmatrix} \Psi^1(x) \\ \Psi^2(x) \\ \Psi^3(x) \end{pmatrix} \tag{7.39}$$

and its *local* gauge transformations are multiplications by the 3×3 unitary, unimodulary matrices of the color SU(3) group:

$$\Psi'(x) = \underbrace{\exp(i\alpha_a(x)T_a)}_{\in\, SU(3)} \Psi(x) = S\Psi(x), \tag{7.40}$$

where the sum in the exponent runs over the 8 generators T_a of $SU(3)$ and $\alpha_a(x), a = 1, \cdots 8$ are functions of x. Postulating the above form of gauge transformations of quark fields, the existence of 3 quark colors requires then introduction of 8 gauge fields A_μ^a, describing 8 **colored gluons** and conveniently represented by a column matrix

$$\vec{A}_\mu(x) \equiv \begin{pmatrix} A_\mu^1(x) \\ \vdots \\ A_\mu^8(x) \end{pmatrix}. \tag{7.41}$$

---

[4] Recall that within the class of covariant gauges only in the Landau gauge does the field A_μ really satisfy the equation (7.25) with $f(A_\mu)$ given in (7.27).

[5] In DIS, discussed in Chapter 5, "Q^2" is just the four-momentum transfer squared Q^2.

There is an alternative representation of the gluon octet by the 3×3 matrix defined as

$$A_\mu(x) \equiv A_\mu^a(x) T^a, \tag{7.42}$$

which is widely used, see for instance [17]. For each color index a we can define the tensor of field strength

$$F_a^{\mu\nu}(x) \equiv \frac{\partial A_a^\nu(x)}{\partial x_\mu} - \frac{\partial A_a^\mu(x)}{\partial x_\nu}. \tag{7.43}$$

Taking into account these technical complications and following step by step the chain of considerations sketched above for QED, we write down the QCD Lagrangian in the form: [6]

$$\mathcal{L}^{QCD} = -\frac{1}{4}\vec{F}_{\mu\nu}\vec{F}^{\mu\nu} + \overline{\Psi}(\mathrm{i}\,\slashed{\partial} - m_q)\Psi + g\overline{\Psi}\gamma_\mu \vec{T}\Psi\vec{A}^\mu, \tag{7.44}$$

where the scalar product of vectors is a shorthand for the sum over the color index a of the gluon fields A_a and the generators T_a. It is straightforward to show that the scalar product in the third term of (7.44) *is* invariant under the simultaneous *global* (i.e. with α_a independent of x) gauge transformations (7.40) of the quark field and corresponding *global* rotations of gluon fields under the **adjoint** representation of SU(3):

$$\vec{A}'_\mu = \exp\left(\mathrm{i}\alpha_a F_a\right)\vec{A}_\mu, \quad (F_a)_{bc} \equiv \frac{1}{\mathrm{i}}f_{abc}. \tag{7.45}$$

In order to distinguish this new SU(3) group from the "old" SU(3) related to quark **flavors**, I shall denote the former $SU_c(3)$.

Contrary to QED, the Lagrangian (7.44) is, however, not the end of the story. The problem, absent in QED, is that for the infinitesimal *local* gauge transformations of nonabelian gluon fields, defined in analogy with QED as

$$\left(A_\mu^a\right)' \equiv A_\mu^a + \alpha_b f_{bac} A_\mu^c + \frac{1}{g}\frac{\partial\alpha^a(x)}{\partial x^\mu} \Rightarrow A'_\mu = SA_\mu S^{-1} + \frac{\mathrm{i}}{g}S\partial_\mu S^{-1} \tag{7.46}$$

the last term, made from the gradients of $\alpha_a(x)$, cancels in the definition of the tensors $F_{\mu\nu}^a$ as in QED, but the color rotations of the first term *do not*, making the gluonic kinetic energy term $F_{\mu\nu}^a F_a^{\mu\nu}$ $SU_c(3)$ noninvariant! The reason for that is the *noncommutativity* of color rotations and the remedy simple: add to the r.h.s. of (7.43) the term which *compensates* this noninvariance, i.e. introduce

$$G_a^{\mu\nu} \equiv \frac{\partial A^{a\nu}(x)}{\partial x_\mu} - \frac{\partial A^{a\mu}(x)}{\partial x_\nu} + gf_{abc}A_b^\mu A_c^\nu \tag{7.47}$$

and use $G_{\mu\nu}^a$ instead of $F_{\mu\nu}^a$ in (7.44). Writing $G_{\mu\nu}^a G^{a\mu\nu}$ explicitly in terms of A_μ^a we find, besides the expected kinetic term $\vec{F}_{\mu\nu}\vec{F}^{\mu\nu}$, the following two additional terms, which represent the **selfinteraction of three and four gluons**

$$gf_{abc}\left[\left(\frac{\partial A_\nu^a}{\partial x^\mu} - \frac{\partial A_\mu^a}{\partial x^\nu}\right)A_b^\mu A_c^\nu + \left(\frac{\partial A^{a\mu}}{\partial x_\nu} - \frac{\partial A^{a\nu}}{\partial x_\mu}\right)A_{b\mu}A_{c\nu}\right] + g^2 f_{abc}f_{ade}A_b^\mu A_c^\nu A_{d\mu}A_{e\nu}. \tag{7.48}$$

This feature has profound consequences, which will be discussed in detail in the following chapter. Note that already the requirement of global nonabelian gauge transformation (7.40) implies the presence of the third term on the r.h.s. of (7.47) and thus leads to gluon selfinteraction.

In comparison with QED, the quantization of the full QCD Lagrangian brings another complication, related to the appearance in the perturbation theory of the so called **Faddeev–Popov ghosts**. In quantum gauge field theories these objects *must* be introduced in certain gauges to guarantee the consistency of the quantization procedure, in particular the unitarity of the theory. From the point of view of Feynman diagram calculations, ghosts behave as *scalar* particles, coupling to gauge bosons, but appearing only in *propagators*. Ghosts are in principle present in QED as well, but only in certain very special gauges, but are absent in the important classes of covariant and axial gauges. In QCD ghosts are still absent in axial gauges, but *do appear* in the class of covariant ones. For this reason the axial gauges are particularly suitable for the construction and probabilistic interpretation of the ladder diagrams, used in the derivation of evolution equations for parton distribution functions, discussed in next chapter.

---

[6] As the gluons are "flavor blind", including arbitrary number of quark flavors is simple: the full Lagrangian is a sum of the second and third terms of (7.44), each corresponding to one particular flavor, to which *one* gluon kinetic term is added.
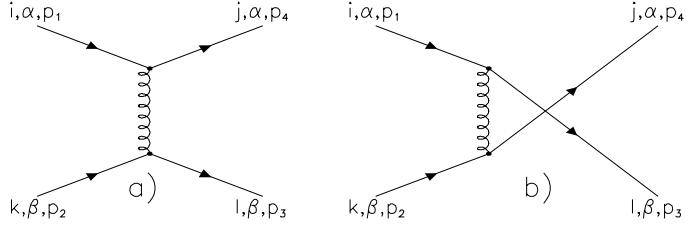
Figure 7.1: Lowest order Feynman diagrams for quark-quark elastic scattering.

7.5 Elementary calculations

In this Section several examples of simplest processes calculable in perturbative QCD are discussed. The main technical complications stemming from the color degree of freedom and gluon selfinteraction vertices are pointed out and cross-sections of various parton-parton subprocesses compared.

7.5.1 Quark-quark scattering

Consider the elastic scattering of two quarks with momenta, flavors and colors as indicated in Fig. 7.1 (with Greek and Latin letters labeling flavors and colors respectively), averaged over the spins and colors of quarks. Working in the Feynman gauge we find for the spin (λ) and color (c) averaged square of the corresponding matrix element

$$\langle |M_{if}|^2 \rangle = \frac{1}{9}\frac{1}{4} \sum_{\lambda,c} \left(|M_t|^2 + |M_u|^2 + 2M_t^* M_u \right), \tag{7.49}$$

where the sum runs over spins λ and colors c of all the four quarks and the Mandelstam variables s, t, u are defined as

$$s \equiv (p_1 + p_2)^2, \quad t \equiv (p_1 - p_4)^2, \quad u \equiv (p_1 - p_3)^2. \tag{7.50}$$

The invariant amplitudes M_t and M_u, corresponding to diagrams in Fig.6.1a and Fig.6.1b, are given as

$$M_t = \frac{\mathrm{i}g^2}{t} \left(T_{ji}^a T_{lk}^a \right) \left[\overline{u}_\alpha(p_4)\gamma^\mu u_\alpha(p_1) \right] \left[\overline{u}_\beta(p_3)\gamma_\mu u_\beta(p_2) \right], \tag{7.51}$$

$$M_u = \frac{\mathrm{i}g^2}{u} \left(T_{li}^a T_{jk}^a \right) \left[\overline{u}_\beta(p_3)\gamma^\mu u_\alpha(p_1) \right] \left[\overline{u}_\alpha(p_4)\gamma_\mu u_\beta(p_2) \right]. \tag{7.52}$$

In writing these expressions use was made of the fact that for colored gluons their propagator contains additional "color" Kronecker δ_{ab} factor (see Appendix). Considering first the square of M_t we find (*no* sum over t below)

$$\frac{1}{9}\frac{1}{4}\frac{g^4}{t^2} A_t B_t, \tag{7.53}$$

where the fractions $1/9$ and $1/4$ come from the color and spin averaging in the initial state. The final form the factor A_t, containing all the color factors, reads

$$A_t \equiv T_{ji}^a T_{lk}^a \underbrace{T_{ji}^{b*}}_{T_{ij}^b} \underbrace{T_{lk}^{b*}}_{T_{kl}^b} = \underbrace{T_{ji}^a T_{ij}^b}_{Tr(T_a T_b)} \underbrace{T_{lk}^a T_{kl}^b}_{Tr(T_a T_b)} = \frac{1}{4}(\delta_{ab})^2 = 2, \tag{7.54}$$

while working out bispinor traces is standard:

$$B_t \equiv \mathrm{Tr}\left(\not{p}_4 \gamma_\mu \not{p}_1 \gamma_\nu \right) \mathrm{Tr}\left(\not{p}_3 \gamma^\mu \not{p}_2 \gamma^\nu \right) = 32\left((p_1 p_2)(p_3 p_4) + (p_1 p_3)(p_2 p_4) \right) = 8(u^2 + s^2). \tag{7.55}$$

Putting all the relevant factors together yields

$$\langle |M_t|^2 \rangle = \frac{2}{9} g^4 \frac{2(u^2 + s^2)}{t^2}, \tag{7.56}$$
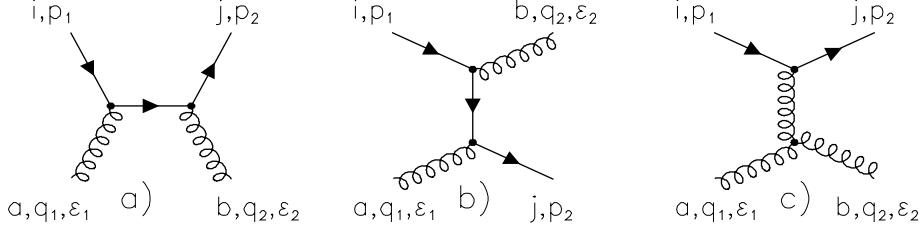
Figure 7.2: Lowest order Feynman diagrams for quark-gluon elastic scattering.

where the factor 2/9 comes from color factors. Proceeding similarly for the other two terms in (7.49) we get

$$\langle |M|^2 \rangle = \left(\frac{2}{9} \right)_c g^4 \left[\frac{2(u^2 + s^2)}{t^2} + \delta_{\alpha\beta} \frac{2(t^2 + s^2)}{u^2} - \delta_{\alpha\beta} \left(-\frac{1}{3} \right)_c \frac{4s^2}{tu} \right], \tag{7.57}$$

where the subscript "c" denotes factors coming from color traces. We see that the effect of color is twofold. It changes the overall *magnitude* of the quark-quark interaction, but also *suppresses* the relative contribution of the interference term.

7.5.2 Quark-gluon scattering

This is the simplest process where the gluon selfinteraction vertex does appear. The appropriate Feynman diagrams are shown in Fig. 7.2 and the corresponding invariant amplitudes, given as [7]

$$M_s = \frac{-ig^2}{s} \left(T^a_{li} T^b_{jl} \right) \left[\overline{u}^j(p_2) \, \not{\epsilon}_2 (\not{p}_1 + \not{q}_1) \, \not{\epsilon}_1 u^i(p_1) \right], \tag{7.58}$$

$$M_u = \frac{-ig^2}{u} \left(T^a_{jk} T^b_{ki} \right) \left[\overline{u}^j(p_2) \, \not{\epsilon}_1 (\not{p}_1 - \not{q}_2) \, \not{\epsilon}_2 u^i(p_1) \right], \tag{7.59}$$

where the polarization vectors ϵ_1, ϵ_2 describe the initial and final state gluons and all the colors and momenta upon which the amplitudes M_s, M_u depend were suppressed. Except for the presence of the color matrices, the above expressions are exactly the same as in QED and so is also the evaluation of Dirac traces. What is new in this channel is the contribution of the diagram in Fig. 7.2c, which contains the 3-gluon vertex

$$M_t = \frac{g^2}{t} \left[f_{abc} (T^c)_{ij} \right] C^{\lambda\mu\nu}(q_1 - q_2, -q_1, q_2) \epsilon_{1\mu} \epsilon_{2\nu} \left[\overline{u}^j(p_2) \gamma_\lambda u^i(p_1) \right]. \tag{7.60}$$

Inserting into (7.60) the explicit form of $C^{\lambda\mu\nu}$ as given in the Appendix, we get

$$M_t = \frac{g^2}{t} \left[f_{abc} (T_c)_{ji} \right] \tag{7.61}$$
$$\left\{ [\overline{u}(p_2) \, \not{\epsilon}_1 u(p_1)] (2q_1 - q_2) \epsilon_2 - [\overline{u}(p_2)(\not{q}_1 + \not{q}_2) u(p_1)] (\epsilon_1 \epsilon_2) + [\overline{u}(p_2) \, \not{\epsilon}_2 u(p_1)] (2q_2 - q_1) \epsilon_1 \right\}.$$

The evaluation of traces of Dirac matrices involved in this calculation is tedious, but essentially the same as in QED and is therefore omitted. What is new with respect to QED are the factors due to color matrices. Using the formulae from the Appendix, they can be easily worked out. As an example consider the color trace in the spin and color average of $|M_s|^2$

$$\frac{1}{3} \frac{1}{8} T^a_{li} T^b_{jl} T^{b*}_{jk} T^{a*}_{ki} = \frac{1}{24} \text{Tr}(T^a \underbrace{T^b T^b}_{(4/3)} T^a) = \frac{2}{9} \tag{7.62}$$

and similarly for M_u. The factors $\frac{1}{3}$ and $\frac{1}{8}$ result from averaging over three colors of initial quark and eight of initial gluon. The evaluation of an analogous color factor in the case of M_t involves another kind of color

---

[7]Flavor index is left out as the results are flavor independent.

traces:

$$\frac{1}{3}\frac{1}{8}f_{abc}\left(T^c\right)_{ji}f_{abd}\left(T^d\right)^*_{ji} = \frac{1}{24}\overbrace{f_{abc}f_{abd}}^{3\delta_{cd}}\underbrace{\mathrm{Tr}\left[T^cT^d\right]}_{(1/2)\delta_{cd}} = \frac{1}{2}. \tag{7.63}$$

The full result for the spin and color average of $|M_t|^2$ is listed in the Table 2 of subsection 7.5.4. Two aspects of this calculation merit a comment:

- In the course of summing over polarizations of initial and final gluons we needed the expression for the tensor

$$\sum_\lambda \epsilon_\mu(\lambda, q)\epsilon^*_\nu(\lambda, q), \tag{7.64}$$

where the sum runs over all polarizations of the gluon with momentum q. This sum *depends* on the gauge we work in and so do the results of the individual contributions of squares and interference terms of M_s, M_u, M_t. The full results is, however, *independent* of this choice. Within the Feynman gauge, i.e. summing over all four gluon polarizations, the result for (7.64) is just $-g^{\mu\nu}$. Summing over the *physical, transverse* polarizations of the gluon, we find (Exercise 23)

$$\sum_\lambda \epsilon_\mu(\lambda, q)\epsilon^*_\nu(\lambda, q) = -g_{\mu\nu} + \frac{q_\mu p_\nu + q_\nu p_\mu}{qp}, \tag{7.65}$$

where the fourvector p is defined as $p \equiv (q_0, -\vec{q})$.

- The division of the full invariant amplitude M_{if} into three M_s, M_t, M_u, corresponding to three diagrams in Fig. 7.2, *is not* gauge invariant and therefore has *no physical* meaning!

In the preceding Section I argued that the requirement of local gauge invariance leads to existence of gluon selfinteraction. It is illuminating to see how this mechanism operates on the level of Feynman diagram calculations. Consider, for instance, the sum $M_s + M_u$ of the contributions corresponding to the diagrams in Fig. 7.2a and 7.2b, i.e. those which don't contain the 3–gluon vertex and which have their close analogies in QED. The decoupling of unphysical, longitudinal gluons implies that the full result for the quark–gluon scattering amplitude must vanish if either the incoming or outgoing gluon is longitudinal. Assuming the second possibility, setting $\epsilon_2 = \lambda q_2$ and denoting the appropriate amplitudes $M_i(\lambda), i = s, t, u$ we find that the sum

$$M_s(\lambda) + M_u(\lambda) = ig^2\lambda\underbrace{\left[T^a, T^b\right]_{ji}}_{\neq 0}\left[\overline{u}(p_2)\,\not\epsilon_1 u(p_1)\right] = -g^2\lambda f_{abc}\left(T_c\right)_{ji}\left[\overline{u}(p_2)\,\not\epsilon_1 u(p_1)\right] \tag{7.66}$$

does not vanish, due to nonabelian character of QCD. It is easy to check that for the same longitudinal polarization $\epsilon_2 = \lambda q_2$ the t–channel amplitude

$$M_t(\lambda) = -\lambda g^2 f_{abc}T^c_{ji}\left[\overline{u}^j(p_2)\,\not\epsilon_1 u(p_1)\right] \tag{7.67}$$

compensates (7.66) provided

$$\left[T^a, T^b\right] = if_{abc}T^c, \tag{7.68}$$

which, indeed, the matrices T^a satisfy!

7.5.3 Gluon-gluon scattering

In QED the scattering of light on light is possible only in higher orders of perturbation theory via the box diagram in Fig. 7.3d and thus only thanks to the coupling of photons to charged particles. In QCD gluons can interact **directly**, even in absence of quarks, via the three and four gluon couplings (which means already at order g^2 in amplitude compared to order e^4 of the box diagram in Fig. 7.3d). Thus contrary to QED, there is a nontrivial theory called **gluodynamics**, which describes *only* the gluons and their selfinteractions. In fact, most of the calculations in lattice gauge theory [108] are of this kind as fermions are notoriously difficult to put on the lattice. The result of the evaluation of the diagrams in Fig. 7.3a-c is given in the next subsection.
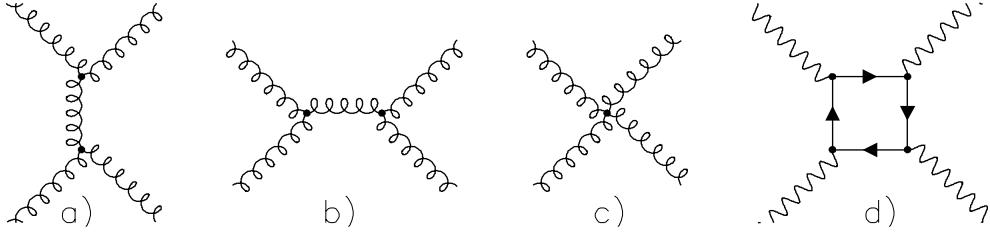
Figure 7.3: Lowest order Feynman diagrams for gluon–gluon elastic scattering in QCD (a-c), and for $\gamma - \gamma$ scattering in QED, (d). Neither momenta nor direction of the particles are shown. In the box diagram d), any charged particle can circulate around the loop.

7.5.4 Comparison of different parton subprocesses

In the following Table the results for the spin and color averaged invariant amplitudes $\langle |\ M_{if}\ |^2 \rangle$, corresponding to the simplest two–body processes in QCD are listed. The normalization is such that

$$\frac{\mathrm{d}\sigma}{\mathrm{d}t} = \frac{1}{16\pi s^2} \langle |M|^2 \rangle. \tag{7.69}$$

Process	$\dfrac{\langle |\ M\ |^2 \rangle}{g^4}$
$q_\alpha q_\beta \to q_\alpha q_\beta$	$\frac{2}{9}\left[\frac{2(s^2+u^2)}{t^2} + \left(\frac{2(t^2+s^2)}{u^2} - \frac{1}{3}\frac{4s^2}{ut} \right)\delta_{\alpha\beta} \right]$
$q_\alpha \overline{q}_\beta \to q_\alpha \overline{q}_\beta$	$\frac{2}{9}\left[\frac{2(s^2+u^2)}{t^2} + \left(\frac{2(t^2+u^2)}{s^2} - \frac{1}{3}\frac{4u^2}{st} \right)\delta_{\alpha\beta} \right]$
$qg \to qg$	$\left[\left(1 - \frac{us}{t^2}\right) - \frac{4}{9}\left(\frac{s}{u} + \frac{u}{s}\right) - 1 \right]$
$gg \to q\overline{q}$	$\frac{1}{6}\left[\frac{u}{t} + \frac{t}{u} \right] - \frac{3}{4}\left[1 - \frac{ut}{s^2} \right] + \frac{3}{8}$
$q\overline{q} \to gg$	$\frac{64}{9}M(gg \to q\overline{q})$
$gg \to gg$	$\frac{8}{9}\left[-\frac{33}{4} - 4\left(\frac{us}{t^2} + \frac{ut}{s^2} + \frac{st}{u^2} \right) \right] - \frac{9}{16}\left[45 - \left(\frac{s^2}{ut} + \frac{t^2}{us} + \frac{u^2}{ts} \right) \right]$

In the CMS of incoming particles (which are assumed massless) the Mandelstam variables s, t, u are related to the scattering angle ϑ^* as follows

$$t = -2E^{*2}(1 - \cos\vartheta^*) \Rightarrow \mathrm{d}t = 2E^{*2}\mathrm{d}\cos\vartheta^*. \tag{7.70}$$

Numerical comparison of these cross-sections for fixed s as functions of $\cos\vartheta^*$ is shown in Fig. 7.4. We see that:

- In all processes with gluon exchange in the t-channel the contribution of the corresponding Feynman diagram clearly dominates in the forward direction.

- Gluon–gluon channel gives by far the biggest cross-section in the whole angular range. This, however, *doesn't* imply that the gluon–gluon channel gives also the biggest contribution to the cross–sections of hard collisions of hadrons!

- The processes in which gluons can be exchanged in the t-channel are very steep at small angles (i.e. for $\cos\vartheta^* \to 1$) where the diagram corresponding to this exchange dominates the full cross-section. An example is given in Fig. 7.4b, where the three contributions to the quark-gluon cross-section are displayed separately, together with their sum. Note the marked difference between the behavior at small and large angles ϑ^*, where the exchange of quarks in the u-channel also leads to divergence, which, however, is weaker than that at small angles.
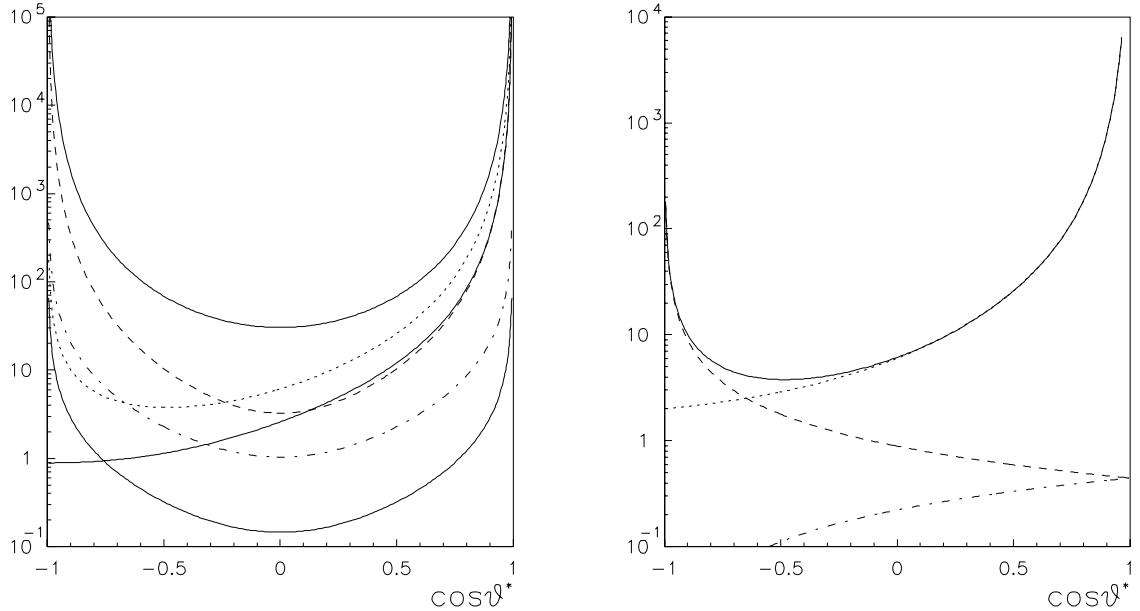
Figure 7.4: a): $\langle | M |^2 \rangle / g^4$ plotted for different 2 body processes as given in the Table below. The curves correspond, in decreasing order of their values for $\cos \vartheta = 0.5$, to $gg \to gg$, $qg \to qg$, $q\overline{q} \to q\overline{q}$, $qq \to qq$, $q\overline{q} \to gg$, $gg \to q\overline{q}$;b): Individual contributions of M_s (dash-dotted line), M_u (dashed line) and M_t (dotted line) to the full $qg \to qg$ cross-section (solid line)

7.6 Exercises

1. Show that longitudinal photon carries zero energy and is thus equivalent to nothing.

2. Construct interaction term between the Dirac fermion field Ψ and electromagnetic field which doesn't use the potential $A_\mu(x)$ but couples it directly the gauge invariant, observable tensor $F^{\mu\nu}$. Discuss physical interpretation of such an interaction.

3. Carry out explicitly the Fourier transformation of the equation (7.23).

4. Show that the expressions (7.31) and (7.35) are solutions of (7.29) and (7.34) respectively.

5. Show that the term $G^a_{\mu\nu} G^{\mu\nu}_a$ with $G^a_{\mu\nu}$ given in (7.47) is locally gauge invariant.

6. Use (7.48) to derive in a heuristic way the Feynman rules for 3 and 4 gluon vertex.

7. Carry out in detail the evaluation of the quark-gluon elastic cross-section.

8. Show in detail (7.66) and (7.67).

Chapter 8

Ultraviolet renormalization – basic ideas and techniques

8.1 *Infinities in perturbative calculations

This Chapter contains a guide through the ideas and techniques of ultraviolet [1] renormalization of perturbative QED and QCD. I shall argue that renormalization procedure is not merely a technique of removing unpleasant ultraviolet infinities, but also, and actually first of all, the matter of an **effective description of quantum phenomena**. This viewpoint will be emphasized in connection with the appearance of various ambiguities of the renormalization procedure. As misunderstanding in these matters often stems from confusion in the terminology I shall try to distinguish the inherent and unavoidable ambiguities of the renormalization procedure from those coming from the necessity to adopt some definitions, conventions etc..

As the essence of the UV renormalization is common to QED and QCD we shall first recall the situation in QED, where the basic quantity governing the strength of electromagnetic interactions of charged particles are their **electric charges**. [2] In QED the Lagrangian (7.18), when properly quantized, leads to the perturbation theory, in which physical quantities are expressed as power expansions in α. It turns out, however, that at higher orders the coefficients of these expansions, calculated according to the standard Feynman rules, come out formally infinite. These infinities are related to the appearance of the **loops** and result from integration over the unconstrained loop momenta. Integrating over the loop momenta we actually encounter two basic kinds of divergences:

A: **Ultraviolet divergences** (UV), coming from integration over the *large* values of the loop momenta.

B: **Mass divergences**, coming from integration over the region of *small* virtualities. These small virtualities appear in two different situations:

- For vanishingly small energy and momentum of the virtual particles. These so called **infrared** (IR) singularities occur for the massless photon, independently of the electron mass.
- Whenever two of the three particles in the QED vertex e-γ-e become parallel to each other. This can happen only if both the electron and photon are massless and thus these *parallel* singularities are, strictly speaking, absent in QED. There is, however, a trace of them in the presence of potentially large logarithms of the type $\alpha^k \ln^k(Q/m_e)$ (where Q is some external momentum) at k–th order of perturbation theory.

Both types of mass singularities can be regularized by introduction of a fictitious mass of the photon.

In this Chapter we shall analyze the ultraviolet divergences which lead to the important concept of the **renormalized electric charge**. The mass singularities are dealt with in a completely different way, discussed in Chapter 9.

---

[1] The adjective "ultraviolet" will be dropped in the following.

[2] It is convenient to express electric charges of all particles in units of the positron charge e. Instead of e^2 it is common to use the so called **fine structure constant** $\alpha \equiv e^2/4\pi$, which, in classical physics, equals $1/137$.

After the fundamentals of QED had been laid down by Dirac, Heisenberg, Pauli and others at the end of twenties, it was expected that perturbative expansions, i.e. the expansions in powers of $\alpha \equiv e^2/4\pi$, make sense as do similar expansions in quantum mechanics. [3] It was, however, soon realized that the expansion coefficients, evaluated according to formal rules of perturbation theory for *fixed, finite* α involve UV divergent integrals. This problem has been resolved in late forties by Feynman, Schwinger, Tomonaga and Dyson who formulated a procedure called **renormalization** which removes these infinities by **redefining** the couplant, masses and wave functions. Reformulating the perturbation theory as expansions in the so called **renormalized couplant**, they showed that all the infinities which originally appeared in the coefficients can be absorbed in these renormalized quantities, making the coefficients of this new expansion *finite*. It is interesting that Dirac himself had never accepted this procedure and considered until his death the renormalization artificial and unphysical. Although modern physics owes Dirac much, it seems that in this last point he was wrong. Renormalization has grown up from an admittedly suspicious method of handling infinities into a powerful tool of calculations, common to many branches of physics. In 1982 the Nobel prize for physics went to Kenneth Wilson for his fundamental contributions to the development and application of the renormalization theory. One of the areas of physics where the idea of renormalization, born originally in quantum field theory, proved extremely fruitful is the theory of phase transitions. An excellent account of this connection can be found in [115]. Before illustrating these statements on a simple example, let me stress that the UV renormalization procedure consists of *two* distinct steps:

A: the **regularization** in which the formally divergent integrals are *regularized*, i.e. basically *cut off* at high values of the loop momenta. There are numerous regularization techniques, for instance,

- simple cut-off,
- Pauli–Villars and its modifications, [2],
- analytical regularization, [116],
- dimensional regularization, [117,118],

each of them having its own advantages and shortcomings. In the following example we shall use the Pauli–Villars technique, primarily because it allows a clear physical interpretation of the renormalized electric charge while maintaining gauge invariance. As the intuitive understanding of the emergence of the scale dependence of the renormalized electric charge and renormalized mass is very helpful it is in my view better to start with a more "physical" regularization technique like the Pauli–Villars one. In multiloop calculations this technique is, however, rather cumbersome and so all current calculations are carried out within the **dimensional regularization**, which is much simpler. Moreover, in nonabelian gauge theories, like QCD, only the dimensional regularization preserves gauge invariance. The price paid for calculational simplicity of dimensional regularization is that the renormalization scale, on which the renormalized electric charge does depend, enters in a rather artificial way and its physical interpretation is therefore rather obscure. An introduction to this technique is given in Section 8.3.

B: the **renormalization**, when the divergent terms are absorbed in the newly defined **renormalized** quantities (couplant, mass and wave functions).

8.2 *Renormalization in QED

8.2.1 Electric charge renormalization

Let us now consider the scattering of an electron on a muon. In the lowest order of perturbative QED, the so called **one photon approximation**, this process is described by the Feynman diagram in Fig. 8.1a, while the order $\mathcal{O}(\alpha^2)$ corrections correspond to diagrams in Figs. 8.1b-e. The fermion circulating around the loop of Fig. 8.1b can be any charged fermion [4] but we shall analyze in detail the contribution of an electron loop only. Contributions of other elementary charged fermions are, apart from a possible difference in the magnitude of their electric charges, the same. Throughout this section we shall work in the Feynman covariant gauge, i.e. set $\alpha_G = 1$ in (7.31).

---

[3]Instead of the expression "coupling constant" for the electric or other charges, I shall in the following often use the term **couplant**, which is more appropriate for the quantity which is actually *not a constant*.

[4]We consider electromagnetic interactions of charged fermions only. For discussion of electromagnetic interactions of charged bosons, pseudoscalar as well as vector, see [2].
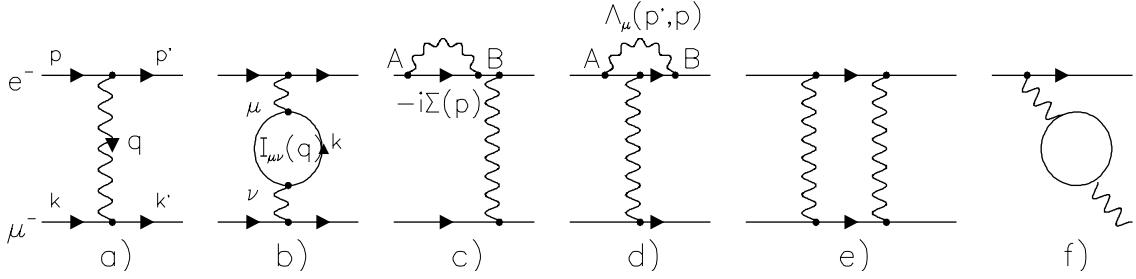
Figure 8.1: Feynman diagrams for electron–muon scattering in a leading (a) and next–to–leading order (b-d). The loop in c) can equally well be on any other of the four fermion legs while that in d) can circumvent the lower vertex as well.

Let us denote by e_B and call **bare electric charge** the quantity standing in each vertex of the Feynman diagrams of Fig. 8.1. In the Feynman gauge the leading order (LO) diagram gives a contribution to the invariant amplitude M_{if} which can be written as follows (see Fig. 8.1 for notation)

$$\underbrace{[\overline{u}(p')(-i\gamma_\mu)u(p)]\,[\overline{u}(k')(-i\gamma_\nu)u(k)]}_{W_{\mu\nu}(k,p,q)}\underbrace{\left(-\frac{ig_{\mu\nu}}{q^2}\right)}_{D_{\mu\nu}(q)}e_B^2, \tag{8.1}$$

where we have separated out the square of the bare electric charge and $D_{\mu\nu}$, the free photon propagator in the Feynman gauge. The contribution of the diagram in Fig. 8.1b has a similar structure

$$W_{\mu\nu}(k,p,q)\left(\frac{-i}{q^2}\right)I_{\mu\nu}(q,m_B)\left(\frac{-i}{q^2}\right)e_B^2, \tag{8.2}$$

where the tensor

$$I_{\mu\nu}(q,m_B)\equiv(-1)\int\frac{d^4k}{(2\pi)^4}\text{Tr}\left[(-ie_B\gamma_\mu)\frac{i}{\slashed{k}-m_B+i\epsilon}(-ie_B\gamma_\nu)\frac{i}{\slashed{k}-\slashed{q}-m_B+i\epsilon}\right] \tag{8.3}$$

has, as a consequence of the gauge invariance, the following general structure

$$I_{\mu\nu}(q,m_B)=(-g_{\mu\nu}q^2+q_\mu q_\nu)I(q^2,m_B) \tag{8.4}$$

with $I(q^2,m_B)$ containing the divergent integral. The quantity m_B is the so called **bare electron mass**, which throughout this subsection will be treated as a fixed number, although it also must be renormalized. This will be done in the next subsection. Note the factor -1 in (8.3) which comes from the closed electron loop as a result of the Pauli principle.

Naive counting of the powers of k in the integrand of (8.3) suggests *quadratic* divergence of the integral, but because of (8.4) the integrand does not behave like $1/k^2$, where k is the loop momentum, but rather like q^2/k^4, thereby causing only the *logarithmic* divergence of $I(q^2,m_B)$. In the following we shall moreover drop the term in (8.4) proportional to $q_\mu q_\nu$ as it vanishes after contraction with the tensor $W_{\mu\nu}$ in (8.2). The regularization of the diagram in Fig. 8.1b uses four technical ingredients:

1. **Schwinger parameterization**

$$\frac{i}{\slashed{k}-m+i\epsilon}=\frac{i(\slashed{k}+m)}{k^2-m^2+i\epsilon}=(\slashed{k}+m)\int_0^\infty dze^{iz(k^2-m^2+i\epsilon)} \tag{8.5}$$

for both fermion propagators forming the closed loop adds two more integration variables z_1,z_2.

2. Changing the original loop variable k to

$$l\equiv k-\frac{qz_2}{z_1+z_2}=k-q+\frac{qz_1}{z_1+z_2} \tag{8.6}$$

130

and using the following identities to perform integrals over d^4l

$$\int \frac{d^4l}{(2\pi)^4} \left[1, l_\mu, l_\mu l_\nu\right] e^{il^2(z_1+z_2+i\epsilon)} = \frac{1}{16\pi^2 i(z_1+z_2)^2} \left[1, 0, \frac{ig_{\mu\nu}}{2(z_1+z_2)}\right]. \tag{8.7}$$

3. Using the identity

$$1 = \int_0^\infty \frac{dt}{t} \delta\left(1 - \frac{z_1+z_2}{t}\right) \tag{8.8}$$

to trade one of the integrations over z_1 or z_2 for that over t.

4. The identity

$$\int_0^\infty \frac{dx}{x} \left(e^{iax} - e^{ibx}\right) = \ln \frac{b}{a}. \tag{8.9}$$

Performing the first two steps we get (for details see [2])

$$I_{\mu\nu}(q) = \frac{-ie_b^2}{4\pi^2} \int_0^\infty dz_1 \int_0^\infty \frac{dz_2}{(z_1+z_2)^2} \exp\left[i\left(q^2 \frac{z_1 z_2}{z_1+z_2} - (m_b^2 - i\epsilon)(z_1+z_2)\right)\right] J_{\mu\nu}(q), \tag{8.10}$$

where the tensor $J_{\mu\nu}$ has two terms

$$J_{\mu\nu}(q) \equiv 2(g_{\mu\nu}q^2 - q_\mu q_\nu) \frac{z_1 z_2}{(z_1+z_2)^2} + g_{\mu\nu} \left[\frac{-i}{z_1+z_2} - \frac{q^2 z_1 z_2}{(z_1+z_2)^2} + m_B^2\right]. \tag{8.11}$$

The first of them satisfies the transversality condition $q_\mu I^{\mu\nu}(q, m_B) = 0$ that follows from gauge invariance, while the second does not. However, a closer examination of the integral standing by $g_{\mu\nu}$ shows that it actually *vanishes*! Using in the first, transverse, term of (8.11), the identity (8.8) and carrying out the integral over z_1 or z_2, we get the following (still UV divergent!) expression for the scalar function $I(q^2)$

$$I(q^2, m_B) = \frac{ie_B^2}{2\pi^2} \int_0^1 dz\, z(1-z) \int_0^\infty \frac{dt}{t} \exp\left[it\left(q^2 z(1-z) - m_B^2 + i\epsilon\right)\right], \tag{8.12}$$

where the original logarithmic singularity of the four–dimensional integral over d^4k has been transformed into the same type of logarithmic singularity over the variable t.

After these preparatory considerations we can formulate the essence of the Pauli–Villars technique in the regularization of the logarithmically divergent integrals of the type (8.12). It rests in the replacement

$$I(q^2, m_B) \Rightarrow \overline{I}(q^2, m_B, M) \equiv I(q^2, m_B) - I(q^2, M), \tag{8.13}$$

where the subtraction is understood to be carried using the identity (8.9). i.e. first subtracting the integrands and then performing the integration. Clearly, M acts as the UV regulator, which should at the end of the renormalization procedure be sent to infinity. An important point to note is that because we use the *fermion* mass M in the subtracted term $I(q^2, M)$, this regularization technique *preserves gauge invariance*, which is of crucial importance for the proof of the full renormalizability of the theory.

Using (8.9) the regularized expression (8.13) equals

$$\overline{I}(q^2, m_B, M) = \frac{ie_B^2}{2\pi^2} \int_0^1 dz\, z(1-z) \ln\left(\frac{M^2}{m_B^2 - q^2 z(1-z)}\right). \tag{8.14}$$

Isolating the logarithmically divergent term this can be rewritten as

$$-i\overline{I}(q^2, m_B, M) = \frac{e_B^2}{12\pi^2} \ln \frac{M^2}{\mu^2} - \frac{e_B^2}{2\pi^2} \int_0^1 dz\, z(1-z) \ln\left(\frac{m_B^2 - q^2 z(1-z)}{\mu^2}\right). \tag{8.15}$$

The integral over the dimensionless parameter z is the trace of the original integration over the loop momentum k. Note that for any nonzero q^2 the expression (8.14) is *regular* for $m_B \to 0$ and can thus be used even for massless fermions. [5] In (8.15) we have introduced a new dimensional scale μ, which scales the cut–off

---

[5]Only for $q^2 = 0$ does the fermion mass act as an regulator of *parallel* singularities (see section 6.5 for details).

parameter M in the singular term $\ln(M^2/\mu^2)$, but obviously $\overline{I}(q^2, m_B, M)$ is in fact *independent* of μ! In conventional treatments, like [2], $\mu = m_B$ is set from the beginning and no new scale μ is introduced, but there is no serious justification for this assumption. It is reasonable for low q^2 processes, but as we shall see even in QED there are situations when this choice would be very inappropriate.

Once the method of regularization has been chosen we can add the contributions of the diagrams in Fig. 8.1a,b obtaining (dropping the tensor $W_{\mu\nu}$)

$$-\frac{\mathrm{i}g_{\mu\nu}}{q^2}e_B^2\left[1 - \frac{e_B^2}{12\pi^2}\ln\frac{M^2}{\mu^2} + \frac{e_B^2}{2\pi^2}\int_0^1 dz z(1-z)\ln\left(\frac{m_B^2 - q^2 z(1-z)}{\mu^2}\right)\right]. \qquad (8.16)$$

The crucial step of the renormalization procedure is to rewrite (8.16) as a product of two terms

$$-\frac{\mathrm{i}g_{\mu\nu}}{q^2}e_B^2\underbrace{\overbrace{\left[1 - \frac{e_B^2}{12\pi^2}\ln\frac{M^2}{\mu^2}\right]}^{Z_3^{-1}(A, e_B^2, M, \mu)}}_{\equiv\, e_R^2(A,\mu)}\left[\;1 + \frac{e_B^2}{2\pi^2}\underbrace{\int_0^1 dz z(1-z)\ln\left(\frac{m_B^2 - q^2 z(1-z)}{\mu^2}\right)}_{C(A, q^2, \mu, m_B)}\;\right], \qquad (8.17)$$

where we have absorbed UV divergence in the newly defined **renormalized electric charge** $e_R^2(A,\mu)$ [6] with the rest of the original $\mathcal{O}(e_B^2)$ term left in the *finite* contribution $C(A, q^2, \mu, m_B)$. The quantity Z_3 defined in (8.17) will be used later on.

In terms of this new parameter (8.17) can be written as

$$-\frac{\mathrm{i}g_{\mu\nu}}{q^2}e_R^2(A,\mu)\left[1 + \frac{e_B^2}{2\pi^2}C(A, q^2, \mu, m_B)\right] \Rightarrow -\frac{\mathrm{i}g_{\mu\nu}}{q^2}e_R^2(A,\mu)\left[1 + \frac{e_R(A,\mu)}{2\pi^2}C(A, q^2, \mu, m_B)\right], \qquad (8.18)$$

where the second expression, differing from the the first one merely by the replacement of the bare charge e_B^2 with the renormalized charge e_R^2 in the brackets, results if the renormalization procedure is carried out to order $\mathcal{O}(e_B^6)$. The above expression is particularly simple in two regions of q^2 where the integral over z in (8.17) can be performed analytically. For large $Q^2 \equiv -q^2$, i.e. for $-q^2/m_B^2 \gg 1$ we find

$$-\frac{\mathrm{i}g_{\mu\nu}}{q^2}e_R^2(A,\mu)\left[1 + \frac{e_R^2(A,\mu)}{12\pi^2}\left(\ln\frac{-q^2}{\mu^2} + \frac{5}{3}\right)\right], \qquad (8.19)$$

while for $-q^2/m_B^2 \ll 1$ we have, on the other hand,

$$-\frac{\mathrm{i}g_{\mu\nu}}{q^2}e_R^2(A,\mu)\left[1 + \frac{e_R^2(A,\mu)}{12\pi^2}\ln\frac{m_B^2}{\mu^2} + \frac{e_R^2(A,\mu)}{60\pi^2}\frac{-q^2}{m_B^2}\right]. \qquad (8.20)$$

In (8.17) we have isolated the singularity of the loop integral in the term proportional to $\ln(M^2/\mu^2)$ and included it in the definition of the renormalized electric charge $e_R^2(\mu)$. This procedure is, however, by no means unique as we can include in $e_R^2(\mu)$ arbitrary finite terms as well. For instance we can separate the integral over z in (8.16) into two parts as follows

$$\underbrace{\frac{e_B^2}{2\pi^2}\int_0^1 \mathrm{d}z z(1-z)\ln\frac{M^2}{m_B^2 + \mu^2 z(1-z)}}_{1 - Z_3^{-1}(B, e_B^2, M, m_B, \mu)} + \underbrace{\frac{e_B^2}{2\pi^2}\int_0^1 \mathrm{d}z z(1-z)\ln\frac{m_B^2 + \mu^2 z(1-z)}{m_B^2 + (-q^2)z(1-z)}}_{C(B, q^2, \mu, m_B)} \qquad (8.21)$$

and include in $e_R^2(\mu) \equiv e_B^2 Z_3^{-1}$ (denoted $e_R^2(B,\mu)$) [7] the whole first term, which can be written as

$$1 - Z_3^{-1}(B, e_B^2, M, m_B, \mu) = \frac{e_B^2}{12\pi^2}\ln\frac{M^2}{\mu^2} - \frac{e_B^2}{2\pi^2}\int_0^1 \mathrm{d}z z(1-z)\ln\left(\frac{m_B^2}{\mu^2} + z(1-z)\right). \qquad (8.22)$$

---

[6]The label "A" etc. will serve to distinguish between different definitions of the renormalized electric charge.

[7]The renormalized charge $e_R^2(B, \mu, m_B)$ depends therefore also on the mass m_B. This dependence will not, however, be explicitly written out.

This definition is motivated by the requirement that the finite correction $C(B, q^2, \mu)$ *vanishes* for $\mu^2 = -q^2$. In other words for $\mu^2 = -q^2$ all the effects of the electron loop correction are included in $e_R^2(B, \mu)$. For large μ, i.e. when $\mu^2 \gg m_B^2$, we find

$$1 - Z_3^{-1}(B) \to \frac{e_B^2}{12\pi^2} \left(\ln \frac{M^2}{\mu^2} + \frac{5}{3} \right), \qquad (8.23)$$

which differs from the previous case only by the presence of the additional constant $5/3$ in the definition of $e_R^2(B, \mu)$, while in the region $\mu \ll m_B$ we find a very different behavior of $Z_3^{-1}(B)$

$$1 - Z_3^{-1}(B) \to \frac{e_B^2}{12\pi^2} \ln \frac{M^2}{m_B^2} - \frac{e_B^2}{60\pi^2} \cdot \frac{\mu^2}{m_B^2}. \qquad (8.24)$$

For $\mu \to 0$, $e_R^2(B, \mu)$ therefore *approaches a finite value* $e_R^2(B, 0)$

$$e_R^2(B, \mu, m_B) = e_B^2 \left(1 - \frac{e_B^2}{12\pi^2} \ln \frac{M^2}{m_B^2} + \frac{e_B^2}{60\pi^2} \frac{\mu^2}{m_B^2} \right) \to e_B^2 \left(1 - \frac{e_B^2}{12\pi^2} \ln \frac{M^2}{m_B^2} \right) \equiv e_R^2(B, 0). \qquad (8.25)$$

The basic difference between the definitions A and B of the renormalized electric charge $e_R(\mu)$ concerns the way the effects of electron mass are taken into account. For massless electron they obviously coincide. In the case of the definition A *all the dependence* on m_B resides in the finite correction $C(A, q^2, \mu, m_B)$. Note in particular the presence in (8.20) of the term proportional to $\ln(m_B/\mu)$, which would be arbitrarily large and therefore dangerous for $\mu \to 0$. On the other hand $e_R^2(A, \mu)$ contains no information on m_B and therefore behaves exactly in the same way as for the massless electron. We shall therefore refer to $e_R^2(A, \mu)$ as **mass independent** definition of renormalized charge. Setting $\mu^2 = -q^2$ and approaching the limit $q^2 \to 0$ we see that while the expansion parameter $e_R^2(A, -q^2) \to 0$ [8], the coefficient function $C(A, q^2, \mu^2 = -q^2, m_B)$ on the contrary diverges in this limit due to the presence of the term $\ln(-q^2/m_B)$ and so do in fact all higher order coefficients as well.

In the case of the definition B the situation is different as $e_R^2(B, \mu, m_B)$ absorbs part of the dependence on m_B and is therefore called **mass dependent** renormalized charge. In the literature it is usually called MOM–like definition. For $\mu \gg m_B$ it behaves, apart from the additional constant $5/3$ in (8.23), as in the case A, but below m_B the character of the μ–dependence changes dramatically, leading to the constant limit $e_R^2(B, 0)$. Moreover, if we choose $\mu^2 = -q^2$, as is the usual practice [9], all the mass effects are included in $e_R^2(B, \mu^2 = -q^2)$, which has a finite limit for $q^2 \to 0$ and so has the coefficient function $C(B, q^2, \mu^2 = -q^2, m_B)$! It should, however, be emphasized that *in principle* both of these definitions of $e_R(\mu)$ are *equally legal*.

The way we introduced μ also suggests its physical interpretation as the *lower bound* of the logarithmically divergent integral of the type

$$\int_\mu^M \frac{\mathrm{d}k}{k} = \ln \frac{M}{\mu}. \qquad (8.26)$$

In view of the Heisenberg uncertainty relations the value of μ thus determines *how much* of the charge associated with the vacuum polarization is included in the definition of the renormalized charge $e_R(\mu)$, which can thus roughly be interpreted as the charge (original plus the induced one) inside a sphere of the radius $r = 1/\mu$ around the center of an electron. The remaining, nonsingular part of the integral over the loop momentum k in (8.4) provides a *finite* $\mathcal{O}(e_B^2)$ correction $C(q^2, \mu, m_B)$ to the leading term.

There are several other aspects of the above construction which merit a comment:

- The box diagram in Fig. 8.1e is *finite* and doesn't contribute to the definition $e_R^2(\mu)$.

- In all above manipulations the bare mass m_B of the loop electron was kept fixed, but the renormalization procedure actually concerns the masses as well. If also masses are renormalized, as discussed in the next subsection, m_B should everywhere be replaced by **renormalized mass** $m_R(\mu)$ of the electron.

---

[8] $e_R^2(A, \mu)$ as defined in (8.17) turns negative at small μ and eventually blows up to $-\infty$ as $\mu \to 0$. However, taking into account higher order terms makes the couplant $e_R^2(A, \mu)$ indeed vanish at $\mu = 0$. This point will be discussed in detail in subsection 7.4.2.

[9] The reasons for this choice are outlined in the comments below.

- The arbitrary scale parameter μ appears in both the renormalized *couplant* $e_R^2(\mu)$ and the coefficient function $C(q^2, \mu, m_B)$. The way we introduced it implies that it is an *unphysical* parameter and consequently the results for physical quantities *cannot* depend on its value. This natural requirement is the essence of the **renormalization group** (RG), introduced in the early fifties by Gell-Mann, Low, Bogoljubov and others. The RG expresses nothing else but the **internal consistency** of renormalized perturbation theory. Many of the considerations concerning practical applications of perturbative calculations are based on these consistency considerations, while the often difficult evaluations of Feynman diagrams provide an input to them in the form of certain RG invariants. This will be demonstrated on specific examples in QCD.

- The mechanism of cancelation of the μ–dependence in the expansion parameter $e_R^2(\mu)$ and the finite coefficient function $C(q^2, \mu, m_B)$ works in such a way that although the *full* sum of perturbation expansion *is μ-independent*, its approximation to order $e_R^4(\mu)$ (in fact, to any finite one) *is not*! This observation is the first signal of the *inevitable ambiguities* which appear when renormalized QED and QCD are considered to finite orders.

- In any practical calculation we therefore *have to choose* some value of μ! I shall discuss this question in great detail later, but it is clear that if the finite correction term, containing in both definitions the logarithm $\ln(-q^2/\mu^2)$ is to be reasonably "small" we should choose $\mu^2 \propto -q^2$.

- All the above steps and claims are simple and more or less obvious at the lowest order in which we have so far worked. What is nontrivial is to prove that this procedure can be carried out systematically to *all orders* of perturbation theory, so that infinities at all orders can be absorbed in the redefined renormalized couplant, masses and wave functions.

- Although the intermediate steps in the renormalization procedure depend on the specific kind of regularization technique used, the structure of the final results *does not!* There will always be some free scale parameter denoted μ, on which the redefined couplant as well as the new expansion coefficients will depend [10]. Once the renormalized perturbation expansions are expressed in terms of these free parameters and the mentioned invariants, they carry no trace of the regularization technique used.

In this way we have succeeded in absorbing the divergent term, containing the logarithm $\ln(M^2/\mu^2)$, in the definition of the renormalized couplants $e_R^2(A, \mu)$ or $e_R^2(B, \mu)$

$$e_R^2(A, \mu) \equiv e_B^2 \left[1 - \frac{e_B^2}{12\pi^2} \ln \frac{M^2}{\mu^2} \right] \tag{8.27}$$

and analogously for $e_R^2(B, \mu)$. In the final step of the renormalization procedure the limit $M \to \infty$ should be taken in (8.27). For *fixed* bare charge e_B^2 this is, however, impossible to do in a well–defined manner. This problem is usually circumvented by the reference to the fact that the bare charge e_B^2 appears *always* in combination with the divergent logarithm $\ln(M/\mu)$, so that we can forget about their separate existence and consider the renormalized electric charge e_R as the basic QED parameter. Although this approach provides a well–defined algorithm for the evaluation of finite coefficients of perturbative expansions, it is clearly unsatisfactory on physical grounds.

In fact many theorists, prominent among them Landau and his school, had attempted to give this procedure mathematically well–defined and physically sound basis by addressing the following question: can one define such a dependence $e_B^2(M)$ of the bare charge on the cut-off M, which would compensate in (8.27) the term proportional to $\ln M$, coming from the divergent integral, and allow the construction of the finite limit for the renormalized charge $e_R^2(\mu)$ when $M \to \infty$? As the inverse of M provides the cut-off in coordinate space we could think of $e_R^2(M)$ as the charge inside a sphere of the radius $1/M$ and the limiting procedure $M \to \infty$ would then correspond to constructing a limit of fields generated by spatially extended charges, shrinking to a point. Although this approach was basically sound, the answer they found in QED was *negative* and thus discouraging.

To see where the problem comes from in QED and why the situation is so *different* in QCD let us first improve a little bit the approximation we have so far worked in. Instead of a single loop in Fig. 8.1b consider the sum of terms corresponding to the subset of all Feynman diagrams shown in Fig. 8.2a. Because the

---

[10]At higher orders further free parameters do appear as well, as discussed in subsection 7.4.1.
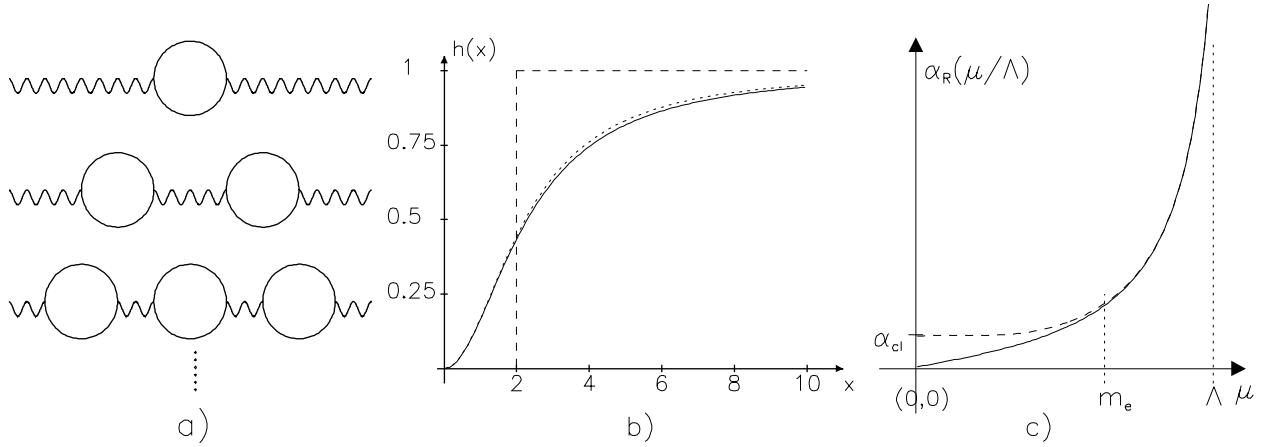
Figure 8.2: a): Subset of Feynman diagrams which gives rise to a geometric series (8.28); b): the shape of the function $h(x)$ together with its "step" approximation (dashed line) and the analytical approximation $x^2/(5+x^2)$ (dotted curve); c): the sketch of the dependence of $\alpha_R(\mu)$ on μ for both A and B definitions.

bubbles are connected via single photon lines and each separately yields exactly the same result as the bubble in Fig. 8.1b, summing the mentioned subset of Feynman diagrams leads to a *geometric series*, the leading term of which is just (8.27). This series is easily summed with the result

$$\alpha_R(\mu) = \alpha_B(M) \left[1 + \alpha_B(M) \left(\beta_0 \ln \frac{\mu}{M} + \delta(x) \right) + \cdots \right] = \frac{\alpha_B(M)}{1 - \alpha_B(M) \left(\beta_0 \ln(\mu/M) + \delta(x) \right)}, \qquad (8.28)$$

where $\beta_0 = 2/3\pi$ and $\delta(x)$ is given in the above defined cases A,B as

$$\delta(A, x) = 0, \qquad (8.29)$$

$$\delta(B, x) = \frac{2}{\pi} \int_0^1 dz \, z (1 - z) \ln \left(\frac{1}{x^2} + z(1-z) \right), \quad x \equiv \frac{\mu}{m_B}. \qquad (8.30)$$

Note that for finite μ and in the limit $m_B \to 0$, i.e. for $x \to \infty$, $\delta(B, x) \to -5/3$ and thus

$$e_R^2(B, \mu) \equiv e_B^2 Z_3^{-1}(B) = e_B^2 \left(1 + \beta_0 \alpha_B \left(\ln \frac{\mu^2}{M^2} - \frac{5}{3} \right) \right) \qquad (8.31)$$

has a finite, μ–independent limit for vanishing m_B, as already noted in (8.23). This reflects the fact that massless fermions contribute to the renormalized $e_R^2(\mu)$ at all scales μ with the same strength. On the other hand for fixed μ and in limit $m_B \to \infty$ (8.30) diverges like $\ln m_B/\mu$. This looks strange and unphysical as one would (correctly, indeed) expect that infinitely heavy electrons should not have any influence on the theory at finite energy scales μ. The remedy to this problem is essentially the same as the remedy to the ultraviolet divergence of each of the terms in (8.28) above and relies on its **resummation**. As mentioned already above, by resumming all the bubbles in Fig. 8.2a the resulting renormalized $\alpha_R(B, \mu, m_B)$ has the property we wish it to have, i.e. a heavy electron *decouples* from it in the limit $m_B \to \infty$.

A closely related feature of (8.28) concerns its behavior as $\mu \to 0$. While for the mass–independent definition of the renormalized couplant (8.28) implies $\alpha_R(A, \mu) \to 0$ [11], for the mass–dependent definition the situation is different. Keeping the UV cut–off parameter M fixed we find

$$\beta_0 \ln \frac{\mu}{M} + \delta(B, x) \to \beta_0 \ln \frac{m_B}{M}, \qquad (8.32)$$

which implies that for $m_B \neq 0$ the infrared limit of $\alpha_R(B, \mu)$ is determined by the *mass* m_B and coincides with that of massless electron at the scale $\mu = m_B$! This is a manifestation of the fact that in QED (and

---

[11]Note how the incorporation of the whole chain of fermion bubbles has dramatically changed the character of the small μ behaviour of $\alpha_R(A, \mu)$, obtained with only one bubble: instead of $-\infty$ from (8.27), we get *zero* from (8.28)!

in general in theories without confinement of elementary fermions) the IR limit of the gauge coupling is determined essentially by the fermion masses.

Inverting (8.28) we can for both definitions express the bare couplant $\alpha(M)$ in terms of the renormalized one

$$\alpha_B(M) = \frac{\alpha_R(\mu)}{1 - \alpha_R(\mu)\left(\beta_0 \ln(M/\mu) + \delta(x)\right)}. \tag{8.33}$$

The problem with (8.33) is that for any *finite* renormalized couplant α_R the bare couplant $\alpha_B(M)$ blows up to infinity at some *finite* value of M, corresponding to the pole of this expression. In the fifties the appearance of this so called **Landau singularity** had marred all attempts of Landau and his school to formulate QED in a mathematically consistent way. This, in turn had led to a temporary loss of interest in the lagrangian formalism of the local quantum field theory. An interesting recollection of these attempts from one of protagonists can be found in [119]. A ghost of this singularity appears in the theory even if we forget about the bare couplant and work with $\alpha_R(\mu)$ only. Eq. (8.28) implies that for *any* dependence of the bare couplant $\alpha_B(M)$ on M which has a *finite* limit as $M \to \infty$ the renormalized couplant $\alpha_R(\mu)$ *vanishes*!

A careful reader must have noted one inconsistency in my writing $\alpha_B(M)$ and $\alpha_R(\mu)$. In both cases a dimensionless quantity is expressed as a function of a dimensional one, which is possible only if the latter enters in the ratio with some dimensional parameter, which will be denoted Λ. [12] This parameter has nothing to do with the cut-off and represents a *fundamental scale parameter*, which appears in the theory entirely due to the renormalization procedure. This phenomenon is called **dimensional transmutation** and is typical for theories with dimensionless couplants. We should thus write more correctly

$$\alpha_R = \alpha_R(\mu/\Lambda), \quad \alpha_B = \alpha_B(M/\Lambda). \tag{8.34}$$

Although the appearance of the fundamental parameter Λ is an inevitable consequence of the renormalization procedure, its *numerical* value is not fixed by these considerations and must be determined from comparison with experimental data. The standard way how to define Λ *unambiguously* is to write down and solve the differential equations which follow from (8.28) and (8.33). For the definition A we have simply

$$\frac{\mathrm{d}\alpha_R(A, \mu/\Lambda)}{\mathrm{d}\ln\mu} = \beta_0 \alpha_R^2(A, \mu/\Lambda) \Rightarrow \alpha_R(A, \mu/\Lambda) = \frac{1}{\beta_0 \ln(\Lambda/\mu)}, \tag{8.35}$$

$$\frac{\mathrm{d}\alpha_B(M/\Lambda)}{\mathrm{d}\ln M} = \beta_0 \alpha_B^2(M/\Lambda) \Rightarrow \alpha_B(M/\Lambda) = \frac{1}{\beta_0 \ln(\Lambda/M)}, \tag{8.36}$$

while for B the equation for $\alpha_R(B,\mu)$ is more complicated

$$\frac{\mathrm{d}\alpha_R(B, \mu/\Lambda)}{\mathrm{d}\ln\mu} = \left(\beta_0 + \frac{\mathrm{d}\delta(B,x)}{\mathrm{d}\ln\mu}\right)\alpha_R^2 = \beta_0 \alpha_R^2(B, \mu/\Lambda) \underbrace{\int_0^1 \mathrm{d}z \frac{6x^2 z^2(1-z)^2}{1 + x^2 z(1-z)}}_{h(x)}. \tag{8.37}$$

The explicit evaluation of the integral defining $h(x)$ yields

$$h(x) = 1 - \frac{6}{x^2} + \frac{12}{x^3\sqrt{4+x^2}}\ln\frac{\sqrt{4+x^2}+x}{\sqrt{4+x^2}-x} \doteq \frac{x^2}{5+x^2}, \tag{8.38}$$

where the last, approximate, equality is a very accurate approximation to the exact form of $h(x)$ in the whole interval $x \in (0,\infty)$, as demonstrated in Fig. 8.2b. We see that $h(x)$ approaches unity for $x \gg 1$ but vanishes like x^2 as $x \to 0$ [13]. Consequently the solutions of (8.37) are essentially the same in A and B in the first region, but differ substantially in the second, where $\alpha_R(A, \mu/\Lambda) \to 0$ while $\alpha_R(B, \mu/\Lambda)$ flattens to a constant value, as shown schematically by the dashed line in Fig. 8.2c.

In both cases the scale parameter Λ labels the elements of an infinite set of solutions of these equations. Note that after taking the derivative of (8.28) with respect to $\ln\mu$ the cut–off M has completely disappeared

---

[12]This is true even in the case of $e_R^2(B,\mu)$, where one might attempt to use m_B for this purpose. However, at large μ this mass cannot play any role and one therefore needs other dimensional quantity, as in the case A.

[13]The function and its so called "step" approximations, will be exploited extensively in section 8.4.2 when discussing the influence of quark mass thresholds on the running QCD couplant.

from the resulting expression (8.35) and (8.37)! Taking (8.35) or (8.37) as **definitions** of the renormalized couplant $\alpha_R(\mu/\Lambda)$ is the *crucial* step of the renormalization procedure.

As the renormalized couplant α_R is supposed to be a positive definite number it makes sense only for $\mu < \Lambda$ and should be trusted only for $\mu \ll \Lambda$. So even if we start from (8.35) we encounter severe problems when we approach short distances where the renormalized couplant diverges. It does so just at $\mu = \Lambda$ and we can, knowing α_R at some μ, evaluate Λ from the any of the expressions

$$\Lambda = \mu \exp\left(\frac{1}{\beta_0 \alpha_R(A, \mu/\Lambda))}\right) = M \exp\left(\frac{1}{\beta_0 \alpha_B(M/\Lambda))}\right). \tag{8.39}$$

For the case B the same relations hold for $x \gg 1$, while for $x \ll 1$ (8.39) are replaced by more complicated formulae. As indicated above we can alternatively determine Λ from the knowledge of the bare couplant α at some cut–off M. So far the dependence of the renormalized couplant $\alpha_R(\mu/\Lambda)$ on the scale μ may appear essentially as a consequence of mathematical considerations. However, it has also a very simple and intuitive physical interpretation. Roughly speaking the renormalized charge $e_R^2(\mu/\Lambda)$ can be interpreted as the charge contained inside the sphere of the radius $r = 1/\mu$ around the "core" of an electron. In other words the renormalized charge can be viewed as the **effective charge** at a distance defined by the scale μ. This point of view forms the basis for the modern formulation of the renormalization theory.

I close this recollection with a remark concerning the large distance behavior of QED. For the process under study this corresponds to the limit $q^2 \to 0$. As argued above the use of $\alpha_R(B, \mu)$ as an expansion parameter is appropriate even in this region as it gives a finite value for the contribution of the sum of diagrams in Fig. 8.1a,b, equal simply to $e_R^2(B, 0)$. Including the (finite) contribution of the sum of other three diagrams in this figure leads to the final result proportional to $e_R^2(B, 0)(1 + r_1 e_R^2(B, 0))$, where r_1 is some number. Recalling the way $e_R(B, \mu)$ was introduced, it is obvious that we could add to the quantity B and simultaneously subtract from $C(B, q^2, \mu, m_B)$ in (8.21) such finite term that the resulting r_1 actually *vanishes*! Denoting the corresponding renormalized charge $e_R(C, \mu)$, the $q^2 \to 0$ limit of the electron–muon scattering amplitude, evaluated using this definition of renormalized electric charge, would be proportional to $e_R^2(C, 0)$, with *no higher order corrections*. Repeating this procedure for the case of the Compton scattering on an electron leads to another renormalized charge at $\mu = 0$, which I denote $e_R^2(D, 0)$. According to the conventional definition of electric charge, this latter quantity coincides with the **fine structure constant** $\alpha_{cl} \equiv (e_R^2(D, 0))^2/4\pi = \frac{1}{137}$.

8.2.2 *Photon wave function renormalization

In the preceding subsection we saw how to get rid of the UV infinities associated with the basic electron loop in Fig. 8.1b by absorbing them in the renormalized electron charge. In that discussion the photon was virtual and the resulting effects were shared equally by both vertices on each side of the loop. The same loop can, however, appear also on any external photon line, representing the incoming or outgoing *real* photon. What will change in this case? Not very much! Imagine that a real photon is radiated, as depicted in in Fig. 8.1f. Note that the apparent singularity (for the real photon $Q^2 = 0$) of the propagator $1/Q^2$, connecting the upper vertex with the electron bubble is canceled by Q^2 coming from the bubble itself. Compared to the case of the virtual photon we merely have

- to replace in (8.14) the second photon propagator, multiplying the function $\overline{I}(Q^2, m_B, M)$, by the photon polarization vector ϵ_μ,

- set $Q^2 = 0$ everywhere. For finite electron mass m_B this can be done safely in (8.14) as well as in the functions $C(A, Q^2, \mu, m_B)$ in (8.17) or $C(B, Q^2, \mu, m_B)$ in (8.21), describing finite $O(\alpha)$ corrections. Moreover, by a suitable choice of the renormalization scale μ ($\mu = m_B$ for $C(A, Q^2 = 0, \mu, m_B)$ and $\mu = 0$ for $C(B, Q^2 = 0, \mu, m_B)$ both these finite corrections can be made to vanish.

There is nevertheless one essential difference from the case of the virtual photon, where each of the vertices connected with the exchanged photon in Fig. 8.1b absorbed one factor $\sqrt{Z_3}$. In Fig. 8.1f there is, on the other hand, no second vertex where to absorb one of these $\sqrt{Z_3}$. This might suggest to incorporate the whole renormalization factor Z_3 in the lowest order matrix element. But if we imagine that the emitted electron will eventually interact at some distant point, it is clearly more consistent to use only "half" of the bubble contribution, i.e. $\sqrt{Z_3}$, to renormalize the upper vertex in Fig. 8.1f, exactly as in the case of the

virtual photon, leaving the second $\sqrt{Z_3}$ to be included in the description of the interaction of the emitted (and in general external) photon. This procedure is clearly equivalent to including the whole contribution of the bubble in Fig. 8.1f and dividing the result by $\sqrt{Z_3}$. Formally the same effect is achieved by introducing the **renormalized photon field** $A_\nu(x, \mu)$

$$A_R^\nu(x, \mu) \equiv Z_3^{-1/2}(M, \mu) A_B^\nu(x, M) \tag{8.40}$$

and considering the matrix elements of this renormalized operator. Note that for this procedure to be mathematically well–defined, the bare photon field operator $A_B^\nu(x, M)$ must, similarly as the bare electron charge $e_B(M)$, depend on the UV cut–off parameter M.

8.2.3 *Renormalization of the electron mass

In this subsection I shall introduce and briefly discuss the concept of the renormalized, i.e. "running" electron mass. I shall work within the same regularization technique as in the preceding subsection, but will skip all technical details, which can again be found in [2].

In the case of the electron mass the emergence of infinities in the evaluation of the loop in the diagram of Fig. 8.1c should not surprise at all, as similar problem does arise already at the classical level! Just try to answer the following simple question: what is the total energy (mass) associated with the static electric field of a pointlike electron? The answer is: infinity! The reason for that is trivial, as the density of the electrostatic energy of a classical pointlike electric charge is proportional to $\vec{E}^2(r)$, and $\vec{E}(r)$, the corresponding electric field strength at the distance r from that pointlike electron, is in turn proportional to $1/r^2$. The integral over the energy density therefore *diverges linearly at short distances*:

$$\int_{r_0}^\infty \mathrm{d}\vec{r}\,\vec{E}^2(\vec{r}) \propto \int \frac{r^2 \mathrm{d}r \mathrm{d}\Omega}{r^4} \propto \frac{1}{r_0}, \tag{8.41}$$

where I have introduced the lower integration cut-off r_0 to regularize the integral. Note that there is, on the other hand, no problem with large distances, where the integral (8.41) converges. Thus already at the classical level the full mass m_{full} of a pointlike electric charge is infinite and requires therefore some kind of redefinition. However, forgetting for the moment about the divergence of m_{full} at short distances, it is obvious that even in classical electrodynamics it is quite useful to introduce the concept of the "effective" electron mass $m_{\text{eff}}(r)$, as the mass contained in the sphere of the radius r around the core of an electron. While the full mass corresponds to the limit $r \to \infty$ and keeps its basic meaning, it is the effective mass $m_{\text{eff}}(r)$ which plays, even in classical electrodynamics, the decisive role when we look deep into the field surrounding the pointlike charge. As far as the electron mass is concerned, quantum electrodynamics does not therefore introduce any new concept but merely changes the form of the dependence of $m_{\text{eff}}(r)$ on r. Instead of the powerlike dependence, $m_{\text{eff}}(r, r_0) \propto (1/r_0 - 1/r)$, characterizing the classical results (8.41), quantum electrodynamics leads, as will be shown below, to a weaker *logarithmic* dependence of the effective mass on r.

In QED the basic Feynman diagram giving the electron self-energy is the loop diagram of Fig. 8.1c. Denoting the corresponding contribution as $-\mathrm{i}\Sigma(p)$, where p is the momentum of the electron, we find

$$-\mathrm{i}\Sigma(p) \equiv -e_B^2 \int \frac{\mathrm{d}^4 k}{(2\pi)^4} \frac{(-\mathrm{i})}{k^2 - \lambda^2 + \mathrm{i}\epsilon} \gamma_\nu \frac{(\mathrm{i})}{\not{p} - \not{k} - m_B + \mathrm{i}\epsilon} \gamma^\nu \tag{8.42}$$

and the analysis similar to that of the preceding subsection gives

$$\Sigma(p, M, M_B, \lambda) \equiv \frac{\alpha_B}{2\pi} \int_0^1 \mathrm{d}z [m_B(z+1) + (\not{p} - m_B)(z-1)] \ln\left(\frac{M^2 z + \lambda^2(1-z) - p^2 z(1-z)}{m_B^2 z + \lambda^2(1-z) - p^2 z(1-z)}\right), \tag{8.43}$$

where m_B is the "bare" electron mass, M is again the <u>ultraviolet</u> cut-off and λ is the <u>infrared</u> regulator, introduced in order to avoid infrared divergence in the limit of an on-shell electron, i.e. for $\not{p} \to m_B$.

The matrix $\Sigma(p, M)$ can be written as a sum of the two terms

$$\Sigma(p, M, m_B, \lambda) = m_B \Sigma_1(p, M, m_B, \lambda) + (\not{p} - m_B)\Sigma_2(p, M, m_B, \lambda), \tag{8.44}$$

where

$$\Sigma_1(p, M, M_B, \lambda) \equiv \frac{\alpha_B}{2\pi} \int_0^1 dz(z+1) \ln\left(\frac{M^2 z + \lambda^2(1-z) - p^2 z(1-z)}{m_B^2 z + \lambda^2(1-z) - p^2 z(1-z)}\right), \tag{8.45}$$

$$\Sigma_2(p, M, m_B, \lambda) \equiv \frac{\alpha_B}{2\pi} \int_0^1 dz(z-1) \ln\left(\frac{M^2 z + \lambda^2(1-z) - p^2 z(1-z)}{m_B^2 z + \lambda^2(1-z) - p^2 z(1-z)}\right). \tag{8.46}$$

As we shall see, Σ_1 renormalizes the bare electron mass, yielding the **renormalized electron mass**, m_R, while Σ_2 renormalizes the bare electron field ψ_B. To see this in little bit more detail, let us first sum the whole set of diagrams corresponding to any number of photon loops on the basic electron leg, similarly as we summed Fig. 8.2a the set of diagrams with strings of electron bubbles. The series is again geometric [14]

$$\frac{i}{\not{p} - m_B}\left[1 + (-i\Sigma(p, M))\frac{i}{\not{p} - m_B} + (-i\Sigma(p, M))\frac{i}{\not{p} - m_B}(-i\Sigma(p, M))\frac{i}{\not{p} - m_B} + \cdots\right] \tag{8.47}$$

and sums to the full bare fermion propagator

$$S_B(p, M) \equiv \frac{i}{\not{p} - m_B - \Sigma(p, M)} = \frac{1}{[1 - \Sigma_2(p, M)]}\frac{i}{[\not{p} - m_B(1 + \Sigma_1(p, M)/(1 - \Sigma_2(p, M)))]}. \tag{8.48}$$

In perturbation theory each of the functions Σ_1, Σ_2 is given as an expansion in the bare couplant α_B, starting at the order α_B. In the lowest order the bare couplant α_B can, as in the preceding subsection, be everywhere replaced with the renormalized one, α_R. Moreover, in the lowest order we can drop Σ_2 in the second part of the denominator in (8.48) and write

$$S_B(p, M) = \frac{i[1 - \Sigma_2(p, M)]^{-1}}{[\not{p} - m_B(1 + \Sigma_1(p, M))]}. \tag{8.49}$$

In the lowest order the effects of the loop in Fig. 8.1c are thus twofold:

- bare electron mass is multiplied by a divergent integral contained in Σ_1,

- there is an overall UV divergent factor $(1 - \Sigma_2)^{-1}$. [15]

To get rid of these infinities we split (8.43), in analogy to (8.17), into two parts

$$\Sigma(p, M) = \frac{\alpha_R}{2\pi} \int_0^1 dz[m_B(z+1) + (\not{p} - m_B)(z-1)] \ln\frac{M^2 z + \lambda^2(1-z) - p^2 z(1-z)}{\mu^2} +$$
$$\frac{\alpha_R}{2\pi} \int_0^1 dz[m_B(z+1) + (\not{p} - m_B)(z-1)] \ln\frac{\mu^2}{m_B^2 z + \lambda^2(1-z) - p^2 z(1-z)}, \tag{8.50}$$

where the scale parameter μ has been again introduced to describes the freedom in the separation of $\Sigma(p, M)$ into a divergent and a finite parts. Note that the UV cut-off parameter M appears in the first integral of (8.50) only, while the second is finite. For $M \to \infty$ only $M^2 z$ in the numerator of the logarithm of the first term in (8.50) contributes. After integration it can be rewritten as

$$m_B \underbrace{\frac{\alpha_R}{2\pi}\left[\frac{3}{2}\ln\frac{M^2}{\mu^2} - \frac{5}{4}\right]}_{\equiv \tilde{\Sigma}_1(M/\mu)} - (\not{p} - m_B)\underbrace{\frac{\alpha_R}{2\pi}\left[\frac{1}{2}\ln\frac{M^2}{\mu^2} - \frac{3}{4}\right]}_{\equiv -\tilde{\Sigma}_2(M/\mu)}. \tag{8.51}$$

Note that the tilded functions $\tilde{\Sigma}_i$, introduced in (8.51), depend beside the cut-off M (and m_B, λ) also on the scale μ, while the original Σ_i are, on the contrary, functions of M and p (and m_B, λ). Let me emphasize that the finite terms in the definition of the functions $\tilde{\Sigma}_i$ are essentially *arbitrary*, as were analogous finite terms in the definition of the renormalized electric charge.

---

[14]In the following the dependence of Σ_i on m_B and λ will be suppressed.

[15]In all manipulations with perturbative expressions like $1 - \Sigma_2(p, M, m_B, \lambda) = 1 + O(\alpha_B) \ln M^2$ the UV cut-off M is held fixed and standard rules for operations with power series are used, despite the fact that M, and thus the whole $O(\alpha_B)$ term, diverges for $M \to \infty$. For instance if $\Sigma_2 \to +\infty$ as $M \to \infty$, $1/(1 + \Sigma_2)$ does not vanish, but goes to $-\infty$! This rule reflects the fact that the renormalization procedure is carried out order by order.

In order to get rid of the divergent logarithms $\ln(M/\mu)$ we first replace the bare electron mass m_B with the renormalized, scale dependent mass $m_R(\mu)$, defined as

$$m_R(\mu) \equiv \left(1 + \tilde{\Sigma}_1(M/\mu)\right) m_B(M) \equiv Z_m^{-1}(M/\mu) m_B(M), \tag{8.52}$$

where the so-called **electron mass renormalization factor** Z_m has been introduced in analogy to the photon renormalization factor Z_3 in (8.31). The dependence of the bare mass $m_B(M)$ on the cut–off M is explicitly written out as in the case of the bare electron charge. The renormalized electron mass thus absorbs the infinity in the denominator of (8.49).

As far as the infinity hidden in Σ_2 is concerned, the renormalization of the bare electron mass does not help and the only way out is to multiply (8.48) by some expression that cancels the divergence of the term $1/(1 - \Sigma_2(p,M))$. This is achieved by introducing the **electron wave function renormalization factor**

$$Z_{2F}(\mu, M) \equiv \frac{1}{1 - \tilde{\Sigma}_2(\mu, M)}, \tag{8.53}$$

in terms of which the bare electron propagator S_B is expressed as [16]

$$S_B(p, M) = \frac{\mathrm{i}\left[Z_{2F}(\mu, M) + \Sigma_2(p, M) - \tilde{\Sigma}_2(\mu, M)\right]}{\left[\not{p} - m_R(\mu)\left(1 + \Sigma_1(p, M) - \tilde{\Sigma}_1(\mu, M)\right)\right]} \tag{8.54}$$

and dividing S_B by the divergent factor Z_{2F}. In this way one can define the so–called **renormalized electron propagator**.

$$S_R(p, \mu) \equiv Z_{2F}^{-1}(\mu, M) S_B(p, M), \tag{8.55}$$

which is *finite* and which can be rewritten as

$$S_R(p, \mu) = \frac{\mathrm{i}\left[1 + \Sigma_2(p, M) - \tilde{\Sigma}_2(\mu, M)\right]}{\left[\not{p} - m_R(\mu)\left(1 + \Sigma_1(p, M) - \tilde{\Sigma}_1(\mu, M)\right)\right]}. \tag{8.56}$$

Note that the UV divergences cancel, as was our design, in both differences $\Sigma_i(p, M) - \tilde{\Sigma}_i(\mu, M)$, which represent finite $O(\alpha_B)$ corrections to the renormalized electron propagator.

The multiplication in (8.55) can be interpreted as a result of the replacement (renormalization) of the bare electron field wave function $\psi_B(x)$ (which has so far been denoted simply as ψ) according to the prescription

$$\psi_R(x, \mu) \equiv Z_{2F}^{-1/2}(M, \mu) \psi_B(x, M) \tag{8.57}$$

in much the same way as we did renormalize the photon wave function in the preceding subsection. In practice it means that after evaluation of the contribution of the the bubble in Fig. 8.1c, the result must be divided by $\sqrt{Z_{2F}}$ for each external electron. One must be careful to take into account that the loop in Fig. 8.1d can be on both incoming and outgoing electron lines. In the above equation I have again written out explicitly the dependence of the renormalized and bare wave functions on the renormalization scale μ and the cut–off M. [17]

Having defined in (8.52) the renormalized electron mass, we can now play with it in the same way as we did with the analogous expression (8.25) for the renormalized electric charge, i.e. take its derivative with respect to μ and see what results:

$$\frac{\mathrm{d}m_R(\mu)}{\mathrm{d}\ln\mu} = -m_B(M)\frac{3}{2\pi}\alpha_R(\mu) = -m_R(\mu)\frac{3}{2\pi}\alpha_R(\mu) \quad \Rightarrow \quad \frac{1}{m_R(\mu)}\frac{\mathrm{d}m_R(\mu)}{\mathrm{d}\ln\mu} = -\frac{3}{2\pi}\alpha_R(\mu), \tag{8.58}$$

where again the replacement of m_B with m_R in the second equality is legal in the lowest order and follows from (8.52). Note that as in the case of an analogous relation (8.35) the dependence on the UV cut-off M has now completely disappeared. Carrying out the renormalization program to all orders leads to a power series

---

[16] The dependences of Z_{2F}, S_B and S_R and on m_B and λ will be suppressed in the following.

[17] In principle ψ_R depends also on other parameters, analogous to the higher order β–function coefficients c_i, introduced in subsection 8.4.1.

in $\alpha_R(\mu)$ on the r.h.s. of (8.58), but the main feature of (8.58), i.e. finite expansion coefficients, persists. The above equation has analogous physical interpretation as eq. (8.35) and shares with it also the property that beyond the leading orders (the first one in (8.58)), all higher order coefficients on the r.h.s. of the second eq, of (8.58) are *ambiguous*.

In (8.58) I have also for the first time in this subsection written m_B explicitly as a function of the cut-off M. As in the case of the bare couplant, this is unavoidable if the renormalization procedure is to have a clear physical interpretation. Recalling the explicit form of $\alpha_R(\mu)$ (8.35) it is then straightforward to solve (8.58) to the leading order:

$$m_R(\mu) = \overline{m} \left(\alpha_R(\mu)\right)^{-3/(2\pi\beta_0)}, \tag{8.59}$$

where $\overline{m}$ is a free parameter with the dimension of mass, which, together with Λ in $\alpha_R(\mu/\Lambda)$, specify the renormalized mass $m_R(\mu, \Lambda, \overline{m})$. At short distances, but still away from the Landau singularity (i.e., for $r > 1/\Lambda_{\mathrm{QED}}$), $m_R \propto (\ln(r\Lambda))^{9/4}$, i.e. the renormalized mass behaves as mentioned at the beginning of this subsection. At large distances, i.e. small μ, the behavior of $m_R(\mu)$ depends crucially on the choice of the renormalization procedure for the electric charge. For the physically motivated MOM–like definition "B" of $\alpha_R(B,\mu)$, $m_R(\mu)$ goes to a <u>finite</u> value $\overline{m}(137)^{9/4}$. The parameter $\overline{m}$ can thus be expressed in terms of the total classical electron mass m_e simply as $\overline{m} = m_e(1/137)^{9/4}$.

8.2.4 *Vertex renormalization

The loop in Fig. 8.1d differs from that of Fig. 8.1c merely by the fact that the virtual photon, emitted in both diagrams at point A, is reabsorbed in Fig. 8.1d at point B after the electron absorbed the exchanged photon, while in Fig. 8.1c before. The **vertex function** associated with the upper vertex in Fig. 8.1d

$$\Lambda_\mu(p',p) \equiv (\mathrm{i}e)^2 \int \frac{\mathrm{d}^4k}{(2\pi)^4} \frac{(-\mathrm{i})}{k^2 - \lambda^2 + \mathrm{i}\epsilon} \gamma_\nu \frac{\mathrm{i}}{\not{p}' - \not{k} - m_B + \mathrm{i}\epsilon} \gamma_\mu \frac{\mathrm{i}}{\not{p} - \not{k} - m_B + \mathrm{i}\epsilon} \gamma_\nu \tag{8.60}$$

is both UV and IR divergent and we have introduced the fictitious photon mass parameter λ to regulate the later. $\Lambda_\mu(p',p)$ represents the $O(\alpha)$ correction to the basic e–γ–e vertex γ_μ and can be written as

$$\Lambda_\mu(p',p) = W(m_B, \lambda, M)\gamma_\mu + \Lambda_\mu^c(p',p), \tag{8.61}$$

where the function $W(m_B, \lambda, M)$ contains the UV divergence, while Λ_ν^c represent the UV finite correction term, conveniently normalized in such a way that $\Lambda_\mu^c(p,p) = 0$. Using the identity

$$\frac{\partial}{\partial p^\mu} \frac{1}{\not{p} - m} = -\frac{1}{\not{p} - m} \gamma_\mu \frac{1}{\not{p} - m}, \tag{8.62}$$

which is a straightforward consequence of the algebra of Dirac matrices, we find

$$\Lambda_\mu(p,p) = -\frac{\partial \Sigma(p)}{\partial p^\mu}. \tag{8.63}$$

The explicit evaluation of the derivative in (8.63) gives

$$\frac{\partial \Sigma(p,M)}{\partial p^\mu} = \gamma_\mu \Sigma_2(p,M) + m_B \frac{\partial \Sigma_1(p,M)}{\partial p^\mu} + (\not{p} - m_B) \frac{\partial \Sigma_2(p,M)}{\partial p^\mu}. \tag{8.64}$$

The last two terms cancel when sandwiched between the external bispinors $\overline{u}(p'), u(p)$ and the second term, containing the derivative $\partial \Sigma_1 / \partial p^\mu$, is finite. We thus have

$$\overline{u}(p)\Lambda_\mu(p,p)u(p) = W(m_B, \lambda, M)\overline{u}(p)\gamma_\mu u(p) = -\overline{u}(p)\gamma_\mu u(p)\Sigma_2(p) - m_B \frac{\partial \Sigma_1(p)}{\partial p^\mu} \overline{u}(p)u(p)$$

$$= \left(Z_{2F}^{-1}(\mu, M) - 1 + \tilde{\Sigma}_2(\mu, M) - \Sigma_2(p, M)\right) \overline{u}(p)\gamma_\mu u(p) - m_B \frac{\partial \Sigma_1(p,M)}{\partial p^\mu} \overline{u}(p)u(p). \tag{8.65}$$

After simple manipulations and using the Gordon decomposition we get

$$W(m_B, \lambda, M) = \left(Z_{2F}^{-1}(\mu, M) - 1\right) + \underbrace{\tilde{\Sigma}_2(\mu, M) - \Sigma_2(p, M) - 2m_B^2 \frac{\partial \Sigma_1(p,M)}{\partial p^2}}_{\text{finite part}}, \tag{8.66}$$

where the UV divergent part resides exclusively in the first term $Z_{2F}^{-1} - 1$. The UV divergent part of the vertex correction function $\Lambda_\mu(p', p)$ is thus uniquely related to the electron wave function renormalization. Because the loop in Fig. 8.1d circumnavigates the basic QED e–γ–e vertex it would be natural to include this vertex correction, beside the vacuum polarization correction, in the definition of the renormalized electron charge. However, this is not necessary as the relation (8.66) guarantees that the UV divergent part of the vertex correction cancels the UV divergent part of the external electron wave function renormalization factor. This will be shown in more detail in the next subsection.

Note, moreover, that the UV divergence coming from the loop in Fig. 8.1d does not require introduction of yet another renormalization factor and is isolated in the same $Z_{2F}(\mu, M)$, introduced in connection with the infinities of electron selfenergy diagram in Fig. 8.1c. This remarkable and crucial fact is a consequence of (8.62). However, we could proceed with evaluation of Λ_μ disregarding this identity and introduce at the beginning of the previous subsection another renormalization factor $Z_1(\mu, m)$ which would then absorb all UV divergencies of Fig. 8.1c

$$\Lambda_\mu(p', p) = \left(Z_1^{-1} - 1\right)\gamma_\mu + \text{finite parts.} \tag{8.67}$$

The identity (8.66) would then imply that singular parts of Z_1 and Z_2 <u>must</u> be the same and as we <u>can</u> assume the same even for finite ones, we would conclude $Z_1(\mu, M) = Z_2(\mu, M)$.

At this stage I interrupt the description of the renormalization procedure within the old Pauli–Villars type of regularization and go on to introduce the elements of the dimensional regularization. I do this before explaining the technique of **counterterms**, as this systematic method of carrying out the renormalization program at arbitrary order is difficult to formulate in a consistent manner within the Pauli–Villars regularization.

8.3   Elements of dimensional regularization

Dimensional regularization (DR) has been invented in 1972 independently in two papers [117, 118] as a way of surmounting serious problems of the conventional regularization methods, in particular in gauge theories. It has rapidly gained prominence, due to its two properties: *technical simplicity*, crucial in complicated multiloop calculations, and explicit *gauge invariance*. In the present time it is practically the only method used in all sorts of perturbative calculations. Its calculational advantages outweigh the main drawback of the renormalization procedure based on this method of regularization, namely the lack of clear physical interpretation of the ensuing dependence of the renormalized charge (electric in QED and colour in QCD) on the renormalization scale. As, however, the mathematical structure of the resulting perturbative expansions in powers of the renormalized charge is independent of the regularization method employed, we can use the DR for calculations while retaining the interpretation on the renormalization scale that follows from the conventional regularizations in four–dimensional space–time. I shall first outline the basic idea of the method on a simple example and then present elements of the mathematical formalism of DR.

8.3.1   A simple example

Let us consider first the case of a theory with a single scalar field $\phi(x)$ with selfinteraction term $\lambda\phi^4$ in four–dimensional Minkowski space–time. The corresponding lagrangian reads

$$\mathcal{L}_\phi \equiv \frac{1}{2}\left(\partial_\mu\phi(x)\right)^2 - m^2\phi^2 - \frac{\lambda}{4!}\phi^4(x), \ \ \lambda > 0. \tag{8.68}$$

The simplest one–loop diagram, displayed in Fig. 8.3b, describes order λ^2 correction to the leading order diagram for the basic binary scattering process, depicted in Fig. 8.3a. Recalling that in λ^4 theory the basic four–particle vertex is associated with the factor $-i\lambda$ we find for the amplitude corresponding to digram in Fig. 8.3b

$$I(p, m) \equiv \frac{\lambda^2}{2} \int \frac{\mathrm{d}^4 k}{(2\pi)^4} \frac{1}{[k^2 - m^2 + i\epsilon]} \frac{1}{[(k-p)^2 - m^2 + i\epsilon]}, \tag{8.69}$$

where $p \equiv p_1 + p_2$ is the total momentum of incoming particles. The basic idea of DR is based on the observation that while in four dimensions the integral in (8.69) is logarithmically divergent, it converges in

$n = 1, 2, 3$ dimensions, due to the fact that in n dimensions $\mathrm{d}^n k \propto k^{n-1}$. In $n = 1, 2, 3$ dimensions and in rest frame of $p = (M, 0, 0, 0)$ (8.69) equals

$$I(p, n) = \frac{\lambda^2}{2(2\pi)^4} \int \mathrm{d}k_0 \int \mathrm{d}^{n-1}k \frac{1}{\left[k_0^2 - \vec{k}^2 - m^2 + \mathrm{i}\epsilon\right]} \frac{1}{\left[(k_0 - M)^2 - \vec{k}^2 - m^2 + \mathrm{i}\epsilon\right]}. \qquad (8.70)$$

The integrand depends merely on the magnitude of the $(n-1)$–dimensional vector $\vec{k}$ and the corresponding integral over $\mathrm{d}^{n-1}k$ can therefore be conveniently performed introducing the polar coordinates in n dimensions ($r \equiv \sqrt{\vec{k}^2}$)

$$\mathrm{d}^n k = r^{n-1} \mathrm{d}r \sin^{n-2}\theta_{n-1} \sin^{n-3}\theta_{n-2} \cdots \sin\theta_2 \mathrm{d}\theta_2 \cdots \mathrm{d}\theta_1, \qquad (8.71)$$

where $0 \leq \theta_1 \leq 2\pi$ is the "azimuthal" angle, while all the other angles $\theta_i \in (0, \pi)$ are generalizations of the familiar polar angles in three space dimensions. Applying (8.71) to (8.70) and denoting $\omega \equiv \sqrt{\vec{k}^2}$ we get

$$I(p, n) = \frac{\lambda^2}{2(2\pi)^4} \int \mathrm{d}k_0 \int_0^\infty \mathrm{d}\omega \omega^{n-2} \frac{1}{[k_0^2 - \omega^2 - m^2 + \mathrm{i}\epsilon]} \frac{1}{[(k_0 - M)^2 - \omega^2 - m^2 + \mathrm{i}\epsilon]} W(n), \qquad (8.72)$$

where $W(n)$ contains the integral over the angular coordinates

$$W(n) \equiv \int_0^{2\pi} \mathrm{d}\theta_1 \int_0^\pi \mathrm{d}\theta_2 \sin\theta_2 \int_0^\pi \mathrm{d}\theta_3 \sin^2\theta_3 \cdots \int_0^\pi \mathrm{d}\theta_{n-2} \sin^{n-3}\theta_{n-2}. \qquad (8.73)$$

The angular part of the integral over a function $f(x)$ depending only on the modulus $r = \sqrt{x^2}$ of the n–dimensional vector x, can be carried out using the standard formulae

$$\int_0^\pi \mathrm{d}\theta \sin^m \theta = \frac{\sqrt{\pi}\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m+2}{2}\right)} \quad \Rightarrow \int \mathrm{d}^n x f(r) = \frac{2\pi^{n/2}}{\Gamma(\frac{n}{2})} \int \mathrm{d}r r^{n-1} f(r) \qquad (8.74)$$

with the result

$$I(p, n) = \frac{\lambda^2}{2(2\pi)^4} \frac{2\pi^{(n-1)/2}}{\Gamma(\frac{n-1}{2})} \int \mathrm{d}k_0 \int_0^\infty \mathrm{d}\omega \omega^{n-2} \frac{1}{[k_0^2 - \omega^2 - m^2 + \mathrm{i}\epsilon]} \frac{1}{[(k_0 - M)^2 - \omega^2 - m^2 + \mathrm{i}\epsilon]}. \qquad (8.75)$$

Note, that while for integer $n = 4$ (8.75) reduces to (8.69), the former has a well–defined meaning in fact for <u>all real</u> $n \in (1, 4)$! With $I(p, n)$ defined on an open interval we can continue it analytically to the whole complex n plane. There must be some singularity at $n = 4$, but its nature is not obvious at first glance. Explicit evaluation shows that this singularity is a simple pole!

To continue $I(p, n)$ below $n = 1$ we rewrite it using l times a simple per partes integration and taking into account that $\partial/\partial\omega = 2\omega\partial/\partial\omega^2$

$$
\begin{aligned}
I(p, n) \;=\; & \frac{\lambda^2}{2(2\pi)^4} \frac{2\pi^{(n-1)/2}}{\Gamma(\frac{n-1}{2} + l)} \\
& \int \mathrm{d}k_0 \int_0^\infty \mathrm{d}\omega \omega^{n-2+2l} \left(-\frac{\partial}{\partial\omega^2}\right)^l \frac{1}{[k_0^2 - \omega^2 - m^2 + \mathrm{i}\epsilon]} \frac{1}{[(k_0 - M)^2 - \omega^2 - m^2 + \mathrm{i}\epsilon]}, \quad (8.76)
\end{aligned}
$$
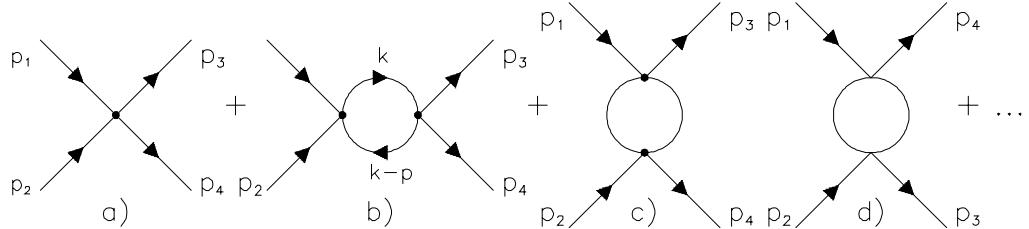
Figure 8.3: Lowest order Feynman diagrams for the elastic scattering of two scalar particles with mass m.

which converges for all integer $1 - 2l < n < 4$. The essence of the above method of analytical continuation is to rewrite $I(p, n)$ in a way that coincides with the original form (8.75) in the region $n \in (1, 4)$ but has a well–defined meaning in a larger interval. The operation of per partes integration does not change the behaviour of the integrand at large ω, but improves the convergence property at the origin $\omega = 0$, thereby extending the convergence og (8.76) to lower values of n. To continue analytically above $n = 4$ we use another trick and namely insert the identity

$$1 = \frac{1}{2}\left(\frac{\mathrm{d}k_0}{\mathrm{d}k_0} + \frac{\mathrm{d}\omega}{\mathrm{d}\omega}\right) \tag{8.77}$$

into (8.75). Performing the integrals over k_0 and ω again using per partes we arrive at the formula

$$I(p, n) = \frac{1}{2}(-n + 6)I(p, n) + I'(p, n) \quad \Rightarrow I(p, n) = \frac{2}{n-4}I'(p, n), \tag{8.78}$$

where

$$I'(p, n) = \frac{\lambda^2}{2(2\pi)^4}\frac{2\pi^{(n-1)/2}}{\Gamma(\frac{n-1}{2})}\int \mathrm{d}k_0 \int \mathrm{d}\omega\, \omega^{n-2} u(k_0, \omega, M), \tag{8.79}$$

$$u(k_0, \omega, M) \equiv \tag{8.80}$$

$$\frac{2m^2}{(k_0^2 - m^2 - \omega^2 + \mathrm{i}\epsilon)^2\left[(k_0 - M)^2 - \omega^2 - m^2 + \mathrm{i}\epsilon\right]} + \frac{2(k_0 - M)M + 2m^2}{(k_0^2 - m^2 - \omega^2 + \mathrm{i}\epsilon)\left[(k_0 - M)^2 - \omega^2 - m^2 + \mathrm{i}\epsilon\right]^2}.$$

Note that $I'(p, n)$ exists for all $1 < n < 5$ and thus (8.78) implies that $I(p, n)$, considered as a function of complex n, has a simple pole at $n = 4$!

So far we have discussed the basic property of the integral $I(p, n)$ in n dimensions without its explicit evaluation. To do that we proceed as follows. Let us consider the diagram in Fig. 8.3c, where the bubble appears in the t–channel. First we combine the two propagators in (8.69) into one, using the identity

$$\frac{1}{ab} = \int_0^1 \mathrm{d}x \frac{1}{[ax + b(1 - x)]^2}, \tag{8.81}$$

which is a special case of **Feynman parametrization** discussed in next subsection. For (8.69) it gives

$$I(p, n) = \frac{\lambda^2}{2}\int \frac{\mathrm{d}^n k}{(2\pi)^n}\int_0^1 \frac{\mathrm{d}x}{[(k^2 - m^2)x + ((k - p)^2 - m^2)(1 - x)]^2} = \frac{\lambda^2}{2}\int_0^1 \mathrm{d}x \int \frac{\mathrm{d}^n l}{(2\pi)^n}\frac{1}{[l^2 - a^2(x) - \mathrm{i}\epsilon]^2}, \tag{8.82}$$

where we have changed the order of integrations and subsituted

$$l \equiv k - p(1 - x), \quad a^2(x) \equiv m^2 - p^2 x(1 - x), \quad p^2 \equiv (p_1 - p_3)^2. \tag{8.83}$$

The inner integral in (8.82) can be performed using the formulae of subsection 8.3.3, yielding

$$I(p, n) = \frac{\lambda^2}{4}\frac{\mathrm{i}}{(4\pi)^2}(4\pi)^\varepsilon \Gamma(\varepsilon)\int_0^1 \mathrm{d}x\, [a^2(x)]^{-\varepsilon}, \quad \varepsilon \equiv 2 - \frac{n}{2}. \tag{8.84}$$

Two aspects of the formula (8.84) are worth noticing:

- The result is not a real, but for arbitrary p always a complex number!

- It diverges for all even integers $n \geq 4$, but except for them is finite. In other words, $I(p, n)$, considered as a function of n in the complex plane is a meromorphic function with simple poles at $n = 4, 6, 8 \cdots$.

Before proceeding further it is crucial to realize that in general n dimensions the coupling λ is no longer dimensionless, as in four dimensions, but has dimension equal to $4 - n = 2\varepsilon$. The same dimension has also the function $I(p, n)$. For n close to the physical value $n = 4$ (8.84) we can use expansions of the gamma function $\Gamma(\varepsilon)$ and $(4\pi)^\varepsilon$

$$\Gamma(\varepsilon) = \frac{1}{\varepsilon}\Gamma(1 + \varepsilon) = \frac{1}{\varepsilon} - \gamma_E + \sum_{n=2}^{\infty}\frac{(-\varepsilon)^{n-1}}{n!}\zeta(n), \tag{8.85}$$

$$(4\pi)^\varepsilon = 1 + \varepsilon\ln 4\pi + \cdots, \tag{8.86}$$

where $\zeta(n)$ is the Riemann zeta function and $\gamma_E = 0.5772 \cdots$ the Euler number. We cannot, however, expand in a similar way directly $a^{-2\varepsilon}$ because a has dimension 2 (in energy units). Physically well–defined expansion requires introducing some **mass parameter** μ that provides a scale for the expansion of the logarithm

$$\left(\frac{a^2(x)}{\mu^2}\right)^{-\varepsilon} = 1 - \varepsilon \ln \frac{a^2(x)}{\mu^2} + \cdots. \tag{8.87}$$

Inserting $1 = \mu^{2\varepsilon}\mu^{-2\varepsilon}$ into (8.84) we get

$$I(p,n) = \mathrm{i}\frac{\lambda^2}{2}\frac{1}{(4\pi)^2}\mu^{-2\varepsilon}\left[\frac{1}{\varepsilon} - \gamma_E + \ln 4\pi - \int_0^1 \mathrm{d}x \ln \frac{a^2(x)}{\mu^2} + \mathcal{O}(\varepsilon)\right]. \tag{8.88}$$

This new parameter μ is an integral part of the dimensional regularization. From the way it has been introduced, it is clear that $I(p,n)$ *does not depend* on it, provided we keep all the terms in expansion (8.87). A truncated version of (8.84), like the first two terms in the Laurent series in (8.88), which is the only part of (8.87) that survives in the limit $\varepsilon \to 0$, still contains μ despite the absence of the factor $\mu^{-2\varepsilon}$ to cancel this dependence. This apparent inconsistency disappears in the process of defining the renormalized coupling, which becomes μ dependent. This procedure is sketched below. Summing the contributions of diagrams in Fig. 8.3a,b we get (from now on λ is replace with λ_B to indicate its later interpretation as **bare** coupling)

$$-\mathrm{i}\lambda_B\left[1 - \frac{\lambda_B}{2}\frac{\mu^{-2\varepsilon}}{(4\pi)^2}\left(\frac{1}{\varepsilon} - \gamma_E + \ln 4\pi - \int_0^1 \mathrm{d}x \ln \frac{a^2(x)}{\mu^2}\right)\right], \tag{8.89}$$

which to the order considered can be rewritten as

$$-\mathrm{i}\underbrace{\lambda_B\left[1 - \frac{\lambda_B}{2}\frac{\mu^{-2\varepsilon}}{(4\pi)^2}\left(\frac{1}{\varepsilon} - \gamma_E + \ln 4\pi\right)\right]}_{\equiv \mu^{2\varepsilon}\lambda_R(\mu)}\underbrace{\left[1 + \frac{\lambda_B}{2}\frac{\mu^{-2\varepsilon}}{(4\pi)^2}\int_0^1 \mathrm{d}x \ln \frac{a^2(x)}{\mu^2}\right]}_{\text{finite terms}}, \tag{8.90}$$

where the **renormalized coupling** $\lambda_R(\mu)$ is defined in terms of the bare coupling λ_B as follows:

$$\lambda_R(\mu) \equiv \mu^{-2\varepsilon}\lambda_B\left[1 - \lambda_B\frac{\mu^{-2\varepsilon}}{32\pi^2}\left(\frac{1}{\varepsilon} - \gamma_E + \ln 4\pi\right)\right]. \tag{8.91}$$

The bare coupling λ_B has dimension 2ε and is **held fixed** when μ is varied. The renormalized coupling $\lambda_R(\mu)$, on the other hand, is dimensionless, and absorbs the singular term $1/\varepsilon$, eventually plus some finite terms, like $-\gamma_E + \ln 4\pi$ in (8.91). This latter choice is, however, a matter of convention only and any finite term is equally legal. The convention defined in (8.91) is called $\overline{\mathrm{MS}}$, that corresponding to vanishing finite terms is called MS (minimal subtraction). As the scale μ appears explicitly only in the factor $\mu^{-2\varepsilon}$ one might at first glance conclude that it disappears in the limit $\varepsilon \to 0$, making thus λ_R μ–independent. This, however, is not the case due to the presence of the singular term $1/\varepsilon$. Evaluating the derivative of $\lambda_R(\mu)$ with respect to $\ln \mu$ we find a finite result as $\varepsilon \to 0$

$$\frac{\mathrm{d}\lambda_R(\mu)}{\mathrm{d}\ln \mu} = -2\varepsilon\lambda_B\mu^{-2\varepsilon} + \frac{1}{8\pi^2}\lambda_B^2\mu^{-4\varepsilon} + \mathcal{O}(\lambda_B^3) = -2\varepsilon\lambda_R + \frac{1}{16\pi^2}\lambda_R^2 + \mathcal{O}(\lambda_R^3) \equiv \beta_\phi(\lambda_R, \varepsilon). \tag{8.92}$$

The first term, proportional to λ_B, vanishes in the limit $\varepsilon \to 0$, but *must be kept* if the renormalization procedure is carried out in terms of renormalized quantities and so called **counterterms** (see next section). The second, finite, term comes from the product

$$-\lambda_B^2\frac{1}{32\pi^2}\frac{1}{\varepsilon}\frac{\mathrm{d}\mu^{-4\varepsilon}}{\mathrm{d}\ln \mu} = -\lambda_B^2\frac{1}{32\pi^2}\frac{1}{\varepsilon}(-4\varepsilon)\mu^{-4\varepsilon} = \frac{\lambda_R^2}{8\pi^2}. \tag{8.93}$$

Replacing λ_B^2 in (8.92) with λ_R^2 is legal as these two expressions start to differ first at order λ_B^3. The equation (8.92) determines the implicit dependence of the renormalized coupling $\lambda_R(\mu)$ on μ, which cancels the explicit dependence on μ of the finite terms in (8.90). This cancellation holds, however, only for physical quantities, like the cross–sections and only if perturbation theory is summed to all orders. The important point to realize is that the leading order β–function coefficient $1/16\pi^2$, obtained within the dimensional

145

regularization in (8.93), is actually independent of the regularization method used in its derivation! For instance, in the conventional Pauli–Villars type of regularization we have instead of (8.91)

$$\lambda_R(\mu) \equiv \lambda_B \left[1 - \lambda_B \frac{1}{32\pi^2} \left(\ln \frac{M^2}{\mu^2} + \text{finite terms} \right) \right], \tag{8.94}$$

where M is the large cut–off to be sent to infinity. In this case the value $1/16\pi^2$ comes from taking the derivative with respect to $\ln \mu$ of the explicit logarithmic term $\ln(M/\mu)$ and has a clear physical interpretation, discussed in Section 8.2.

8.3.2 Feynman parametrization

The formula (8.81) is the simplest case of a general expression for the combination of arbitrary number of scalar propagators of the same type

$$\frac{1}{a_1 a_2 \cdots a_n} = (n-1)! \int_0^1 dx_1 \cdots \int_0^1 dx_n \delta \left(\sum_{i=1}^n x_i - 1 \right) \frac{1}{[x_1 a_1 + \cdots + x_n a_n]^n}. \tag{8.95}$$

For the combination of 2,3 and 4 propagators with general powers we have

$$\frac{1}{a^\alpha b^\beta} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 dx \frac{x^{\alpha-1}(1-x)^{\beta-1}}{[xa + (1-x)b]^{\alpha+\beta}}, \tag{8.96}$$

$$\frac{1}{a^\alpha b^\beta c^\gamma} = \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} \int_0^1 dx\, x \int_0^1 dy \frac{u_1^{\alpha-1} u_2^{\beta-1} u_3^{\gamma-1}}{[u_1 a + u_2 b + u_3 c]^{\alpha+\beta+\gamma}}, \tag{8.97}$$

where

$$u_1 = xy, \quad u_2 = x(1-y), \quad u_3 = 1-x,$$

$$\frac{1}{a^\alpha b^\beta c^\gamma d^\delta} = \frac{\Gamma(\alpha+\beta+\gamma+\delta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)\Gamma(\delta)} \int_0^1 dx\, x^2 \int_0^1 dy\, y \int_0^1 dz \frac{u_1^{\alpha-1} u_2^{\beta-1} u_3^{\gamma-1} u_4^{\delta-1}}{[u_1 a + u_2 b + u_3 c + u_4 d]^{\alpha+\beta+\gamma+\delta}}, \tag{8.98}$$

and

$$u_1 = 1-x, \quad u_2 = xyz, \quad u_3 = x(1-y), \quad u_4 = xy(1-z).$$

Another function encountered in evaluation of loop diagrams is the Euler function $B(x,y)$ defined as

$$\int_0^1 dx\, x^\alpha (1-x)^\beta = B(\alpha,\beta) = \frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)}. \tag{8.99}$$

In all the above formulae we have used the standard definition of the Γ-function

$$\Gamma(z) \equiv \int_0^\infty t^{z-1} e^{-t} dt, \tag{8.100}$$

which is related to another widely used ψ-function

$$\psi(z) \equiv \frac{\Gamma'(z)}{\Gamma(z)}, \quad \psi(1) = -\gamma_E. \tag{8.101}$$

8.3.3 Momentum integrals and related formulae

The basic formulae needed for evaluation of one loop diagrams are (a is a real number) the following

$$I(m,r) \equiv \int \frac{d^n k}{(2\pi)^n} \frac{(k^2)^r}{[k^2 - a^2 + i\varepsilon]^m} = \frac{i}{(4\pi)^{n/2}} (-1)^{r-m} (a^2)^{r-m+n/2} \frac{\Gamma(r+\frac{n}{2})\Gamma(m-r-\frac{n}{2})}{\Gamma(\frac{n}{2})\Gamma(m)} \tag{8.102}$$

$$\int \frac{d^n k}{(2\pi)^n} \frac{k_\mu k_\nu}{[k^2 - a^2 + i\varepsilon]^m} = \frac{1}{n} g_{\mu\nu} \int \frac{d^n k}{(2\pi)^n} \frac{k^2}{[k^2 - a^2 + i\varepsilon]^m}, \tag{8.103}$$

$$\int d^n k \frac{1}{k^2 + i\varepsilon} = 0. \tag{8.104}$$

These formulae can be derived

quantity	dimension
fermions	$(n-1)/2 = 3/2 - \varepsilon$
gauge bosons	$(n-2)/2 = 1 - \varepsilon$
gauge coupling	$(4-n)/2 = \varepsilon$

Table 8.1: Dimensions of the fields and the couplant in QED and QCD.

- using the so called **Wick rotation** write $I(m,r)$ as an integral over the momentum in <u>Euclidean</u> space (where $k_E^2 = \sum_{i=1}^{n} k_i^2$)

$$I(m,r) = \mathrm{i}(-1)^{r-m} \int \mathrm{d}^n k_E \frac{\left(k_E^2\right)^r}{[k_E^2 + a^2 - \mathrm{i}\varepsilon]^m}, \qquad (8.105)$$

- working out the angular integral according to (8.74) and

- evaluate the remaining one–dimensional integral by means of per partes for integer n or using formula 2.2.9.7 of [135]:

$$\int_0^\infty \mathrm{d}x \frac{x^{n-1+2r}}{[x^2 + a^2]^m} = \frac{1}{2} \left(a^2\right)^{r-m+n/2} \frac{\Gamma(r + \frac{n}{2})\Gamma(m - r - \frac{n}{2})}{\Gamma(m)}. \qquad (8.106)$$

8.3.4 Dimensional regularization in QED and QCD

Similarly as in the $\lambda\phi^4$ theory the way basic fields enter the QED lagrangian (7.18) determines their dimensions, as well as that of the gauge couplant, in n–dimensional space. They are listed in the Table 7.1 and are the same in QED and QCD. In QED and QCD there is additional complication in n dimensions, related to the extension of the Clifford algebra of γ matrices into n–dimensional space–time. There is some ambiguity how to do it, but this ambiguity has no physical consequences. In order to generalize the Clifford algebra, defined by the anticommutation relations,

$$\{\gamma_\mu, \gamma_\nu\} = 2g_{\mu\nu}, \ \ g_{\mu\nu} = 0 \ \ \text{for } \mu \neq \nu, \ \ g_{00} = 1, \ \ g_{ii} = -1 \ \ \text{for } i = 1, 2, \cdots n - 1 \qquad (8.107)$$

to n dimension it is useful to start with integer n, where (8.107) makes sense mathematically and we can ask whether there is a **realization** of this Clifford algebra by some mathematical objects. The answer is not trivial even for integer n and has two parts:

- **even $n = 2k$**: there are n matrices of the dimension 2^k satisfying (8.107) and the analogue of γ_5 in 4 dimensions is defined as

$$\gamma^* \equiv \mathrm{i}\gamma^0\gamma^1 \cdots \gamma^{n-1}, \qquad (8.108)$$

- **odd $n = 2k + 1$**: one starts with even number of dimensions $n = 2k + 2$ and takes $n - 1 = 2k + 1$ of the associated γ matrices, which obviously satisfy (8.107). This procedure is, however, not entirely trivial as it might seem at first glance, but it can be carried out.

For integer n the Clifford algebra can thus be realized by matrices of the dimension ([x] denotes the whole part of a real number x) $2^{[(n+1)/2]}$, with the following properties:

$$\begin{aligned}
\mathrm{Tr}\,\mathbf{1} &= 2^{[(n+1)/2]}, \\
\mathrm{Tr}\,\gamma_\mu\gamma_\nu &= 2^{[(n+1)/2]}g_{\mu\nu}, \\
\mathrm{Tr}\,\gamma_\mu\gamma_\nu\gamma_\alpha\gamma_\beta &= 2^{[(n+1)/2]}\left(g_{\mu\nu}g_{\alpha\beta} - g_{\mu\alpha}g_{\nu\beta} + g_{\mu\beta}g_{\nu\alpha}\right), \\
\gamma_\mu\gamma^\mu &= n, \\
\gamma_\mu\gamma^\alpha\gamma^\mu &= (2 - n)\gamma^\alpha, \\
\gamma_\mu\gamma^\alpha\gamma^\beta\gamma^\mu &= 4g^{\alpha\beta} + (n - 4)\gamma^\alpha\gamma^\beta, \\
\gamma_\mu\gamma^\alpha\gamma^\beta\gamma^\delta\gamma^\mu &= -2\gamma^\delta\gamma^\beta\gamma^\alpha + (4 - n)\gamma^\alpha\gamma^\beta\gamma^\delta.
\end{aligned} \qquad (8.109)$$

The trace of an odd number of γ matrices vanishes in n dimensions as well. In the above relations n must obviously be an integer, but after the trace reductions and momentum integrations are done the resulting

147

expressions will make sense for **arbitrary real n**. This is the essence of dimensional regularization. In QED the contribution of the basic vacuum polarization loop in Fig. 8.1b has in n dimensions the form similar to that of (8.69):

$$I_{\mu\nu}(q, n) = -e_B^2 \int \frac{\mathrm{d}k^n}{(2\pi)^n} \frac{L_{\mu\nu}(q, k, n)}{[k^2 - m^2][(k-q)^2 - m^2]}. \tag{8.110}$$

Note that contrary to the analogous quantity $I(p, n)$ in $\lambda\phi^4$ theory (8.110) is <u>dimensionless</u>. The numerators in (8.110) are easily recombined using the Feynman parametrization exactly as in (8.82)-(8.83) while the numerator equals

$$L_{\mu\nu}(k, q, n) = 4 \left[2k_\mu k_\nu - (k_\mu q_\nu + k_\nu q_\mu) - (k^2 - kq - m^2)g_{\mu\nu} \right] \tag{8.111}$$

$$= 8x(1-x) \left[q^2 g_{\mu\nu} - q_\mu q_\nu \right] + 4g_{\mu\nu} \left[\left(\frac{2}{n} - 1 \right) l^2 + (m^2 - q^2 x(1-x)) \right]. \tag{8.112}$$

Using the formula (8.102) it is straightforward to show that the second term in (8.112) vanishes, similarly as in the Pauli–Villars regularization technique, after integration over the momentum l, yielding

$$I_{\mu\nu}(q, n) = -e_B^2 \left[q^2 g_{\mu\nu} - q_\mu q_\nu \right] \int_0^1 \mathrm{d}x 8x(1-x) \int \frac{\mathrm{d}l^n}{(2\pi)^n} \frac{1}{[l^2 - a^2 + \mathrm{i}\varepsilon]^2}, \tag{8.113}$$

where $a^2(x, q^2) = m^2 - q^2 x(1-x)$. Performing the momentum integrals we get

$$I_{\mu\nu}(q, n) = \left[q^2 g_{\mu\nu} - q_\mu q_\nu \right] I(q^2) \tag{8.114}$$

where the scalar function $I(q^2)$ is given as

$$I(q^2, \varepsilon) \equiv -\mathrm{i}\mu^{-2\varepsilon} \frac{e_B^2}{2\pi^2} (2\pi)^\varepsilon \Gamma(\varepsilon) \int_0^1 \mathrm{d}x x(1-x) \left(a(x, q) \right)^{-\varepsilon}. \tag{8.115}$$

Proceeding further as in $\lambda\phi^4$ theory by introducing the scale μ etc. we get, keeping only terms that do not vanish in the limit $\varepsilon \to 0$

$$I(q, \varepsilon) = -\mathrm{i}\mu^{-2\varepsilon} \frac{e_B^2}{12\pi^2} \left[\frac{1}{\varepsilon} - \gamma_E + \ln 4\pi - 6 \int_0^1 \mathrm{d}x x(1-x) \ln \frac{(m^2 - q^2 x(1-x))}{\mu^2} \right]. \tag{8.116}$$

The sum of the diagrams in Fig. 8.1a,b thus equals (without the fermion bispinors and omitting the term $q_\mu q_\nu$ which vanishes after sandwiching between them)

$$\left(-\frac{\mathrm{i}}{q^2} \right) e_B^2 \left[1 + q^2 g_{\mu\nu} I(k, q, \varepsilon) \left(-\frac{\mathrm{i}}{q^2} \right) \right] =$$

$$\left(-\frac{\mathrm{i}}{q^2} \right) e_B^2 \left(1 - \frac{e_B^2 \mu^{-2\varepsilon}}{12\pi^2} \left[\frac{1}{\varepsilon} - \gamma_E + \ln 4\pi - 6 \int_0^1 \mathrm{d}x x(1-x) \ln \frac{m^2 - q^2 x(1-x)}{\mu^2} \right] \right). \tag{8.117}$$

Eq. (8.117) has similar structure as (8.89) and leads to introduction of the renormalized electric charge

$$e_R^2(A, \mu) \equiv e_B^2 \mu^{-2\varepsilon} \left(1 - \frac{e_B^2 \mu^{-2\varepsilon}}{12\pi^2} \left[\frac{1}{\varepsilon} - \gamma_E + \ln 4\pi \right] \right). \tag{8.118}$$

Taking the derivative of (8.118) with respect to $\ln \mu$ we get in the limit $\varepsilon \to 0$, analogously to (8.92),

$$\frac{\mathrm{d}\alpha_R(A, \mu)}{\mathrm{d}\ln \mu} \equiv \beta_{\mathrm{QED}}(\alpha_R, \varepsilon) = -2\varepsilon \alpha_R(A, \mu) + \beta_0 \alpha_R^2(A, \mu), \quad \alpha \equiv \frac{e^2}{4\pi}, \quad \beta_0 = \frac{2}{3\pi}, \tag{8.119}$$

where, as in the case of the $\lambda\phi^4$ theory, the linear term, proportional to ε, has to be retained. In $\lambda\phi^4$ theory, QED as well as QCD (actually quite in general) we can therefore write β–functions in $n = 4 - 2\varepsilon$ dimensions as

$$\beta_\varepsilon(x, \varepsilon) = -2\varepsilon x + \beta(x), \tag{8.120}$$

where $\beta(x)$ are β–functions in 4 dimensions. Moreover, the equation for the renormalized coupling is exactly the same as that obtained using the Pauli–Villars regularization. So long as the regularization technique is consistent with the basic requirements of relativistic and gauge invariance, the properties of *physical* renormalized quantities should not depend on its choice.

8.3.5 *Renormalization at one loop level and the technique of counterterms

In the preceding subsections we have shown explicitly at one loop how in QED all UV divergencies can be absorbed in the definition of the **renormalized**

- electron charge $e_R^2(\mu)$,

- electron mass $m_R(\mu)$,

- electron and photon fields $A_R^\mu(x, \mu)$ and $\psi_R(x, \mu)$,

which replaced their "bare" analogues appearing in the original QED lagrangian (7.18). In terms of these quantities all physical observables become finite. At one loop this claim follows from the previous four subsections if we add the contributions from diagrams in Fig. 8.1a-d, with the lower vertex amputated. In terms of the renormalization factors Z_1, Z_{2F}, Z_3, Z_m, and writing out only the singular parts of the respective contributions we had from

Fig. 8.1a: $-\mathrm{i}e_B\gamma_\mu$,

Fig. 8.1b: $-\mathrm{i}e_B\gamma_\mu(Z_3 - 1)$,

Fig. 8.1c: $-\mathrm{i}e_B\gamma_\mu\left(1 - Z_{2F}^{-1}\right)$,

Fig. 8.1d: $-\mathrm{i}e_B\gamma_\mu\left(Z_1^{-1} - 1\right)$.

In summing the above expressions we have to keep in mind that in perturbation theory each renormalization factor Z_i has the form $Z_i = 1 + \mathcal{O}(e_B^2)$ so that to the leading order in e_B^2 the sum of divergent factors should be written as

$$\left[1 + (Z_3 - 1) + 2\left(1 - Z_{2F}^{-1}\right) + \left(Z_1^{-1} - 1\right)\right] = \frac{\left[1 + \left(Z_1^{-1} - 1\right)\right]\left[1 + (Z_3 - 1)\right]}{\left[1 + \left(Z_{2F}^{-1} - 1\right)\right]^2} = \frac{Z_3 Z_{2F}^2}{Z_1}, \qquad (8.121)$$

where the factor 2 in front of $(1 - Z_{2F}^{-1} - 1)$ reflects the fact that the loop in Fig. 8.1c can be on the outgoing electron leg as well. Multiplying this sum with $1/\sqrt{Z_3}$ for each external (from the point of view of the upper vertices in Fig. 8.1) photon and with $1/\sqrt{Z_{2F}}$ for each external electron line we end up with

$$-\mathrm{i}e_B\gamma_\mu\sqrt{Z_3}Z_{2F}Z_1^{-1} = -\mathrm{i}e_B\gamma_\mu\sqrt{Z_3} = -\mathrm{i}e_R\gamma_\mu, \qquad (8.122)$$

where we have now used the identity $Z_1 = Z_{2F}$. The diagram in Fig. 8.1c contributes also to the renormalization (redefition) of the electron mass, but this divergence can be ignored if we write electron propagators everywhere in Fig. 8.1 with the renormalized mass instead of the bare one.

The renormalization procedure outlined in this and preceding subsections is physically motivated and reasonably transparent at one loop level but its complexity grows dramatically at higher orders, where diagrams with more loops must be taken into account. In a systematic way to deal with UV divergencies at all orders of perturbation theory we proceed in a sequence of steps:

- Bare quantities are expressed in terms of renormalized ones, by introducing for all of them appropriate renormalization factors

$$\psi_B \equiv \sqrt{Z_{2F}}\psi_R, \qquad (8.123)$$
$$A_B^\mu \equiv \sqrt{Z_3}A_R^\mu, \qquad (8.124)$$
$$m_B \equiv Z_m m_R, \qquad (8.125)$$
$$\alpha_B \equiv Z_\alpha \alpha_R. \qquad (8.126)$$

Each of the factors Z_i can be expanded in terms of the renormalized coupling α_R as

$$Z_i(M/\mu) = 1 + \alpha_R\left((\text{singular terms}) + \text{finite terms}\right) + \mathcal{O}(\alpha_R^2), \qquad (8.127)$$

where the explicit expressions for the singular term depend on the choice of regularization. It is here where the properties of the dimensional regularization become crucial for a consistent formulation of the renormalization to *all* orders. In DR the general structure of the counterterms reads

$$Z_i(M/\mu) = 1 + \alpha_R \left(\frac{\gamma_1^{(1)}}{\varepsilon} + \gamma_0^1 \right) + \alpha_R^2 \left(\frac{\gamma_2^{(2)}}{\varepsilon^2} + \frac{\gamma_1^{(2)}}{\varepsilon} + \gamma_0^2 \right) + \cdots, \tag{8.128}$$

where all the $\gamma_j^{(i)}$ are *finite* numbers, calculable in perturbation theory. It turns out that except for the lowest coefficients $\gamma_0^{(i)}$, all those standing by the inverse powers of ε, and thus defining *singular* terms, are *uniquely defined* by the requirements that they *cancel* the UV singularities encountered in calculating the loops. On the other hand all the $\gamma_0^{(i)}$'s can be chosen completely *arbitrarily*. In this approach we thus work from the very beginning exclusively with *renormalized* quantities and introduce the counterterms to cancel all the UV infinities coming from loops.

- These counterterms formally appear when the original bare lagrangian (7.18) is rewritten in terms of the renormalized quantities

$$\mathcal{L}^{\text{QED}} = -\frac{1}{4} Z_3 F_R^{\mu\nu} F_{\mu\nu}^R + Z_{2F} \overline{\Psi}_R (\mathrm{i}\, \slashed{\partial} - Z_m m_R) \Psi_R + \sqrt{Z_\alpha} \sqrt{Z_3} Z_{2F} e_R \overline{\Psi}_R \gamma_\mu \Psi_R A_R^\mu \tag{8.129}$$

and then reorganized in the following equivalent way

$$\mathcal{L}^{\text{QED}} = -\frac{1}{4} F_R^{\mu\nu} F_{\mu\nu}^R + \overline{\Psi}_R (\mathrm{i}\, \slashed{\partial} - m_R) \Psi_R + e_R \overline{\Psi}_R \gamma_\mu \Psi_R A_R^\mu + \mathcal{L}^{\text{counterms}}, \tag{8.130}$$

where

$$\mathcal{L}^{\text{count}} \equiv -\frac{1}{4}(Z_3 - 1) F_R^{\mu\nu} F_{\mu\nu}^R + (Z_{2F} - 1)\overline{\Psi}_R \mathrm{i}\, \slashed{\partial} \Psi_R - (Z_{2F} Z_m - 1) m_R \overline{\Psi}_R \Psi_R +$$
$$\left(\sqrt{Z_\alpha} \sqrt{Z_3} Z_{2F} - 1 \right) e_R \overline{\Psi}_R \gamma_\mu \Psi_R A_R^\mu. \tag{8.131}$$

- Each of the so called **counterterms** in (8.131) is then treated as additional interaction term.

8.4   *Renormalization in QCD

With all basic concepts of the renormalization procedure already introduced, I shall now concentrate on those of its features which are new in QCD. Apart from technical complications connected with the color degree of freedom, the basic novel features are related to the existence of gluon *selfinteraction*. The crucial one concerns the sign of the derivative of the renormalized color couplant $\alpha_s \equiv g^2/4\pi$ with respect to the scale μ. To illustrate this difference let us return to the physical quantity:

$$R(Q) \equiv \frac{\sigma(Q, \mathrm{e^+e^-} \to \text{hadrons})}{\sigma(Q, \mathrm{e^+e^-} \to \mu^+\mu^-)} = \overbrace{\left(3 \sum_i e_i^2 \right)}^{\text{QPM}} (1 + r(Q)), \tag{8.132}$$

introduced already in the Section 6. In QCD the simple QPM formula (6.11) is just the leading order term, while QCD provides corrections, given by diagrams like those in Fig. 8.4b,c,d and contained in the quantity $r(Q)$. This correction can be expressed as a power expansion in the couplant $a(\mu) \equiv \alpha_s/\pi$ [18]

$$r(Q) = a(\mu/\Lambda) \left[r_0 + r_1(\mu/Q) a(\mu/\Lambda) + r_2(\mu/Q) a^2(\mu/\Lambda) + \cdots \right]; \quad r_0 = 1 \tag{8.133}$$

The equation determining the μ–dependence of $a(\mu/\Lambda_{QCD})$ reads [19]

$$\frac{\mathrm{d}a(\mu, c_i)}{\mathrm{d}\ln\mu} = -ba^2(\mu, c_i) \left[1 + c a(\mu, c_i) + c_2 a^2(\mu, c_i) + \cdots \right], \tag{8.134}$$

---

[18]The choice of $a(\mu)$ instead of $\alpha_s(\mu)$ as the expansion parameter is motivated by the fact that with this convention $r_0 = 1$ in (8.133) and many other processes. I shall drop the subscript "R" to denote the renormalized quantities in QCD, but shall keep the subscript "B" to denote the bare ones.

[19]I use the same letter Λ to denote the basic scale parameter in both QCD as in QED. They are, of course, quite different, but as from now on only QCD will be discussed, I shall tacitly mean Λ_{QCD} under it.
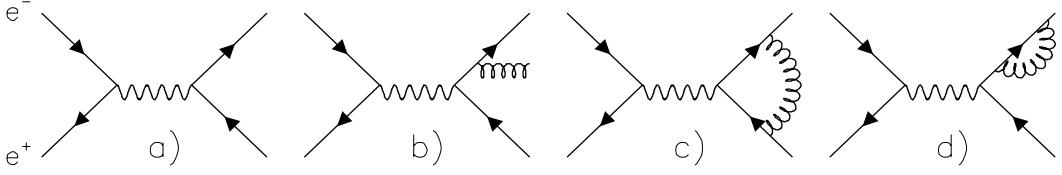
Figure 8.4: Examples of Feynman diagrams describing the e^+e^- annihilation to hadrons in the QPM (a) and the lowest order of perturbative QCD (b-d).

where for brevity of this (and the following) formulae the QCD parameter Λ is always understood to scale μ and the subscript "R" standing for "renormalized" is dropped. Whenever I shall talk about the "bare" couplant I shall explicitly say so. The r.h.s. of (8.134), considered as a function of a, is called the β-**function** of QCD. In the case of massless quarks, to which this discussion is restricted, and keeping the renormalized gauge parameter α_G appearing in the gauge fixing term (7.26) fixed [20], all the coefficients $b, c, c_i; i \geq 2$ are pure numbers. The first two of them are *uniquely* determined by the numbers of quark flavors (n_f) and colors (N_c)

$$ b = \frac{11N_c - 2n_f}{6}; \quad c = \frac{51N_c - 19n_f}{22N_c - 4n_f}. \tag{8.135} $$

In realistic QCD, $N_c = 3$ and $n_f \leq 6$ and consequently the coefficient b on the r.h.s. of (8.134) is *positive*! Herein lies the basic difference from QED, where, on the contrary, $b < 0$.

The fact that $b > 0$ and thus the leading term on the r.h.s. of (8.134) is *negative* implies that $a(\mu/\Lambda) \to 0$ as $\mu \to \infty$. This phenomenon, called **asymptotic freedom** is a fundamental feature of QCD and can be traced back to the selfinteraction of gluons. As a result the behavior of $a(\mu)$ changes dramatically compared to the situation in QED (Fig. 8.5b). In QED there are unsurmountable problems in defining the theory at *short* distances, while at long ones QED smoothly joins the classical electrodynamics. In QCD the situation is reversed: there are no difficulties at *short* distances, where perturbation theory can be safely applied, but perturbation theory becomes inadequate at *large* ones. This is clear if we take the leading order (LO) solution of (8.134), i.e. the approximation in which only the first term on its r.h.s. is taken into account:

$$ a(\mu/\Lambda) = \frac{1}{b\ln(\mu/\Lambda)}. \tag{8.136} $$

We conclude that in QCD the parameter Λ provides the *lower* and not, as in QED, the *upper* bound on the meaningful values of μ! Graphically the situation is sketched in Fig. 8.5b, which could be obtained from Fig. 8.2b by transformation $\mu \to 1/\mu$. This reversal of the role of short and long distances is not merely formal, but has profound consequences on the possibility to define QCD on the **lattice**. There is a widespread consensus among the theorists working in constructive field theory that only **asymptotically free** gauge theories, like QCD, can be formulated in a mathematically rigorous way as *local* field theories (see [129] for discussion of this point). In asymptotically free theories the cut-off dependence of QCD bare couplant $a_B(M/\Lambda)$ leads to a *well–defined* limit

$$ a(\mu/\Lambda) = \frac{1}{b\ln(\mu/\Lambda)} \xrightarrow[\mu \to \infty]{} 0, \tag{8.137} $$

which is a crucial ingredient in the construction of the continuum limit of lattice approximations. There are several further aspects of the relations (8.133) and (8.134) which merit a comment:

- Except for the first two coefficients b, c in (8.134) *all* the higher order ones $c_i; i \geq 2$ are *arbitrary* real numbers. The relation (8.134) should thus be considered as a **definition** of the couplant $a(\mu)$.

- The arbitrariness in $c_i; i \geq 2$ doesn't, however, mean that the full QCD results depend on their choice. Besides $a(\mu, c_i)$ also the expansion coefficients r_k *depend* on $c_i, i \leq k$ and these dependences cancel each other, provided we calculate (8.133) to *all* orders. More on this point in the next subsection.

---

[20]This is legal to do, see [139] for discussion of this subtle point. Roughly speaking, the bare gauge parameter can be varied in such a way, that the renormalized gauge parameter remains constant.
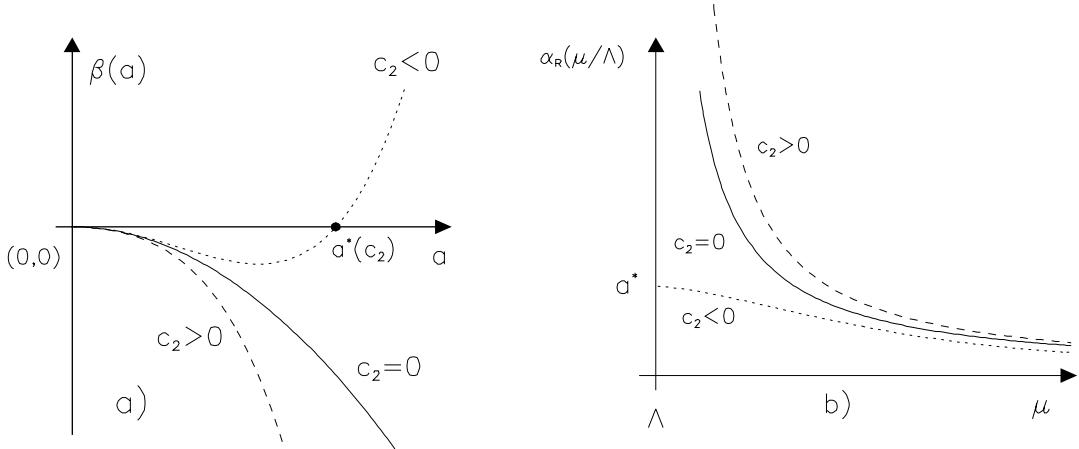
Figure 8.5: The shape of QCD β–function for three different cases of c_2 (a), and the behavior of the corresponding couplants as functions of the scale μ.

- Taking into account the next-to-leading order (NLO) term in (8.134) the equation for $a(\mu)$ can no longer be solved analytically, but the implicit equation

$$b \ln \frac{\mu}{\Lambda} = \frac{1}{a} + c \ln \frac{ca}{1 + ca} \tag{8.138}$$

can easily be solved numerically. It is also quite common to use the first two terms [21]

$$a(\mu/\Lambda) = \frac{1}{b \ln(\mu/\Lambda)} - \frac{c}{b^2} \frac{\ln((b/c)\ln(\mu/\Lambda))}{\ln^2(\mu/\Lambda)} + \cdots \tag{8.139}$$

in the expansion of the exact solution of (8.138) in powers of the inverse logarithm $1/\ln(\mu/\Lambda)$.

- The reverse side of the phenomenon of asymptotic freedom concerns the behavior of $a(\mu/\Lambda)$ at large (compared to $1/\Lambda$) distances. While the *ambiguous* higher order terms in (8.134) are irrelevant at *short* distances (positive powers of $a(\mu)$ die out for $\mu \to \infty$), they become crucially important at *large* ones, where the QCD couplant $a(\mu)$ *diverges* if only the first two (unique) terms of the β–function are taken into account. However, this argument is not very convincing because by an appropriate choice of the higher order parameters $c_i; i \geq 2$ we can obtain essentially *any* behavior of $a(\mu)$ at large distances (see Fig. 8.5). For instance, choosing $c_2 < 0; c_i = 0, i \geq 3$ we get, instead of the divergence at $\mu = \Lambda$, a *finite* IR limit of $a(\mu/\Lambda)$ as $\mu \to 0$! Physical relevance of this phenomenon, called "freezing" of the couplant, is discussed in the next subsection.

- One aspect of (8.134), which is a frequent source of confusion is the following. To solve this equation to any finite order we first have to fix the free parameters $c_i; i \geq 2$ (they may be set to 0, but a choice *must* be made). [22] The specification of all the free coefficients $c_k, k \geq 2$ defines the so called **renormalization convention** (RC). However, even for fixed coefficients c_i, i.e. fixed r.h.s. of (8.134), there is an *infinite* set of its solutions. The specification of these solutions can be done in various ways, for instance also by means of the parameter Λ introduced above. The theory doesn't prefer any of these solutions, but when going from one solution $a^{(1)}(\mu/\Lambda_1)$, corresponding to Λ_1, to another, $a^{(2)}(\mu/\Lambda_2)$, corresponding to Λ_2, the values of all expansion coefficients $r_k; k \geq 1$ in (8.133) vary as well. For instance, for the lowest order coefficient r_1 we find

$$r_1(\mu/\Lambda_1) = r_1(\mu/\Lambda_2) + b \ln \frac{\Lambda_2}{\Lambda_1}. \tag{8.140}$$

---

[21]Note that the parameter Λ, as introduced at the LO in (8.136), as well as at the NLO in (8.138), satisfies the condition $a(\mu = \Lambda) = \infty$. This, however, is not necessary and we could *define* Λ in many other ways.

[22]Or taking into account all terms in the β–function expansion, but demanding that they define an analytic function.

Fixing the free coefficients $c_i, i \geq 2$ and selecting one of the solutions to (8.134) *defines* the so called **renormalization scheme** (RS) At the NLO only the latter specification is needed and thus to this order the RS each is fully specified by the associated Λ_{RS}. [23] Note also, that only *after* this RS has been chosen does the specification of the scale μ *uniquely* determine the value of the couplant $a(\mu/\Lambda_{\mathrm{RS}})$ and the coefficients r_k! The relation (8.140) is an example of a **consistency condition** QCD perturbation theory *must* satisfy to make sense. I shall return to this subject in the next subsection.

For fixed c_i we can in principle vary *both* the choice of the solution to (8.134), labeled by the Λ parameter, and the scale μ, but it is clear that this is redundant and that we can select once and for all the former, i.e. *fix* the corresponding Λ_{RS} and vary only the scale μ (and at higher orders also the coefficients $c_i, i \geq 2$). In making this choice physical insight, experience and intuition are essential. The point is that while the full sum (if we can define it, see subsection 8.4.3) of (8.133) must be independent of μ, any *finite* order approximation to (8.133) *violates* this invariance! Basically the dependence on the scale μ is a measure of our ignorance of uncalculated higher order terms. The more sensitive are the results to the variation of μ, the more important these higher order terms probably are. But there is a practical problem how to *quantify* the "strength" of such dependence, i.e. how to determine the "allowed" range of μ and the coefficients $c_i, i \geq 2$. The conventional wisdom, based on years of experience with perturbative calculations gives the following crude recipe: the scale μ should be chosen equal to the **typical physical scale of the analyzed process**. Although basically correct, this statement should be taken with reservation as we know from the preceding paragraph that the specification of μ *does not fix* the renormalized couplant $a(\mu)$, the latter depending on the choice of the RS as well. In the case of (8.133) we could take for such a typical scale the square root of the CMS energy, in DIS of Chapter 5 either $\sqrt{Q^2}$ or $\sqrt{W^2}$ and for the Drell-Yan dilepton pair production the invariant mass of the dilepton pair.

Now we must address the following question: in which circumstances are perturbative expansions in QCD reliable? Clearly, the smaller the expansion parameter $a(\mu, \mathrm{RS})$, the better, but quantitatively the meaning of the term "small" depends also on the magnitude of the expansion coefficients $r_k(Q/\mu, \mathrm{RS})$. The applicability of perturbative QCD is thus governed by the somewhat loosely formulated demand that the terms $r_k(Q/\mu, \mathrm{RS})a^k(\mu, \mathrm{RS})$ are small compared to the leading term r_0. Setting μ equal to the typical scale Q of the process, the expansion coefficients r_k become numbers, depending only on the choice of the RS, and the magnitude of the higher order corrections is then governed by the value of the couplant at the scale Q, i.e., by $a(Q, \mathrm{RS})$. In asymptotically free theories *small* couplant requires *large* value of the scale μ and so processes which should be reliably described in perturbative QCD are the **hard scattering processes**, characterized by a *large* typical mass scale Q (short distances). For the above examples it means large primary energy, momentum transfer squared, or dilepton mass. In such circumstances QCD corrections can then be expected to be small enough not to destroy the remarkable success of the QPM.

The practical question how far down with μ (i.e. to how large distances) we can go before running into troubles with perturbation expansions is difficult to answer quantitatively. However, we expect the behavior of the couplant in the infrared region to be related to the phenomenon of color confinement. Disregarding the phenomenon of "freezing" of QCD couplant $a(\mu)$ at small μ in certain RS, it is clear that the same effect which leads to the asymptotic freedom also causes the couplant to *increase* at large distances! Although we cannot believe perturbation theory in this region it is certainly an indication that QCD has at least a chance to describe *both* the success of QPM in hard scattering processes and the color confinement!

8.4.1   *Higher orders, IR fixed points and consistency of perturbation theory.

At higher orders of perturbative QCD new free parameters $c_i; i \geq 2$ appear in the QCD β–function. The dependence of the couplant $a(\mu, c_i)$ on these parameters is determined by differential equations similar in form and meaning to (8.134):

$$\frac{\mathrm{d}a(\mu, c_i)}{\mathrm{d}c_i} \equiv \beta_i = -\beta(a) \int_0^a \frac{bx^{i+2}}{(\beta(x))^2}\mathrm{d}x, \tag{8.141}$$

where the generalized β–functions $\beta_i(a)$ can, according to the above relation, be expressed in terms of the "basic" β–function $\beta(a)$ and thus don't contain any new free parameters. The internal consistency of

---

[23]The variation of Λ_{RS}, corresponding to the change of the RS and implying *correlated* variation of the couplant $a(\mu/\Lambda_{\mathrm{RS}})$ *and* the coefficients r_k, must be distinguished from the variation of the *numerical value* of Λ_{RS} in any given RS. In the latter case the couplant $a(\mu/\Lambda_{\mathrm{RS}})$ varies, but the coefficients r_k do not.

perturbation expansions like (8.133) requires that the *full sum* of $r(Q)$ must be independent of the choices of μ and $c_i, i \geq 2$, i.e.

$$\frac{\mathrm{d}r(Q)}{\mathrm{d}\ln\mu} = \frac{\mathrm{d}r(Q)}{\mathrm{d}c_i} = 0; \; i \geq 2 \tag{8.142}$$

must hold. In the framework of perturbation theory these conditions imply the following recurrence relations between perturbative coefficients r_k

$$\frac{\mathrm{d}r_k(Q/\mu, c_i; i \leq k)}{\mathrm{d}b\ln\mu} = kr_{k-1} + c(k-1)r_{k-2} + c_2(k-2)r_{k-3} + \cdots \tag{8.143}$$

with the boundary condition $r_{-1} = 0$. Similar relations can be written for the derivatives with respect to c_i's. These relations can be solved iteratively. Exploiting the relations (8.143) and (8.141) for c_2 we get

$$\begin{aligned}
r_1(Q/\mu) &= b\ln\frac{\mu}{Q} + r_1(\mu = Q) = b\ln\frac{\mu}{\Lambda} - \rho(Q/\Lambda), \\
r_2(Q/\mu, c_2) &= \rho_2 - c_2 + (r_1 + c/2)^2
\end{aligned} \tag{8.144}$$

and similarly for higher order coefficients $r_k, k \geq 3$. The quantities ρ and ρ_2 (and generally at k–th order ρ_k) are **RG invariants**, i.e. quantities independent of μ and c_i. There is, however, one important difference between ρ and ρ_k. All of them are *dimensionless* quantities, but while the latter are *pure numbers*, calculable in perturbation theory, the former is actually a *function* of the ratio Q/Λ. It is only through the invariant ρ that the Q–dependence enters the theoretical formulae. As the numerical values of the Λ–parameter in a given RS can only be determined by comparison with experimental data, ρ cannot be calculated. It can, however, be expressed as

$$\rho = b\ln\frac{\mu}{\Lambda_{\mathrm{RS}}} - r_1(Q/\mu, \mathrm{RS}) = b\ln\frac{Q}{\Lambda_{\mathrm{RS}}} - r_1(1, \mathrm{RS}), \tag{8.145}$$

i.e. as a function of Λ in a particular RS and the corresponding value of r_1. Note that the RG invariance of ρ means not only its independence of μ, but also of the choice of the RS. To evaluate (8.145) we need Λ_{RS} and r_1 in some RS, but the difference $b\ln(Q/\Lambda_{\mathrm{RS}}) - r_1(1, \mathrm{RS})$ is *independent* of the chosen RS.

The relations (8.144) suggest, and explicit solution of (8.143) confirm, that in general the coefficients r_k are polynomials of order k in the variable $\ln(\mu/Q)$, which can be very large, if μ is too much different from Q. This observation, based on general RG considerations, is in fact the main argument for setting $\mu \approx Q$. On the other hand, it is also evident that this conclusion is only of semiquantitative nature, as the magnitude of r_k depends also on the so called 'finite' term $r_k(1, \mathrm{RS})$. The N–th order partial sum of the perturbation expansion (8.133) thus has the following general form

$$r^N(Q) \equiv \sum_{k=0}^{N-1} r_k a^{k+1} = \mathcal{F}(Q/\mu, c_i, \rho_i; i \leq k). \tag{8.146}$$

The freedom of the choice of the coefficients $c_i; i \geq 2$ means that away from the region of asymptotic freedom (i.e. $a = 0$), the behavior of $\beta(a)$ is *completely arbitrary*. In other words there is no such thing like a "**QCD** β–**function**", as there are no convincing physical arguments for a particular set of $c_i's$. Despite this, there are still many texts which discuss the "possible" behavior of QCD β-function to all orders of perturbation theory. The underlying motivation for these considerations is the erroneous assumption, that the ambiguity in the definition of $\beta(a)$ concerns only its perturbative coefficients c_i, but that when $\beta(a)$ is summed to *all orders*, this sum *is* unique. Such an invariance must hold for physical quantities, but not for the β–function itself as the latter is not a physical quantity but serves merely to *define* the couplant $a(\mu, c_i)$. This is demonstrated in Fig. 8.5 for the simplest case that only the first nonuniversal coefficient c_2 is taken into account. The three different choices of c_2 (zero, positive and negative) lead to fundamentally different behavior of $a(\mu)$ at large distances (small μ):

- For $c_2 = 0$ the couplant, blows up to infinity at $\mu = \Lambda$.

- Similarly for $c_2 > 0$ which only makes the dependence even steeper.

- For $c_2 < 0$ the couplant "freezes" at large distances at the finite value $a^*(c_2)$, given by the zero of the corresponding β–function $\beta(a)$, called the **infrared fixed point**.

The ambiguity of the β–function implies that these fixed points can be introduced for free and thus cannot have a direct physical meaning! Were there any plausible arguments why a particular set of coefficients c_i, defining the β–function with an IR fixed point, should be preferred, we could set $\mu = Q$ and conclude that the IR limit of any finite order approximation

$$\lim_{Q \to 0} r(Q) = a^*(\mathrm{RS}) \left(1 + r_1(1, \mathrm{RS})a^*(\mathrm{RS}) + r_2(1, \mathrm{RS})a^{*2}(\mathrm{RS}) + \cdots\right) \tag{8.147}$$

is finite, i.e. QCD is **IR stable**! We would be happy with this conclusion, but, unfortunately, there are no obvious reasons why β–functions with IR fixed point should be an appropriate choice of the RS in the IR region. I am convinced that the question of the IR behavior of perturbation theory is closely related to the summability of perturbative expansions, which is another outstanding problem of QCD perturbation theory, discussed in subsection 8.4.3.

8.4.2 Quark mass thresholds in the running α_s

In this section I shall discuss the importance of the proper treatment of quark mass effects in the QCD running couplant. I shall describe the approximation in which the "light" quarks u, d and s are considered massless, while the c, b, t quarks remain massive.

As multiloop calculations with massive quarks are notoriously very difficult, all higher order phenomenological analyses use calculation with a fixed **effective number** n_f of massless quarks, depending on the characteristic scale of the quantity. In order to relate two regions of different effective numbers of massless quarks the approximate matching procedure developed in [130] is commonly used. It should be emphasized that this procedure concerns only the mass effects that can be absorbed in the renormalized couplant. At higher orders there are, however, mass effects that remain in the expansion coefficients even after the effects of heavy quarks have been absorbed in the suitably defined running couplant.

In massive QCD even the lowest order β–function coefficient, b, depends on the chosen RS. While in the mass–independent RS, [24] like the so called MS, or $\overline{\mathrm{MS}}$ ones, $b(\overline{\mathrm{MS}}) = 11/2 - n_f/3$, in the mass–dependent RS [25] b becomes a nontrivial function of the scale μ and the quark masses [131]. The result (8.37) for the renormalized electric charge $\alpha(\mathrm{MOM}, \mu)$ in QED, can be directly taken over into QCD as the fermion loops of Fig. 8.2 are the same in both theories. In QCD, however, there are still other contributions to the renormalized QCD couplant $a = \alpha_s/\pi$. Instead of $b = \pi\beta_0 = 2/3$ we then have [26]

$$b(\mu/m_i) = \frac{11}{2} - \frac{1}{3}\sum_i h_i(x_i), \quad x_i \equiv \frac{\mu}{m_i}, \tag{8.148}$$

where the sum runs over all the quarks considered, m_i are the corresponding renormalized quark masses [27] and the threshold function $h(x)$ is given exactly as in QED, i.e. in (8.38). Recall that the last, approximate, equality in (8.38) is a very accurate approximation to the exact form of $h(x)$ in the whole range $x \in (0, \infty)$ [131] (see Fig. 8.2b), allowing a simple treatment of the quark mass thresholds at the LO. There is no analogous calculation of next β–function coefficient, c, for massive quarks and thus no possibility to repeat the present analysis at the NLO. This is one of the reasons why most of the phenomenological analyses use the so called "step" approximation in which at any value of μ one works with a finite effective number of massless quarks, which changes discontinuously at some **matching points** μ_i and which is the same for all β–function coefficients. Consequently $b(n_f)$ becomes effectively a function of μ, which is discontinuous at these matching points, as shown in the next figure. In this section the accuracy of this approximation and its dependence on the choice of the matching points will be investigated in some detail.

Intuitively it is clear that the matching points should be proportional to the masses of the corresponding quarks, $\mu_i \equiv \kappa m_i$. The free parameter κ, allowing for the variation of the proportionality factor, turns out to be quite important for the accuracy of the step approximation. The matching procedure then consists in the following relations at the matching points μ_i (the number in the superscript defines the corresponding effective number of massless quarks)

$$a_{app}^{\mathrm{LO},3}(\kappa m_c/\Lambda^{(3)}) = a_{app}^{\mathrm{LO},4}(\kappa m_c/\Lambda^{(4)}) \Rightarrow \Lambda^{(4)} = \Lambda^{(3)} \left(\frac{\Lambda^{(3)}}{\kappa m_c}\right)^{1/3b(4)}, \tag{8.149}$$

---

[24]Corresponding to the renormalized couplant $\alpha(A, \mu)$ in QED as introduced in Section 8.2.1.

[25]Exemplified by $\alpha(B, \mu)$ in QED.

[26]Notice that quark loop contributions to the coefficient b have the same sign in QED and QCD.

[27]For the purpose of this discussion quark masses are regarded as constants, although they, too, "run".
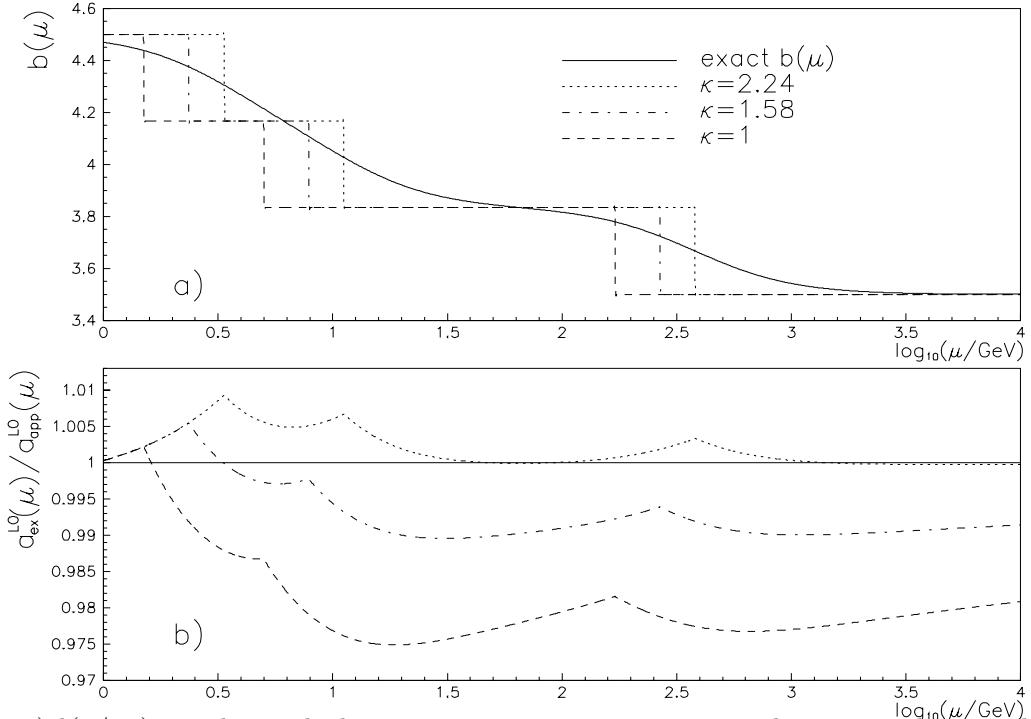
Figure 8.6: a) $b(\mu/m_i)$ together with three step approximations, corresponding to matching at the points $\mu_i = \kappa m_i; i = c, b, t$ with $\kappa = 1, 1.58, 2.24$; b) the ratio R_a for the same three approximations.

$$a_{app}^{\mathrm{LO},4}(\kappa m_b/\Lambda^{(4)}) \;\; = \;\; a_{app}^{\mathrm{LO},5}(\kappa m_b/\Lambda^{(5)}) \Rightarrow \Lambda^{(5)} = \Lambda^{(4)} \left(\frac{\Lambda^{(4)}}{\kappa m_b} \right)^{1/3b(5)}, \tag{8.150}$$

$$a_{app}^{\mathrm{LO},5}(\kappa m_t/\Lambda^{(5)}) \;\; = \;\; a_{app}^{\mathrm{LO},6}(\kappa m_t/\Lambda^{(6)}) \Rightarrow \Lambda^{(6)} = \Lambda^{(5)} \left(\frac{\Lambda^{(5)}}{\kappa m_t} \right)^{1/3b(6)}. \tag{8.151}$$

Note that each of the intervals of fixed effective number of massless quarks is necessarily associated with a different value of the Λ–parameter, $\Lambda^{(n_f)}$. The resulting dependence $a(\mu/m_i)$ is thus continuous at each of the matching points, but its derivatives at these points are discontinuous, reflecting the discontinuity of the step approximations of $b(\mu/m_i)$. The main advantage of this procedure is that it can be easily extended to any finite order.

To gain a quantitative estimate of the errors resulting from the approximate treatment of quark thresholds described above, let us solve the equation 8.134 to the leading order and with explicit mass dependence as given in (8.148)

$$\frac{\mathrm{d}a(\mu)}{\mathrm{d}\ln\mu} = -a^2 \left(\frac{11}{2} - \frac{1}{3} \sum_{i=1}^{6} h(x_i) \right), \tag{8.152}$$

using the approximate expression (8.38) for $h(x)$

$$a(\mu) = \frac{1}{\left(\dfrac{11}{2} - \dfrac{3}{3} \right) \ln \dfrac{\mu}{\Lambda^{(3)}} - \dfrac{1}{3} \displaystyle\sum_{i=c,b,t} \ln \dfrac{\sqrt{\mu^2 + 5m_i^2}}{\sqrt{(\Lambda^{(3)})^2 + 5m_i^2}}}, \tag{8.153}$$

where the fraction $\frac{3}{3}$ comes from the sum over the three massless quarks u, d, s and $\Lambda^{(3)}$ is the corresponding Λ–parameter appropriate to 3 massless quarks. What concerns the heavy quarks c, b, t I take in the following $m_c = 1.5$ GeV, $m_b = 5$ GeV, $m_t = 170$ GeV. The distinction between the "light" and "heavy" quarks is given essentially by the relative magnitude of m_i and Λ, the later being defined by the condition $5m_i^2 \gg \Lambda^2$. For the above values of m_c, m_b, m_t this condition is very well satisfied. Consequently, for $\mu \ll m_i, i = c, b, t$

156

(8.153) approaches smoothly $a^{\rm LO}$ for $n_f=3$, while for $\mu \gg m_i$ it goes to

$$a(\mu) = \cfrac{1}{b(3)\ln\dfrac{\mu}{\Lambda^{(3)}} + \dfrac{1}{3}\ln\left(\dfrac{\sqrt{5}m_c}{\Lambda^{(3)}}\dfrac{\sqrt{5}m_b}{\Lambda^{(3)}}\dfrac{\sqrt{5}m_t}{\Lambda^{(3)}}\right)} = \cfrac{1}{b(6)\ln\dfrac{\mu}{\Lambda^{(6)}}}, \qquad (8.154)$$

where the parameter

$$\Lambda^{(6)} \equiv \Lambda^{(3)}\left(\frac{\Lambda^{(3)}}{\sqrt{5}m_c}\right)^{\frac{1}{3b(6)}}\left(\frac{\Lambda^{(3)}}{\sqrt{5}m_b}\right)^{\frac{1}{3b(6)}}\left(\frac{\Lambda^{(3)}}{\sqrt{5}m_t}\right)^{\frac{1}{3b(6)}} = \Lambda^{(6)}(\kappa=\sqrt{5}) = \left(\frac{1}{\sqrt{5}}\right)^{\frac{1}{b(6)}}\Lambda^{(6)}(\kappa=1) \quad (8.155)$$

coincides (show!) with $\Lambda^{(6)}$ defined via the subsequent application of the matching relations (8.149)–(8.151) for $\kappa = \sqrt{5} \doteq 2.24$. Although from the point of view of the matching procedure any κ is in principle equally good, the choice $\kappa = \sqrt{5}$ will be shown to be in some sense the best choice. Note that the relation between $\Lambda^{(6)}$ and $\Lambda^{(3)}$ depends nontrivially on κ. In Fig. 8.6 μ dependence of the ratio

$$R_a(\mu,\kappa) \equiv \frac{a_{\rm ex}^{\rm LO}(\mu)}{a_{\rm app}^{\rm LO}(\mu,\kappa)} \qquad (8.156)$$

between the above exact solution (8.153) and the approximate expressions obtained by the above described matching procedure is plotted for three values of $\kappa = 1, \sqrt{5/2}, \sqrt{5}$ and $\Lambda^{(3)} = 200$ MeV. In order to secure that the exact and approximate solutions coincide at the scale $\mu = 1$ GeV, the value of $\Lambda^{(3)}$ used in the approximate solutions was rescaled by the factor 1.004 with respect to $\Lambda^{(3)}$ in (8.153). Consequently, any deviation from unity is purely an effect of the approximate treatment of the heavy quark thresholds. There are several conclusions that can be drawn from Fig. 8.6.

There is hardly any sign of the steplike behavior of the function $h(x)$ in the region of the c and b quark thresholds and only a very unpronounced indication of the plateau between the b and t quark thresholds. For any κ the steplike approximations are poor representations of the exact $h(x)$ primarily due to the rather slow approach of $h(x)$ to unity as $x \to \infty$. Fig. 8.6b also demonstrates the importance of the proper choice of κ. While the step approximation corresponding to the conventional choice $\kappa = 1$ underestimates the true $h(x)$ in the whole interval, and would do so even when some smoothing were applied, $\kappa = \sqrt{5}$ gives clearly much better approximation and intersects the exact $h(x)$ just at the point where we "feel" the thresholds should be.

The approximate solution based on the matching procedure defined in (8.149)–(8.151) is significantly better immediately *below* the matching point than above it and is worst at about $5m_{\rm match}$. This reflects the fact that the function $h(x)$ vanishes very fast (like x^2) at zero but approaches unity much slower. For instance in the M_Z range the effect of *both* b and c quark thresholds are much more important than that of the top quark, although $M_Z/m_c \doteq 60$, $M_Z/m_b \doteq 18$, while $M_Z/m_t \approx 1/2$!

The effect of varying κ is quite important, in particular with respect to the c and b quark thresholds. In general $\kappa > 1$ improves the approximation above the matching point, but worsens it immediately below it. The choice $\kappa = \sqrt{5}$, suggested by the asymptotic behavior of (8.154), is superior practically in the whole considered interval $\mu \in (1, 10^4)$ and leads there an excellent (on the level of 0.1%) agreement with the exact solution. On the contrary the conventional choice $\kappa = 1$ leads to a rather large deviation from the exact result, which exceeds 2% in most of this region. It turns out that the effects of smooth thresholds are of the same magnitude as those of the next–to–next–to–leading (NNLO) corrections to the β–function in $\overline{\rm MS}$ RS and must therefore be taken into account, whenever the latter is considered.

The influence of the choice of κ on the results of the matching procedure has recently been investigated [132] within the approach introduced in [133]. Although this matching procedure is more sophisticated than that of [130] and applicable at any order, it relates couplants corresponding to different numbers of massless quarks and does not concern the effects of smooth quark mass thresholds.

8.4.3 *Summation of QCD perturbation expansions

Do QCD perturbation expansions converge? If yes, how fast and if not, how to deal with the their divergence? These crucial questions must be answered before QCD becomes really a theory. It was argued by Dyson [138] already 40 years ago that the answer to the first question is *negative*. His arguments have since been critically

reanalyzed by several other people, [139, 140] and it was pointed out that his conclusions actually don't concern the *convergence* of perturbation expansions but rather the possibility to *recover* from these expansions the "exact" results. [28] The recognition of fundamental difference between "summing" of perturbation expansions and recovering from them the "exact" results paved the way for recent efforts [139] to use renormalization group to sum perturbation expansions for physical quantities. [29] This Section contains an outline of these ideas.

Before discussing them let me briefly recall the situation in QED, as the problem of summing perturbation expansions is *not* a special feature of QCD, but exists QED as well. Moreover, as we shall see, the situation in QED is in *principle* even worse than in QCD. On the other hand in *practice* the divergence of QED perturbation expansions is much less disturbing for reasons related to the fact that the magnitude of QED couplant $\alpha = \frac{1}{137}$ is *very small*. So taking into account that the few lowest order coefficients of expansions in powers of α, calculated so far, are numbers of the order of unity, the **apparent** convergence of these lowest order approximants is very fast. Secondly, despite the fact that the expansions are expected to diverge, this should happen only at *very large* orders, starting roughly at the order $1/\alpha \doteq 137$. Both of these circumstances make the divergence problem in QED of essentially only academic interest. The situation changes at very short distances, where the QED renormalized couplant $\alpha(\mu)$ blows to infinity, but there QED cannot be trusted anyway. A couple of lowest order terms, which behave orderly, can thus be understood as an **effective theory** appropriate to distances large compared to $1/\Lambda_{\text{QED}}$.

In QCD the situation is quite different, essentially because the value of QCD couplant α_s, is much *larger*. For instance, at scales appropriate for the process (6.1) at the CMS energy between 10 and 40 GeV, where most of experiments have been performed, $\alpha_s \in (0.3, 0.6)$. Recalling the values of the first two known [30] coefficients r_k in the expansion (8.133), $r_1 = 1.41$, $r_2 = -12.7$, we find that for the typical value $\alpha_s/\pi = 0.1$ the relative magnitudes of the first three known terms in this expansion

$$r(Q) = a(1 + r_1 a + r_2 a^2 + \cdots) = a(1 + 0.14 + (-0.13) + \cdots) \tag{8.157}$$

indicate that in QCD there are *practical* problems with the convergence of perturbative expansions already at fairly low orders, making the task of summing these expansions of utmost phenomenological importance. In addition, due to the selfinteractions of gluons, the number of Feynman diagrams at each order of perturbation theory is much larger than in QED, which leads in general to larger coefficients r_k.

The current state of our understanding of convergence properties of perturbation expansions in QED and QCD is roughly the following. The original arguments of Dyson (as well as some more recent ones) provide strong evidence that scattering amplitudes of physical processes, considered as functions of renormalized couplant (α in QED or α_s in QCD) in the complex plane have a cut along the negative semiaxis from some negative value up to zero. This excludes the possibility, that they possess *convergent* Taylor expansion. However, it does not imply that perturbation expansions *have* to diverge, as the nonanalytic terms may be simply *invisible* in perturbation theory. Nevertheless, it is generally expected that these expansions are, indeed, divergent, if considered in fixed renormalization scheme and that in both theories they are only **asymptotic** expansions of the exact results! Before proceeding further, let me recall the meaning of asymptotic series. [31]

Imagine a function $F(x)$ of a complex variable x, which at some point x_0 has the property that all *right* derivatives $F^{(n)} \equiv dF(x)/dx$ at that point exist and are finite. Then the *formal* series

$$\sum_{n=0}^{\infty} r_n(x - x_0)^n \tag{8.158}$$

is said to be **asymptotic** to $F(x)$ at $x = x_0$ from right if the following limit holds for *any* finite N:

$$\lim_{x \to x_0^+} \left| \frac{F(x) - \sum_{k=0}^{N} r_k(x - x_0)^k}{r_N(x - x_0)^N} \right| = 0. \tag{8.159}$$

---

[28] Unfortunately, in QCD there are no mathematically well–defined equations, like the Schrödinger equation in quantum mechanics, which could be solved *exactly* and this exact solution then compared to any approximate one, for instance that from perturbation theory. So speaking about "exact" results in QCD we speak about something we actually don't know.

[29] From now on I shall omit, when speaking about the summation of QCD perturbation expansions, the quotation marks.

[30] These numbers correspond to the so called $\overline{\text{MS}}$ renormalization scheme. See discussion later in this section.

[31] The following definition of asymptotic series, though not the most general one, is quite sufficient for our purposes.

For any function $F(x)$ which has a convergent Taylor series, this series represents also the asymptotic expansion of $F(x)$. The condition (8.159) is, however, more general and expresses only the fact that the "remnant" of the partial sum $\sum_{k=0}^{N} r_k x^k$ (i.e. the numerator in the above relation) is negligible compared to the last included term *in the limit* $x \to x_0^+$. This by itself does not guarantee the convergence of the series (8.158) and tells nothing about the function $F(x)$ in other points than x_0 and from other directions than positive semiaxis. In (8.159) first the limit $x \to x_0^+$ for any finite partial sum from $k=0$ to $k=N$ is taken and only then can N be sent to infinity. For series convergent at x_0 the numerator *itself* vanishes in the limit $N \to \infty$ at this point. For divergent asymptotic series it doesn't but is small with respect to $r_N x^N$, which, however, for any $x \neq x_0$ diverges as $N \to \infty$.

Each function $F(x)$ has at most one series which is asymptotic to it at a given point and from a given direction. The opposite is, however, *not* true and there is an infinite number of functions sharing the *same* asymptotic series. All these functions have the *same* value at $x = x_0$, but they may be *arbitrarily* far apart away from x_0. As an example consider the following sets of functions ($F(x)$ is assumed to be a function with an asymptotic expansion at $x_0 = 0^+$)

$$f(x) = F(x) + \exp(-1/x)(1 + \cdots), \tag{8.160}$$

$$f(x) = F(x) + \exp(-1/x^2)(1 + \cdots), \tag{8.161}$$

$$f(x) = F(x) + \exp\left(\frac{1 - \exp(1/x)}{x}\right)\ln x, \tag{8.162}$$

which have all the right derivatives at $x = 0^+$ vanishing and so share together with $F(x)$ the same asymptotic expansion. The terms added to $F(x)$ are **invisible** in perturbation theory and provide examples of **nonperturbative** contributions to the exact, but unknown, results. In all three above sets of functions the behavior at the origin $x = 0$ is a complicated essential singularity, which is invisible if we approach it from right. The third example, moreover, has the cut along the negative x-axis, proportional to $\exp(-1/|x|)$, exactly of the kind implied by Dyson's arguments, but $F(x)$ may still have a convergent Taylor series!

The previous examples show that the task of summing divergent series is **inherently ambiguous**. Notice that even for convergent series there is an infinite set of functions which have this series as asymptotic, and which thus share the same value at the point of asymptoticity, but are arbitrarily different at $x \neq x_0$. Although in this latter case the conventional sum of the convergent series plays an exceptional role among them, we must state clearly and unambiguously what our task is. The example of the third set of functions above shows that the conventional sum of a convergent series needn't always be the correct choice. In the case of divergent series, no such preferred sum does exist and all possible sums should be considered at the same footing. In distinguishing between the **summability** of a given QCD perturbation expansion and **recoverability** from this series of the exact result, we should keep in mind that while the first task concerns exclusively the series itself, the latter one involves also the exact (but unknown) function and cannot thus be solved without some *additional information* about it. Dyson's arguments prevent ambiguous recoverability of exact results from perturbation expansions, but tells us nothing about the *convergence* of these expansions themselves.

Various algorithms have been developed to sum divergent series, prominent among them the familiar Borel technique. The basic information needed for any attempt to sum a given asymptotic series concerns the large order behavior of its coefficients. Consider the following type of the large order behavior of the expansion coefficients r_k

$$r_k \propto A(\delta)^k k^l k!. \tag{8.163}$$

For $\delta < 0$ the corresponding series is **Borel summable**, for $\delta > 0$ it is not. The Borel summation technique is based on the following identity for the conventional sum of a *convergent* series with coefficients r_k

$$\sum_{k=0}^{\infty} r_k a^{k+1} = \int_0^\infty \exp(-u/a) \underbrace{\left(\sum_{k=0}^{\infty} \frac{r_k}{k!} u^k\right)}_{\text{Borel transform}} \mathrm{d}u. \tag{8.164}$$

The basic idea of this technique is to use (8.164) *as a definition* of the **Borel sum**, denoted BS, of the divergent series on its l.h.s.. This generalization is nontrivial due to the fact that even if the coefficients r_k define divergent series, those of the Borel transform, $r_k/k!$, may define *convergent* one. Provided the integral

in (8.164) is **unique**, (8.164) provides a natural definition of the sum. Consider two examples

$$\mathrm{BS}\left(\sum_{k=0}^{\infty}(-1)^k k! a^{k+1}\right) = \int_0^{\infty} \frac{\exp(-u/a)}{1+u}\mathrm{d}u, \tag{8.165}$$

$$\mathrm{BS}\left(\sum_{k=0}^{\infty} k! a^{k+1}\right) = \int_0^{\infty} \frac{\exp(-u/a)}{1-u}\mathrm{d}u. \tag{8.166}$$

In the first case the Borel sum exists for all $a > 0$ and can be intuitively understood as a consequence of subtle cancelations between large numbers of opposite sign. In the second example, typical for both QED and QCD, Borel sum *doesn't exist* due to the pole in the integrand of (8.166) at $u = 1$. This single pole is no obstacle to *defining* the integral in (8.166), but it implies its **nonuniqueness**. A standard, though by far not the only possible, way of defining it is to use the **principle value prescription**, defined as

$$\mathrm{PV}\int_0^{\infty} \frac{\exp(-u/a)}{1-u}\mathrm{d}u = \lim_{\epsilon \to 0}\left[\int_0^{1-\epsilon} \frac{\exp(-u/a)}{1-u}\mathrm{d}u + \int_{1+\epsilon}^{\infty} \frac{\exp(-u/a)}{1-u}\mathrm{d}u\right]. \tag{8.167}$$

There are many other plausible definitions, like nonsymmetrical integration, circumventing the singularity in the complex plane etc.. All of them share the same perturbation expansion, but they may differ arbitrarily in the "invisible" (in perturbative expansions) terms. In other words, the various possible sums differ in the way various **nonperturbative** contributions are included. If we knew in detail the analytical properties of the exact result in the vicinity of $\alpha_s = 0$, we could use this information in combination with perturbation expansion, to uniquely reconstruct it. Unfortunately, this information is not available and we have to rely on a few lowest order perturbative coefficients and some semiquantitative knowledge of certain nonperturbative effects. In renormalized field theories like QED and QCD, the situation is further complicated by the RS dependence of

- the couplant (α or α_s),

- the free coefficients $c_i, i \geq 2$ of the corresponding β-function,

- the coefficients r_k of the perturbation expansions like (8.133).

It is crucial to realize that in renormalized field theories perturbation theory **doesn't yield one expansion**, but the whole infinite set of them, each corresponding to a particular renormalization scheme! These expansions differ, but are related through **consistency conditions**. All of these expansions are in principle on the same footing and thus summing perturbation expansion in QCD actually means working with this infinite set of expansions.

We should keep in mind that the perturbation expansion of either the β-function or for $r(Q)$, but *not both*, can be made convergent just by the appropriate choice of the RS, and so any statement concerning the divergence of perturbation expansions must be accompanied by the specification of the latter. There are two opposite possibilities in this respect:

- Set $c_i = 0, i \geq 2$. In this so called 't Hooft RS [141] the β-function is the third order polynomial, the couplant thus well-defined and all the problems concern the convergence properties of perturbation expansions of physical quantities, like (8.133).

- Set $r_k; k \geq 1$, defining the so called "effective charges" (EC) approach of ref. [143], in which $r(Q) = a^{\mathrm{EC}}$ by definition. In this case all the problems concern the definition of the expansion parameter $a^{\mathrm{EC}}(\mu)$, as now $\beta^{\mathrm{EC}}(a)$ is given by a divergent expansion in a^{EC}.

In RS which are *calculationally* simple both of the discussed expansion are (presumably) divergent, making the situation even less clear. Conceptually it is certainly better to select one of the RS in which the β–function is an analytic function of a and concentrate on the divergence properties of perturbation expansions of physical quantities.

There is, however, another possible strategy how to address the problem of the divergence of perturbation expansions. This approach emphasizes the fact that our basic task is to construct a sequence of finite order approximants

$$r^N(Q, \mathrm{RS}(N)) = \sum_{k=0}^{N-1} r_k(\mathrm{RS}(N))a^{k+1}(\mathrm{RS}(N)) \tag{8.168}$$

and investigate the limit

$$r^\infty(Q) \equiv \lim_{n\to\infty} r^{(N)}(Q, \mathrm{RS}(N)).$$ (8.169)

In the above expression I have explicitly written out the possible dependence of the chosen renormalization scheme RS=$\{\mu, c_i\}$ on the integer N in both the coefficients r_K and the couplant a. There is nothing wrong in selecting at each N a different RS and in several popular approaches [139,143] the selected RS does, indeed, depend on N. But can we gain anything by playing with the RS dependence if we know that the full sum should be independent of it? The answer, maybe surprising, is positive. The point is that the consistency conditions discussed in subsection 6.3.3 are formal in the sense that they guarantee unique result under two circumstances

1. perturbation series in all considered RS are *convergent*,

2. the RS is *fixed* when we carry out the limit (8.169).

If we drop the second restriction, we find that the limit (8.169) can, by suitable choice RS(N), be *finite* even for *divergent* series, but depends on details of the limit in (8.169). I shall demonstrate these general statements on two simple examples of perturbation series (8.165) and (8.166). Imagine we live in an imaginary QCD world where, for simplicity, $c = 0$ and assume that the expansions hold in some RS with the couplant a_0. Let us work in the RS where $c_i = 0, i \geq 2$ so that the scale μ is the only remaining free parameter and the couplant has a very simple dependence (8.136) on it. Under these simplifying circumstances the recurrence relations (8.143) have a simple solution

$$r_k(\tau) = \sum_{j=0}^{k} \tau^j r_{k-j}(0) \binom{k}{j}; \quad a_0 = \frac{1}{\tau_0}.$$ (8.170)

Let us now assume the following dependence of τ on the order N of the partial sum in (8.168):

$$\tau(N) \equiv \tau_0 + \chi N = \frac{1}{a_0} + \chi N,$$ (8.171)

where χ is an *arbitrary* real parameter, characterizing the "speed" of the change of RS with N. Substituting (8.171) into (8.170), evaluating the finite sum in (8.168) and carrying out the limit in (8.169) we find [144] in the general case

$$r(\chi, a_0) = \int_0^{1/\chi} \exp(-u/a_0) \sum_{k=0}^{\infty} \frac{r_k(\tau_0)}{k!} u^k \mathrm{d}u,$$ (8.172)

which for the "toy" series in (8.165) and (8.166) reduces to the following compact expressions

$$r(\chi, a_0) \equiv \int_0^{1/\chi} \frac{\exp(-u/a)}{1 \pm u} \mathrm{d}u.$$ (8.173)

Obviously, the Borel sum corresponds to the trivial case $\chi = 0$. For the sign alternating series the shape of $r(\chi, a)$ as a function of the couplant a and for a few values of χ is displayed in Fig. 8.7a. In Fig. 8.7b the same is done for the series with coefficients $r_k = k!$, i.e. of the kind expected in QCD and QED. Three important conclusions can be drawn from these figures:

- The sums defined by the limiting procedures *depend* on the value of χ.

- In the case of Borel summable series all values of χ are allowed and the Borel sum is just a limit

$$\mathrm{BS} = \lim_{\chi\to0} r(\chi)$$ (8.174)

  The Borel sum as well as all $r(\chi)$ are **monotonously** increasing functions of the couplant a_0.

- For factorially divergent nonalternating, and thus Borel nonsummable series (8.166), only $\chi > 1$ are allowed for (8.173) to make sense. As in the preceding case $r(\chi)$ are monotonous functions of a_0. This latter property of the sum of series with *positive*, factorially increasing coefficients is not respected by the Principle value prescription of the integral in (8.173), which is displayed for comparison in Fig. 8.7b as well, and which for a above roughly 2.7 becomes even negative!

161
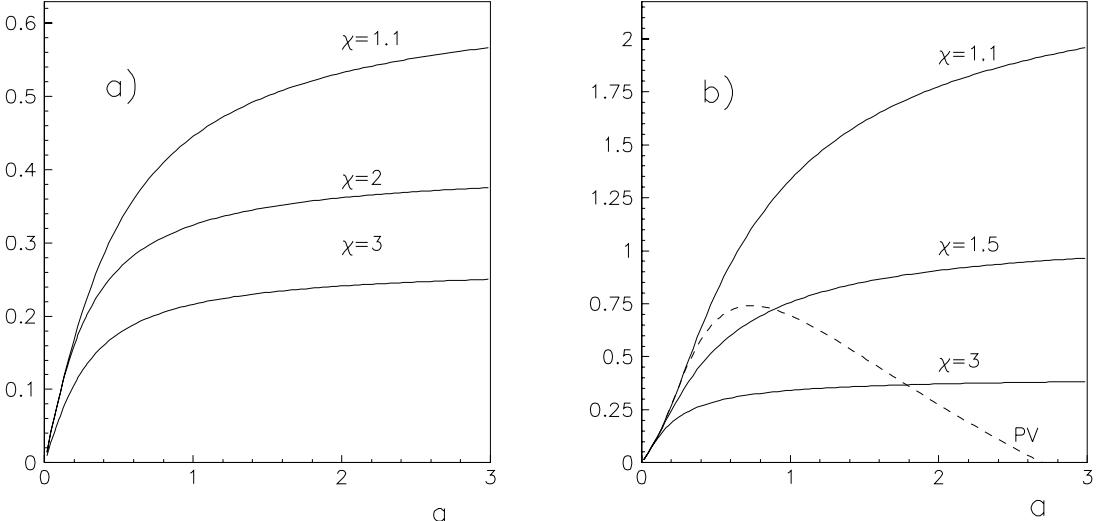
Figure 8.7: The dependence of $r(\chi, a_0)$ on a_0 for sign alternating (a), and sign constant (b), factorially divergent series.

The above exercise with the toy examples and the introduction of the specific N-dependence of the selected RS(N) should merely be taken as an indication that there is no principal problem to get a finite limit for the sequence of finite order approximants even in the case of badly divergent series. For this strategy to be of real phenomenological relevance two questions have to be answered:

1. What is the large order behavior of the coefficients r_k in some well–defined RS (like the 't Hooft RS)? This information is necessary to see whether a given summation method works in realistic QCD. This question can in principle be answered within perturbation theory itself, but in practice the answer is still eluding. Extensive discussion of this point is given in [145].

2. Which χ to choose. Here information from **outside** of perturbation theory is indispensable. Let me emphasize that the above construction using the explicit dependence (8.171) is just an example and one can imagine many other ways of relating the summation of perturbation expansions to nonperturbative effects.

8.5 Exercises

1. Show that the second term in (8.11) does, indeed, vanish.

2. Taking into account that for μ of the order of electron mass $\alpha_R(\mu/\Lambda) = 1/137$, estimate the value of Λ in QED as given in (8.39).

3. Discuss the behavior of the exact solution to the NLO equation (8.138) in the limit $\mu \to \Lambda$. Where does it blow up? How does it differ from the behavior of the LO solution?

4. Show that the simultaneous changes of $a(\mu)$ and r_1 with μ cancel to the order a^2 in (8.133).

5. Justify (8.61).

6. Prove (8.62).

7. Prove (8.141) starting from the basic relation (8.134).

8. Extract the value of Λ corresponding to the RS where $r_1 = 1.411$, using the NLO approximation to (8.133) and the fact that for $Q = 35$ GeV its experimental value equals 0.05.

9. Express the IR fixed point $a^*(c_2)$ as a function of c_2 taking into account the first three terms in the expansion of $\beta(a)$.

10. Derive (8.170) and prove (8.172).

11. Discuss the results of the limiting procedure (8.173) for geometric series with $r_k = 1$.

Chapter 9

Mass singularities and jets

In the previous Chapter we discussed the problems resulting from singular behavior of QED and QCD at *short* distances and outlined the remedy to these problems: the renormalization procedure. In this Chapter we shall see that gauge theories with massless gauge particles [1] have problems at *long* distances as well. The way the associated infinities are treated is principally different from the treatment of ultraviolet infinities and naturally leads to the concept of **jet**. We shall see how this concept emerges from theoretical considerations on the level of partons and how it naturally reflects the restrictions imposed on of experimental observations by their finite resolution power. We shall also address the important and nontrivial question of why and how the concept of jet, developed theoretically on the level of *partons*, has anything to do with experimentally defined jets of *hadrons*. This connection is crucial for many applications of QCD to hard scattering processes. All this will be illustrated on the theoretically as well as experimentally simplest example of three jet production in e^+e^- annihilations.

9.1 *Mass singularities in perturbation theory

In quantum field theories like QED and QCD the evaluation of loops and/or integrals over momenta of final state particles frequently leads to singularities due to the vanishing mass of either the gauge bosons or fermions. These **mass singularities** are related to the behavior of gauge theory at *large* distances, in contrast to the UV singularities discussed in the previous section, which came from *short* ones. The treatment of these singularities doesn't involve redefinition of electric or color charges, but requires subtle considerations concerning the definition of **physically distinguishable** states. Let us illustrate the basic idea on the following simple process:

$$e^+e^- \to q\bar{q}G, \tag{9.1}$$

where a single gluon is radiated from the $q\bar{q}$ pair produced in e^+e^- annihilations according to the Feynman diagram in Fig. 8.4b. The corresponding spin and color averaged cross section depends on two of the three dimensionless fractions x_i ($E_i, i = 1, 2, 3$ are the CMS energies of the outgoing quark, antiquark and gluon respectively)

$$0 \le x_i \equiv \frac{2E_i}{\sqrt{s}} \le 1; \quad \sqrt{s} = E_1 + E_2 + E_3 \Rightarrow x_1 + x_2 + x_3 = 2, \tag{9.2}$$

or, equivalently, on the three scaled invariant masses y_{ij} of pairs of two partons, related to the fractions x_i,

$$y_{12} \equiv \frac{(p_1 + p_2)^2}{s} = 1 - x_3; \quad y_{13} \equiv \frac{(p_1 + p_3)^2}{s} = 1 - x_2; \quad y_{23} \equiv \frac{(p_2 + p_3)^2}{s} = 1 - x_1, \tag{9.3}$$

for which the kinematical constraint in (9.2) implies $y_{12} + y_{13} + y_{23} = 1$. For *massless* quarks and gluons

$$\frac{d\sigma}{dx_1 dx_2} = \sigma_0 \frac{\alpha_s}{2\pi} C_F \frac{x_1^2 + x_2^2}{(1-x_1)(1-x_2)} \Leftrightarrow \frac{d\sigma}{dy_{13} dy_{23}} = \sigma_0 \frac{\alpha_s}{2\pi} C_F \frac{(1-y_{23})^2 + (1-y_{13})^2}{y_{13} y_{23}}, \tag{9.4}$$

where $C_F = 4/3$ and $\sigma_0 = 12\pi\alpha^2 e_q^2/3s$ denotes the familiar Born cross–section for the production of a $q\bar{q}$ pair with electric charge e_q and three colors. The singularity at $x_1 = 1$ occurs for $y_{23} = 0$ and corresponds

---

[1]That is gauge theories *without* spontaneous symmetry breaking.
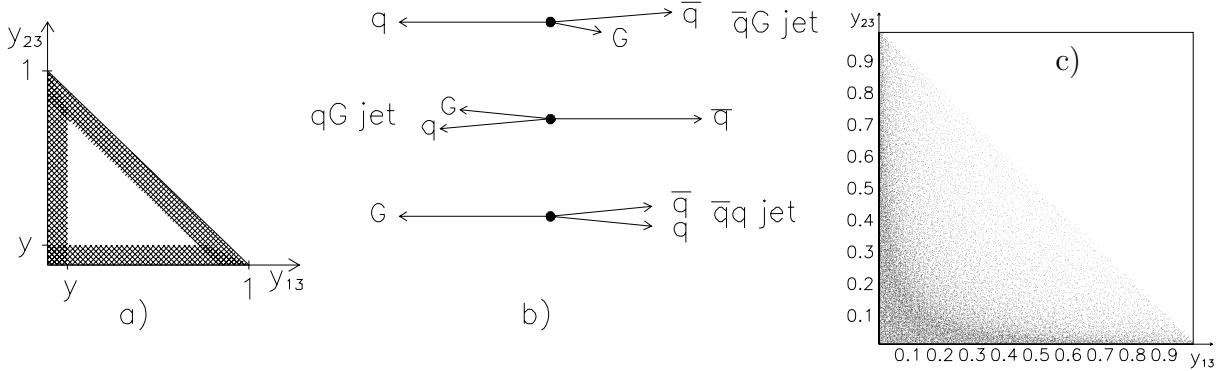
Figure 9.1: Kinematics of the $q\overline{q}G$ final state (a), typical configurations of the momenta corresponding to qG or $\overline{q}$G jets (b) and scatterplot of events generated according to the cross–section (9.4).

to the configuration where the radiated gluon is parallel to the outgoing antiquark, while that at $x_2 = 1$ occurs when this gluon is parallel to the outgoing quark. The double singularity at $x_1 = x_2 = 1$ (i.e., $x_3 = 0$) corresponds to the limiting case when the energy of the emitted gluon vanishes. Introducing Q^2 to denote the invariant mass of the combination of an antiquark and gluon, $Q^2 = s(1 - x_1)$, we can write in the collinear limit $x_1 \to 1$, which is equivalent to $Q^2/s \to 0$,

$$\frac{d\sigma}{dQ^2 dx_2} = \frac{1}{s}\frac{d\sigma}{dx_1 dx_2} = \sigma_0 \frac{\alpha_s}{2\pi}\frac{1}{Q^2}C_F \frac{x_1^2 + x_2^2}{1 - x_2} \xrightarrow[x_1 \to 1]{} \sigma_0 \frac{\alpha_s}{2\pi}\frac{1}{Q^2}C_F \frac{1 + x_2^2}{1 - x_2} = \sigma_0 \frac{\alpha_s}{2\pi}\frac{1}{Q^2}P_{qq}^{(0)}(x_2). \tag{9.5}$$

Note that in the collinear limit $x_1 \to 1$ the variable x_2 represents a fraction of the momentum of the original antiquark carried by it after the emission of the gluon. Interchanging $x_1 \Leftrightarrow x_2$ analogous approximation can be written for the region where the gluon is (almost) colinear with the outgoing quark. The so called splitting function $P_{qq}^{(0)}(x)$, which describes the emission of a gluon from a quark or an antiquark in the collinear region (with minor modification discussed in Section 9.2.2), does not depend on the process in which the quark or antiquark are produced. As shown in the next Section, this and related splitting functions play a central role in the formalism of perturbative QCD.

For the three parton final state the available kinematical region in variables y_{ij} corresponds to the interior of the triangle in Fig. 9.1a, where the bands define the three regions where one of the y_{ij} is small. Typical configurations corresponding to these three regions are shown in Fig. 9.1b. Note that in the region of small y_{12}, i.e. when the $q\overline{q}$ pair is close in phase space, the cross–section is *not* singular. This is illustrated in Fig. 9.1c, where a scatterplot corresponding to the cross–section (9.4) is displayed.

The infinite cross-section for the production of an unphysical state is clearly meaningless, but even a finite, but arbitrarily large probability for the production of real soft gluons, causes problems. The first step toward their treatment is, as for the UV divergences, some kind of *regularization*. This can be done in a variety of ways. For instance, the IR singularities can be avoided by assigning a nonzero mass m_g to gluons. Although this is particularly simple and physically transparent method of IR regularization it has a serious drawback as it violates the sacred principle of gauge invariance. We can use it, but must check that after removing the regularization the results are gauge invariant. Moreover, as evaluation of Feynman diagrams with massive particles becomes enormously complicated already at two loops, this method is impractical at higher orders. It turns out that the dimensional regularization (which maintains gauge invariance in any dimension) can be used to regularize IR singularities as well (in this case, however, for $\varepsilon > 0$!) and is calculationally superior to any other regularization method. We shall use it as well, but first demonstrate the origin and treatment of the mass singularities using the gluon mass, or better, the dimensionless parameter $\beta \equiv m_g^2/Q^2$ as the IR regulator. For general $\beta \neq 0$ the cross-section for the process (9.1) reads [149]

$$\frac{d\sigma}{dx_1 dx_2} = \sigma_0 \frac{2\alpha_s}{3\pi}\frac{1}{(1 - x_1)(1 - x_2)}\left[x_1^2 + x_2^2 + \beta\left(2(x_1 + x_2) - \frac{(1 - x_1)^2 + (1 - x_2)^2}{(1 - x_1)(1 - x_2)}\right) + 2\beta^2\right]. \tag{9.6}$$

The singularities, both infrared and parallel, are screened off due to the fact that also the kinematical bounds on x_1, x_2 are now functions of β

$$0 \le x_1 \le 1 - \beta; \quad 1 - \beta - x_1 \le x_2 \le 1 - \frac{\beta}{1 - x_1}. \tag{9.7}$$

165

The above expression is symmetric in x_1, x_2 and can easily be integrated over one of these fractions giving

$$\frac{d\sigma}{dx} = \sigma_0 \frac{\alpha_s}{2\pi} \frac{4}{3} \left[\frac{1+x^2}{1-x} \ln \frac{x(1-x)}{\beta} - \frac{3}{2} \frac{1}{1-x} + \frac{1}{2}(x+1) + \beta \frac{2-x}{(1-x)^2} + \frac{1}{2}\beta^2 \frac{1}{(1-x)^3} \right], \qquad (9.8)$$

where $0 \le x \le 1 - \beta$ and only terms which contribute in the limit $\beta \to 0$ were retained. Contrary to the double differential cross–section (9.6), which has a meaning even for $\beta = 0$, (9.8) blows to infinity for $\beta \to 0$. We shall now demonstrate how the inclusion of virtual gluon emissions cures this problem.

The point is that instead of radiating a *real* gluon, the quark can radiate a *virtual* one, which will eventually recombine, as shown in Fig. 8.4d, with the parent quark, or with an accompanying antiquark (Fig. 8.4c), producing a real $q\overline{q}$ pair. In both cases the virtual gluon can be arbitrarily close to its mass-shell and thus propagate to arbitrarily large distances. Integration over the loops of Fig. 8.4c,d therefore also leads to mass singularities. It is evident that the interference term between the lowest order QPM diagram in Fig. 8.4a and the loop diagrams in Fig. 8.4c,d is of the same order α_s as the square of the diagram in Fig. 8.4b, describing the real gluon emission

$$\sigma_{\text{virt}} = \sigma_0 \frac{\alpha_s}{2\pi} \frac{4}{3} \left[-\ln^2 \beta - 3\ln \beta + \frac{\pi^2}{3} - \frac{7}{2} \right] + \mathcal{O}(\alpha_s^2). \qquad (9.9)$$

This term contributes obviously only for $x = 1$ and we can write

$$\frac{d\sigma_{\text{virt}}}{dx} = \sigma_0 \frac{\alpha_s}{2\pi} \frac{4}{3} \left[-\ln^2 \beta - 3\ln \beta + \frac{\pi^2}{3} - \frac{7}{2} \right] \delta(1-x). \qquad (9.10)$$

The crucial observation is now the following: as the gluon with zero energy is the same as no gluon at all, we should, for massless gluons, add the *negative* infinity from virtual correction to the cross section $\sigma(e^+e^- \to q\overline{q})$ to the *positive* infinity coming from the integral over the real gluon emission. To carry out this cancelation in a mathematically well-defined way, the regulator $\beta \ne 0$ was introduced. To work out this sum, several rather technical steps have to be carried out:

1. Although the last two terms in (9.8) are proportional to β and β^2, they *cannot* be neglected in the limit $\beta \to 0$. It can easily be shown that they go to $\frac{5}{4}\delta(1-x)$.

2. Realizing that
$$\delta(1-x) = \delta(1-x-\beta) + \beta \delta'(1-x) \qquad (9.11)$$

and taking into account that in (9.10) the dependence on β comes only in powers of $\ln \beta$, we can replace $\delta(1-x)$ in (9.10) with $\delta(1-x-\beta)$.

3. We now recall the definition of the so called "+" distribution

$$[f(x)]_+ \equiv \lim_{\beta \to 0} \left(f(x)\theta(1-x-\beta) - \delta(1-x-\beta) \int_0^{1-\beta} f(y) dy \right), \qquad (9.12)$$

where $\theta(x)$ and $\delta(x)$ are the usual step and Dirac δ-functions. Note that for functions singular at $x = 1$ the second term subtracts at $x = 1$ the whole (possibly divergent) integral over this interval. We shall need the following explicit results

$$\left[\frac{1}{1-x} \right]_+ = \frac{1}{1-x}\theta(1-x-\beta) + \ln\beta \delta(1-x-\beta), \qquad (9.13)$$

$$\left[\frac{1+x^2}{1-x} \right]_+ = \frac{1+x^2}{[1-x]_+} + \frac{3}{2}\delta(1-x), \qquad (9.14)$$

$$\frac{1+x^2}{[1-x]_+} = \frac{1+x^2}{1-x}\theta(1-x-\beta) + 2\ln\beta\delta(1-x-\beta), \qquad (9.15)$$

$$(1+x^2)\left[\frac{\ln(1-x)}{1-x} \right]_+ = \frac{(1+x^2)\ln(1-x)}{1-x}\theta(1-x-\beta) + 2\ln^2\beta\delta(1-x-\beta). \qquad (9.16)$$

166

4. Having all the necessary formulae at hand we can now regroup the individual terms in the sum of (9.8) and (9.10) in such a way that the limit $\beta \to 0$ can be carried out and we get

$$\frac{d\sigma}{dx} \equiv \frac{d\sigma_{\text{real}}}{dx} + \frac{d\sigma_{\text{virt}}}{dx} = \sigma_0 \frac{\alpha_s}{2\pi} \frac{4}{3} \left\{ \ln\frac{1}{\beta} \left[\frac{1+x^2}{1-x}\right]_+ + \right. \tag{9.17}$$

$$\underbrace{\left. (1+x^2)\left[\frac{\ln(1-x)}{1-x}\right]_+ + \frac{1+x^2}{1-x}\ln x - \frac{3}{2}\frac{1}{[1-x]_+} + \frac{1+x}{2} + \left(\frac{\pi^2}{3} - \frac{9}{4}\right)\delta(1-x) \right\}}_{\text{constant terms}} .$$

5. Finally, integrating (9.17) over x we get

$$\sigma_{\text{tot}} \equiv \sigma_{\text{real}} + \sigma_{\text{virt}} = \sigma_0 \frac{\alpha_s}{\pi}. \tag{9.18}$$

We could get the last result also by adding the integral of (9.6), equal to

$$\sigma_{\text{real}}(Q) = \sigma_0 \frac{\alpha_s}{2\pi} \frac{4}{3}\left[\ln^2\beta + 3\ln\beta - \frac{\pi^2}{3} + 5\right], \tag{9.19}$$

to (9.9) and sending $\beta \to 0$. Note that the terms proportional to $\ln\beta$ come from parallel as well as IR region of gluon momenta, while the double logarithm $\ln^2\beta$ comes entirely from the IR region. In both cases the virtuality of the quark or antiquark emitting the gluon is proportional to m_g^2 and thus vanishes with it. Physically it means that virtual quarks or antiquarks propagate to distances of the order $1/m_g$ before radiating the gluon and going to their mass shell. Remarkably, if we add the real and virtual gluon emission cross-sections we get a *finite* result for the total cross–section (9.18) and a well–defined expression for the inclusive spectrum (9.17).

9.2 *Kinoshita–Lee–Nauenberg theorem

This example of cancelation of soft and parallel singularities is not merely a coincidence, but is a consequence of the **Kinoshita–Lee–Nauenberg theorem**:

Consider the general scattering process $A \to B$ between states A, B of massive particles. It may happen that the scattering matrix $|S_{AB}|$ has mass singularities but if we sum the squares of scattering amplitudes

$$\sum_{D(A),D(B)} \left|S_{D(A)D(B)}\right|^2 \tag{9.20}$$

over the sets $D(A), D(B)$ of states degenerate with A,B, the sum has no mass singularities, i.e. is finite even for massless particles.

By "degenerate" states of particles we mean all states of particles which have the same conserving quantum numbers as well as the same total four–momentum. The crucial aspect of this concept is the fact that degenerate states may have different number and composition of particles. For instance, in QED, the states of an electron and an electron accompanied by a photon with zero energy are obviously degenerate. However, even a 10 GeV electron and a *parallel* pair of a 5 GeV electron with a 5 GeV photon would be degenerate in the limit of vanishing electron mass. In real QED electrons are massive and thus the only real singularity comes from the soft photon region. Nevertheless even in these circumstances the KLN theorem may be useful, when the experimental resolution δ of energy or transverse momentum measurements (related to gluon emission angle) is poor in comparison with the electron mass m_e. In such a situation we can safely set $m_e = 0$ and invoke the KLN theorem to calculate the cross–sections for the particular process with or without additional photon with energy (or transverse momentum) *bigger* than the resolution δ.

As will be discussed in detail in the next Chapter, the KLN theorem is even more important in QCD, where it provides an indispensable ingredient of the QCD-improved QPM. Here I want to point out one particular consequence of this theorem in DIS and compare it to the case of e^+e^- annihilations. Consider the lowest order QCD corrections to the QPM in DIS, i.e. the one gluon emission from incoming or outgoing quarks as shown in Fig. 9.2a. Similarly to the gluon emission in e^+e^- annihilations, the integration over
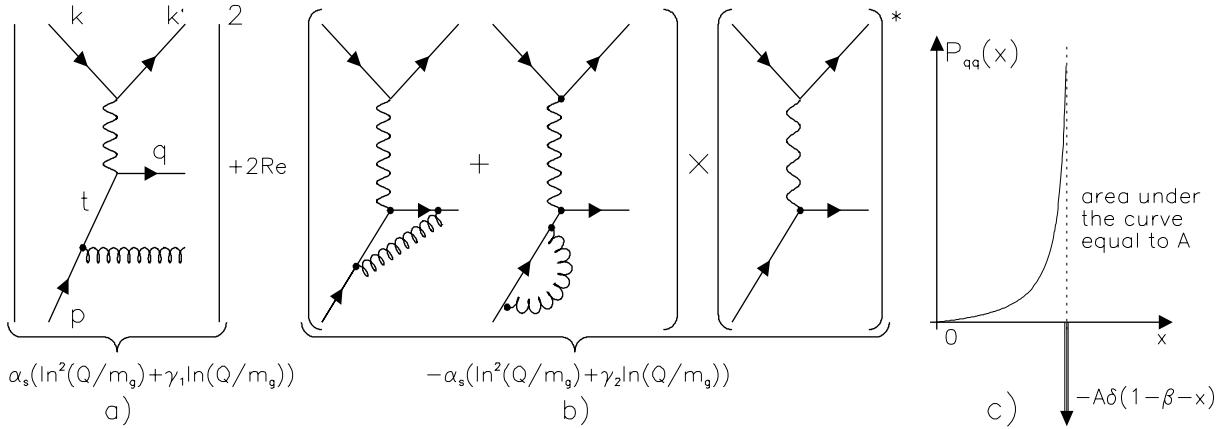
Figure 9.2: The application of KLN theorem to DIS (a-b) and the shape of the resulting branching function $P_{qq}(z)$ (c). t in a) denotes virtuality of the intermediate quark. Note that the double logarithms, coming from the soft gluon emission, cancel in the sum of real and virtual contributions, while the single logarithms do not and give the term $\ln(Q^2/m_g^2) = \ln(1/\beta)$ in (9.17).

the energies and angles of the emitted gluon leads to soft and parallel singularities. If regulated by the fictitious gluon mass m_g these singularities become again proportional to $\ln \beta, \ln^2 \beta$ as in (9.19). Keeping $x \equiv -q^2/2pq$ and $y \equiv (qp)/(kp)$ fixed, but integrating over all other variables describing the quark level subprocess e + q → e + q + g, we find (for derivation see Section 7.2) that diagram in Fig. 9.2a contributes

$$\frac{\mathrm{d}\sigma^{\mathrm{real}}}{\mathrm{d}x\mathrm{d}y} = \left[\frac{4\pi\alpha^2 xs}{Q^4} \frac{1 + (1-y)^2}{2} \right] \frac{\alpha_s}{\pi} \left(\frac{4}{3} \frac{1+x^2}{1-x} \ln \frac{1}{\beta} + f(x,\beta) \right), \tag{9.21}$$

where $s = (k+p)^2$, x can be interpreted as the fraction of incoming quark momentum, carried by it after the gluon emission and the function $f(x,\beta)$ has a *finite* limit for $\beta \to 0$. The singular term proportional to $\ln 1/\beta$, results, as in (9.17), from integration over the angle of the emitted gluon. The interference term between the diagrams in Fig. 9.2b contributes again a negative divergent term

$$\frac{\mathrm{d}\sigma^{\mathrm{virt}}}{\mathrm{d}x\mathrm{d}y} = \left[\frac{4\pi\alpha^2 xs}{Q^4} \frac{1 + (1-y)^2}{2} \right] \frac{4\alpha_s}{3\pi} \left(-\ln^2 \beta - 3\ln\beta - \frac{7}{2} - \frac{2\pi^2}{3} \right) \delta(1-x) \tag{9.22}$$

where the δ-function $\delta(1-x)$ expresses the fact that virtual gluon emission contributes only for $x = 1$. Adding the above two contributions involves the same kind of manipulations as for the process (9.1) and leads to the following result

$$\frac{\mathrm{d}\sigma_{\mathrm{real}}}{\mathrm{d}x\mathrm{d}y} + \frac{\mathrm{d}\sigma_{\mathrm{virt}}}{\mathrm{d}x\mathrm{d}y} = \left[\frac{4\pi\alpha^2 xs}{Q^4} \frac{1 + (1-y)^2}{2} \right] \left(\frac{\alpha_s}{\pi} \underbrace{\frac{4}{3} \left[\frac{1+x^2}{1-x} \right]_+}_{P_{qq}^{(0)}(x)} \ln \frac{1}{\beta} + \tilde{f}(x,\beta) \right). \tag{9.23}$$

The shape of this $P_{qq}^{(0)}(z)$ is sketched in Fig. 9.2c. This so called Altarelli–Parisi "splitting function" has the important property

$$\int_0^1 P_{qq}^{(0)}(z)\mathrm{d}z = 0, \tag{9.24}$$

which is a straightforward consequence of the definition (9.12). Physical interpretation of this relation will be discussed in the next chapter.

Note that even after the addition of virtual contributions the sum (9.23) contains the singular term proportional to $P_{qq}^{(0)}(z)\ln(1/\beta)$ which comes from *parallel* emission of gluons. This singularity can be removed if

we incoherently add, as required by the KLN theorem, another cross–section, corresponding to the scattering process

$$e^- + q + G \rightarrow e^- + q \tag{9.25}$$

in which the *incoming* gluon and quark are *parallel*. Although the KLN theorem guarantees this cancelation its realization is by no means simple, as the evaluation of the cross–section for the $3 \rightarrow 2$ process (9.25) requires the knowledge of two-parton distribution function inside the nucleon. There is, however, an alternative and simpler way of getting rid of this remaining parallel singularity, which will be discussed in the next chapter under the name **factorization of parallel singularities**.

9.3 *Introduction into the theory of jets

Cancellation of the mass singularities in the sum (9.18) solves the problem of the divergences of separate terms σ_{real} and σ_{virt}, but we may be interested in more detailed properties of the final state. Although we cannot distinguish gluon with zero energy from no gluon at all, we can in principle distinguish these states if the gluon has a *finite* energy, however small it might be. Or in terms of the scaled invariant masses y_{ij}, experiments can never measure exactly $y_{ij} = 0$, but can go in principle to arbitrarily small $y_{ij} > 0$. Theoretically meaningful and experimentally answerable question then concerns the cross–section for producing three parton in the kinematical region outside the singularity, for instance for $y_{ij} \geq y$, where the cut–off parameter $y > 0$ defines this region. The integral of the real gluon emission cross–section (9.21) is then finite, but diverges as $y \rightarrow 0$, reflecting the fact that as $y \rightarrow 0$ it includes more and more of the immediate vicinity of the singularity.

Another, complementary, question concerns the cross–section for producing either a $q\overline{q}$ pair, or a $q\overline{q}G$ final states in the region where $y_{13} < y$ or $y_{23} < y$. The integral over this region *always* contains also the contribution from the virtual corrections to the $q\overline{q}$ final state and therefore *decreases* as $y \rightarrow 0$. In this ways we are naturally led to the y–dependent definition of the **jet**. At the order α_s jet is simply either a single parton or a pair of partons (i, j) with the scaled invariant mass satisfying $y_{ij} \leq y$. The **physically measurable** two–jet final state is then either a $q\overline{q}$ pair or $q\overline{q}G$ final state, in which the gluon is close to either q or $\overline{q}$. In the latter case the jet momentum is defined as the vector sum $p_{\text{jet}} \equiv p_i + p_j$ of the pair (i, j). [2]

After this verbal description we shall show how this procedure works at the lowest order in α_s. In this demonstration I shall work with *massless* quarks within the dimensional regularization instead of the gluon mass regulator of the previous section. [3] The formulae presented below are taken from an excelent review article [136], where all details can be found. To the order α_s and in $n = 4 - 2\varepsilon$ dimensions, the contribution to 2–jet cross–section coming from $q\overline{q}$ final state reads [136]

$$\sigma_{2\text{jet}}(q\overline{q}) = \frac{\sigma^{(2)}}{1 - \varepsilon} \left[1 - \varepsilon + C_F \frac{\alpha_s(\mu)}{2\pi} \left(\frac{4\pi\mu^2}{Q^2} \right)^{\varepsilon} A_1 + \mathcal{O}(\alpha_s^2) \right], \tag{9.26}$$

where $\sigma^{(2)}$ defined as

$$\sigma^{(2)} \equiv \left(\frac{4\pi\mu^2}{Q^2} \right)^{\varepsilon} \frac{\Gamma(2-\varepsilon)}{\Gamma(2-2\varepsilon)} \sigma_0, \quad \sigma_0 \equiv \frac{4\pi\alpha^2}{3Q^2} \left(3 \sum_f e_f^2 \right), \tag{9.27}$$

is the Born cross–section for e^+e^- annihilation into a $q\overline{q}$ pair in n dimension, $\zeta_2 \equiv \pi^2/6$ and the contribution from virtual corrections is determined by the coefficient

$$A_1 = \frac{\Gamma(1+\varepsilon)\Gamma^2(1-\varepsilon)}{\Gamma(1-2\varepsilon)} \left(-\frac{2}{\varepsilon^2} - \frac{1}{\varepsilon} + 6\zeta_2 - 5 + 3\zeta_2\varepsilon - 8\varepsilon \right). \tag{9.28}$$

where $\zeta_2 = \pi^2/3$. Note that ε can be sent to zero everywhere except A_1 and that the cut–off parameter y is absent from (9.26). Integrating the cross–section corresponding to the $q\overline{q}G$ final state over the region $y_{ij} \leq y$ we get

$$\sigma_{2\text{jet}}(q\overline{q}G) = \frac{\sigma^{(2)}}{1 - \varepsilon} \left[C_F \frac{\alpha_s(\mu)}{2\pi} \left(\frac{4\pi\mu^2}{Q^2} \right)^{\varepsilon} B_1 + \mathcal{O}(\alpha_s^2) \right], \tag{9.29}$$

---

[2]There are subtleties with this definition at higher orders, which are briefly mentioned below.
[3]The regularization is not needed for the integral over the emitted gluon momentum in the nonsingular region $y_{ij} \geq y$.
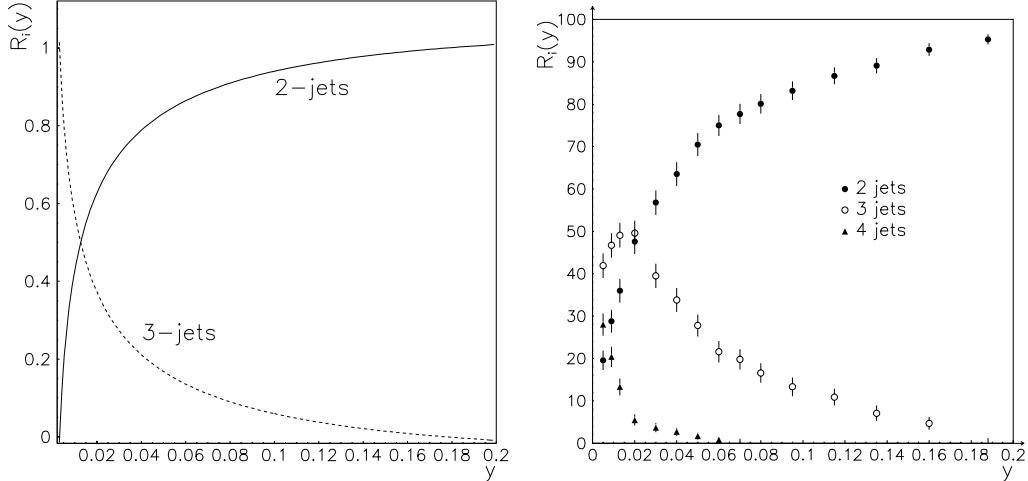
Figure 9.3: Two and three jet rates $R_i(y)$ according to formulae (9.31) and (9.32) (a) and experimental data on 2–4 jet rates in e^+e^- annihilations as measured by the OPAL Collaboration at LEP.

where

$$B_1 = \frac{\Gamma(1+\varepsilon)\Gamma^2(1-\varepsilon)}{\Gamma(1-2\varepsilon)}\left(\frac{2}{\varepsilon^2} + \frac{1}{\varepsilon} + 4 - 4\zeta_2 - 3\ln y - 2\ln^2 y\right). \qquad (9.30)$$

Comparing (9.26) and (9.29) we see that the singular terms $1/\varepsilon^2$ and $1/\varepsilon$ enter A_1 and B_1 with *opposite* signs and thus *cancel* in the physical 2–jet cross–section $\sigma_{2\mathrm{jet}}$, given as their sum:

$$\sigma_{2\mathrm{jet}}(y) = \sigma_0\left[1 + \frac{\alpha_s}{2\pi}C_F\left(-2\ln^2 y - 3\ln y + 2\zeta_2 - 1\right) + \mathcal{O}(\alpha_s^2)\right]. \qquad (9.31)$$

The $\mathcal{O}(\alpha_s)$ 3–jet cross–section, resulting from integration of (9.29) in the complementary region $y_{ij} > y$ equals

$$\sigma_{3\mathrm{jet}}(y) = \sigma_0\left[\frac{\alpha_s}{2\pi}C_F\left(2\ln^2 y + 3\ln y - 2\zeta_2 + 5/2\right) + \mathcal{O}(\alpha_s^2)\right]. \qquad (9.32)$$

Note the different character of the y dependence in (9.31) compared to (9.32): while the first is a decreasing function of y it is vice versa for the latter! These dependences have intuitively clear interpretation: the smaller y the more gluon radiation is "resolved" and counted as 3–jet final state. Summing the above two expressions we get $\sigma_{\mathrm{tot}}^{(2)} \equiv \sigma_0(1 + \alpha_s/\pi)$, the total cross–section of e^+e^- annihilation to partons at the order $\mathcal{O}(\alpha_s)$, which by construction is y–independent. In Fig. 9.3a the so called **jet rates** $R_i \equiv \sigma_{ij\mathrm{et}}(y)/\sigma_0(1 + \alpha_s/\pi)$ are plotted as a function of y for $i = 2, 3$. Although the decrease of $\sigma_{2\mathrm{jet}}(y)$ with decreasing y has clear physical interpretation, at about $y \doteq 0.003$ the $\mathcal{O}(\alpha_s)$ approximation (9.31) turns negative and simultaneously the 3–jet rate exceeds unity! These unphysical properties of jet rates result from the fact that at very small y the negative $\mathcal{O}(\alpha_s)$ virtual corrections to 2–jet rate overwhelm even the (positive) leading order Born term! To cure this problem the cross–sections (9.31) and (9.32) have to be considered to higher orders in α_s.

In Fig. 9.3b the two, three and four jet rates, measured in the OPAL experiment at LEP, are shown. The general trend of data agrees well with the theoretical predictions: as y decreases more and more jets pop up, their cross–section first rapidly increasing, peaking at some y and finally decreasing to make place for even more jets. Quantitative comparison of these data with theoretical predictions is, however, more complicated (see next two sections for discussion of some of these points).

9.3.1 Jet algorithms

In e^+e^- annihilations into 2 or 3 final state partons the definition of jets is simple and almost unambiguous. The situation turns more complicated if

- more partons in final states are taken into account,

170

- other process (like hadron–hadron or electron–proton collisions) are analyzed,

- consistency of the theoretical definition of the jet is to be guaranteed,

- jet definition should be applicable to observable hadrons as well.

The last problem will be discussed in the next subsection. Here I shall mention some problems and complications stemming from the presence of more partons in the final state. I shall outline the basic idea of the so called **jet algorithms** and briefly describe some of them.

For many parton final states there are two different ways how to define jets. The first is based on the concept of **clustering** which means repeated combining two massless partons into one, again massless, parton. As at higher orders perturbative calculations are available for massless quarks only, the grouping of two partons of an n–parton final state should yield again a massless parton. This implies that we *cannot* simply add parton four–momenta, as for noncollinear pairs this would inevitably result in a nonzero mass of the recombined pair. This caused us no headaches in the previous section, where we combined only one pair, but in general case this is a real problem. To keep partons after pairing massless one either sums the three–momenta and adjust the energy (thus violating energy conservation) or, vice versa, sums the energies and adjust the longitudinal momentum of the pair (thus violating its conservation). These violations are usually tiny, but they are, nevertheless, there. In the generalization of the procedure described in the previous section, the distance of a pair of partons is given by their scaled invariant mass $y_{ij} \equiv 2E_i E_j (1 - \cos\theta_{ij})/Q^2$, where θ is the relative polar angle of the partons. One starts with any parton, recombines it with its nearest neighbourgh in y_{ij} if their distance satisfies $y_{ij} \leq y$ and proceeds untill all the remaining pairs have $y_{ij} > y$. This defines the so called **JADE** algorithm. A closely related, but theoretically superior, is the so called **Durham** jet algorithm, which replaces m_{ij} with the "transverse" distance $d_{ij} \equiv 2\min(E_i^2, E_j^2)(1 - \cos\theta_{ij})$.

In the second approach we do not pair partons, but look for groups of partons that are within given "distance" from their "center". The relevant distance can be defined in various ways.

- In the **Sterman–Weinberg cone** jet algorithm angular separation defines this distance. The jet momentum and energy are then given as sums of momenta and energies of all particles that lie with a given angle θ from their center. The jet energy must furthermore exceed certain minimal value ϵ.

- In the **CDF cone** jet algorithm the "distance" is defined by the variable

$$R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}, \quad \eta \equiv -\ln(\tan\theta/2), \tag{9.33}$$

where ϕ is the azimuthal angle. In this case the transverse energy E_T, η and ϕ of the jets are given as weighted sums

$$E_T^{\text{jet}} \equiv \sum_i E_T^i, \tag{9.34}$$

$$\eta^{\text{jet}} \equiv \sum_i \frac{E_T^i}{E_T^{\text{jet}}} \eta^i, \tag{9.35}$$

$$\phi^{\text{jet}} \equiv \sum_i \frac{E_T^i}{E_T^{\text{jet}}} \phi^i, \tag{9.36}$$

where the sums run over all particles within the jet. Each of the jet algorithms has its advantages and shortcomings. Although there is no "best" jet algorithm, it is true that for different processes different algorithms are theoretically preferable. The JADE algorithm has originally been invented for e^+e^- annihilations, and is still being used there. In lepton–hadron and hadron–hadron collisions the CDF cone algorithm is theoretically superior.

9.3.2 Jets of partons vs jets of hadrons

In the previous subsection jets were constructed out of sets of partons. But hadrons, not partons, are observed in experiment. So for the concept of jet to be of any physical relevance, two conditions must be met:

- it must be possible to formulate the jet algorithm for the hadrons as well,

- we must understand *quantitatively* the relation between the properties of partonic and hadronic jets.

While there is no problem to meet the first condition, the second is much tougher. Imagine, for instance, that for a given value of the jet resolution parameter y we evaluated theoretically the differential cross–section $\mathrm{d}\sigma^{\mathrm{theor}}/\mathrm{d}p_T$. The experimentalist can do the same with observed hadrons and thus measure $\mathrm{d}\sigma^{\mathrm{exp}}/\mathrm{d}p_T$. Should we compare these two distributions for the same y, or not? And under which conditions can we do it? These are complicated questions that have no simple answers. The only way how to address them is to construct some **models** of hadronization (recall the simplest one of them, the independent fragmentation model of Section 5), study this relation within each model separately and then compare the corresponding results among as many hadronization models as possible. To summarize the current understanding of this relation one can say that the higher transverse momentum of the jets, the smaller the differences between the properties of partonic and hadronic jets.

An alternative way how to compare theoretical predictions with experimental data is to include as much as possible of the theoretical calculations at the level of partons within the so called **event generators**, Monte Carlo programs that mix perturbative QCD calculation of partonic collisions with models of hadronizations. In this case experimental jets are compared not with partonic jets but also with hadronic jets, simulated within these event generators. Although this comparison is in some sense more appropriate, it, on the other hand, suffers from the impossibility to include in the MC event generators all theoretical subtleties necessary for the really proper definition of jets. In the next version of this text a short introdution into the basic ideas of some of the most frequently used MC event generators will be included. For the moment I refer the reader to a review [146].

9.4   Exercises

1. Calculate explicitly (9.4), (9.6).

2. Derive the kinematical bounds (9.7).

3. Prove that for quite general $f(z)$ the definition of "+" distribution as given in (9.12) is equivalent to the following one
$$\int_0^1 [f(z)]_+ g(z)\mathrm{d}z \equiv \int_0^1 f(z)(g(z) - g(1))\mathrm{d}z.$$

4. Show that from definition
$$\int_0^1 [f(z)]_+ = 0.$$

Chapter 10

QCD improved quark-parton model

In this chapter I shall show *how* the naive QPM, introduced in Chapter 4, is modified within the framework of perturbative QCD (PQCD) and *why* this modification is such that the basic concepts of the QPM maintain their meaning and are useful even in a theory which simultaneously aspires to describe the confinement of colored partons. I shall carefully distinguish the effects which *can* be calculated in perturbation theory and in which partons are treated essentially as massless observable particles, from those where the color confinement plays a crucial role and perturbation theory is inapplicable.

10.1 General framework

The framework of QCD improved QPM model is sketched in Fig.10.1. It generalizes the basic features of Fig. 5.7, but applies to any "hard" collision of two incoming particles,

$$A + B \to F \tag{10.1}$$

producing some, at the moment unspecified, final state F. The term "hard" denotes processes, which are dominated by *short distance interactions*. The colliding particles may be hadrons, leptons or gauge bosons. In the SM leptons and gauge bosons are considered elementary and thus coincide with their constituents (with distribution functions given by the δ-function $\delta(1-x)$). The description of such processes in perturbative QCD (PQCD) retains the basic QPM strategy of dividing the space–time evolution of the collision into three distinct stages, but introduces some important modifications.

A: The initial evolution is represented by the distribution functions $D_{a/A}(x_1, M, \mathrm{FS}_1)$, $D_{b/B}(x_2, M, \mathrm{FS}_2)$ of partons a, b [1] inside beam particles A,B. These distribution functions are determined by interactions of partons at *large* distances and are thus incalculable in PQCD. They depend, besides the momentum fraction x, also on the so called **factorization scale** and **factorization scheme** [2]. Although incalculable in PQCD, we shall later discuss equations governing their dependence on the factorization scale M.

B: The hard scattering of a given pair of partons a, b, producing partons c, d [3]

$$a + b \to c + d \tag{10.2}$$

is described by parton level cross–section

$$\sigma_{ab \to cd}(s, x_1, x_2, p_c, p_d, \mu, M_1, \mathrm{FS}_1, M_2, \mathrm{FS}_2), \tag{10.3}$$

where $s = (p_a + p_b)^2$ is the square of total CMS energy of the colliding partons a, b, μ is the *hard scattering scale*, familiar from Chapter 8 and in general *different* from the factorization scales M_i, and

---

[1] For brevity of expression the term "parton" will in these considerations mean also the incoming and outgoing leptons and photons.

[2] These concepts are defined later in this section.

[3] In principle we should consider also the more general $2 \to n$ parton level subprocesses. The generalization of the following considerations to such cases is straightforward.
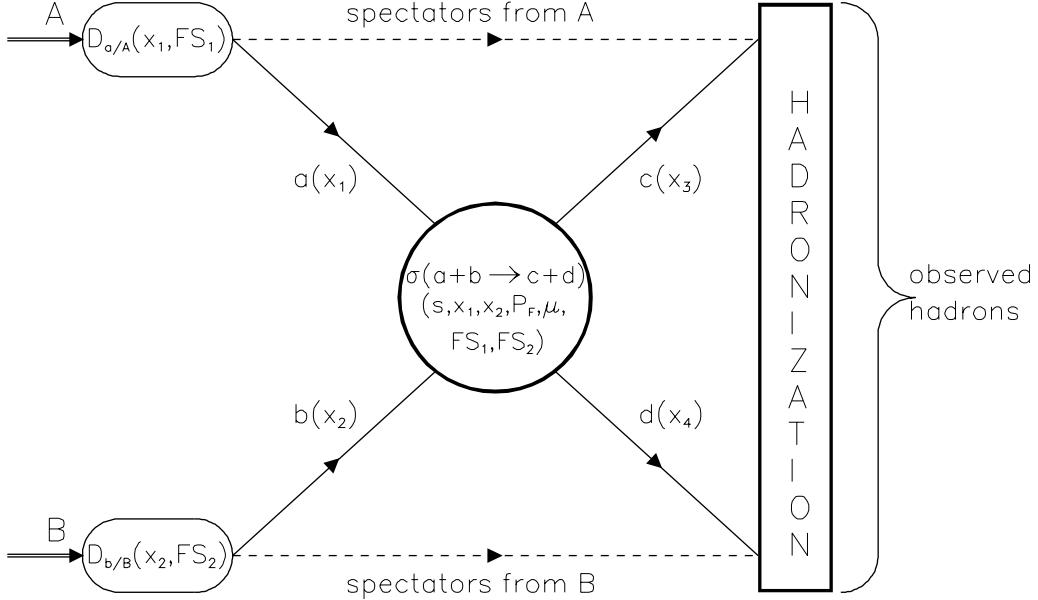
Figure 10.1: The general scheme of QCD improved QPM

FS_1, FS_2 are factorization schemes defining the distribution functions $D_{a/A}, D_{b/B}$. The hard scattering cross–section (10.3) *is calculable* in PQCD, provided the process is dominated by short distances, which in practice means that some measure of "hardness" is large compared to the mass of proton. The QPM provides the lowest order approximation to (10.3) which is then systematically improved in PQCD.

C: The hadronization of the partonic state produced in (10.2) remains the least understood stage of the collision. PQCD is *not applicable* here and various sorts of models must be employed. In the **independent fragmentation model** of Chapter 5, the fragmentation functions $D_{h/p}(z, M, \text{FS})$ acquire, similarly to parton distribution functions inside hadrons, the dependence on the M and FS.

In this chapter only the first two stages will be discussed. The reader interested in hadronization models should turn to reviews [146, 147]. The general formula for the evaluation of cross–sections of the hard scattering processes (10.1) in PQCD is basically the same as in the QPM. Schematically it reads

$$\sigma(A + B \rightarrow F + \text{anything}) = \tag{10.4}$$

$$\sum_{abcd} \int\int \mathrm{d}x_1 \mathrm{d}x_2 D_{a/A}(x_1, M_1) D_{b/B}(x_2, M_2) \sigma_{ab \rightarrow cd}(s, x_1, x_2, p_c, p_d, M_1, M_2) \otimes \underbrace{D_{hadr}(p_c, p_d, P_F)}_{\text{model dependent}},$$

where the sum runs over all parton combinations which can lead to the required final hadronic state F and the cross–section $\sigma_{ab \rightarrow cd}$ contains all appropriate δ–functions expressing the overall momentum conservation on the partonic level. The symbol of convolution $\otimes$ stands for further integrations over the momenta of final state partons p_c, p_d, which lead to the final *hadronic* state F. The initial partonic distribution functions depend on the (in general different) factorization scales M_1, M_2 as well as on the corresponding factorization conventions, discussed in detail in next section. In this section the latter dependence will not be written out explicitly. The factorization scales M_i are, similarly to the hard scale μ, unphysical and the cross–sections of physical processes must, if evaluated to all orders, be independent of them.

The following discussion is divided into two parts. First we shall consider the interaction of partons *with their own chromodynamic field.* Here the calculations can be carried out in a close analogy to QED and, indeed, the basic idea of **partons within partons** has been borrowed from the old papers of Weizsäcker and Williams on QED [148]. This is the true arena of PQCD, contrary to the much more complicated task of describing the interactions of **partons within hadrons**.
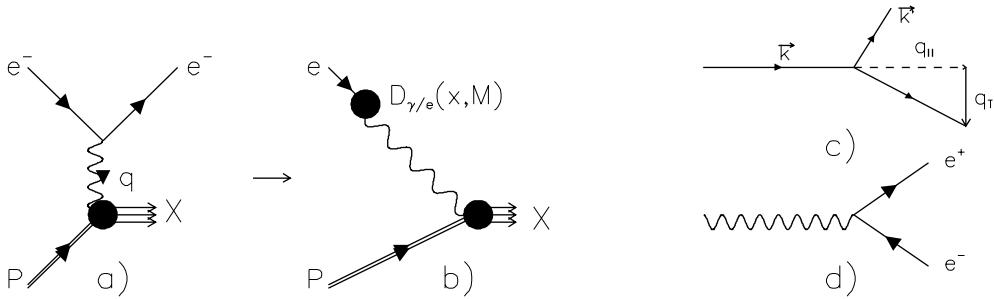
Figure 10.2: The relation between cross–sections of ep (a) and γp (b) interactions in low Q^2 region. Kinematics of the branching $e^- \to e^- + \gamma$ (c) and the vertex describing the branching $\gamma \to e^+ e^-$ (d).

10.2 Partons within partons

Let me start with a detailed discussion of the so called **Weizsäcker-Williams approximation** (WWA) in QED, as it contains the basic idea of "partons within partons" picture, but avoids the conceptual problem of parton confinement in hadrons, and is also technically simpler.

10.2.1 Equivalent photon approximation

Consider again the interaction of an electron with a proton, but now in the region of *small Q^2*, where this interaction cannot be described by an incoherent sum of cross-sections of interactions on individual partons. As in this subsection we are not concerned with structure of the proton, but rather with that of the beam electron, we stay within the perturbative QED and concentrate on the upper vertex in Fig.10.2. The basic idea of the WWA is to interpret the process

$$e + p \to e + \text{anything} \qquad (10.5)$$

as the emission from the beam electron of a nearly real photon followed by its interaction with the target proton, and correspondingly express the total cross–section $\sigma_{\text{tot}}(\text{ep})$ in terms of the total cross–section $\sigma_{\text{tot}}(\gamma \text{p})$ of the *real* photon–proton collision

$$\gamma + p \to \text{anything}. \qquad (10.6)$$

This step is based on the assumption that for small photon virtuality (mass) its interaction with the target proton should be well approximated by that of the real γ. The measure of the "smallness" of the photon virtuality is not directly Q^2, but rather its ratio Q^2/W^2, where $W^2 \equiv s_{\gamma p}$. In physical terms small virtuality means that in the overall ep CMS the photon in Fig. 10.2 behaves as *nearly real and parallel* to the beam electron, which thus looks like being accompanied by the "beam" of photons with certain momentum distribution function. The determination of this distribution function proceeds in several steps:

1. First we write down the expression for the cross–section of the process (10.5), for instance in the rest frame of the proton with mass M_p (this choice is not essential),

$$d\sigma = \frac{(2\pi)^4}{|\vec{v}|} \frac{e^4}{2E2M} L^{\mu\nu} W_{\mu\nu}(q,p) \frac{d^3 k'}{(2\pi)^3 2E_{k'}} \frac{1}{Q^4} = \frac{(2\pi)^4 e^4}{2E2M} L^{\mu\nu} W_{\mu\nu}(q,p) \frac{d^4 q}{(2\pi)^3 Q^4} \delta(k'^2), \qquad (10.7)$$

where $L^{\mu\nu}$ is the lepton tensor introduced in Section 4, and $W_{\mu\nu}(q,p)$ is the hadronic tensor associated with the lower vertex in Fig. 10.2a. Neglecting, as always in this text, the effects of finite electron mass, we set in the second part of (10.7) $|\vec{v}| = 1$. The hadronic tensor $W_{\mu\nu}$ is related to DIS structure functions F_i of (5.46), but for low Q^2 it cannot be expressed in terms of parton distribution functions.

2. Similarly we express the total cross–section of the process (10.6) as

$$\sigma_{tot}(\gamma \text{p}) = \frac{(2\pi)^4 e^2}{2E_\gamma 2M} \left(-\frac{1}{2} g^{\mu\nu} \right) W_{\mu\nu}(Q^2 = 0, pq), \qquad (10.8)$$

where $W_{\mu\nu}(Q^2 = 0)$ is the same hadronic tensor as in (10.7) and $1/2$ results from the averaging over the spin states of the incoming photon.

3. Assuming Lorentz invariance and parity conservation, the most general form of the *symmetric* hadronic tensor $W_{\mu\nu}(Q^2)$ for *unpolarized* electron–proton scattering has the form

$$W_{\mu\nu} = C_1(Q^2, pq)g_{\mu\nu} + C_2(Q^2, pq)p_\mu p_\nu + C_3(Q^2, pq)q_\mu q_\nu + C_4(Q^2, pq)\left(p_\mu q_\nu + p_\nu q_\mu\right), \qquad (10.9)$$

where $C_i(Q^2, pq)$ are functions of Q^2 and pq which contain all information about the structure of the unpolarized proton. [4] In the following the dependence on pq will not be explicitly written out.

4. The requirement of *gauge invariance* imposes further restriction on this tensor

$$q^\mu W_{\mu\nu}(Q^2) = 0 \Rightarrow q_\nu \left[C_1 + C_3 q^2 + C_4(pq)\right] + p_\nu \left[C_2(pq) + C_4 q^2\right] = 0, \qquad (10.10)$$

which implies that only two of the four functions C_i are independent

$$C_4(Q^2) \;=\; -C_1(Q^2)\frac{1}{(pq)} - C_3(Q^2)\frac{q^2}{(pq)}, \qquad (10.11)$$

$$C_2(Q^2) \;=\; C_1(Q^2)\frac{q^2}{(pq)^2} + C_3(Q^2)\frac{q^4}{(pq)^2}. \qquad (10.12)$$

5. Assuming that the functions C_i are *regular* at $Q^2 = 0$ (there are no reasons why they shouldn't) only $C_1(Q^2 = 0)$ contributes for the real photon to the contraction

$$-\frac{1}{2}g^{\mu\nu}W_{\mu\nu}(Q^2 = 0)) = -C_1(Q^2 = 0). \qquad (10.13)$$

6. For the virtual photon only the terms proportional to C_1, C_2 contribute after contraction with the leptonic tensor $L^{\mu\nu}$, defined in (5.19)

$$L^{\mu\nu}W_{\mu\nu} = 2\left[-2C_1(Q^2)(kk' - 2m^2) + C_2(Q^2)\left(2(kp)(k'p) - M^2(kk')\right)\right]. \qquad (10.14)$$

7. Recalling basic kinematical relations of Section 4.2, taking into account (10.12) and keeping in (10.14) the first two leading terms in Q^2, we get

$$
\begin{aligned}
L^{\mu\nu}W_{\mu\nu} &= -2C_1(Q^2)Q^2\left[1 + 2\left(\frac{kp}{qp}\right)^2 - 2\left(\frac{kp}{qp}\right) - \frac{2m^2}{Q^2}\right] \\
&= -2C_1(Q^2)Q^2\left[\frac{1 + (1-y)^2}{y^2} - \frac{2m^2}{Q^2}\right],
\end{aligned}
\qquad (10.15)
$$

where $y \equiv qp/kp$ has a physical interpretation as the fraction of incoming electron energy, carried away by the exchanged virtual, nearly parallel, photon.

8. What remains is to express $\delta(k'^2)\mathrm{d}^4k'$ in terms of y and the transverse (with respect to the incoming electron) momentum q_T of the emitted (virtual) photon. For general $Q^2 \neq 0$ this relation is more complicated, but in the so called **collinear kinematics**, i.e. for $q_T \ll q_\parallel$ (see Fig. 10.2c) we have [5]

$$\int \frac{\mathrm{d}^3k'}{2E_{k'}} = \int \delta(k'^2)\mathrm{d}^4q = \int \frac{1}{4E_{k'}}\mathrm{d}\phi\mathrm{d}q_\parallel \mathrm{d}q_T^2 \doteq \int \frac{\pi}{2(1-y)}\mathrm{d}y\mathrm{d}q_T^2 \doteq \frac{\pi}{2}\int \mathrm{d}y\mathrm{d}Q^2, \qquad (10.16)$$

where we have integrated over the azimuthal angle ϕ and in the last equality used the approximation $y = q_\parallel/E_k$, hence the last equality. In the collinear kinematics we also have

$$Q^2 = 4E_k E_{k'}\sin^2(\vartheta/2) \doteq \frac{q_T^2}{1-y}, \qquad (10.17)$$

i.e. the photon virtuality is, for fixed y, proportional to q_T^2.

---

[4]The invariant pq is related to the total energy squared W^2 in γp rest frame as $W^2 = 2pq - Q^2 + m_p^2$.
[5]For unpolarized scattering the cross–section is ϕ–independent.

9. Inserting the preceding two expressions together with (10.15) into (10.7) we see that one power of Q^2 coming from the photon propagator is canceled by the Q^2 in (10.15), leading to the following result for the total cross–section of ep collision at *low Q^2*:

$$\sigma_{\mathrm{ep}}(S, Q^2_{max}) = \iint dy \frac{dq_T^2}{q_T^2} \left[\frac{\alpha}{2\pi} \left(\frac{1 + (1-y)^2}{y} - \frac{2m^2 y}{Q^2} \right) \right] \sigma_{\gamma\mathrm{p}}(yS), \qquad (10.18)$$

where the integral is taken over those values of y, q_T satisfying the condition $Q^2 < Q^2_{max}$. This is the mentioned **Weizsäcker-Williams approximation**.

There are several aspects of this formula which merit a comment.

- Note that one power of y in the denominator of (10.15) has been used to convert electron energy E_k to E_γ, the energy of the virtual photon, appearing in (10.6).

- The expression (10.18) naturally suggests to interpret the function

$$\frac{\alpha}{2\pi} \frac{1}{Q^2} \left[\frac{1 + (1-y)^2}{y} - \frac{2m^2 y}{Q^2} \right] \qquad (10.19)$$

as the **probability to find a photon with virtuality $\mathbf{Q^2}$ and momentum $\mathbf{y}P$** inside an electron of momentum P. The crucial aspect of the WWA is the question in which region of y and q_T, or Q^2, it can be trusted.

- Although formally of the order $\mathcal{O}(m^2)$ the second term in (10.19) gives actually a **finite** contribution after integration over Q^2 (or q_T^2) in some interval (Q^2_{min}, Q^2_{max})

$$\sigma_{\mathrm{ep}}(S, Q^2_{max}, Q^2_{min}) \doteq \int dy \underbrace{\frac{\alpha}{2\pi} \left(\left[\frac{1 + (1-y)^2}{y} \right] \ln \frac{Q^2_{max}}{Q^2_{min}} - 2m^2 y \left[\frac{1}{Q^2_{min}} - \frac{1}{Q^2_{max}} \right] \right)}_{f_{\gamma/e}(y, Q^2_{max}, Q^2_{min})} \sigma_{\gamma p}(yS),$$

$$(10.20)$$

where the lower limit Q^2_{min} follows from kinematics

$$Q^2_{min} = \frac{m^2 y^2}{1-y}. \qquad (10.21)$$

Considering Q^2_{max}, relabelled again simply as Q^2, as a free parameter, we finaly get

$$\begin{aligned} f_{\gamma/e}(y, Q^2) &= \frac{\alpha}{2\pi} \left[\frac{1 + (1-y)^2}{y} \ln \frac{Q^2(1-y)}{m^2 y^2} - \frac{2(1-y)}{y} + O(m^2/Q^2) \right] \\ &= \frac{\alpha}{2\pi} \left[\frac{1 + (1-y)^2}{y} \ln \frac{Q^2}{m^2} + \frac{1 + (1-y)^2}{y} \ln \frac{1-y}{y^2} - \frac{2(1-y)}{y} \right]. \end{aligned} \qquad (10.22)$$

What happened is that m^2 from the matrix element got canceled by $1/m^2$ coming from integration over the phase space, giving in the end finite contribution to $f_{\gamma/e}(y, Q^2)$. This function is then interpreted as the **distribution function of photons, carrying fraction y** of electron energy and **having virtuality up to $\mathbf{Q^2}$**.

- The crucial feature of the WWA is the presence of the *logarithmic* term $\ln(Q^2/m^2)$, which is due to the fact that the matrix element (10.15) is proportional to Q^2.

- Beside the distribution function of photons inside an electron we can also introduce the complementary distribution function describing the probability of finding electrons inside an electron

$$f_{e/e}(y, Q^2) = \delta(1-y) + \frac{\alpha}{2\pi} \left[\frac{1 + y^2}{1-y} \right] \ln \frac{Q^2}{m^2}, \qquad (10.23)$$

where the δ–function corresponds to the case of no photon radiation. To include the contribution of virtual diagrams in one needs merely to replace the square brackets in the above expression for $f_{e/e}$

by the "+" distribution. This modification is, however, crucial for the probability interpretation of $f_{e/e}(x, Q^2)$ as it implies

$$\int_0^1 f_{e/e}(x, Q^2)\mathrm{d}x = 1 \tag{10.24}$$

and thus guarantees that the total number of electrons (i.e. the number of electrons minus the number of positrons) inside an electron is *conserved* (and equal to 1 for a physical electron).

- The same ideas can be applied also to the case that the beam particle is a **real photon** which can split, as sketched in Fig. 10.2d, into a e^+e^- pair. The distribution function of electrons (or positrons) inside a photon, carrying fraction y of its energy and having virtuality up to Q^2, is given as

$$f_{e/\gamma}(y, Q^2) = \frac{\alpha}{2\pi} \left[y^2 + (1-y)^2 \right] \ln \frac{Q^2}{m^2} \tag{10.25}$$

with the same provision about the neglected terms as for previous case. Notice that there is *no* IR singularity in (10.25) and consequently no "+" distribution in this expression.

The approximations made in the process of derivation of the WWA restrict its validity to low Q^2 region, but how small should Q^2 actually be to guarantee a required accuracy depends on the process in the lower vertex of Fig. 10.2ab, in which the virtual photon interacts with the target particle, as well as on the kinematical region considered. Roughly speaking we expect the WWA to be good approximation whenever the photon virtuality is small compared to values of basic kinematical variables describing its interaction with the target particle. For the case of the total cross–section in (10.7) this meant that we required $Q^2/s_{\gamma p} = Q^2/yS \ll 1$, for the **photoproduction of jets** in electron–proton collisions, which is currently under intensive study at HERA electron–proton collider at DESY, it means that Q^2 should be much smaller than the transverse momentum squared of the produced jets.

In summary, the essence of the WWA is contained in the **branching functions**

$$P_{\gamma e}(x) = \frac{1 + (1-x)^2}{x}, \tag{10.26}$$

$$P_{ee}(x) = \left[\frac{1 + x^2}{1-x} \right]_+, \tag{10.27}$$

$$P_{e\gamma}(x) = x^2 + (1-x)^2, \tag{10.28}$$

all of which reflect the basic QED vertex eγe. For processes dominated by low Q^2 of the exchanged photon the incoming electron behaves as a beam of nearly real electrons and photons, described by distribution functions $f_{\gamma/e}, f_{e/e}$. Similarly, a single photon can be viewed as a photon accompanied by a beam of nearly real electrons and positrons, distributed in the photon according to the function $f_{e/\gamma}(x, Q^2)$. Finally let me emphasize two points. First, due to the fact that all QED branching functions appear always in the product with QED couplant α, the "admixture" of photons inside an electron and electrons in a photon are *small effects*. Secondly, these effects are entirely due to the interaction of electrons and photons with *their own* electromagnetic field.

10.2.2 Branching functions in QCD

Most of this subsection is a straightforward generalization of the basic idea of the WWA, with the substitutions electron $\rightarrow$ quark and photon $\rightarrow$ gluon and taking into account the 3–gluon vertex, which leads to a novel feature of QCD: the $P_{gg}(x)$ branching function. Its physical interpretation is the same as for other branchings. Due to the color degree of freedom the QCD branchings do acquire additional n_f dependent numerical factors: [6]

$$P_{qq}^{(0)}(x) = P_{\bar{q}\bar{q}}^{(0)}(x) = \frac{4}{3} \left[\frac{1 + x^2}{1-x} \right]_+, \tag{10.29}$$

---

[6] These factors depend in general also on the number of colors. The following expressions correspond to $N_c = 3$. The superscript "(0)" has been introduced in view of the latter discussion of higher order corrections to this picture.

$$P_{Gq}^{(0)}(x) \;=\; P_{G\overline{q}}^{(0)}(x) = \frac{4}{3}\left[\frac{1+(1-x)^2}{x}\right], \tag{10.30}$$

$$P_{qG}^{(0)}(x) \;=\; P_{\overline{q}G}^{(0)}(x) = \left[\frac{x^2+(1-x)^2}{2}\right], \tag{10.31}$$

$$P_{GG}^{(0)}(x) \;=\; 6\left\{\left[\frac{x}{1-x}\right]_+ + \frac{1-x}{x} + x(1-x) + \left(\frac{33-2n_f}{36}-1\right)\delta(1-x)\right\}. \tag{10.32}$$

These branching functions are *the same* for all quark flavors $q_i, \overline{q}_i$ and therefore general symbols $q, \overline{q}$ were used to subscript them. Note that in $P_{GG}^{(0)}(x)$ we again encounter the "+" distribution. As the QCD couplant α_s is much larger than α, the effects of parton branching are correspondingly more important as well. As in QED, they are due to the interaction of quarks and gluons with *their own* chromodynamic field. Note that the n_f–dependent term in the $g \to g$ branching function is nonvanishing only for $x = 1$ as it comes from the virtual quark loop correction to the gluon propagator.

10.2.3 *Multigluon emission and Sudakov formfactors of partons

In the Section 9.2 we saw how the integration over the transverse momentum of the gluon emitted in Fig. 9.2a from the target quark, together with the virtual corrections coming from Fig. 9.2b, yields the contribution to the cross–section of the process $e^- + q \to e^- + q + g$ which has the structure

$$\sigma \propto \alpha_s\left(P_{qq}^{(0)}(x)\ln\frac{Q^2}{m_g^2} + f(x)\right) \tag{10.33}$$

The function $f(x)$, which is process dependent, contains terms like $\ln x, \ln(1-x)$ and so is potentially large for x parametrically close to 0 or 1. For large Q^2 and provided we stay away from these dangerous regions this term may, however, be neglected, and we are left with the first, logarithmic term, which is *process independent* and *unique*. This result can be simply generalized taking into account multigluon emission from a single quark. We only need to keep in mind that the logarithmic terms come from the integration over the transverse momentum around the singularity at $q_T = 0$. In the case of two gluon emission the leading term proportional to $\alpha_s^2\ln^2(Q^2/m_g^2)$ comes from integration over the region of **strongly ordered** virtualities t_1, t_2 in Fig. 10.3a

$$|t_2| \le \epsilon\,|t_1|, \qquad |t_1| \le \epsilon Q^2, \tag{10.34}$$

where ϵ is some small parameter introduced to quantify the meaning of "strongly ordered". As we shall see the coefficients in front of the leading logarithms *do not* depend on this parameter. Omitting the branching functions and concentrating on the transverse momenta of radiated gluons we find that the contribution of the diagram with two radiated gluons is proportional to (we define $\tau = |t|$ for all virtualities)

$$\alpha_s^2(\mu)\int_{m_g^2}^{\epsilon Q^2}\frac{\mathrm{d}\tau_1}{\tau_1}\int_{m_g^2}^{\epsilon\tau_1}\frac{\mathrm{d}\tau_2}{\tau_2} = \alpha_s^2(\mu)\int_{m_g^2}^{\epsilon Q^2}\frac{\mathrm{d}\tau_1}{\tau_1}\ln\frac{\epsilon\tau_1}{m_g^2} = \alpha_s^2(\mu)\left[\frac{1}{2}\ln^2\frac{Q^2}{m_g^2} + f_1(\ln\epsilon)\ln\frac{Q^2}{m_g^2} + f_2(\ln\epsilon)\right], \tag{10.35}$$

where $f_i(\ln\epsilon)$ are simple polynomials in $\ln\epsilon$. As advertised, the coefficient $(1/2)$ by the leading, here double, logarithm is *independent* of ϵ, while the nonleading terms do depend on it. This calculation can obviously be generalized to an arbitrary number n of emitted gluons with the same result: the coefficient A_n in front of the term with the same power of α_s and $\ln Q^2$ (the so called "leading log" (LL)) in the expression for the cross–section

$$\sigma^{(n)} \propto \alpha_s^n\left(\underbrace{A_n\ln^n\left(\frac{Q^2}{m_g^2}\right)}_{\text{LL}} + \underbrace{B_n\ln^{n-1}\left(\frac{Q^2}{m_g^2}\right)}_{\text{NLL}} + \cdots + f_n(x)\right) \tag{10.36}$$

comes entirely from the region (10.34) of strongly ordered virtualities, is ϵ–independent and equals $1/n!$.

In writing (10.36) we should also specify the scale μ to be used as an argument in $\alpha_s(\mu)$. This task is, however, far from trivial and I shall merely outline the arguments behind the claim that the "best" scale to be used in the vertices of the ladder of Fig. 10.3a, where the incoming quark of virtuality τ_{i+1} emits a quark with virtuality τ_i and a real gluon, is $\mu^2 = \tau_i$, i.e., μ^2 should be identified with the highest virtuality of all partons interacting in this vertex. This implies that the scales of α_s are different at different vertices
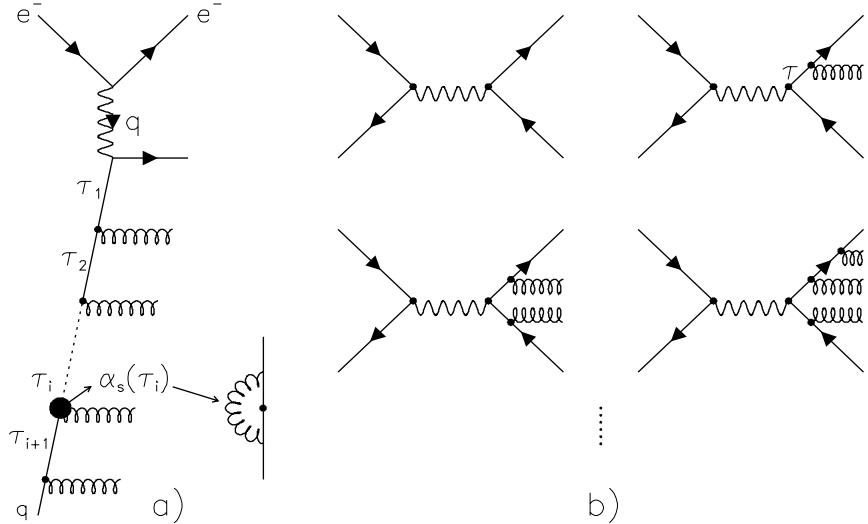
Figure 10.3: Multigluon emission from incoming parton leg (a) in eq collision (a), and from outgoing quarks in e$^+$e$^-$ →q$\overline{\text{q}}$ annihilation (b).

along the ladder and *increase* as we move along it from the proton up to the qγq vertex. This feature will be crucial for the considerations of the next subsection.

The derivation of this feature is based on the incorporation of radiative corrections. Consider the correction to the emission of the first gluon in Fig. 10.3a, brought about by the emission and subsequent reabsorption of another gluon. The *ultraviolet* renormalization of this loop leads to the appearance of logarithms of the type $\ln(\mu/\tau_i)$ where μ is the argument of $\alpha_s(\mu)$ at the loop vertices . These potentially large UV logarithms can be avoided by setting μ of the order of τ_i. The fact that it is the largest virtuality which plays a decisive role in this argument is intuitively understandable. In the case of strongly ordered virtualities the smaller one in each ladder link can be neglected and thus the vertex effectively describes the interaction of one off mass–shell and two massless partons. As in the mentioned logarithm μ enters scaled by something describing the kinematics of the vertex, the maximal virtuality at that vertex is essentially the scale available.

The task of resumming the series (10.36) is done for us by the so called **evolution equations** of parton distribution functions, but before coming to them let me mention another important consequence of multi-parton emission. Consider again the e$^+$e$^-$ annihilation into quarks as discussed in connection with the KLN theorem in Section 6.4. According to (9.19) the probability of emitting a single *real gluon* is proportional to $\ln^2\beta$, while the radiative corrections (virtual emissions) give negative contributions, which *cancel* both single and double logs of β. By the KLN theorem this mechanism operates at each order of α_s. It is clear that the probability of multiple gluon emission grows with the order of perturbation expansion as there are more powers of $\ln^2\beta$. As a result the probability of *no* gluon emission must accordingly be suppressed. This important phenomenon is expressed quantitatively in the so called **Sudakov formfactors** of quarks and gluons.

Imagine that the quark (or antiquark) produced in e$^+$e$^-$ annihilations has a positive virtuality τ [7] and assume for simplicity that τ is small with respect to the CMS energy squared $s = 4E^2$. Kinematical bounds on the fraction z of the parent quark energy, carried by the daughter quark after the gluon emission in Fig. 10.3b are given as

$$z_{min} = \frac{1}{2}\left[1 - \sqrt{1 - \tau/E^2}\right] \doteq \frac{\tau}{s}; \;\; z_{max} = \frac{1}{2}\left[1 + \sqrt{1 - \tau/E^2}\right] \doteq 1 - \frac{\tau}{s}. \tag{10.37}$$

The time–like Sudakov formfactor gives the probability that there will be *no gluon radiation* with the

_____

[7]In these considerations quarks can be massless as the IR and parallel singularities will be regulated by the lower bound on the virtuality τ.
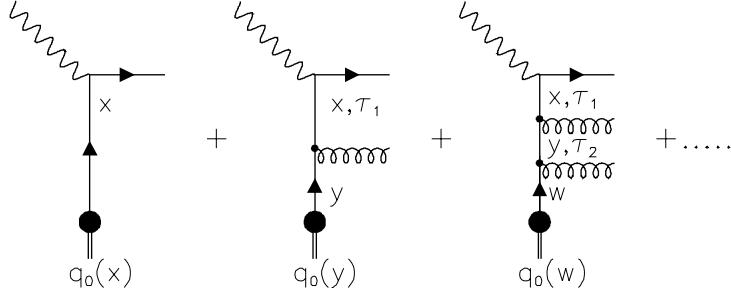
Figure 10.4: Graphical representation of the definition of "dressed" parton distribution functions inside a given hadron for the nonsinglet quark distribution function. τ_i denotes absolute value of the virtuality of a given intermediate state.

virtuality of the outgoing off mass–shell quark between τ_0 and τ and the fraction z between z_{min} and z_{max}

$$S(\tau_0, \tau, E) \equiv \exp\left\{ -\int_{\tau_0}^{\tau} \frac{\mathrm{d}\tau'}{\tau'} \int_{z_{min}}^{z_{max}} \frac{\alpha_s(\sqrt{\tau'})}{2\pi} P_{qq}^{(0)}(z)\mathrm{d}z \right\}, \tag{10.38}$$

where the double integral in the exponent gives the probability that a gluon *will* be radiated from the quark with $\tau \in (\tau_0, \tau)$ and will carry energy fraction $1 - z \in (1 - z_{max}, 1 - z_{min})$. This expression is closely analogous to the usual nondecay probability of an unstable system. Neglecting the dependence of $\alpha_s(\sqrt{\tau})$ on τ the integrals in the exponent of (10.38) can be performed and we get

$$S(\tau_0, \tau, E) \approx \exp\left[-\frac{\alpha_s}{2\pi} \frac{4}{3} \ln\left(\frac{\tau}{\tau_0}\right) \ln\left(\frac{4E^2}{\tau_0}\right) + \mathcal{O}(\ln \tau_0) \right]. \tag{10.39}$$

Notice that the exponent contains a *double log* of τ_0 and that the Sudakov formfactor *vanishes* as the lower limit on considered virtualities τ_0 goes to zero.

10.3 Partons within hadrons

10.3.1 Evolution equations at the leading order

In all preceding QCD calculations quarks and gluons were treated in the same way as leptons, i.e. as free particles before and after the collision, despite the experimental evidence that quarks and gluons exist merely *inside hadrons* and that they behave like free particles only if probed at short distances. The evolution equations we are now going to derive take this fact into account and simultaneously allow us to get rid of the remaining parallel logs in (10.36). The basic idea of the derivation, graphically represented in Fig. 10.4 is very simple and reminiscent of the UV renormalization of electric or color charges.

The derivation starts with the introduction of the concept of **primordial** (sometimes also called "bare") distribution functions of partons inside hadrons, $q_0(x), G_0(x)$, which depend on x only and are interpreted in the sense of the QPM of Chapter 5. Let us discuss first the case of the so called **nonsinglet** (NS) quark distribution functions, which at the leading order coincide with the valence ones, introduced in Chapter 5. This implies that we need to consider only the effects of multiple gluon emissions off the primordial quarks, described by distribution function $q_{NS,0}(x)$, as sketched in Fig. 10.4. Summing up the contributions of the ladders in Fig. 10.4 we define the **renormalized**, or **"dressed"**, NS quark distribution function

$$q_{NS}(x, M) \equiv q_{NS,0}(x) + \int_x^1 \frac{\mathrm{d}y}{y} \left[P_{qq}^{(0)}\left(\frac{x}{y}\right) \int_{m^2}^{M^2} \frac{\mathrm{d}\tau_1}{\tau_1} \frac{\alpha_s(\tau_1)}{2\pi} \right] q_{NS,0}(y) + \tag{10.40}$$

$$\int_x^1 \frac{\mathrm{d}y}{y} \int_y^1 \frac{\mathrm{d}w}{w} \int_{m^2}^{M^2} \frac{\mathrm{d}\tau_1}{\tau_1} \int_{m^2}^{\tau_1} \frac{\mathrm{d}\tau_2}{\tau_2} \frac{\alpha_s(\tau_1)}{2\pi} \frac{\alpha_s(\tau_2)}{2\pi} \left[P_{qq}^{(0)}\left(\frac{x}{y}\right) P_{qq}^{(0)}\left(\frac{y}{w}\right) \right] q_{NS,0}(w) + \cdots,$$
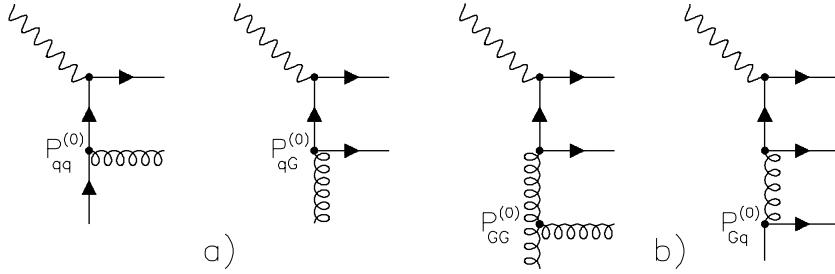
Figure 10.5: Graphical representation of the kernels of the evolution equations for the quark (a) and gluon(b) distribution functions inside the proton.

where the newly introduced scale M has the meaning of *maximal virtuality* of the quark, which interacts with the photon in the upper, QED, vertex of Fig. 10.4a and coincides with the factorization scale introduced at the beginning of this Chapter.

The way we get rid of the parallel singularities implicitly present in (10.40) closely resembles the step from bare to the renormalized electric charge in QED (or color charge in QCD). Taking the derivative of both sides in (10.40) with respect to the scale M we find

$$\frac{dq_{NS}(x, M)}{d\ln M^2} = \frac{\alpha_s(M)}{2\pi} \int_x^1 \frac{dy}{y} P_{qq}^{(0)}\left(\frac{x}{y}\right) \underbrace{\left[q_{NS,0}(y) + \int_y^1 \frac{dw}{w} \int_{m^2}^{M^2} \frac{d\tau_2}{\tau_2} \frac{\alpha_s(\tau_2)}{2\pi} P_{qq}^{(0)}\left(\frac{y}{w}\right) q_{NS,0}(w) + \cdots\right]}_{\equiv q_{NS}(y, M)},$$

(10.41)

i.e.

$$\frac{dq_{NS}(x, M)}{d\ln M^2} = \frac{\alpha_s(M)}{2\pi} \int_x^1 \frac{dy}{y} P_{qq}^{(0)}\left(\frac{x}{y}\right) q_{NS}(y, M) = \frac{\alpha_s}{2\pi} \int dz \int dy P_{qq}^{(0)}(z) q_{NS}(y) \delta(x - yz) \equiv \frac{\alpha_s}{2\pi} P_{qq}^{(0)} \otimes q_{NS},$$

(10.42)

which is the so called **Altarelli-Parisi evolution equation** [150] for the dressed NS quark distribution function $q_{NS}(x, M)$. Its derivation is based on two essential ingredients:

- Strong ordering of virtualities in the ladder of Fig. 10.3a.

- The fact that the argument of $\alpha_s(\mu)$ at each vertex of this ladder is given by the upper (i.e. the largest) virtuality. As a consequence the only dependence on M appears in the upper bound on the integration over the largest virtuality τ_1 in each ladder of Fig. 10.4.

Note that there is no longer any trace of parallel logarithms implicitly in (10.40). They have completely disappeared in the process of taking derivative with respect to $\ln Q^2$! Moreover, this equation effectively resums the LL series (10.40). Both of these features closely resemble the transition from eq. (8.28) to (8.35).

A careful reader may ask why in the above evolution equation dy is scaled by y, i.e. appears as dy/y. By definition the generic branching function $P(z)$ describes the distribution of daughter particles carrying fraction z of their mother's particle momentum. There is a simple relation between particle densities in the intervals of z and x, where x is the daughter particle momentum fraction (see Fig. 10.4) when $y \equiv x/z$ is held *fixed*: $d\ln x = d\ln z$. The contribution to the density $D(x)$ of particles having *after* the branching x in the interval $(x, x + dx)$ and coming from the interval $(y, y + dy)$ of the primordial momentum y is equal to

$$x = yz \Rightarrow dx = ydz \Rightarrow D(x) \equiv \frac{dN}{dx} = \frac{1}{y}\frac{dN}{dz} = \frac{1}{y}D(z).$$

(10.43)

So far only the effects of multigluon emission on the nonsinglet quark distribution functions have been considered. For the case of general quark, antiquark and gluon distribution functions other branchings, introduced in the subsection 10.2.2, contribute as well. For instance, the branching function $P_{qG}^{(0)}(z)$, describing the probability to find a quark (of any flavor) inside the gluon, also contributes to the LL evolution

equations for dressed [8] quark/antiquark distribution functions

$$\frac{dq_i(x,M)}{d\ln M} = \frac{\alpha_s(M)}{\pi}\left[\int_x^1 \frac{dy}{y}P_{qq}^{(0)}\left(\frac{x}{y}\right)q_i(y,M)+\int_x^1 \frac{dy}{y}P_{qG}^{(0)}\left(\frac{x}{y}\right)G(y,M)\right], \qquad (10.44)$$

$$\frac{d\overline{q}_i(x,M)}{d\ln M} = \frac{\alpha_s(M)}{\pi}\left[\int_x^1 \frac{dy}{y}P_{qq}^{(0)}\left(\frac{x}{y}\right)\overline{q}_i(y,M)+\int_x^1 \frac{dy}{y}P_{qG}^{(0)}\left(\frac{x}{y}\right)G(y,M)\right], \qquad (10.45)$$

where the gluon distribution function $G(x,M)$ satisfies similar evolution equation

$$\frac{dG(x,M)}{d\ln M} = \frac{\alpha_s(M)}{\pi}\left[\sum_{i=1}^{n_f}\int_x^1 \frac{dy}{y}P_{Gq}^{(0)}\left(\frac{x}{y}\right)(q_i(y,M)+\overline{q}_i(y,M))+\int_x^1 \frac{dy}{y}P_{GG}^{(0)}\left(\frac{x}{y}\right)G(y,M)\right]. \qquad (10.46)$$

Graphical representation of these coupled evolution equations is displayed in Fig. 10.5. The factorization scale M, plays a role similar to that of the renormalization scale μ. In both cases these scales emerged in the process of replacement of the "bare" quantities (color couplant and parton distribution functions respectively) with their dressed counterparts, and in both cases they also lead to unavoidable *ambiguities*. There is, however, an important difference between them, as

- the renormalization scale μ emerged in the process of *ultraviolet* renormalization, i.e. concerns *short* distance properties of the theory, while

- the factorization scale M has been introduced to deal with *parallel* singularities, i.e. concerns *large* distances.

In the above evolution equations M is interpreted merely as the upper bound on parton virtualities included in the definition of dressed parton distribution functions, without specification of its relation to kinematic variables of any physical process. In the case of the DIS it is common to set $M^2 = Q^2$ and thus include in the dressed parton distribution functions gluon emission even if it is actually far from parallel. This is legal to do as there is no sharp dividing line between "small" and "large" virtualities, but for large virtualities the pole term $1/t$, which, after integration, leads to $\ln(Q^2/m^2)$, no longer dominates the exact matrix element for the gluon radiation. In the LL approximation there are no quantitative arguments in flavour of this, or any other, choice of M and *all* values of M are in *principle* equally good.

To solve the AP evolution equations requires the specification of the boundary conditions, which contain information on the x–dependence of parton distribution functions at some *arbitrary* initial scale M_0. The choice of this initial scale is a delicate matter as the results of solving the AP evolution equations truncated to any finite order *do depend* on the choice of M_0, but it is a common practice to choose M_0 in the range of a few GeV. These boundary conditions *cannot* be calculated from perturbative QCD and must be taken from experimental data. The evolution equations therefore determine merely the M–dependence of parton distribution functions, not their absolute magnitude.

There are various methods of solving these evolution equations, briefly surveyed in Section 10.4. The most straightforward of them solves the system of coupled evolution equations by means of sophisticated numerical algorithms on powerful computers. The standard analysis of experimental data on DIS proceeds in the following steps:

- The initial M_0 is chosen.

- Some parameterization of the boundary condition for parton distribution function $D_i(x,M_0), i = q, \overline{q}, G$, for instance,

$$D_i(x,M_0) = A_i x^{\alpha_i}(1-x)^{\beta_i}(1+\gamma_i x) \qquad (10.47)$$

where $A_i, \alpha_i, \beta_i, \gamma_i$ are free parameters, is chosen.

- These parameters, together with the value of the basic QCD parameter Λ, entering $\alpha_s(M/\Lambda)$, are then varied within some reasonable ranges and for each such set the evolution equations (10.44)-(10.46) are solved. This yields $D_i(x,M)$ at all x and M.

- In this way obtained theoretical predictions are *fitted* to experimental data, allowing thereby the determination of the parameters $\Lambda, A_i, \alpha_i, \beta_i, \gamma_i$ for each flavor of the quarks as well as for the gluon.

---

[8]From now on I drop the adjective "dressed" and write the derivative with respect to $\ln M$ instead of $\ln M^2$.

10.3.2 Moments of structure functions and sum rules

The Altarelli–Parisi evolution equation look particularly simple in terms of moments, defined as

$$f(n) \equiv \int_0^1 x^n f(x) \mathrm{d}x. \tag{10.48}$$

Instead of convolutions in (10.44)-(10.46) we get simple multiplications for the moments

$$\frac{\mathrm{d}q_i(n, M)}{\mathrm{d}\ln M} = \frac{\alpha_s(M)}{\pi} \left(P_{qq}^{(0)}(n) q_i(n, M) + P_{qG}^{(0)}(n) G(n, M) \right), \tag{10.49}$$

$$\frac{\mathrm{d}\overline{q}_i(n, M)}{\mathrm{d}\ln M} = \frac{\alpha_s(M)}{\pi} \left(P_{qq}^{(0)}(n) \overline{q}_i(n, M) + P_{qG}^{(0)}(n) G(n, M) \right), \tag{10.50}$$

$$\frac{\mathrm{d}G(n, M)}{\mathrm{d}\ln M} = \frac{\alpha_s(M)}{\pi} \left(P_{Gq}^{(0)}(n) \sum_{i=1}^{n_f} (q_i(n, M) + \overline{q}_i(n, M)) + P_{GG}^{(0)}(n) G(n, M) \right). \tag{10.51}$$

A particularly simple equation holds for moments of the nonsinglet quark distribution function ($a \equiv \alpha_s/\pi$)

$$\frac{\mathrm{d}q_{\mathrm{NS}}(n, M)}{\mathrm{d}\ln M} = \frac{\alpha_s(M)}{\pi} P_{qq}^{(0)}(n) q_{\mathrm{NS}}(n, M) \ \Rightarrow q_{\mathrm{NS}}(n, M) = A_n \left(\frac{ca(M)}{1 + ca(M)} \right)^{-P_{qq}^{(0)}(n)/b}, \tag{10.52}$$

where A_n are *unknown constants*, which play the same role as the boundary condition (10.47) on the distribution functions and which must also be determined from experimental data. Two features of the above solution are worth noting.

- Due to the fact that, trivially, $P_{qq}^{(0)}(0) = 0$ the integral over q_{NS} is M–*independent*.

- Evaluating the moments of $P_{qq}^{(0)}$ for positive n we easily find

$$P_{qq}^{(0)}(n) = \frac{4}{3} \left[-2S_1(n) + \frac{1}{n+1} + \frac{1}{n+2} - \frac{3}{2} \right], \quad S_1(n) \equiv \sum_{k=1}^n \frac{1}{k} = \psi(n+1) + \gamma_E, \tag{10.53}$$

 which implies $P_{qq}^{(0)}(n > 0)$ is **negative**. This fact is also a straightforward consequence of the definition of the "+" distribution. Using the formulae in Appendix (12.4) the above as well as all other relations for the splitting functions can be generalized to the whole complex plane of the variable n. This provides the basis for one of the methods for solving the evolution equations.

Combining the above evolution equations (10.49)-(10.51) we can write down the evolution equation for sum

$$S(n, M) \equiv \sum_{i=1}^{n_f} (q_i(n, M) + \overline{q}_i(n, M)) + G(n, M) = q(n, M) + \overline{q}(n, M) + G(n, M), \tag{10.54}$$

which reads

$$\frac{\mathrm{d}S(n, M)}{\mathrm{d}\ln M} = \frac{\alpha_s(M)}{\pi} \left[\left(P_{qq}^{(0)}(n) + P_{Gq}^{(0)}(n) \right) (q(n, M) + \overline{q}(n, M)) + \left(2n_f P_{qG}^{(0)}(n) + P_{GG}^{(0)}(n) \right) G(n, M) \right] \tag{10.55}$$

Eq. (10.55) acquires particular importance for $n = 1$, since $S(1, M)$ represents the fractional sum of the momenta carried by all partons inside the proton (or other hadrons). As such it should be equal to unity at *any scale M* and therefore the derivative (10.55) should vanish. This happens provided

$$P_{qq}^{(0)}(1) + P_{Gq}^{(0)}(1) = \int_0^1 \mathrm{d}z z \left(P_{qq}^{(0)}(z) + P_{Gq}^{(0)}(z) \right) = 0, \tag{10.56}$$

$$2n_f P_{qG}^{(0)}(1) + P_{GG}^{(0)}(1) = \int_0^1 \mathrm{d}z z \left(2n_f P_{qG}^{(0)}(z) + P_{GG}^{(0)}(z) \right) = 0. \tag{10.57}$$

Straightforward evaluation of the above integrals using explicit formulae (10.29)-(10.32) for the branching functions shows that in QCD these conditions are, indeed, satisfied.
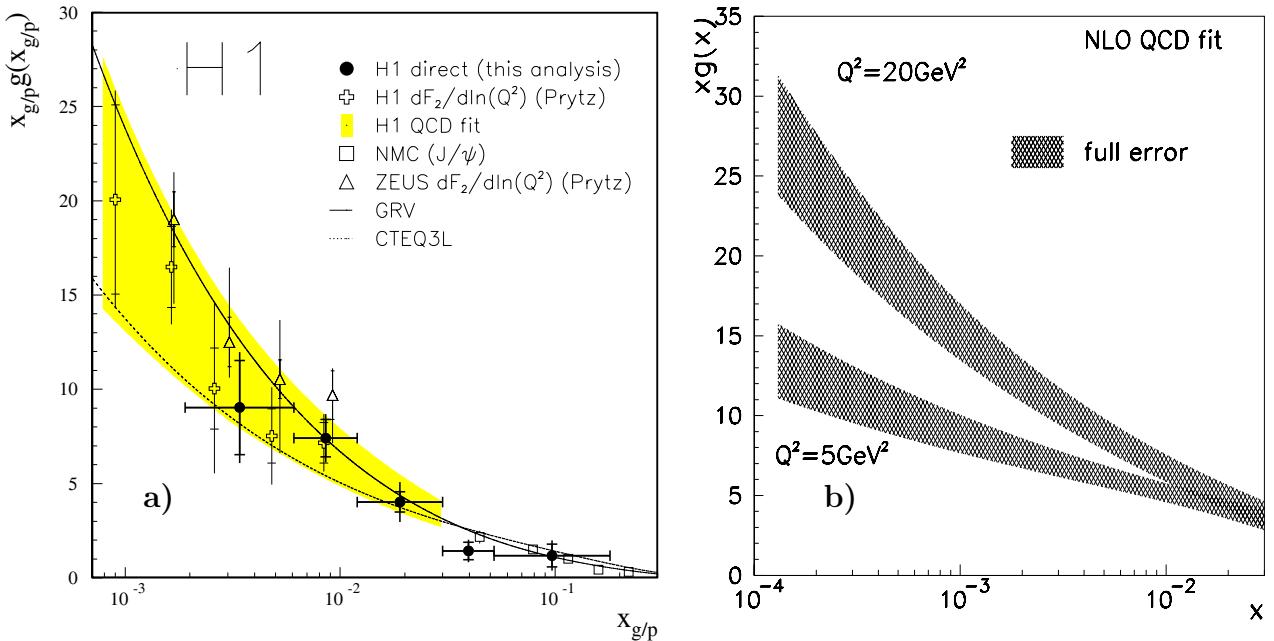
Figure 10.6: a) Gluon distribution functions as measured at HERA an in NMC experiment, using several different methods, including the one based on eq. (10.61) and denoted "Prytz"; b) Gluon density $xG(x, M)$ at two values of the scale $M = Q$, as determined in [152].

10.3.3 Extraction of the gluon distribution function from scaling violations

In deep inelastic lepton–nucleon scattering quark distribution functions are directly measured, while the gluon distribution function enters only indirectly as one of the sources (the other being quark distributions themselves) of scaling violations, described by the evolution equations (10.44) and (10.45). The relative importance of the two branchings $q \to q + G$ and $G \to q + \overline{q}$ depends on x, with former dominant at large x and the latter in small x region, roughly for $x \leq 10^{-2}$. Dropping in small x region the first term in brackets of (10.44) and (10.45), the evolution equations for quark and antiquark distribution functions $q_i(x, M)$, $\overline{q}_i(x, M)$ become identical and read

$$\frac{\mathrm{d}q_i(x, M)}{\mathrm{d}\ln M} = \frac{\mathrm{d}\overline{q}_i(x, M)}{\mathrm{d}\ln M} = \frac{\alpha_s(M)}{\pi} \left[\int_x^1 \frac{\mathrm{d}y}{y} P_{qG}^{(0)} \left(\frac{x}{y} \right) G(y, M) \right]. \tag{10.58}$$

The number of effectively massless quarks to be taken into acount in (10.58) depends on the scale M. For $n_f = 4$, these equations imply the following expression for the derivative of $F_2^{\mathrm{ep}}(x, M)$

$$\frac{\mathrm{d}F_2^{\mathrm{ep}}(x, M)}{\mathrm{d}\ln M} = \frac{\alpha_s(M)}{\pi} \underbrace{\left(2 \sum_{i=1}^{n_f=4} e_i^2 \right)}_{20/9} \int_x^1 \mathrm{d}z \left((x/z)G(x/z) \right) P_{qG}^{(0)}(z). \tag{10.59}$$

Eq. (10.59) is *nonlocal*, in the sense that the value of its left hand side at some x depends on gluon distribution function $G(x)$ in the whole interval $(x, 1)$. In order to get a *local* relation between $\mathrm{d}F_2^{\mathrm{ep}}(x, M)/\mathrm{d}x$ and $G(x)$ Prytz has suggested [151] to expand $H(x/z, M) \equiv (x/z)G(x/z, M)$ around $z = 1/2$, the symmetry point of the branching function $P_{qG}^{(0)}(z) = (z^2 + (1 - z)^2)/2$

$$H(x/z, M) = H(2x, M) + H'(2x, M)(z - 1/2) + H''(2x, M)(z - 1/2)^2/2! + \cdots \tag{10.60}$$

and keeping only the first two terms in (10.60). At low x and for any integrable function the lower integration bound in (10.59) can safely be set to zero. This implies vanishing of the contribution of the second term in

(10.60) and yields

$$\frac{\mathrm{d}F_2^{\mathrm{ep}}(x,M)}{\mathrm{d}\ln M} = \frac{20\alpha_s(M)}{9\pi} 2xG(2x,M) \underbrace{\int_0^1 \mathrm{d}z P_{qG}^{(0)}(z)}_{1/3} = \frac{20\alpha_s(M)}{27\pi} 2xG(2x,M). \tag{10.61}$$

Note that for small x the second term in (10.60), inserted into (10.59), does not contribute after the integration over z. The expansion point has been chosen deliberately at $x_0 = 1/2$ to make this happen. The accuracy of this approximation depends to some extent on the x–dependence of $G(x,M)$, but is typically better than 10 %. The price paid for the simplicity of (10.61) is that in practice the data do not allow us to determine the derivative $\mathrm{d}F_2(x,M)/\mathrm{d}\ln M$ *locally* at each M, but only as some average for the measured range of $M^2 = Q^2$. Consequently we cannot attach in this way extracted gluon distribution function to any *well–defined* scale M. This simple approximate formula has recently been used at HERA to determine average gluon distribution function from scaling violations, see Fig. 10.6a.

One can, however, do better and carry out a complete analysis of data at all x and Q^2 using the full set of coupled AP evolution equations (10.44-10.46). The results of a very recent analysis of this kind made by the H1 Collaboration at HERA [152] is shown in Fig. 10.6b. Note the sizable variation of the gluon distribution function $G(x,Q)$ as Q^2 increases from 5 to 20 GeV!

10.3.4 *Evolution equations at the next-to-leading order

In the LL approximation (10.36) we have kept at each order of α_s only the term with the *same* power of the parallel logarithm $\ln(M^2/m^2)$, coming from the integration over the region of *strongly ordered* parton virtualities. In the **next-to-leading logarithm** (NLL) approximation we extend the integration range by taking into account also the configuration where one of the emissions is not strongly ordered, thereby contributing one power of the parallel log less, i.e. at each order of α_s we include also the terms

$$\alpha_s^k \ln^{k-1}\left(\frac{M^2}{m_g^2}\right). \tag{10.62}$$

The inclusion of these terms entails several novel features but also additional ambiguity.

- The relation between parton distribution functions in the NLL approximation and observable structure functions $F_i(x,Q^2)$ becomes more complicated. This feature is discussed in the next subsection.

- The branching functions P_{ij} become perturbative expansions in powers of the couplant and thus functions of both z and $\alpha_s(M)$

$$P_{ij}(z,\alpha_s(M)) = P_{ij}^{(0)}(z) + \frac{\alpha_s(M)}{\pi} P_{ij}^{(1)}(z) + \cdots. \tag{10.63}$$

  While the LL branching functions $P_{ij}^{(0)}(z)$ correspond directly to basic QCD interaction vertices $qG\overline{q}$ and $3g$, in the NLL approximation, the situation is more complicated. The NLL branching functions $P_{ij}^{(1)}(z)$ are not associated with any single QCD vertex, but come from integration over the transverse momenta of outgoing partons in the whole set of diagrams. For instance the branching function $P_{q_j q_i}^{(1)}$ is associated with both diagrams in Fig. 10.7 and its flavor structure can be decomposed as follows:

$$P_{q_j q_i}^{(1)} = P_{qq}^{(1)V}\delta_{ji} + P_{qq}^{(1)S}. \tag{10.64}$$

- Starting at the NLL approximation we encounter branching functions $P_{\overline{q}_i q_j}$, which describe the probability of finding antiquarks inside quarks and vice versa. They are associated with the diagram in Fig. 10.7a and can be decomposed similarly to (10.64):

$$P_{\overline{q}_j q_i}^{(1)} = P_{\overline{q}q}^{(1)V}\delta_{ji} + P_{\overline{q}q}^{(1)S}, \tag{10.65}$$
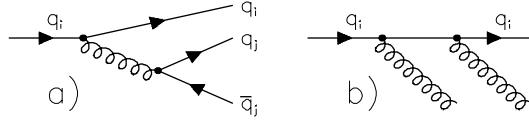
Figure 10.7: Feynman diagrams contributing to the NLO branching functions $P_{q_i q_j}^{(1)}$ and $P_{\overline{q}_i q_j}$.

where the singlet parts in qq and $\overline{q}$q channels, coming from the gluon branching in Fig. 10.7a, *are equal*, while the valence parts, originating from both diagrams, differ

$$P_{qq}^{(1)S} = P_{\overline{q}q}^{(1)S}; \quad P_{qq}^{(1)V} \neq P_{\overline{q}q}^{(1)V}. \tag{10.66}$$

The reason for the latter nonequality is related to the fact that in the case of the diagram in Fig. 10.7a there are, for $i = j$, two *identical* fermions in the final state. The actual techniques (there are several of them) used in evaluation of these branching functions are quite complicated and the reader is referred to a review [154] for details.

- Contrary to the LL branching functions $P_{ij}^{(0)}$, which are *universal*, $P_{ij}^{(1)}(z)$ are in fact *ambiguous*, similarly as the coefficients c_i of the β–function in (8.134). The set of nonuniversal higher order branching functions $P_{ij}^{(k)}(z)$ defines the **factorization scheme**, FS=$\{P_{ij}^{(k)}; k \geq 1\}$, of parton distribution functions, introduced at the beginning of this Chapter. The corresponding evolution equation, below written for the generic nonsinglet distribution function $q_{\mathrm{NS}}(x, M)$,

$$\frac{dq_{\mathrm{NS}}(x, M)}{d \ln M} \equiv \frac{\alpha_s(M)}{\pi} \int_x^1 \frac{dy}{y} q_{\mathrm{NS}}(y, M) \left[P_{\mathrm{NS}}^{(0)} \left(\frac{x}{y} \right) + \frac{\alpha_s(M)}{\pi} P_{\mathrm{NS}}^{(1)} \left(\frac{x}{y} \right) + \cdots \right] \tag{10.67}$$

should therefore be considered as a *definition equation* of $q_{\mathrm{NS}}(x, M)$, similarly as (8.134) is a definition equation of the QCD couplant $\alpha_s(M)$! Factorization schemes that are most frequently used in phenomenological analyses will be specified at the end of the next subsection.

- At the NLL approximation there is a certain subtlety concerning the concept of nonsinglet quark distribution functions. By definition, the NS distribution function $q_{\mathrm{NS}}(x, M)$ does not mix in the evolution equation with any other distribution function, contrary to the general quark and gluon distribution functions, which satisfy *coupled* evolution equations (10.44-10.46). There are, however, two types of quark nonsiglet distribution functions

  1. The **valence** type defined as

$$q_{\mathrm{NS},i}^{(-)} \equiv q_i - \overline{q}_i; \quad i = u, d, s, \cdots, \tag{10.68}$$

  2. The **nonvalence** type

$$q_{\mathrm{NS}}^{(+)} \equiv q_i + \overline{q}_i - \frac{1}{n_f} \sum_{k=1}^{n_f} (q_k + \overline{q}_k), \tag{10.69}$$

  which for $n_f = 2$ reduces to

$$u^{(+)} = \frac{1}{2} \left(u + \overline{u} - d - \overline{d} \right) = \frac{1}{2} \left[(u - \overline{d}) - (d - \overline{u}) \right]. \tag{10.70}$$

In the LL approximation antiquark distribution functions are generated exclusively by gluon splitting and thus we expect $\overline{u} = \overline{d}$. Consequently (10.70) coincides with the difference $\frac{1}{2}(u^{(-)} - d^{(-)})$ of the valence-like NS distributions. In the NLL approximation the $q^{(+)}$ and $q^{(-)}$ NS quark distribution functions evolve according to *different* kernels, expressed as combinations of $P_{qq}^{(1)V}$ and $P_{\overline{q}q}^{(1)V}$:

$$P_{(-)}^{(1)} \equiv P_{qq}^{(1)V} - P_{\overline{q}q}^{(1)V}, \tag{10.71}$$

$$P_{(+)}^{(1)} \equiv P_{qq}^{(1)V} + P_{\overline{q}q}^{(1)V}. \tag{10.72}$$

187

In factorization schemes used in phenomenological analyses $P_{\bar{q}q}^{(1)V}$ is numerically very small and thus the difference between these two types of NS distributions practically negligible. Converted into moments the NLL NS evolution equation (10.67) implies the analogous equation for the moments

$$\frac{\mathrm{d}q_{\mathrm{NS}}(n, M)}{\mathrm{d}\ln M} = \frac{\alpha_s}{\pi} q_{\mathrm{NS}}(n, M) \left[P_{\mathrm{NS}}^{(0)}(n) + \frac{\alpha_s(M)}{\pi} P_{\mathrm{NS}}^{(1)}(n) \right], \tag{10.73}$$

which can easily be solved, yielding

$$q_{\mathrm{NS}}(n, M) = A_n \left[\frac{ca(M)}{1 + ca(M)} \right]^{-P_{\mathrm{NS}}^{(0)}(n)/b} (1 + ca(M))^{-P_{\mathrm{NS}}^{(1)}(n)/bc}, \tag{10.74}$$

where $a(M) \equiv \alpha_s(M)/\pi$. Note that the NLL correction term, containing $P_{NS}^{(1)}(n)$, does not influence the asymptotic behavior of (10.74) as $M \to \infty$, which is therefore *unique*

$$q(n, M) \xrightarrow[M \to \infty]{} A_n \left[ca(M) \right]^{-P_{NS}^{(0)}(n)/b} \tag{10.75}$$

and fully determined by the constants A_n.

10.3.5 *Hard scattering and the factorization theorem

Before closing this section I still want to discuss one important aspect of the definition of dressed distribution functions, concerning the so called **factorization theorem**. This theorem [153] guarantees that all parallel logarithms, uncanceled via the KLN theorem if only **degenerate final states** are summed over, can be systematically, i.e. to all orders and in all processes absorbed (factorized) into dressed parton distribution functions. We saw this mechanism at work in the LL approximation, where it was rather transparent. Its proof to all orders is, however, much more complicated.

Le us again restrict ourselves to the NS channel. The basic idea of factorization is to split the range of virtualities $\tau \equiv |t|$ corresponding to Feynman diagram in Fig. 9.2a into two parts:

- $\tau \leq \tau_0$: the integral over the corresponding (nearly) parallel configurations is *absorbed* in the *dressed* parton distribution functions $q_{\mathrm{NS}}(x, \sqrt{\tau_0})$ and similarly for quark and gluon distribution functions.

- $\tau > \tau_0$: Integration over this range yields *finite* result $C_{\mathrm{NS}}(z, Q/M, \mathrm{FS})$, which is called **hard scattering cross–section** in the NS channel and depends, beside $Q^2, M^2 = \tau_0$ and z also on the FS of parton distribution functions, specified at the NLL by the branching functions $P_{ij}^{(1)}(z)$, and the hard scattering scale μ, which in general is different from the factorization scale M.

The hard scattering cross–section is itself given as perturbation expansion in powers of $\alpha_s(\mu)$

$$C_{\mathrm{NS}}\left(z, \frac{Q}{M}, \mathrm{FS} \right) = \delta_{\mathrm{NS}}^{(k)} \left[\delta(1 - z) + \frac{\alpha_s(\mu)}{\pi} C_{\mathrm{NS}}^{(1)}\left(z, \frac{Q}{M}, \mathrm{FS} \right) + \cdots \right], \tag{10.76}$$

where the superscript "(k)" distinguishes between different NS channels and the quantities $\delta_{\mathrm{NS}}^{(k)}$ include electromagnetic (or weak) couplings of quarks. The leading term in (10.76) corresponds to the LL approximation in which the measurable structure functions are given as quark distribution functions, weighted by squares of their electric charges, maintaining thus the basic QPM relation (5.81). According to the factorization theorem a generic NS structure function $F_{\mathrm{NS}}(x, Q^2)$ can be written as the convolution

$$F_{\mathrm{NS}}(x, Q^2) = \int_x^1 \frac{\mathrm{d}y}{y} q_{\mathrm{NS}}(y, M, \mathrm{FS}) C_{\mathrm{NS}}\left(\frac{x}{y}, \frac{Q}{M}, \mathrm{FS} \right), \tag{10.77}$$

where all the dependence on the external momentum Q resides in the hard-scattering cross–section C_{NS}. Note that both $q_{\mathrm{NS}}(x, M, \mathrm{FS})$ and $C_{\mathrm{NS}}(z, Q/M, \mathrm{FS})$ depend on the factorization scale M, as well as on the factorization convention FS. The factorization theorem not only guarantees that all parallel singularities can be absorbed in q_{NS}, but it also tells us that the dependence on unphysical quantities M and higher order branching functions $P_{ij}^{(k)}$, which define the FS, *cancel* in the convolution (10.77), provided perturbative expansions in (10.76) as well as (10.67) are summed *to all orders*.

For the NS case the cancelation of the dependences of $q_{\rm NS}$ and $C_{\rm NS}$ on the FS is guaranteed by the following relation between the NLL term $C_{\rm NS}^{(1)}$ in (10.76) and the NLL branching function $P_{\rm NS}^{(1)}$

$$C_{\rm NS}^{(1)}\left(z,\frac{Q}{M},{\rm FS}\right)=\delta_{\rm NS}^{(k)}\left(P_{\rm NS}^{(0)}(z)\ln\frac{Q}{M}+\frac{P_{\rm NS}^{(1)}(z)}{b}+k(z)\right), \tag{10.78}$$

which expresses $C_{\rm NS}^{(1)}$ explicitly as a function of $P^{(1)}$ and where $k(z)$ is a function of z, *independent* of M and $P_{\rm NS}^{(1)}$. In terms of moments the relations (10.77) and (10.78) read

$$C_{\rm NS}^{(1)}\left(n,\frac{Q}{M},{\rm FS}\right)=\delta_{\rm NS}^{(k)}\left(P_{\rm NS}^{(0)}(n)\ln\frac{Q}{M}+\frac{P_{\rm NS}^{(1)}(n)}{b}+k(n)\right) \tag{10.79}$$

and

$$F_{\rm NS}(n,Q^2)=q_{\rm NS}(n,M,{\rm FS})C_{\rm NS}\left(n,\frac{Q}{M},{\rm FS}\right). \tag{10.80}$$

A careful reader may be confused by the fact that while the factorization scale M appears in both $q_{\rm NS}$ and $C_{\rm NS}$, the hard scattering scale μ is written in (10.76) only on its r.h.s.. The point is that the renormalization scale μ enters only in the moment we write $C_{\rm NS}$ as perturbative expansion in some RS. The separation of small and large distances, embodied in the convolution (10.77), is, on the other hand, quite general statement, valid *beyond* perturbation theory. The physical quantity $F_{\rm NS}(x,Q^2)$ in (10.77) depends *neither* on the factorization scale M, *nor* on the renormalization scale μ, but while the compensation mechanism for the former involves *both* $q_{\rm NS}(x,M)$ and $C_{\rm NS}(z,Q/M)$, it operates for the former *inside* the hard scattering cross–section $C_{\rm NS}(z,Q/M)$ *only*. This difference is expressed by the fact that while the explicit dependence of the coefficient functions $C_{\rm NS}^{(k)}$ on M starts already at the leading order $C_{\rm NS}^{(1)}$, the dependence on μ appears first for $k=2$. So in the NLL approximation there is no quantitative suggestion what to take for μ, while the explicit dependence (10.78) suggests to set $M \propto Q$. In practical applications one often identifies $\mu = M$, but besides simplifying the analysis there is no convincing argument for this assumption.

Although there is an innumerably infinite set of possible factorization schemes, all phenomenological analyses carried out in the NLL approximation so far have used one of the following two factorization schemes FS, fixed by the specification of the NLL branching function $P_{\rm NS}^{(1)}$:

1. The **"Universal"** FS, in which some of the first theoretical calculations were carried out [154]. In this FS both $P_{\rm NS}^{(1)}$ and $C_{\rm NS}^{(1)}$ are *nonzero*.

2. The **"DIS"** FS, proposed in [155], in which $C_{\rm NS}^{(1)}=0$ *by definition*.

There is yet another, theoretically well–motivated FS, discussed in [156], where, $P_{\rm NS}^{(1)} = 0$ by definition. Compared to the dependence on factorization scale M, phenomenological consequences of the enormous freedom in the choice of FS are largely unexplored.

10.4   Brief survey of methods of solving the evolution equations

In this section I briefly review some of the methods used in solving the evolution equations for quarks and gluon distribution functions. Some of them are now of historical importance only, but I think it is nevertheless useful to be aware of them. I shall not go into details, but after general characterization merely provide references to the original literature. All the methods use the boundary condition in the form of the specification of quark and gluon distribution functions at some initial scale M_0. I shall furthermore concentrate on the case of the non–singlet quark distribution functions. However, all the methods can be extended to include singlet quark and gluon distribution functions as well.

Buras–Gaemers method [77]

This has been historically the first method used in solving the evolution equations in the LO (10.42). It starts by assuming the initial conditions on the nonsinglet, i.e. valence $u_{\rm v}$ and $d_{\rm v}$ quark distribution functions in the form

$$q_{\rm v}(x,M_0)=A^q x^{\eta_1^q}(1-x)^{\eta_2^q}, \tag{10.81}$$

which implies for the moments

$$q_{\mathrm{v}}(n, M_0) = A^q \frac{\Gamma(\eta_1^q + n + 1)\Gamma(\eta_2^q + 1)}{\Gamma(\eta_1^q + n + \eta_2^q + 2)}. \tag{10.82}$$

Using these initial conditions in the evolution equations (10.42) for the moments $q_{\mathrm{v}}(n, M)$ yields explicit analytic expressions for their scale dependence. Note, however, that the simplicity of these explicit solutions is lost when we attempt to translate them back into the x–space. To keep the solutions in x–space simple, Buras and Gaemers therefore assumed that these solutions can be reasonably approximated *at all M* by the form (1.15) in which the coefficients A^q, η_1^q, η_2^q are allowed to depend on M. In particular they assumed the following simple dependence on M ($B(x, y)$ is the Euler function)

$$\eta_i^q(s) = \eta_i^q + \eta_i'^q s, \ i = 1, 2; \quad A^q = \frac{\delta_q}{B(\eta_1^q(s), 1 + \eta_2^q(s))}, \quad s \equiv \ln\left[\frac{\ln(M/\Lambda)}{\ln(M_0/\Lambda)}\right], \tag{10.83}$$

where $\delta_u = 2$ and $\delta_d = 1$. These later conditions guarantee the fundamental QPM sum rules, like the GLS sum rule (5.121). The variable s defined in (10.83) is usually called the "evolution variable". The coefficients $\eta_i^q, \eta_i'^q$ were determined by fitting, for a chosen initial M_0, the approximation based on (10.83) to the exact LO solutions for the first 10 moments. Although this approximative procedure is strictly speaking incompatible with the exact solutions of the evolution equation (10.42), it is extremely simple and was therefore widely used in the early eighties to describe the scaling violations in nucleon structure functions. In view of the accuracy of the data available at that time it was quite adequate. The parameter Λ appearing in (10.82) can not be directly associated with any renormalization scheme and is usually denoted Λ_{LO}. A number of experiments used this method to determine Λ_{LO} from fits to experimental data. Although this method is no longer used in the described form, a modern version thereof is employed by the CTEQ group [158].

Bernstein polynomials method [159]

This method was meant to replace the too crude Buras–Gaemers approach. It employs the expansion in the set of Bernstein polynomials, but is so technical that it has never really taken roots. I mention it for completeness only.

Laguerre polynomials method [160]

The authors of [160] have developed an elegant method of solving the evolution equation (10.42) in an analytical form using the expansion in the set of Laguerre polynomials. This method is in principle arbitrarily accurate, but in practice the expansions must, of course, be truncated. It had been used in the middle eighties by several experimental groups in analyzing their data, but as it is rather cumbersome to implement (it employs many special functions and infinite sums) and also its accuracy is difficult to control it soon gave way to the most straightforward way of solving the evolution equation, i.e. numerical integration of evolution equations.

The numerical integration method [161]

Starting with the paper [161] sophisticated numerical methods of solving the evolution equation have been developed. Due to the immense increase of the speed and capacity of modern computers the original limits of this method have disappeared and now most of the groups use these methods. The numerical routines are usually made available by their authors, so that it is relatively easy and straightforward to use them even for nonexperts and experimentalists.

The Jacobi polynomials method [162, 163]

Despite the overall superiority of the numerical integration methods there is, however, one method of solving the evolution equations, which besides being conceptually very simple, has one crucial advantage over them. This technique is based on the idea [162, 163] of expanding the convolution (10.77) in the set of Jacobi polynomials

$$\Theta_k^{\alpha\beta}(x) \equiv \sum_{j=0}^k c_{kj}^{\alpha\beta} x^j, \tag{10.84}$$

orthogonal on the interval $(0,1)$ with the weight $x^\alpha(1-x)^\beta$

$$\int_0^1 \mathrm{d}x x^\alpha (1-x)^\beta \Theta_k^{\alpha\beta}(x)\Theta_l^{\alpha\beta}(x) = \delta_{kl}. \tag{10.85}$$

The nonsinglet structure function $F_{\mathrm{NS}}(x,Q^2)$ can be expanded

$$F_{\mathrm{NS}}(x,Q^2) = x^\alpha(1-x)^\beta \sum_{k=0}^\infty \Theta_k^{\alpha\beta}(x)a_k^{\alpha\beta}(Q^2) \tag{10.86}$$

in terms of the Jacobi moments $a_k^{\alpha\beta}(Q^2)$, which, in turn, are given as linear combinations of conventional moments $F_{\mathrm{NS}}(j,Q^2)$:

$$a_k^{\alpha\beta}(Q^2) \equiv \int_0^1 \mathrm{d}x F_{\mathrm{NS}}(x,Q^2)\Theta_k^{\alpha\beta}(x) = \sum_{j=0}^k c_{kj}^{\alpha\beta} F_{\mathrm{NS}}(j,Q^2). \tag{10.87}$$

Substituting (10.74) and (10.79) into (10.80) and this subsequently into (10.87) and (10.86), we get an explicit expression for $F_{\mathrm{NS}}(x,Q^2)$ as a function of the unknown constants A_n and parameter the $\Lambda_{\overline{\mathrm{MS}}}$.

Instead of A_n we can also start with some initial distribution, like (10.81), specifying $q_{\mathrm{NS}}(x,M_0)$ at some M_0, and use it to evaluate the constants A_n by means of the expression (10.74). Having determined A_n we can then proceed as outlined before. Technical details of this procedure, like the question of the number of terms that must be taken into account in (10.86) to obtain a required precision etc., are described in [163]. The superiority of this method becomes apparent when we address the problem of transforming parton distribution functions between different factorization schemes [156].

Over the last decade a large number of phenomenological analyses of experimental data have been carried out and parton distribution functions of nucleons, pions and recently also photons, determined. In the recent years several groups of theorists and phenomenologists have been systematically improving these analyses by incorporating new and more precise data as well as by employing more sophisticated theoretical methods. As a result there is now a flourishing market with parton distribution functions, some of which are shown in Fig. 5.9. All these distribution functions are now available, as functions of both the momentum fraction x and the factorization scale M, in the computerized form in the CERN library **PDFLIB**.

The inverse Mellin transformation method [165–167]

As shown above the evolution equations are much simpler in the momentum space, where they allow for the solution in closed analytical form. To get back to the x−space, where the measurements are done, we can use the inverse Mellin transformation which reads in general

$$f(x) = \frac{1}{2\pi\mathrm{i}} \int_C x^{-z} f(z)\mathrm{d}z, \tag{10.88}$$

where the function $f(z)$ of the momentum variable z must be known in the whole complex plane and the integration contour C runs to the right of all singularities of the integrand. Beside the splitting functions also the boundary conditions, like (10.47), on the parton distribution functions must be known for any complex momentum variable z. This restriction limits in principle the applicability of the method but in practice it presents no obstacle.

This relation provided the basis for a very efficient method of solving the evolution equations introduced in [165]. For a recent thorough discussion of various technical aspects of this method see, [166, 167].

10.5   Exercises

1. Argue on physical terms why $P_{GG}^{(0)}(x)$ must contain the "+" distribution.

2. Show explicitly that the integration over the region of unordered virtualities doesn't produce the leading logs.

3. Derive (10.37) and evaluate the integrals in (10.38), neglecting the dependence of α_s on τ.

4. Show what results if the argument μ of $\alpha_s(\mu)$ were the same and equal to M^2 for all the vertices in the ladder of Fig. 10.4a.

5. Show that the NS combinations in (10.69) and (10.68) do not mix under branching.

6. Working in the LL approximation and assuming the parametrization (10.47), express the constants $A(n)$ as functions of the parameters A, α, β, γ, setting for simplicity $\gamma = 0$.

7. Determine the value of the coupling δ_{NS} in (10.76) for the NS structure function defined as $\frac{1}{2}(F_2^{ep} - F_2^{en})$.

8. Derive the WW approximation for the scalar charged particle. Show that the corresponding branching function $P_{\gamma e}(x) \propto \frac{1-x}{x}$.

Chapter 11

Particle interactions as seen in modern detectors

In this concluding chapter I will demonstrate, on several examples from current experiments at DESY, CERN and Fermilab, what kind of observation these experiments provide and how do these measurements corroborate the general picture of the hard scattering processes discussed in previous chapters. In particular, we shall see clear emergence of jets of hadrons, which will be interpreted as "traces" of the underlying parton dynamics. In getting proper understanding of the jets, computer aided visual representation of the data taken by present generation of detectors plays an indispensable role. I believe that the "pictures" taken by the modern detectors are nice by themselves and provide intuitively compelling evidence for some measure of reality of quarks and gluons. Moreover, the concept of jets is by now so well established both theoretically and experimentally that jets are no longer only objects of study, but have become a *tool* for investigating other phenomena. Three examples of this new role of jets, discussed bellow, are

- the discovery of the top quark at Fermilab,

- signals for the substructure of quarks, also at Fermilab, and

- first observation of pair production of W bosons at LEP.

As there is no place in this kind of text for a detailed description of the detectors, I shall mention merely those elements of their structure that are crucial for the interpretation of the measurements. Moreover only the **H1** detector at DESY will be described as the basic structural elements of all modern detectors are essentially the same, differing more in materials used for their construction than in principles of their operation. On a few beautiful examples of real events from this detector I shall illustrate several important properties of jets, discussed in Chapter 9.

11.1   The H1 experiment at DESY

Let me first point out the basic difference of a modern detector, like the H1, compared to the pioneering experiments with the deep inelastic scattering of electrons on nucleons, carried out at SLAC in the late sixties and discussed in Chapter 5. In these earlier detectors, only the scattered electron (or positron) was detected and its momentum analyzed by means of the so called "single arm" spectrometers. The "arm" itself had a narrow aperture, determining the scattering angle of the electron, and could be continuously rotated around the interaction point thereby allowing a precise angular measurement. Inside the arm a stack of electronic devices provided a measurement of the energy of the scattered electron from the curvature of its track in the magnetic field. Knowing the (polar) angle and energy of the scattered electron it is easy to evaluate all the kinematical quantities, describing the inclusive process (5.1). No hadrons were detected in these experiments and, indeed, the study of the hadronic system, accompanying the scattered electron, was not considered particularly interesting at that time.

The situation changed with the advent of QCD, although even in the middle seventies the emphasis of a new generation of experiments, using high energy (up to 280 GeV) beams of muons at CERN and Fermilab,
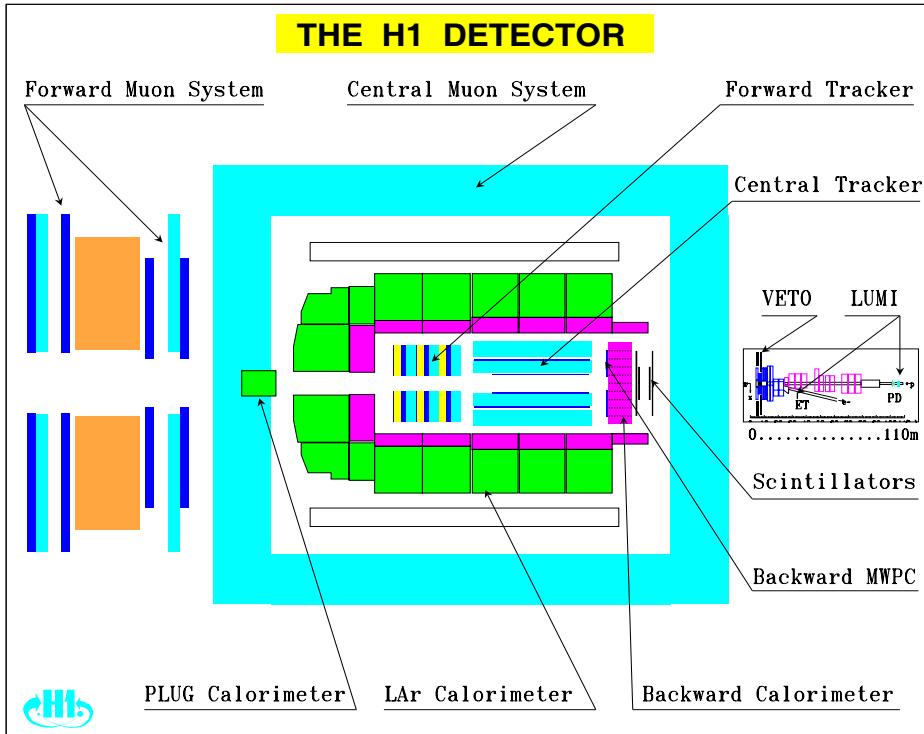
Figure 11.1: The longitudinal cross–section of the **H1 detector** at DESY. The detector is about $10 \times 5 \times 5$ meters large and weighs over 5000 tons.

was on the precise measurement of the scattered muons with as high Q^2 as possible. The first experiment which undertook a serious study of the accompanying hadrons was the EMC Collaboration at CERN. At the present time all new experimental facilities, using stationary targets or colliders, have very good capability to detect and analyze these hadrons, as the investigation of final state hadronic system reveals, among other things, the presence of **jets**, traces of the underlying quark and gluon dynamics. In Fig. 11.1 the schematic layout of the H1 detector is shown. The electrons come from left, the protons from right and collide in the middle of the central tracker. Charged particles leave tracks in the central or forward tracker and all hadrons, charged as well as neutral, deposit part or all of their energies in the electromagnetic and/or hadronic parts of the calorimeter. Particles closer to the beams than about $7°$ disappear in the beam tube and are unobservable in this detector. Its basic structural elements are

Inner trackers: the central and forward trackers are systems of drift chambers providing the tracking of charged particles and for slower particles also their energy measurement.

Calorimeter: the basic principle of calorimeters is simple: particles, charged as well as neutral, passing through some ionizing medium (like liquid argon, various gases etc.) deposit, due to their collisions with atoms of this medium, part or all of their energy therein. This deposited energy can be collected and measured and from the known relation between the collected energy and the primary energy of the passing particle, the latter can be determined. All modern calorimeters are composed of two parts:

- **the inner electromagnetic part**, where electrons and photons deposit practically all their energy and their energy is measured,

- **the outer hadronic part**, which is immediately behind the electromagnetic one and where hadrons deposit a large fraction of their energy.

Muon systems: made of slabs of steel detect the muons (in most of the solid angle) which penetrate without significant attenuation the whole calorimeter. The capability to detect muons is crucial for the observability of certain very important processes, like the production of the J/ψ particle (see below).

The calorimetric measurement of energy has many advantages but, naturally, also some shortcomings. Its errors, which depend on various details of its structure and the primary energy, are given by the following

194
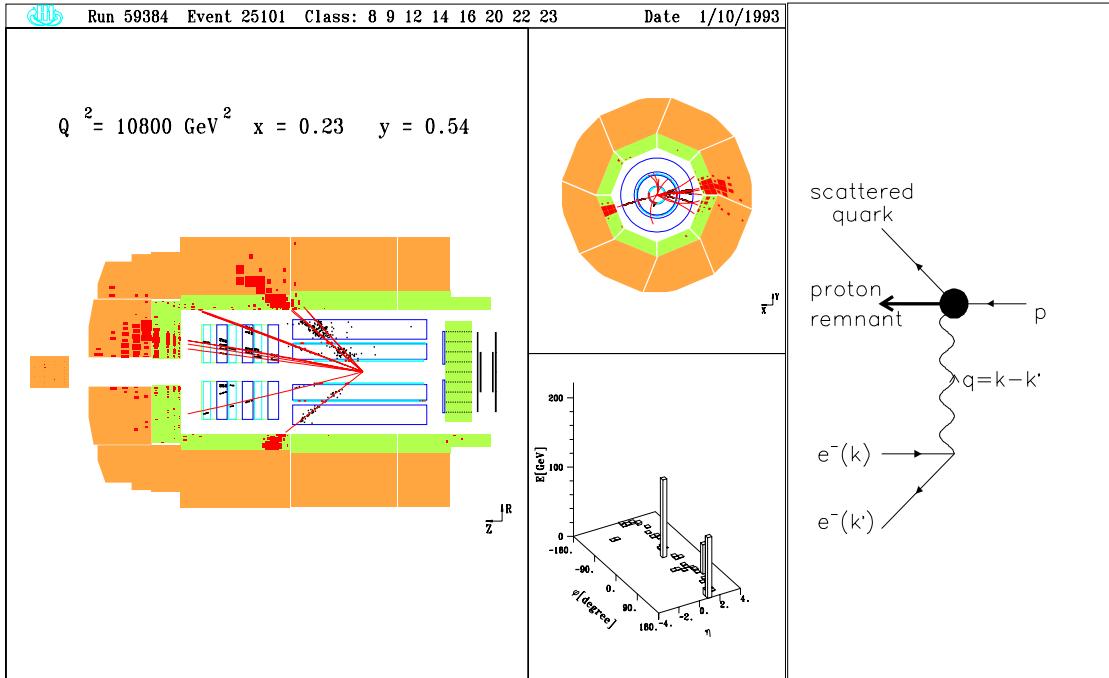
Figure 11.2: A typical DIS event as seen in the H1 detector and the corresponding Feynman diagram describing its mechanism, drawn in a way to reflect the kinematics of the event on the left.

approximate empirical formula

$$\Delta E/E \doteq A/\sqrt{E} + B, \qquad (11.1)$$

where the constants A, B characterize the quality of the detector. An important property of the calorimeters is their angular segmentation. The calorimeters actually measure the "energy flow" in a certain solid angle, in the case of the H1 detector of the order of 10^{-3} steradians. This is crucial for the observability of jets, as will be shown below. The calorimetric measurement is practically the only way how to measure energies of fast particles, which bend only very little in the magnetic field. Moreover it also avoids the problem with the separation of charged tracks when they are close to each other and thus practically indistinguishable in the track chambers.

Out of many results obtained so far with the H1 detector I shall discuss in more detail several phenomena, which illustrate some of the statements made in the chapters on the parton model and QCD.

11.1.1 Manifestation of the jet–like structure of final states in DIS

The fact that calorimeters do not "see" individual particles, but measure only the energy flow into individual angular "towers" might seem at first glance as a major disadvantage compared to the detection of individual particles in track chambers, but it turns out that this is in fact not so. The point is that also the theory (i.e. perturbative QCD) can reliably calculate *only the energy flow of partons* into a definite (but in principle arbitrary) cone (angular or other, see Chapter 9). As emphasized in Chapters 9–10, hadronization, i.e. the conversion of quarks and gluons into observable hadrons, is theoretically so far incalculable and models have to be invented to describe this stage of collisions. The fundamental assumption of all these models is that this hadronization is **soft**, i.e. *does not* significantly distort the energy flow of final state particles produced in some hard collision! In other words they assume that the energy flow calculated on the level of partons is (almost) equal to the energy flow of the measured hadrons. In particular this property gives meaning to the important concept of **jets**.

On the other hand all such more detailed quantities like the number of particles in the final state, their species etc., are much more sensitive to the concrete mechanism of hadronization and are currently
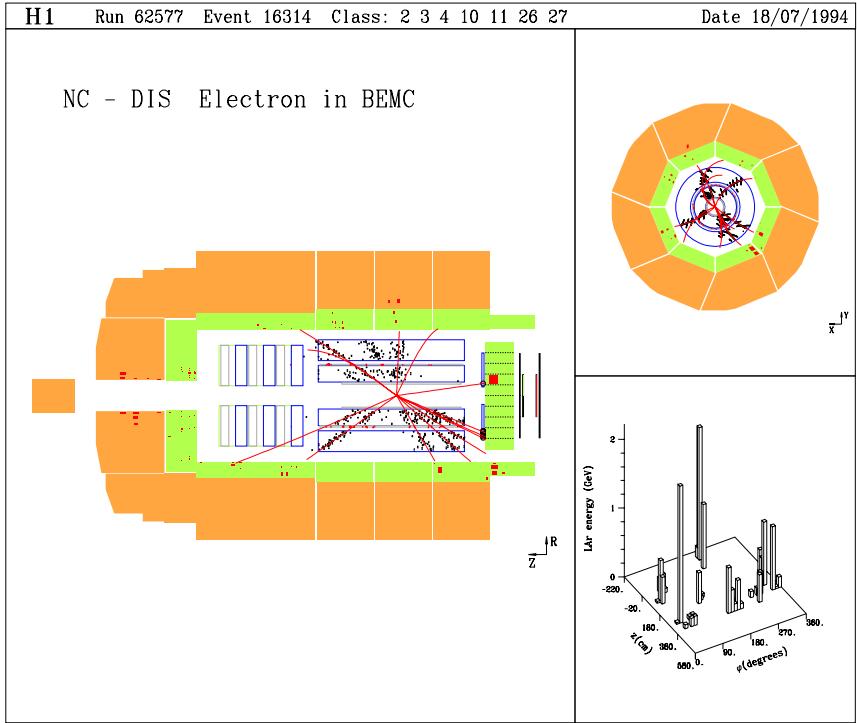
Figure 11.3: An example of an "unjetty" event in DIS scattering.

not calculable in perturbative QCD. The fact that hadronization does, indeed, have this property is nicely illustrated on an example of a typical DIS event in the H1 detector shown in Fig. 11.2. The mechanism of the DIS was discussed at length in Section 5.6 and described in Fig. 5.7. In the right part of Fig. 11.2 this diagram is redrawn in such a way as to reflect the geometry of the collision shown on the left. There we see the computer aided reconstruction of one typical DIS event. It consist of two projections, lateral (left) and azimuthal (upper right), of energy depositions (indicated by dark hatching) in the electromagnetic (inner) and hadronic (outer) parts of the calorimeter. The magnitude of energy depositions, and thus also of the corresponding energy flow, are proportional to the areas of the dark squares. The curves shown in the track chamber represent computer aided interpolation of the tracks of those charged particles whose tracks were completely reconstructed from the hits in the individual chambers.

The reconstructed energy flow as a function of the azimuthal (ϕ) and polar angles (ϑ), or more exactly the so called pseudorapidity $y = -\ln(\tan(\vartheta/2))$, displayed in the so called **lego plot** in lower right part of Fig. 11.2, gives intuitively the best picture of the whole event. In this particularly clean event we clearly identify two pronounced peaks of about 100 GeV each above the "silent" towers in the lego plot, one of them corresponding to the scattered electron and the second to the jet coming from the hadronization of the struck quark. Note that the scattering is so violent ($Q^2 = 10800$ GeV^2), that the electron is actually backscattered in HERA laboratory frame (this is, of course, due in part to the fact that in this frame the proton has much higher energy than the electron)!

In the lateral (but not azimuthal, for reasons see exercise) projection also the proton remnant jet is clearly visible at small angle with respect to the incoming proton. In most DIS events, however, the proton remnant jet is at so small polar angle that most of it disappears in the beam tube. The lego plot gives us also the idea how narrow the jets are. From the lateral projection it is obvious that if not the energy, but rather the particles flow were plotted, the lego plot would look quite different, with many more towers giving appreciable signal.

The clear two jet structure of the typical DIS events provides a convincing evidence for some measure of "existence" of quarks (and gluons). There is simply no other viable explanation for the frequent appearance of such angularly collimated streams of hadrons. The probability that such a configuration will occur incidentally is utterly negligible. Although we don't "see" directly the quarks and gluons as individual free particles in the same way as we "see" electrons etc., the jets are now considered by most particle physicists as unambiguous imprints of the underlying partons. Physical relevance of jets is underscored by the fact that their properties are independent of the process in which they are produced. On the other hand not all events in which jets are produced look as nice as those in Fig. 11.2 and there are events in which no jets are
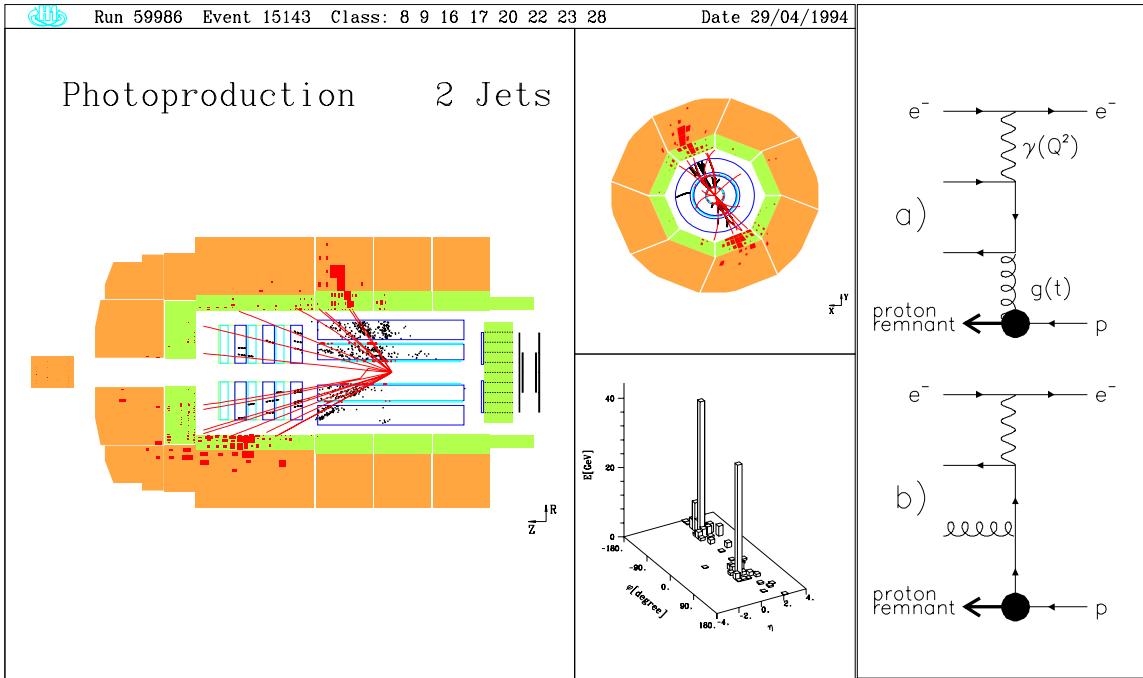
Figure 11.4: Jet photoproduction as seen in the H1 detector and Feynman diagrams describing its mechanism in two possible channels.

present at all. An example of an "unjetty" event is shown in Fig. 11.3. The contrast with the lego plot of Fig. 11.2 is evident: the towers are much smaller, typically of 1 GeV only and there are many of them of about equal height.

11.1.2 Photoproduction of jets

The observation of jet structure of typical events in DIS, illustrated in Fig. 11.2 corroborates the basic idea of the QPM, i.e. the separation of hard collisions into two relatively distinct stages, hard scattering on the parton level and the hadronization, but provides only an indirect evidence for the validity of QCD itself. A simple direct evidence of this kind is provided by the observation of two additional (to the proton remnant) jets in the photoproduction, the so called 2+1 jet events. [1]

The mechanism of the photoproduction of jets in lowest order perturbative QCD is described by the Feynman diagrams of Fig. 11.4, which also displays one clean example of such an event. The term "photoproduction" is a synonym for the kinematical region of *small virtuality* of the exchanged photon in Fig. 11.4. Consequently, there is in fact no principal difference between the deep inelastic ep scattering and the photoproduction in ep collisions. Recall that the measure of "smallness" of the virtuality of the exchanged photon is nothing absolute, but, as emphasized in section 10.2.1, depends on the type and kinematics of the process in lower vertex of Fig. 10.2b. In fact the lower vertices in Fig. 11.4 are just two specific examples of such a process. The condition for the validity of the WW approximation in this case reads $Q^2 \ll 4p_T^2$, where p_T is the magnitude of the transversal (with respect to the incoming virtual photon) momentum of the outgoing partons ($q\bar{q}$ or qG).

The event in Fig. 11.4 contains two clear jets, produced at large angles, and essentially no trace of either the proton remnant jet or the scattered electron. The proton remnant jet disappears, as in most cases, in the beam tube and so does in this event also the scattered electron as small virtuality of the exchanged photon implies small scattering. It is impossible to assign one particular event, like that in Fig. 11.4, to one of the two possible channels ($\gamma + q \rightarrow G + q$ or $\gamma + G \rightarrow q\bar{q}$) it corresponds. We can, however, measure

---

[1]We could take the 2+1 jet events in DIS as well, but as the photoproduction of jets has much higher cross–section, the cleanest 2+1 jet events come from this process. Moreover it provides the simplest example of the application of the Weizsäcker–Williams approximation to specific final states of the general process (10.5).

197
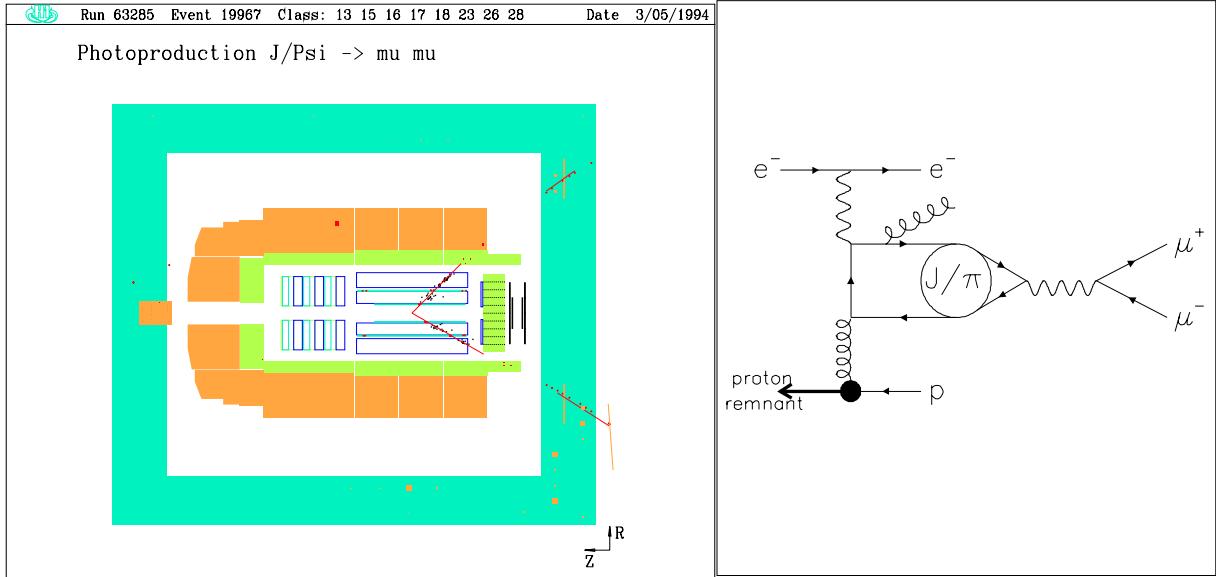
Figure 11.5: J/ψ production in the H1 detector and the mechanism of its production and decay. The big open circle stands for the J/ψ wave function at the origin.

experimentally the distributions of the pseudorapidity y and the transverse momentum p_T of the outgoing jets and compare them with the predictions of perturbative QCD, where both of the channels contribute

$$\frac{\mathrm{d}\sigma}{\mathrm{d}y\mathrm{d}p_T} = \int_0^1 \mathrm{d}x G(x, M)\sigma(\mathrm{e}^- G \to \mathrm{e}^- q\overline{q}; x, y, p_T) + \sum_{i=1}^{2n_f} \int_0^1 \mathrm{d}x q_i(x, M)\sigma(\mathrm{e}^- q \to \mathrm{e}^- q G; x, y, p_T). \quad (11.2)$$

While the individual events, like that of Fig. 11.4 provide clear confirmation for the concept of jets, the comparison of data with QCD predictions (11.2) allows us to go further and determine the gluon distribution function in the proton! Note that this quantity is not directly measurable in DIS as gluons carry no electric charge and $G(x, M)$ enters only via the evolution equations (10.44) for the quark distribution functions. On the other hand, in the case of the quantity (11.2) the gluon distribution function $G(x, M)$ enters directly the appropriate convolution!

11.1.3 Photoproduction of the J/ψ particle

One of the cleanest type of events observed in the H1 detector is the photoproduction of the J/ψ particle, the vector meson interpreted as the bound state of the $c\overline{c}$ system. The production mechanism is the same as in the upper right diagram of Fig. 11.4, i.e. photon–gluon fusion gives rise to a $c\overline{c}$ pair, which can materialize either as a pair of charmed particles, predominantly charmed mesons $D, \overline{D}$, or as the J/ψ particle (or eventually its radial excitations). In the latter case additional gluon has to be radiated to carry off the color of the incoming gluon and thus guarantee the color conservation. As this gluon can be soft, it does not appreciably distort the kinematics of the process.

The whole process is depicted in Fig. 11.5 and its cross–section contains, beyond quantities calculable in perturbative QCD, also the probability that the produced $c\overline{c}$ pair will coalesce into the J/ψ [164]. This latter probability is not calculable in perturbative QCD, but as the J/ψ is to a good approximation nonrelativistic system, it is proportional to the value of the corresponding wave function at the origin, calculable, for instance, in potential models. The event is so clean because both the scattered electron and proton remnant disappear in the beam tube and, moreover, apart from the soft gluon only the pair of muons comes from the photon–gluon fusion. These muons are detected with high efficiency in the outer muons chambers and their momenta measured from the curvature in the magnetic field. This is not difficult to do because these

muons are typically of a few GeV energy and their tracks defined by both the central tracker and the outer muon chambers, as nicely illustrated in Fig. 11.5.

11.1.4   Are all jets alike?

The answer is simple: definitely not! The point is that given exactly the same final states of partons, the hadronization leads to different patterns of hadrons originating from these partons. In some cases the jets are slim, with just a few narrowly collimated hadrons, while in other the same partons may give rise to quite fat jets. If, however, we would look on the whole jet, we might find that its full momentum is the same in both cases atfer all! This is illustrated in Fig. 11.6 on the comparison of two events from the H1 detector at DESY. In both events two jets are produced in the collision of an almost real photon with a parton from the proton. Despite the fact that if one looks on the particles flying in the directions of the jets, one does not see much difference between the two events, if the energy deposited in individual towers of the electromagnetic and hadronic calorimeters are lego plotted the difference is striking: two very slim jets are produced in the upper event while in the lower event one of the jets is pretty fat and the second something in between. However, summing all the towers going into the fat jet we would get about the same energy as in the upper event. Finally in Fig. 11.7 an event with three jets of very different characteristics, slim, moderate and fat one, is shown.

11.2   Top quark discovery at Fermilab

After many years of intensive search in various processes the hunt for the last, sixth, quark to complete the three generations of quarks and leptons has finally been succesful. The CDF Collaboration at Fermilab has found it among the debris produced in proton–proton collisions at 2 TeV, the highest energy attained so far in accelerators. The reason I mention this discovery in connections with jets is the fact that jets played a crucial role in it. Without understanding the properties of jets and without experience with the jet algorithms the top quark would probably still elude its detection.

In proton–proton collisions the top quark is produced by one of two mechanisms: either in light quark annihilation process (upper diagram in Fig. 11.8) or via the gluon–gluon fusion. In kinematical range accessible with the Fermilab accelerator the first process is dominant. In all decays there is a pair of bottom and antibottom quarks in the final states plus various other quarks and/or leptons. Out of many such possible decay modes of the top quark, listed in the table at the right of Fig. 11.8, almost 50 % goes into the decay mode in which there are six jets and in nearly all decays one encounters at least four of them. To find these jets in a myriad of particles originating from the collisions, one needs some reliable signature. In the top decays what helps is that the decay chains of the bottom/antibottom quarks contain relatively frequently electrons or muons with rather large transverse momentum. This, together with the application of jet finding algorithms, led to observation of events like that in Fig. 11.8, where a familiar lego plot shows four nice jets and one muon, corresponding to the decay chain of the $t\bar{t}$ pair described on third line of the table in Fig. 11.8: two jets were identified as jets originated by the b or $\bar{b}$ quarks, while the remaining two come from W^+ decay. From the decay products of the accompanying W^- boson only the negative muon was observed, but as it is not strongly interacting its energy deposition in the calorimetr is very small.

While the lego plot in Fig. 11.8 seems so obvious one might question the need for any jet algorithm in this search. The point is that finding such clear examples among the milions of events collected in CDF experiment is inconcievable without the use of fast and effective jet finding algorithms discussed in the Chapter 9.

11.3   Are quarks composite?

In the spring of 1996 another interesting and, if confirmed, fundamental observation has been made at Fermilab by the same CDF Collaboration that had earlier found the top quark. It is closely related to the crucial question: are quarks truly elementary or actually composed of some still more fundamental constituents? The idea that also quarks have some structure had been quite popular in the early eighties, the subquark constituents carrying names like **preons, rishons** etc.. Due to the lack of experimental support the interest in this question had subsided. The observations at Fermilab may revive it.
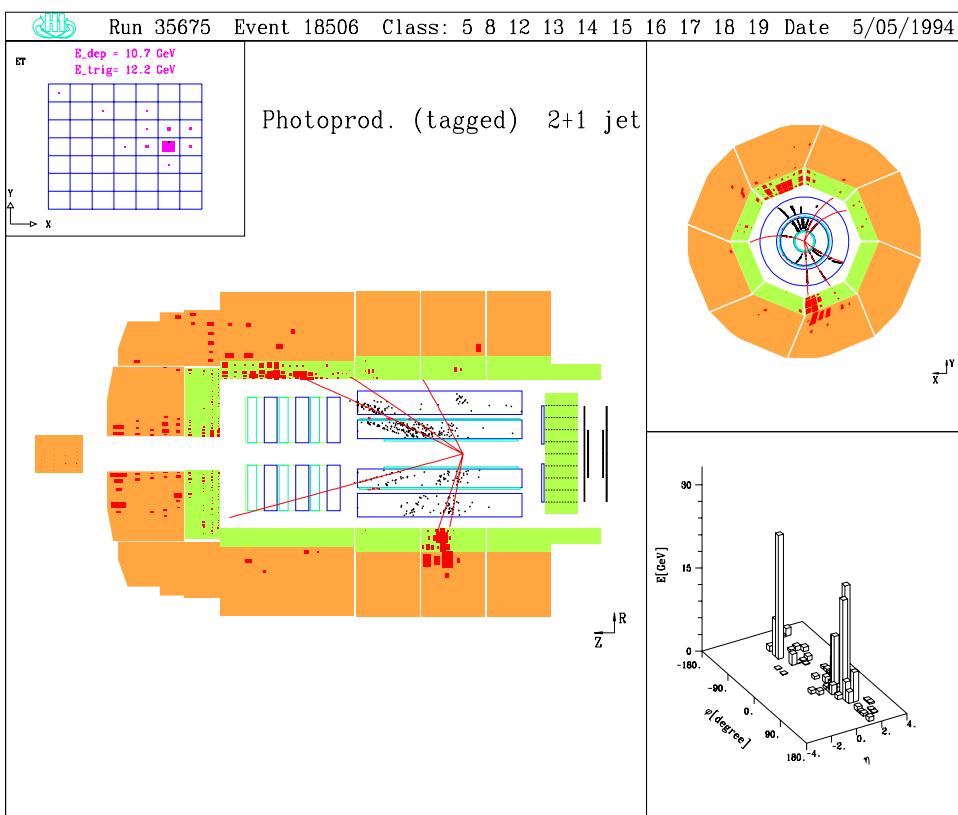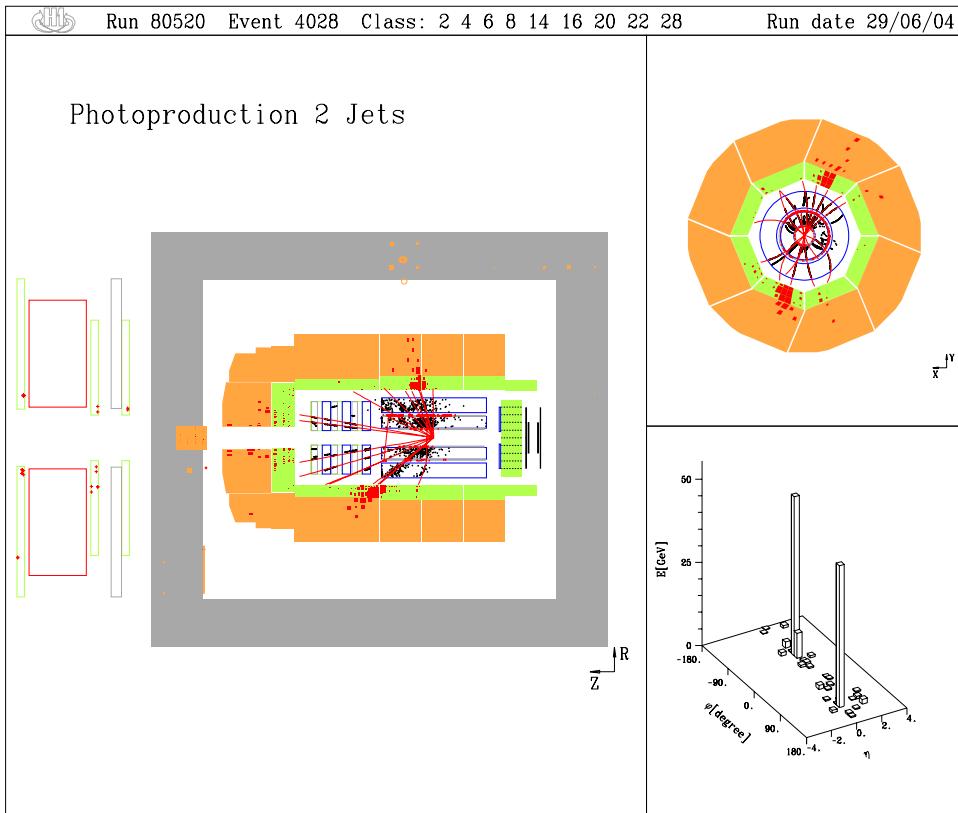
Figure 11.6: Examples of slim and fat jets in H1 detector
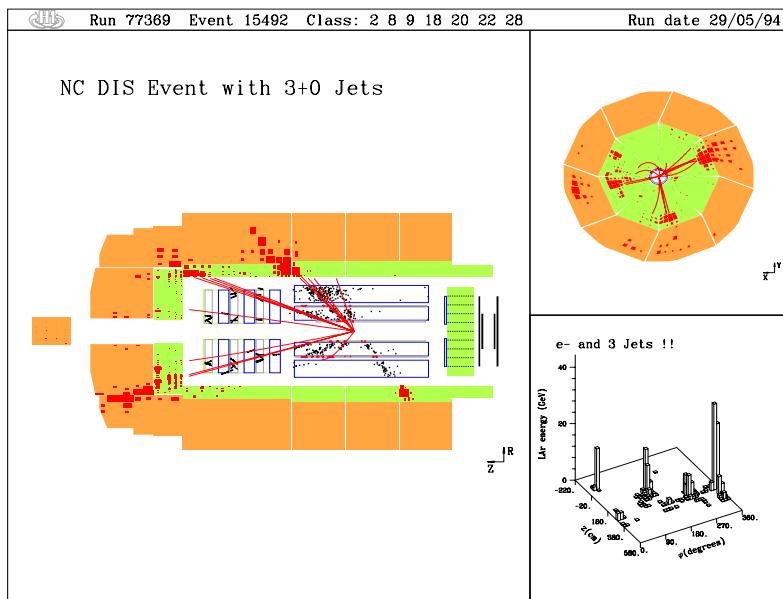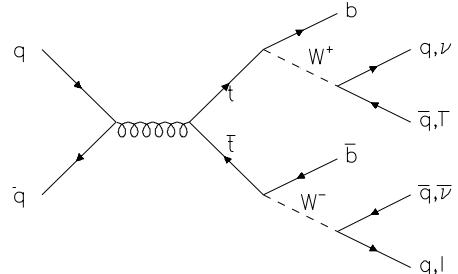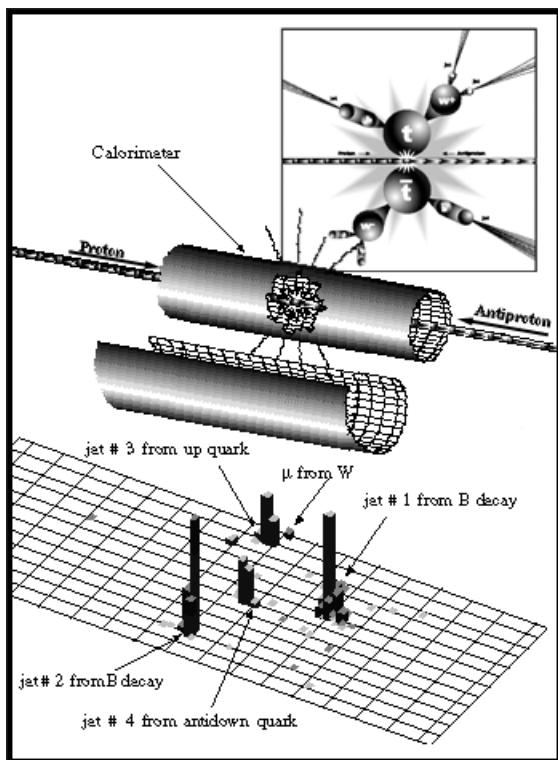
Figure 11.7: Event with three jets of various properties.

Decay mode	Branching ratio
$t\bar{t} \to (q\bar{q}'b)(q\bar{q}'\bar{b})$	36/81
$t\bar{t} \to (q\bar{q}'b)(e\nu\bar{b})$	12/81
$t\bar{t} \to (q\bar{q}'b)(\mu\nu\bar{b})$	12/81
$t\bar{t} \to (q\bar{q}'b)(\tau\nu\bar{b})$	12/81
$t\bar{t} \to (e\nu b)(\mu\nu\bar{b})$	2/81
$t\bar{t} \to (e\nu b)(\tau\nu\bar{b})$	2/81
$t\bar{t} \to (\mu\nu b)(\tau\nu\bar{b})$	2/81
$t\bar{t} \to (e\nu b)(e\nu\bar{b})$	1/81
$t\bar{t} \to (\mu\nu b)(\mu\nu\bar{b})$	1/81
$t\bar{t} \to (\tau\nu b)(\tau\nu\bar{b})$	1/81

Figure 11.8: One of the first events in which the top quark was observed (left), a chain of decays of the $t\bar{t}$ pair (upper right) and Standard Model values of the brachning ratia of all available channels (table at lower right).
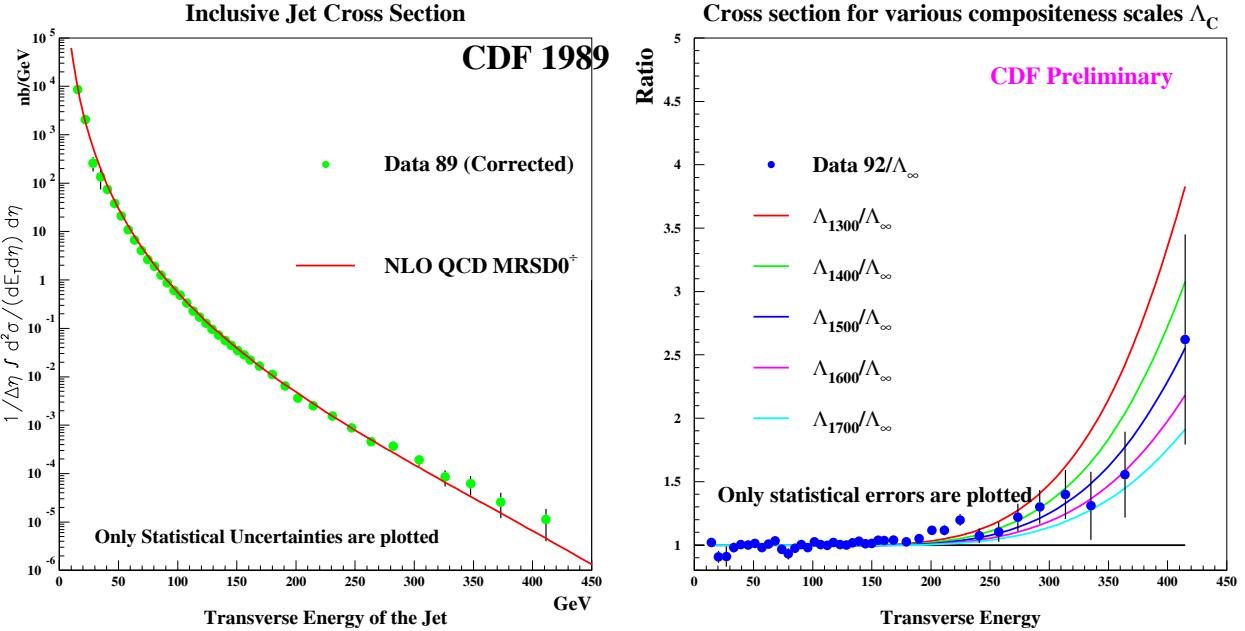
Figure 11.9: One of the first events in which the top quark was observed (left), a chain of decays of the $t\bar{t}$ pair (upper right) and Standard Model values of the brachning ratia of all available channels (table at lower right).

The basic pattern of discoveries of ever smaller constituents of matter remains essentially the same since the discovery of the atomic nucleus: particles are collided and the resulting data compared with the current theoretical understanding of the structure and properties of these particles. The evidence obtained by the CDF Collaboration at Fermilab closely reminds the way Rutheford studied the properties of atoms. Recall that what he observed was an excess of particles scattered at *large* angle, i.e. with *large momentum transfer*. The same has been observed at Fermilab, with the only, but crucial, difference: in the CDF experiment jets with very large momentum transverse and not individual particles, were detected. These jets were produced in collisions of quarks and gluons of the colliding protons according to the general scheme discussed in Chapter 9. CDF compared their data with perturbative QCD, using the current knowledge of parton distribution functions. The plot that made headlines even on CNN and was mentioned also in the daily *Mladá Fronta* **DNES**, is shown in Fig. 11.9a. Note the tiny difference between the data points and the solid curve, describing the theoretical predictions, for jet transverse energy above roughly 300 GeV. The same data, represented in a different way as the ratio between data and theory, is shown in Fig. 11.9b.

Similarly as in the Rutherford experiment, the data were immediatly exploited by theorists who have attempted to interpret the observed deviation as an evidence for **substructure** of the quarks. The curves in Fig. 11.9b correspond to five different values of the so called **compositness scale** Λ, the inverse of which determines the distance at which the substructure becomes observable. As we see the data give a lower bound on this scale somewhere in the region of 1700 GeV. The problem with the CDF evidence and its interpretation is twofold:

- On the experimental side the problem is that the second Fermilab experiment, D0, *does not* see it.

- On the theoretical side, intensive discussion on whether the new CDF data can be reconciled with the other existing data on hard scattering processes or not, goes on.

The dust has not yet settled, but it seems that the Fermilab data by itself will not be able to solve this important problem. In any case the CDF observation has given new impetus to the ongoing **LHC** project,

Run 7395 Event 9137

6 Aug 1996 14:35:21

In run: CV CJ CZ TB PB EB PE EE HT HS
        HP MB ME FD TR TT SC SI SW
Crossing 4 bunchlet 1
Filter bits (600):
 Sum E(cl) > 2
 Min ion EB
 E(cl 1) > 1
 MinIonEE L.R
 PtTrk1 > 400
 PtTrk2 > 250
 Sum Pt > 700
 Pt miss
 HOTLINE
 Spec. trigs
 GPMH
 noTOFMH
Bunchlet Info
 FI1 CV1 CJ1 TB1

Pre: Trigger:16
CV2   TM1
CJ2   TM2
CVOP   TM3
CJOP   TBM1
TB2   o   TBM2
EB2   TBM3
TBOP   CV1
EBOP   J1H
CV1   J2H
CJ1   J3H
CVCJS   TOV2V2
CVCJO   TOVADJ
TB1   CJ1
EB1   TOVCV
TBEBS o   EBTOTHI
TBEBO   TOCJHV
TBCVO o   EBTPHI
TBCVS   EEPRLR
TBCJO   EBWEDGE
TBCJS   TBEBS
EBCVO   EBTOTLO
EBCVS   SWSEGR
EBCJO   SWLOR
EBCJS o   TPTTTO
TEL1   o   TPTTEM
EEL1   o   TPTOEM
TER1   TPTOCL
TEEELS   TPTTB
TEEELCVO   TPTT1
TEEELCVS   TPTT2
TEEELCJO   TPTO1
TEEELCJS   TPTO2
EBPHI   TPEMB
BPCKUP   TPEM1
           TPEM2
           TPTTL
           TPTTR
           TPEML
       o   TPTTCL
       o   TPEMCL
           PRTR_ACT

Ebeam 80.
Nem 20   ot 84.1 E1 37.4 E2 21.9
Ntk 1   ptot 77.9 p1 41.5 p2   8.1

Figure 11.10: One of the first pairs of W^+W^- bosons observed by OPAL detector at LEP.

in which, starting from 2005, protons will be collided with protons at the total center of mass energy of 14 TeV, seven times higher than the energy now available at Fermilab. If quarks are composite, LHC has a great chance to discover it.

11.4   W pair production at LEP

As the final example of jets as tools in discovering new particles and phenomena, I will mention the role of jets in identifying the first events in which pairs of W bosons were produced in e$^+$e$^-$ annihilations at the CERN LEP Collider. The W pair production in e$^+$e$^-$ annihilations represents a crucial test of the Standard Model, as it is sensitive to the presence of the **triple gauge boson vertex**, principal feature of nonabelian gauge theories, like QCD (see Chapter 7 for discussion of this point). In summer 1996 the upgraded LEP reached 161 GeV and soon afterwards all four big experiments (**DELPHI, OPAL, ALEPH, L3**) started to log their first candidates for the W^+W^- production. One of the first candidates for such events, observed in the OPAL detector, is shown in Fig. 11.10. In this lateral view of the detector, two very clear and energetic jets stand out, one flying in the northwest and the other in southeast directions, and one, also very energetic, electron points southwestwards. Kinematical analysis of this particular event shows that there is a large missing energy and momentum as well. The invariant mass of the two jets equals 84.1 GeV, while that of the electron and missing four–momentum gives 77.9 GeV. Taking into account errors of the energy and momentum measurements, these numbers are in excellent agreement with the experimental value $M_W = 80.3 \pm 0.25$ GeV.

Chapter 12

Appendix

Basic conventions and relations used in evaluation of Feynman diagrams in QCD are summarized. In all relations summation over repeated indices is understood.

12.1 Dirac equation and traces of γ matrices

Consider a fermion with spin $\frac{1}{2}$ and mass m. The bispinors $u(p,s), v(p,s)$ and their Dirac conjugates $\overline{u}, \overline{v}$ satisfy the Dirac equations

$$(\not{p} - m)u(p,s) = 0 \quad \Rightarrow \overline{u}(p,s)(\not{p} - m) = 0 \tag{12.1}$$

$$(\not{p} + m)v(p,s) = 0 \quad \Rightarrow \overline{v}(p,s)(\not{p} + m) = 0 \tag{12.2}$$

We adopt the normalization convention

$$\overline{u}(p,s)u(p,s) = 2m \quad \Rightarrow u^{+}(p,s)u(p,s) = 2E \tag{12.3}$$

$$\overline{v}(p,s)v(p,s) = -2m \quad \Rightarrow v^{+}(p,s)v(p,s) = 2E \tag{12.4}$$

The following sums over the spins are frequently needed (i, j denote quark colors)

$$\sum_{\pm s} u_{\alpha}(p,s)\overline{u}_{\beta}(p,s) = (\not{p} + m)_{\alpha\beta} \tag{12.5}$$

$$\sum_{\pm s} v_{\alpha}(p,s)\overline{v}_{\beta}(p,s) = (\not{p} - m)_{\alpha\beta} \tag{12.6}$$

In taking traces we need hermitian conjugates of matrix elements for which

$$[\overline{u}(p',s')\Gamma u(p,s)]^{+} = \overline{u}(p,s)\overline{\Gamma}u(p',s'); \quad \overline{\Gamma} \equiv \gamma^{0}\Gamma^{+}\gamma^{0} \tag{12.7}$$

The basic identity following from anticommutation relations of γ matrices:

$$\not{a}\,\not{b} = ab - \mathrm{i}\sigma_{\mu\nu}a^{\mu}b^{\nu}; \quad \sigma_{\mu\nu} \equiv \frac{1}{2}\left[\gamma_{\mu}, \gamma_{\nu}\right] \tag{12.8}$$

can be used to derive the following relation for traces of γ matrices

$$
\begin{aligned}
\mathrm{Tr}\gamma_{5} &= 0 & (12.9)\\
\mathrm{Tr}\,1 &= 4 & (12.10)\\
\mathrm{Tr}\,(\not{a}\,\not{b}) &= 4ab & (12.11)\\
\mathrm{Tr}\,(\not{a}_{1}\,\not{a}_{2}\,\not{a}_{3}\,\not{a}_{4}) &= 4[(a_{1}a_{2})(a_{3}a_{4}) - (a_{1}a_{3})(a_{2}a_{4}) + (a_{1}a_{4})(a_{2}a_{3})] & (12.12)\\
\mathrm{Tr}\,(\gamma_{5}\,\not{a}\,\not{b}) &= 0 & (12.13)\\
\mathrm{Tr}\,(\gamma_{5}\,\not{a}\,\not{b}\,\not{c}\,\not{d}) &= \mathrm{i}\epsilon_{\alpha\beta\gamma\delta}a^{\alpha}b^{\beta}c^{\gamma}d^{\delta} & (12.14)\\
\gamma_{\mu}\,\not{a}\gamma^{\mu} &= -2\,\not{a} & (12.15)\\
\gamma_{\mu}\,\not{a}\,\not{b}\gamma^{\mu} &= 4(ab) & (12.16)\\
\gamma_{\mu}\,\not{a}\,\not{b}\,\not{c}\gamma^{\mu} &= -2\,\not{c}\,\not{b}\,\not{a} & (12.17)
\end{aligned}
$$

12.2 $SU_c(3)$ matrices and color traces

The generators of $SU_c(3)$ are defined by the commutation relations

$$[T_a, T_b] = \mathrm{i} f_{abc} T_c \tag{12.18}$$

and their product can be expressed as linear combination of a unit matrix $\mathbf{1}$ and the generators T_c

$$T_a T_b = \frac{1}{2} \left[\frac{1}{3} \delta_{ab} \overline{1} + (d_{abc} + \mathrm{i} f_{abc}) T_c \right] \tag{12.19}$$

where d_{abc} is a *fully symmetric* tensor of rank 3. The above relation can be compared to the analogous but simpler one valid for the SU(2) group

$$\sigma_i \sigma_j = \delta_{ij} + \mathrm{i} \epsilon_{ijk} \sigma_k. \tag{12.20}$$

The matrices F_a, D_a defined as

$$(F_a)_{bc} \equiv -\mathrm{i} f_{abc} \quad (D_a)_{bc} \equiv -\mathrm{i} d_{abc} \tag{12.21}$$

satisfy the commutation relations

$$[F_a, F_b] = i f_{abc} F_c, \quad [F_a, D_b] = i f_{abc} D_c \tag{12.22}$$

i.e. F_a form the adjoint representation of $SU_c(3)$. A useful relation between the tensors f_{abc} and d_{abc} is

$$f_{ijm} f_{klm} = \frac{2}{3} \left[(\delta_{ik} \delta_{jl} - \delta_{il} \delta_{jk}) + (d_{ikm} d_{jlm} - d_{ilm} d_{jkm}.) \right] \tag{12.23}$$

Basic traces of color matrices in the fundamental and adjoint representations

$$\mathrm{Tr} T_a = 0, \quad \mathrm{Tr} (T_a T_b) = \frac{1}{2} \delta_{ab} \tag{12.24}$$

$$\mathrm{Tr} (T_a T_b T_c) = \frac{1}{4} (d_{abc} + \mathrm{i} f_{abc}) \tag{12.25}$$

$$\mathrm{Tr} (T_a T_b T_a T_c) = -\frac{1}{12} \delta_{bc} \tag{12.26}$$

$$\mathrm{Tr} F_a = \mathrm{Tr} D_a = 0 \tag{12.27}$$

$$f_{acd} f_{bcd} = 3 \delta_{ab} \Rightarrow \mathrm{Tr} (F_a F_b) = 3 \delta_{ab} \tag{12.28}$$

$$f_{acd} d_{bcd} = 0 \Rightarrow \mathrm{Tr} (F_a D_b) = 0 \tag{12.29}$$

$$d_{acd} d_{bcd} = \frac{5}{3} \delta_{ab} \Rightarrow \mathrm{Tr} (D_a D_b) = \frac{5}{3} \delta_{ab} \tag{12.30}$$

$$\mathrm{Tr} (F_a F_b F_c) = \mathrm{i} \frac{3}{2} f_{abc}, \quad \mathrm{Tr} (D_a D_b D_c) = -\frac{1}{2} d_{abc}, \tag{12.31}$$

$$\mathrm{Tr} (F_a F_b D_c) = \frac{3}{2} d_{abc}, \quad \mathrm{Tr} (D_a D_b F_c) = \frac{3}{2} d_{abc}, \tag{12.32}$$

$$\mathrm{Tr} (F_a F_b F_a F_c) = \frac{9}{2} \delta_{bc} \tag{12.33}$$

Trace of the most general four color matrices in fundamental representation is given as

$$\mathrm{Tr} (T_a T_b T_c T_d) = \frac{1}{12} (\delta_{ab} \delta_{cd} - \delta_{ac} \delta_{bd} + \delta_{ad} \delta_{bc}) +$$

$$\frac{1}{8} (d_{abe} d_{cde} - d_{ace} d_{dbe} + d_{ade} d_{bce}) + \frac{1}{8} (d_{abe} f_{cde} - d_{ace} f_{dbe} + d_{ade} f_{bce}) \tag{12.34}$$

while in the adjoint representation we get

$$\mathrm{Tr} (F_a F_b F_c F_d) = \delta_{ab} \delta_{cd} + \delta_{ad} \delta_{cd} + \frac{3}{4} (d_{abc} d_{cde} - d_{ace} d_{dbc} + d_{ade} d_{bce}) \tag{12.35}$$

12.3 Feynman diagram rules in QCD

Feynman diagram rules for QCD in the class of covariant gauges are listed. Ghost vertices and propagators appear in loops only. The letters i, j, k, l denote quark (triplet) color, a, b, c, d gluon (octet) color and Greek letters space–time coordinates. Small letters p, q, r denote momenta of corresponding lines.

Rules for external particles:

- $u(p, s)$ for each incoming fermion with momentum p and spin s,
- $\overline{v}(p, s)$ for each incoming antifermion with momentum and spin s,
- $\overline{u}(p, s)$ for each outgoing fermion with momentum p and spin s,
- $v(p, s)$ for each outgoing antifermion with momentum p and spin s.
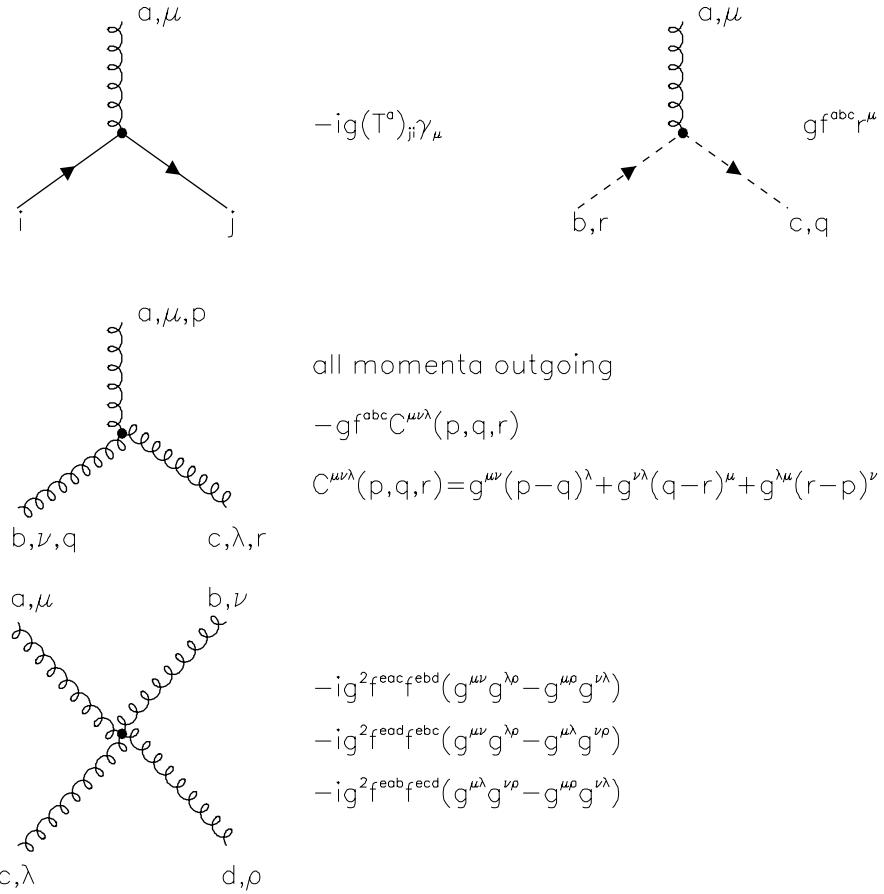
Statistical factors in loops:

- -1 for each closed fermion loop.
- $1/2!$ for each gluon loop with two internal gluons.
- $1/3!$ for each gluon loop with three internal gluons.

Rules for propagators

$$\delta^{ij} \; \frac{i}{(\not{p} - m + i\varepsilon)}$$

$$\delta^{ab} \; \frac{i}{(p^2 + i\varepsilon)} \qquad \text{ghost propagator in loops only}$$

$$\delta^{ab} \; \frac{i}{(p^2 + i\varepsilon)} \left[\!\left[-g_{\mu\nu} + (1 - \alpha_G) \; \frac{p_\mu p_\nu}{(p^2 + i\varepsilon)} \right]\!\right]$$

Rules for vertices

$$-\mathrm{ig}(\mathsf{T}^{\mathsf{a}})_{\mathsf{ji}}\gamma_{\mu}$$

$$\mathsf{g}\mathsf{f}^{\mathsf{abc}}\mathsf{r}^{\mu}$$

all momenta outgoing

$$-\mathsf{g}\mathsf{f}^{\mathsf{abc}}\mathsf{C}^{\mu\nu\lambda}(\mathsf{p},\mathsf{q},\mathsf{r})$$

$$\mathsf{C}^{\mu\nu\lambda}(\mathsf{p},\mathsf{q},\mathsf{r})=\mathsf{g}^{\mu\nu}(\mathsf{p}-\mathsf{q})^{\lambda}+\mathsf{g}^{\nu\lambda}(\mathsf{q}-\mathsf{r})^{\mu}+\mathsf{g}^{\lambda\mu}(\mathsf{r}-\mathsf{p})^{\nu}$$

$$-\mathrm{ig}^{2}\mathsf{f}^{\mathsf{eac}}\mathsf{f}^{\mathsf{ebd}}(\mathsf{g}^{\mu\nu}\mathsf{g}^{\lambda\rho}-\mathsf{g}^{\mu\rho}\mathsf{g}^{\nu\lambda})$$

$$-\mathrm{ig}^{2}\mathsf{f}^{\mathsf{ead}}\mathsf{f}^{\mathsf{ebc}}(\mathsf{g}^{\mu\nu}\mathsf{g}^{\lambda\rho}-\mathsf{g}^{\mu\lambda}\mathsf{g}^{\nu\rho})$$

$$-\mathrm{ig}^{2}\mathsf{f}^{\mathsf{eab}}\mathsf{f}^{\mathsf{ecd}}(\mathsf{g}^{\mu\lambda}\mathsf{g}^{\nu\rho}-\mathsf{g}^{\mu\rho}\mathsf{g}^{\nu\lambda})$$

12.4   Special functions

For evaluation of the inverse Mellin transformation the following special functions defined in the complex plane and their relations are useful.

Gamma function:

$$\Gamma(z) \equiv \int_0^\infty t^{z-1}\mathrm{e}^{-t}\mathrm{d}t \quad \Rightarrow \quad \Gamma(n+1) = n! \tag{12.36}$$

Psi function:

$$\psi(z) \equiv \frac{\mathrm{d}\Gamma(z)/\mathrm{d}z}{\Gamma(z)}, \quad \psi(1) = -\gamma_E \doteq -0.5772, \quad S_1(n) \equiv \sum_{k=1}^{n}\frac{1}{k} = \psi(n+1) + \gamma_E \tag{12.37}$$

Bibliography

[1] S. Weinberg: *The Quantum Theory of Fields*; Volumes I and II, Cambridge University Press, 1995

[2] J. D. Bjorken, S. D. Drell: *Relativistic Quantum Mechanics and Relativistic Quantum Fields*; McGraw–Hill, 1965 (r.p.)

[3] R. P. Feynman: *QED, strange theory of light and matter*; Princeton University Press, 1985 (r.p.)

[4] R.P. Feynman: *Photon-hadron interactions*; Frontiers in Physics, Benjamin, 1972 (r.p)

[5] F. Close: *Introduction to Quarks and Partons*; Academic Press, 1979 (r.p)

[6] H. Georgi: *Lie Algebras in Particle Physics*; Benjamin, 1982

[7] S. Weinberg: *Discovery of Subatomic Particles*; W. Freeman, New York, 1986

[8] R. E. Taylor, J. Friedman, H. W. Kendall: Rev. Mod. Phys. **63** (1991), 573, 596, 615

[9] Ta–Pei Cheng, Ling–Fong Li: *Gauge Theory of elementary particle physics*; Claredon Press, 1984

[10] Yndurain: *Quantum Chromodynamics*; Springer Verlag, New York Berlin Heidelberg Tokyo, 1983 (r.p)

[11] Y. L. Dokshitzer, V. A. Khoze, A. H. Mueller, S. I. Troyan: *Basics of Perturbative QCD*; Editions Frontiers, 1991

[12] G. Altarelli: Physics Reports **C81** (1982), 1

[13] R. D. Field: *Applications of Perturbative QCD*; Addison–Wesley, 1989

[14] K. Huang: *Quarks, Leptons and Gauge Fields*; World Scientific, 1981 (r.p.)

[15] F. Halzen, A. D. Martin: *Quarks and leptons*; John Wiley & Sons, 1984 (r.p.)

[16] Ch. Quigg: *Gauge Theories of the Strong, Weak and Electromagnetic Interactions*; Benjamin, 1983

[17] J. Hořejší: Fundamentals of Electroweak theory; Universitas Carolinae, 2003

[18] A. Pais; Inward bound, Clarendon Press, New York, 1986

[19] J. Hořejší: Historie standardního modelu (in Czech)

[20] H. Geiger, E. Marsden: Phil. Mag. **25** (1913), 604

[21] R. Evans: The atomic nucleus, McGraw-Hill, New York, 1955, p. 18

[22] N. Bohr: Phil. Mag. **26** (1913), 1

[23] N. Bohr: Phil. Mag. **26** (1913), 476

[24] M. Hamermesh: *Group Theory and its applications to physical problems*; Addison–Wesley, 1964

[25] R. Cahn, G. Goldhaber: The experimental Foundations of Particle Physics,

[26] M. Riordan: The Hunting of the Quark, Simon and Schuster, New York, 1987

[27] G. Johnson: Strange Beauty, Random House, 1999

[28] R.M. Barnett, H. Mühry, H. Quinn: The Charm of Strange Quarks, Springer Verlag. New York, 2000

[29] S. Neddermeyer, C. Anderson: Phys. Rev. **51** (1937), 884

[30] R. Thomson et al.: Phys. Rev. **83** (1951), 175

[31] Ch. Peyrou: The role of cosmic rays in the development of particle physics, Proceedings of *Colloquium on History of Particle Physics*, Journal de Physique **43** supplement to No. 12 (1982), C8-7.

[32] M. Gell-Mann: Phys. Rev. **92** (1953), 833

[33] M. Gell-mann: Strangness, Journal de Physique **12** (1982), 395

[34] T. Nakano, K. Nishijima: Prog. Theor. Phys. **10** (1953), 581

[35] K. Nishijima: Prog. Theor. Phys. **12** (1954), 107

[36] W. Fowler et al.: Phys. Rev. **93** (1953), 861

[37] M. Gell-Mann, A. Pais: Phys. Rev. **97** (1955), 1387

[38] M. Gell-Mann, A. Pais: Proc. 1954 Int. Conference on Nuclear and Meson Physics, Glasgow, 1954

[39] M. Gell-Mann: Suppl. Nuovo Cimento **IV** (1956), 848

[40] J. Mehra: Golden Age of Physics, World Scientific, Singapore, 2000

[41] B. Cassen, E. Condon: Phys. Rev. **50** (1936), 846

[42] G. Chew, S. Frautschi: Phys. Rev. Lett. **7** (1961), 394

[43] C. Anderson: Phys. Rev. **43** (1933),

[44] J. Ashkin et al.: Phys. Rev. **93** (1954), 1129

[45] M. Alston et al.: Phys. Rev. Lett. **6** (1961), 300

[46] C. N. Yang, H. Mills: Phys. Rev. **96** (1954), 191

[47] C.N. Yang: Interview in The Mathematical Intelligencer 15/4

[48] S. Sakata: Prog. Theor. Phys. **16** (1956), 686

[49] E. Fermi, C. N. Yang: Phys. Rev. **76** (1949), 1739

[50] Y. Ne'eman: Nucl. Phys. **26** (1961), 222

[51] D. Stonehill et al.: Phys. Rev. Lett. **6** (1961), 624

[52] B. Maglic et al.: Phys. Rev. Lett. **7** (1961), 178

[53] L. Bertanze et al.: Phys. Rev. **9** (1962), 180

[54] A.Pevsner et al.: Phys. Rev. Lett. **7** (1961), 421.

[55] M. Alstom et al.: Phys. Rev. Lett. **5** (1960), 520

[56] N. P. Samios et al: Phys. Rev. **9** (1962),

[57] M. Gell–Mann, Y. Neeman: *The Eightfold Way*, Benjamin, 1964

[58] M. Gell-Mann: CALTECH preprint CTSL-20, 1961

[59] G. Johnson: Strange Beauty, Vintage Books, 1999

[60] M. Gell-Mann: Phys. Rev. **125** (1961), 1067

[61] M. Gell-Mann: Phys. Lett. **B8** (1964), 214

[62] G. Zweig: An SU(3) model for strong interaction symmetry and its breaking, CERN preprint 8419/TH.412

[63] M. Gell-Mann: in the Proceedings of the Conference *QCD - 20 years later*; ed. K. Kastrup and P. Zerwas, World Scientific, 1993, p. 3

[64] G. Zweig: CALT-68-805

[65] D. Lichtenberg and S. Rosen (editors): Developments in the Quark Model of Hadrons, A Reprint Collection, Hadronic Press, 1980, pp. 22-101

[66] V. Barnes et al.: Phys. Rev. Lett. **12** (1964), 204

[67] B. Sakita: Phys. Rev. **136** (1964), B1756

[68] O. Greenberg: Phys. Rev. Lett. **13** (1964), 598

[69] O. Greenberg: Talk at the Wigner Centenial Conference, Pecs, Hungary, July 2002, hep-ph/0212174

[70] A. Pias, in Recent Developments in High Energy Physics, Plenum Press, Neq York, 1980, p. 61.

[71] Y. Nambu: in Proc. Second Coral Gables Conference on Symmetry Principles at High Energy, University of Miami, January 1965

[72] Y. Nambu: in *Preludes in theoretical physics*, eds. A. de-Shalit, H. Feschbach and L. van Hove, North Holland,Amsterdam 1966, p.133

[73] S. Ting: *The discovery of the J particle; A personal recollection*, Čs. čas. fyz. **A28** (1978), 124

[74] B. Richter: *From ψ particles to charm*, 1977 Nobel Lecture, Čs. čas. fyz. **A28** (1978), 323

[75] M. Y. Han, Y. Nambu: Phys. Rev. **139** (1965), B1006

[76] H. Lipkin: Phys. Lett. **45B** (1973), 267

[77] J. Bjorken, S. Glashow: Phys. Lett. **11** (1964), 255

[78] J. Christenson et al.: Phys. Rev. Lett. **25** (1970), 1523

[79] J. Christenson et al.: Phys. Rev. **D 8** (1973), 2016

[80] E. Cazolli et al.: Phys. Rev. lett. **34** (1975), 1125

[81] D. Hom et al.: Phy. Rev. Lett. **36** (1976), 1236

[82] S. Herb et al.: Phy. Rev. Lett. **37** (1976), 1374

[83] D. Hom et al.: Phy. Rev. Lett. **39** (1977), 252

[84] W. Innes et al.: Phy. Rev. Lett. **39** (1977), 1240

[85] F. Abe et al.: Phy. Rev. Lett. **73** (1994), 225

[86] R. Hofstadter, H. Fechter, J. McIntyre: Phys. Rev. **92** (1953), 878

[87] R. McAllister, R. Hofstadter: Phys. Rev. **102** (1956), 851

[88] R. Hofstadter: Rev. Mod. Phys. **28** (1956), 214

[89] T. Janssens et al.: Phys. Rev. **142** (1965), 922

[90] H. Feschbach: Phys. Rev. **84** (1951, 1206

[91] E. Bloom et al.: Phys. Rev. Let. **23** (1960), 930

[92] M. Breidenbach et al.: Phys. Rev. Let. **23** (1960), 1935

[93] G. Miller et al: Phys. Rev. **D5** (1972), 528

[94] J. Bjorken: Phys. Rev. **179** (1969), 1547

[95] J. Bjorken: Phys. Rev. Lett. **16** (1966), 408

[96] K. Gottfried: Phys. Rev. Lett. **18** (1967), 1174

[97] C. Callan, D. Gross: Phys. Rev. Lett. **21** (1968), 311

[98] J. Bjorken, E. Paschos: Phys. Rev. **185** (1969), 1975

[99] H. Georgi, H. Politzer: Phys. Rev. **D14** (1976), 1829

[100] SMC Collab.: Phys. Lett. **B329** (1994), 399

[101] J. Ellis, M. Karliner: hep-ph/9601280

[102] R.P. Feynman, R.D. Field: Nucl. phys. **B136** (1977), 1

[103] C. Peterson et. al.: Phys. Rev. **D27** (1983), 105

[104] P. A. M. Dirac: Directions in Physics, John Wiley & Sons, New York, 1976, pp. 35-37.

[105] K. Wilson: Rev. Mod. Phys. **55** (1983), 583-600.

[106] G. Marchesini et al: Computer Physics Commun. **67** (1992), 465

[107] T. Sjöstrand: Computer Physics Commun. **39** (1986), 367

[108] M. Creutz: *Quarks, Gluons and Lattices*; Cambridge University Press, 1983

[109] M. Shifman, A. Vainstein, V. Zakharov: Nucl. Phys. **B191** (1981), 301

[110] S. Coleman: in the Proceedings of the *International School of Subnuclear Physics*; Erice, 1979

[111] J. Schwinger: *Particles, Souces and Fields*; vol.1,2,3, Addison–Wesley

[112] P. Hasenfratz, J. Kuti: Physics Reports **40C** (1978), 75

[113] E. Shuryak: Physics Reports **115C** (1984), 151

[114] S. Brodsky et al.: SLAC–PUB–5811 (1992)

[115] K. Wilson: Pokroky matematiky, fyziky a astronomie **31** (1986), 1

[116] E. Spear; *Generalized Feynman amplitudes*, Princeton, 1969

[117] G. 't Hooft, M. Veltman; Nucl. Phys. **B44** (1972), 189

[118] C. Bollini, J. Giambiagi; Nuovo Cimento **12B** (1972), 20

[119] V. Beresteckij: Čs. čas. fyz. **A28** (1978), 433

[120] D. Gross: Twenty five years of asymptotic freedom, hep-ph/9809060

[121] D. Gross, F. Wilczek: Phys. Rev. Let. **30** (1973), 1343

[122] D. Politzer: Phys. Rev. Let. **30** (1973), 1346

[123] D. Gross, F. Wilczek: Phys. Rev. **D8** (1973), 3633

[124] D. Gross, F. Wilczek: Phys. Rev. **D9** (1974), 980

[125] H. Georgi, D. Politzer: Phys. Rev. D **D9** (1974), 416

[126] http: //particleadventure.org/particleadventure/other/history/smt.html

[127] H. Fritzsch, M. Gell-Mann, H. Leutwyler: Phys. Lett. **47B** (1973), 365

[128] D. Gross, F. Wilczek: Fermilab preprint NAL-PUB-73/49-THY, July 1973

[129] J. Chýla, R. Kotecký: Čs. čas. fyz. **A36** (1986), 102

[130] W. Marciano: Phys. Rev. **D29** (1984), 580

[131] A. De Rujula, H. Georgi: Phys. Rev. **D13** (1976), 1296

[132] W. Bernreuther: preprint PITHA **94/31**

[133] W. Bernreuther, W. Wetzel: Nucl. Phys. **B197** (1982), 228

[134] J. Chýla: Phys. Lett. **B351** (1995),

[135] A. Prudnikov et al.: *Integraly i rjady*, Nauka, Moskva

[136] G. Kramer: Springer Tracts in Physics, **102** (1984)

[137] H. Bethke, Z. Kunszt, D. Soper: Nucl. Phys. **B370** (1992), 310

[138] F. J. Dyson: Phys. Rev. **85** (1952), 631

[139] P. M. Stevenson: Nucl. Phys. **B231** (1984) 65

[140] N. N. Khuri: Phys. Rev. **D16** (1977), 1754

[141] G. 't Hooft: in *Deeper Pathways in High–Energy Physics*; ed. B. Kursunoglu, Plenum Press, 1977

[142] P. M. Stevenson: Phys. Rev. **D23** (1981), 2916

[143] G. Grunberg: Phys. Rev. **D29** (1984), 2315

[144] J. Chýla: Czech. J. Phys. **42** (1992), 263

[145] J. Chýla, P. Kolář, J. Fischer: Phys. Rev. **D47** (1993), 2578

[146] T. Sjöstrand: In the Proceedings *Z Physics at LEP* ed. R. Kleiss, CERN Yellow Report 89–08, vol. 3

[147] In the Proceedings *Physics at HERA*, ed. W. Buchmüler, G. Ingelman, volume 3

[148] C. F. Weizsäcker: Z. f. Physik **88** (1934), 612

[149] C. Basham, L. Brown, S. Ellis, S. Love: Phys. Rev. **D17** (1978) 2298

[150] G. Altarelli, G. Parisi: Nucl. Phys. **B126** (1977), 298

[151] K. Prytz: Phys. Lett. **B311** (1993), 286

[152] H1 Collab.: **Nucl. Phys. B470** (1996),

[153] H. D. Politzer: Nucl. Phys. **B192** (1984), 493

[154] W. Furmanski, R. Petronzio: Z. f. Physik **C31** (1982), 293

[155] G. Altarelli, R. K. Elllis, G. Martinelli: Nucl. Phys. **B143** (1978), 521

[156] J. Chýla: Phys. Rev. **D48** (1993), 4385

[157] A. Buras, K.J.F Gaemers: Nucl. Phys. **B132** (1978), 249

[158] J. Morfin, W.K. Tung: Z. Physik **C52** (1991), 13

[159] F.J. Yndurain: Phys. Lett. **B74** (1978), 68

[160] W. Furmanski, R. Petronzio: Nucl. Phys. **B195** (1982), 237

[161] A. Devoto, D. Duke, J.F. Owens, R.G. Roberts: Phys. Rev. **D27** (1983), 508
    M. Botje: QCDNUM16: a fast QCD evolution program, Zeus Note 97-066

[162] I. Barker, C. Langensiepen, G. Shaw: Nucl. Phys. **B186** (1981), 61

[163] J. Chýla, J. Rameš: Z. Phys. **C31** (1986), 151

[164] V. Weisskopf, Van Royen: Nuovo Cimento **50A** (1967), 617

[165] M. Glück, E. Reya: Phys. Rev. **D14** (1976), 3034

[166] D. Kosower: Nucl. Phys. **B506** (1997), 439

[167] S. Weinzierl: hep-ph/0203112