Chapter 2
QCD: The Theory of Strong Interactions

2.1 Introduction

This chapter is devoted to a concise introduction to quantum chromodynamics (QCD), the theory of strong interactions [215, 234, 360] (for a number of dedicated books on QCD, see [173], and also [33]). The main emphasis will be on ideas without too many technicalities. As an introduction we present here a broad overview of the strong interactions (for reviews of the subject, see, for example, [29, 30]). Then some methods of non-perturbative QCD will be briefly described, including both analytic approaches and simulations of the theory on a discrete spacetime lattice. Then we shall proceed to the main focus of the chapter, that is, the principles and applications of perturbative QCD, which will be discussed in detail.

As mentioned in Chap. 1, the QCD theory of strong interactions is an unbroken gauge theory based on the colour group $SU(3)$. The eight massless gauge bosons are the gluons g_μ^A and the matter fields are colour triplets of quarks q_i^a (in different flavours i). Quarks and gluons are the only fundamental fields of the Standard Model (SM) with strong interactions (hadrons). The QCD Lagrangian was introduced in (1.28)–(1.31) of Sect. 1.4. For quantization the classical Lagrangian in (1.28) must be extended to contain gauge fixing and ghost terms, as described in Chap. 1. The Feynman rules of QCD are listed in Fig. 2.1. The physical vertices in QCD include the gluon–quark–antiquark vertex, analogous to the QED photon–fermion–antifermion coupling, but also the 3-gluon and 4-gluon vertices, of order e_s and e_s^2 respectively, which have no analogue in an Abelian theory like QED.

Why $SU(N_C = 3)_{\text{colour}}$? The choice of $SU(3)$ as colour gauge group is unique in view of a number of constraints:

- The group must admit complex representations because it must be able to distinguish a quark from an antiquark [214]. In fact, there are meson states made up of $q\bar{q}$ but not analogous qq bound states. Among simple groups, this restricts

© Springer International Publishing AG 2017

G. Altarelli, *Collider Physics within the Standard Model*,
Lecture Notes in Physics 937, DOI 10.1007/978-3-319-51920-3_2

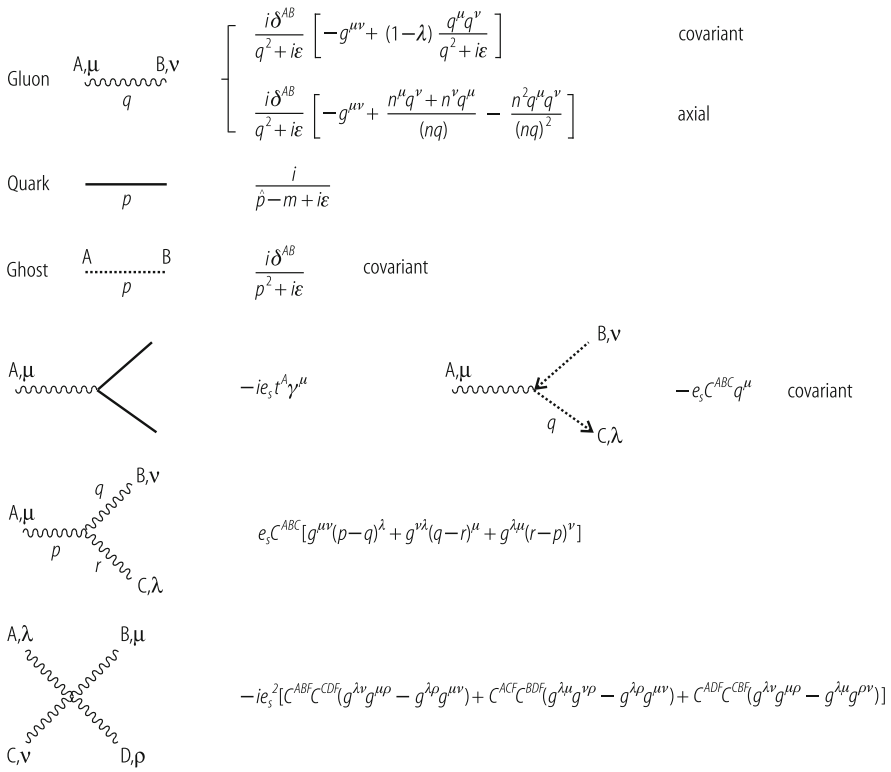

Fig. 2.1 Feynman rules for QCD. *Solid lines* represent the quarks, *curly lines* the gluons, and *dotted lines* the ghosts (see Chap. 1). The gauge parameter is denoted by λ. The 3-gluon vertex is written as if all gluon lines are outgoing

the choice to $SU(N)$ with $N \geq 3$, $SO(4N+2)$ with $N \geq 2$ [taking into account the fact that $SO(6)$ has the same algebra as $SU(4)$], and $E(6)$.

- The group must admit a completely antisymmetric colour singlet baryon made up of three quarks, viz., qqq. In fact, from the study of hadron spectroscopy, we know that the low-lying baryons, completing an octet and a decuplet of (flavour) $SU(3)$ (the approximate symmetry that rotates the three light quarks u, d, and s), are made up of three quarks and are colour singlets. The qqq wave function must be completely antisymmetric in colour in order to agree with Fermi statistics. Indeed, if we consider, for example, a N^{*++} with spin z-component +3/2, this is made up of $(u \uparrow u \uparrow u \uparrow)$ in an s-state. Thus its wave function is totally symmetric in space, spin, and flavour, so that complete antisymmetry in colour is required by Fermi statistics. In QCD this requirement is very simply satisfied by $\epsilon_{abc}q^a q^b q^c$, where a, b, c are $SU(3)_{\text{colour}}$ indices.
- The choice of $SU(N_{\text{C}} = 3)_{\text{colour}}$ is confirmed by many processes that directly measure N_{C}. Some examples are listed here.

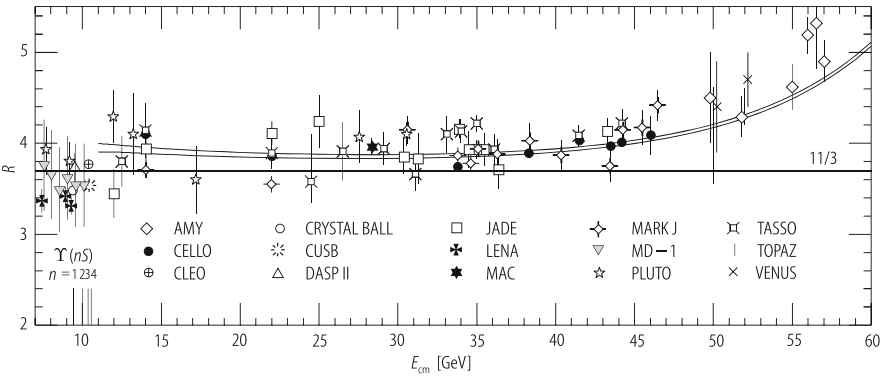

Fig. 2.2 Comparison of the data on $R = \sigma(e^+e^- \to \text{hadrons})/\sigma_{\text{point}}(e^+e^- \to \mu^+\mu^-)$ with the QCD prediction (adapted from [306]). $N_C = 3$ is indicated by the data points above ∼10 GeV (the $b\bar{b}$ threshold) and ∼40 GeV, where the rise due to the Z_0 resonance becomes appreciable

The total rate for hadronic production in e^+e^- annihilation is linear in N_C. More precisely, if we consider $R = R_{e^+e^-} = \sigma(e^+e^- \to \text{hadrons})/\sigma_{\text{point}}(e^+e^- \to \mu^+\mu^-)$ above the $b\bar{b}$ threshold and below m_Z, and if we neglect small computable radiative corrections (which will be discussed in Sect. 2.7), we have a sum of individual contributions (proportional to Q^2, where Q is the electric charge in units of the proton charge) from $q\bar{q}$ final states with $q = u,\ c,\ d,\ s,\ b$:

$$R \approx N_C \left(2 \times \frac{4}{9} + 3 \times \frac{1}{9}\right) \approx N_C \frac{11}{9} . \tag{2.1}$$

The data neatly indicate $N_C = 3$, as can be seen from Fig. 2.2 [306]. The slight excess of the data with respect to the value 11/3 is due to QCD radiative corrections (see Sect. 2.7).

Similarly, we can consider the branching ratio $B(W^- \to e^-\bar{\nu})$, again in the Born approximation. The possible fermion–antifermion ($f\bar{f}$) final states are for $f = e^-$, μ^-, τ^-, d, s (there is no $f = b$ because the top quark is too heavy for $b\bar{t}$ to occur). Each channel gives the same contribution, except that for quarks we have N_C colours:

$$B(W^- \to e^-\bar{\nu}) \approx \frac{1}{3 + 2N_C} . \tag{2.2}$$

For $N_C = 3$, we obtain $B = 11\%$ and the experimental number is $B = 10.7\%$.

Another analogous example is the branching ratio $B(\tau^- \to e^-\bar{\nu}_e\nu_\tau)$. From the final state channels with $f = e^-, \mu^-, d$, we find

$$B(\tau^- \to e^-\bar{\nu}_e\nu_\tau) \approx \frac{1}{2 + N_C} . \tag{2.3}$$

For $N_C = 3$, we obtain $B = 20\%$ and the experimental number is $B = 18\%$ (the lower accuracy in this case is explained by the larger radiative and phase-space corrections, because the mass of τ^- is much smaller than m_W).

An important process that is quadratic in N_C is the rate $\Gamma(\pi^0 \to 2\gamma)$. This rate can be reliably calculated from a theorem in field theory which has to do with the chiral anomaly:

$$\Gamma(\pi^0 \to 2\gamma) \approx \left(\frac{N_C}{3}\right)^2 \frac{\alpha^2 m_{\pi^0}^3}{32\pi^3 f_\pi^2} = (7.73 \pm 0.04)\left(\frac{N_C}{3}\right)^2 \text{eV}\,, \tag{2.4}$$

where the prediction is obtained for $f_\pi = (130.7 \pm 0.37)$ MeV. The experimental result is $\Gamma = (7.7 \pm 0.5)$ eV, in remarkable agreement with $N_C = 3$.

There are many more experimental confirmations that $N_C = 3$. For example, the rate for Drell–Yan processes (see Sect. 2.9) is inversely proportional to N_C.

2.2 Non-perturbative QCD

The QCD Lagrangian in (1.28) has a simple structure, but a very rich dynamical content. It gives rise to a complex spectrum of hadrons, implies the striking properties of confinement and asymptotic freedom, is endowed with an approximate chiral symmetry which is spontaneously broken, has a highly nontrivial topological vacuum structure (instantons, $U(1)_A$ symmetry breaking, strong CP violation which is a problematic item in QCD possibly connected with new physics, like axions, and so on), and an intriguing phase transition diagram (colour deconfinement, quark–gluon plasma, chiral symmetry restoration, colour superconductivity, and so on).

How do we get testable predictions from QCD? On the one hand there are non-perturbative methods. The most important at present is the technique of lattice simulations (for a recent review, see [272]): it is based on first principles, it has produced very valuable results on confinement, phase transitions, bound states, hadronic matrix elements, and so on, and it is by now an established basic tool. The main limitation is from computing power, so there is continuous progress and good prospects for the future.

Another class of approaches is based on effective Lagrangians, which provide simpler approximations than the full theory, valid in some definite domain of physical conditions. Typically at energies below a given scale L, particles with mass greater than L cannot be produced, and thus only contribute short distance effects as virtual states in loops. Under suitable conditions one can write down a simplified effective Lagrangian, where the heavy fields have been eliminated (one says "integrated out"). Virtual heavy particle short distance effects are absorbed into the coefficients of the various operators in the effective Lagrangian. These coefficients are determined in a matching procedure, by requiring that the effective theory reproduce the matrix elements of the full theory up to power corrections.

Chiral Lagrangians are based on soft pion theorems [362] and are valid for suitable processes at energies below 1 GeV (for a recent, concise review, see [212] and references therein). Heavy quark effective theories [178] are obtained by expanding in inverse powers of the heavy quark mass and are mainly important for the study of b and, to lesser accuracy, c decays (for reviews, see, for example, [301]).

Soft-collinear effective theories (SCET) [84], are valid for processes where quarks have energies much greater than their mass. Light energetic quarks not only emit soft gluons, but also collinear gluons (a gluon in the same direction as the original quark), without changing their virtuality. In SCET, the logs associated with these soft and collinear gluons are resummed.

The approach using QCD sum rules [298, 325] has led to interesting results but now appears not to offer much potential for further development. On the other hand, the perturbative approach, based on asymptotic freedom, still remains the main quantitative connection to experiment, due to its wide range of applicability to all sorts of "hard" processes.

2.2.1 Progress in Lattice QCD

One of the main approaches to non-perturbative problems in QCD is by simulations of the theory on a lattice, a technique initiated by K. Wilson in 1974 [366] which has shown continuous progress over the last decades. In this approach the QCD theory is reformulated on a discrete space time, a hypercubic lattice of sites (in the simplest realizations) with spacing a and 4-volume L^4. On each side, there are N sites with $L = Na$. Over the years we have learned how to efficiently describe a field theory on a discrete spacetime and how to implement gauge symmetry, chiral symmetry, and so on (for a recent review see, for example, [272]).

Gauge and matter fields are specified on the lattice sites and the path integral is computed numerically as a sum over the field configurations. Much more powerful computers than in the past now allow for a number of essential improvements. As one is eventually interested in the continuum limit $a \to 0$, it is important to work with as fine a lattice spacing a as possible. Methods have been developed for "improving" the Lagrangian in such a way that the discretization errors vanish faster than linearly in a. A larger lattice volume (i.e., large L or N) is also useful since the dimensions of the lattice should be as large as possible in comparison with the dimensions of the hadrons to be studied. In many cases the volume corrections are exponentially damped, but this is not always the case. Lattice simulation is limited to large enough masses of light quarks: in fact, heavier quarks have shorter wavelengths and can be accommodated in a smaller volume. In general, computations are done for quark and pion masses heavier than in reality, and then extrapolated to the physical values, but at present one can work with smaller quark masses than in the past. One can also take advantage of the chiral effective theory in order to control the chiral logs $\log(m_q/4\pi f_\pi)$ and guide the extrapolation.

A big step that has been taken recently, made possible by the availability of more powerful dedicated computers, is the evolution from quenched (i.e., with no dynamical fermions) to unquenched calculations. In doing this, an evident improvement is obtained in the agreement between predictions and data. For example [272], modern unquenched simulations reproduce the hadron spectrum quite well. Calculations with dynamical fermions (which take into account the effects of virtual quark loops) involve evaluation of the quark determinant, which is a difficult task. Just how difficult depends on the particular calculation method. There are several approaches (Wilson, twisted mass, Kogut–Susskind staggered, Ginsparg–Wilson fermions), each with its own advantages and disadvantages (including the time it takes to run the simulation on a computer). A compromise between efficiency and theoretical purity is needed. The most reliable lattice calculations are today for $2+1$ light quarks (degenerate up and down quarks and a heavier strange quark s). The first calculations for $2+1+1$ including charm quarks are starting to appear.

Lattice QCD is becoming increasingly predictive and plays a crucial role in different domains. For example, in flavour physics it is essential for computing the relevant hadronic matrix elements. In high temperature QCD the most illuminating studies of the phase diagram, the critical temperature, and the nature of the phase transitions are obtained by lattice QCD: as we now discuss, the best arguments to prove that QCD implies confinement come from the lattice.

2.2.2 *Confinement*

Confinement is the property that no isolated coloured charge can exist. One only sees colour singlet particles. Our understanding of the confinement mechanism has much improved thanks to lattice simulations of QCD at finite temperatures and densities (for reviews see, e.g., [85, 162, 199]). For example, the potential between a quark and an antiquark has been studied on the lattice [256]. It has a Coulomb part at short range and a linearly increasing term at long range:

$$V_{q\bar{q}} \approx C_{\mathrm{F}} \left[\frac{\alpha_{\mathrm{s}}(r)}{r} + \cdots + \sigma r \right] , \tag{2.5}$$

where

$$C_{\mathrm{F}} = \frac{1}{N_{\mathrm{C}}} \sum_{\mathrm{A}} t^A t^A = \frac{N_{\mathrm{C}}^2 - 1}{2N_{\mathrm{C}}} \tag{2.6}$$

with N_{C} the number of colours ($N_{\mathrm{C}} = 3$ in QCD). The scale dependence of α_{s} (the distance r is Fourier-conjugate to the momentum transfer) will be explained in detail later. The slope decreases with increasing temperature until it vanishes at a critical temperature T_{C}. Then above T_{C} the slope remains zero, as shown in Fig. 2.3. The value of the critical temperature is estimated to be around $T_{\mathrm{C}} \sim 175\,\mathrm{MeV}$.

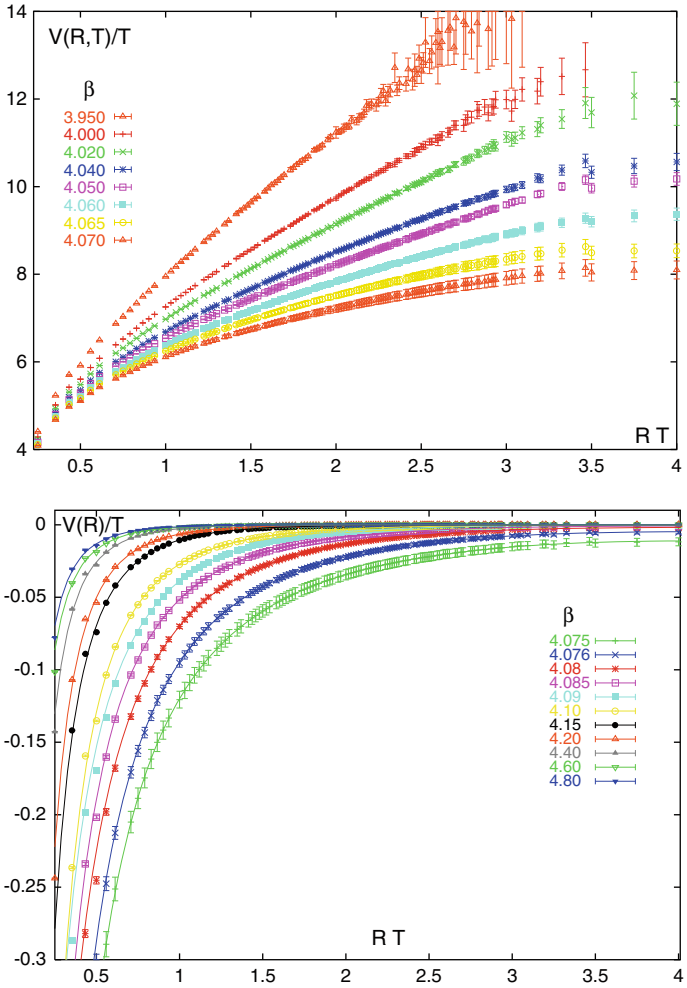

Fig. 2.3 The potential between a quark and an antiquark computed on the lattice in the quenched approximation [256]. The *upper panel* shows that the slope of the linearly rising term decreases with temperature and vanishes at the critical temperature T_C. At $T \geq T_C$ the slope remains at zero (*lower panel*)

The linearly increasing term in the potential makes it energetically impossible to separate a $q\bar{q}$ pair. If the pair is created at one spacetime point, for example in e^+e^- annihilation, and then the quark and the antiquark start moving away from each other in the center-of-mass frame, it soon becomes energetically favourable to create additional pairs, smoothly distributed in rapidity between the two leading charges, which neutralize colour and allow the final state to be reorganized into two jets of colourless hadrons that communicate in the central region by a number of

"wee" hadrons with small energy. It is just like the familiar example of the broken magnet: if you try to isolate a magnetic pole by stretching a dipole, the magnet breaks down and two new poles appear at the breaking point.

Confinement is essential to explain why nuclear forces have very short range while massless gluon exchange would be long range. Nucleons are colour singlets and they cannot exchange colour octet gluons but only colourless states. The lightest colour singlet hadronic particles are pions. So the range of nuclear forces is fixed by the pion mass $r \simeq m_\pi^{-1} \approx 10^{-13}$ cm, since $V \approx \exp(-m_\pi r)/r$.

The phase transitions of colour deconfinement and of chiral restoration appear to happen together on the lattice [85, 162, 199, 272] (see Fig. 2.4). A rapid transition is observed in lattice simulations where the energy density $\epsilon(T)$ is seen to increase sharply near the critical temperature for deconfinement and chiral restoration

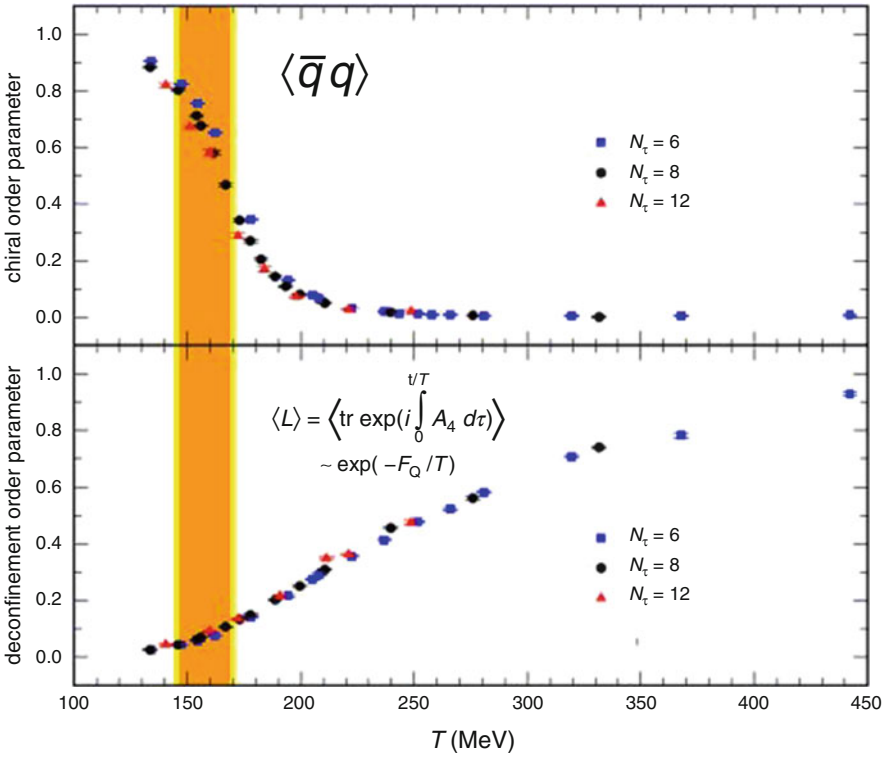

Fig. 2.4 Order parameters for deconfinement (*bottom*) and chiral symmetry restoration (*top*), as a function of temperature [85, 272]. On a finite lattice the singularities associated with phase transitions are not present, but their development is indicated by a rapid rate of change. With increasing temperature, the vacuum expectation value of the quark–antiquark condensate goes from the finite value that breaks chiral symmetry down to zero, where chiral symmetry is restored. In a comparable temperature range, the Wilson plaquette, the order parameter for deconfinement, goes from zero to a finite value. Figure reproduced with permission. Copyright (c) 2012 by *Annual Reviews*

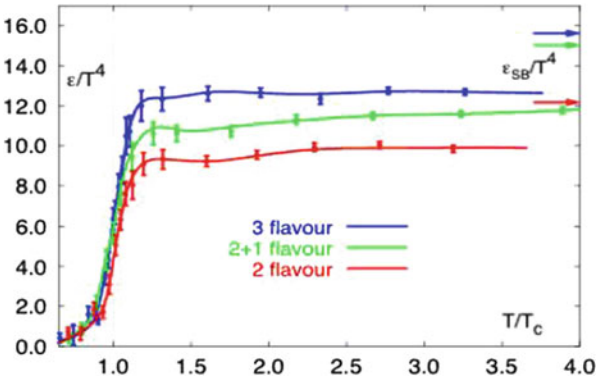

Fig. 2.5 The energy density divided by the fourth power of the temperature, computed on the lattice with different numbers of sea flavours, shows a marked rise near the critical temperature (adapted from [85] and [272]). The *arrows* on top show the limit for a perfect Bose gas (while the hot dense hadronic fluid is not expected to be a perfect gas)

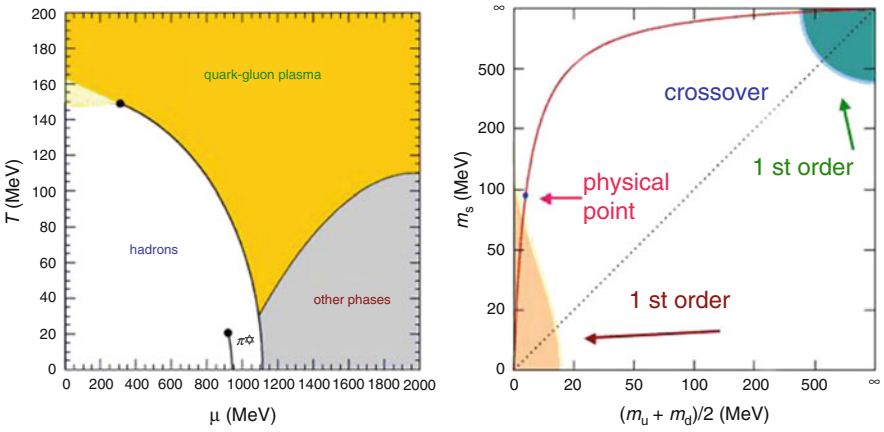

Fig. 2.6 *Left*: a schematic view of the QCD phase diagram. *Right*: on the lattice the nature of the phase transition depends on the number of quark flavours and their masses as indicated [272]. Figure reproduced with permission. Copyright (c) 2012 by *Annual Reviews*

(see Fig. 2.5). The critical parameters and the nature of the phase transition depend on the number of quark flavours n_f and on their masses (see Fig. 2.6). For example, for $n_f = 2$ or $2+1$ (i.e., 2 light u and d quarks and 1 heavier s quark), $T_C \sim 175\,\text{MeV}$ and $\epsilon(T_C) \sim 0.5$–$1.0\,\text{GeV/fm}^3$. For realistic values of the masses m_s and $m_{u,d}$, the two phases are connected by a smooth crossover, while the phase transition becomes first order for very small or very large $m_{u,d,s}$. Accordingly, the hadronic phase and the deconfined phase are separated by a crossover region at small densities and by a critical line at high densities that ends with a critical point. Determining the exact location of the critical point in T and μ_B is an important challenge for theory and

is also important for the interpretation of heavy ion collision experiments. At high densities, the colour superconducting phase is also present, with bosonic diquarks acting as Cooper pairs.

A large investment is being made in heavy ion collision experiments with the aim of finding some evidence of the quark–gluon plasma phase. Many exciting results have been found at the CERN SPS in the past few years, more recently at RHIC and now at the LHC, in dedicated heavy ion runs [296] (the ALICE detector is especially designed for the study of heavy ion collisions).

2.2.3 Chiral Symmetry in QCD and the Strong CP Problem

In the QCD Lagrangian (1.28), the quark mass terms are of the general form $[m\bar{\psi}_{\rm L}\psi_{\rm R}+{\rm h.c.}]$ (recall the definition of $\psi_{\rm L,R}$ in Sect. 1.5 and the related discussion). These terms are the only ones that show a chirality flip. In the absence of these terms, i.e., for $m=0$, the QCD Lagrangian would be invariant under independent unitary transformations acting separately on $\psi_{\rm L}$ and $\psi_{\rm R}$. Thus, if the masses of the $N_{\rm f}$ lightest quarks are neglected, the QCD Lagrangian is invariant under a global $U(N_{\rm f})_{\rm L}\bigotimes U(N_{\rm f})_{\rm R}$ chiral group.

Consider $N_{\rm f}=2$. Then $SU(2)_{\rm V}$ corresponds to the observed approximate isospin symmetry and $U(1)_{\rm V}$ to the portion of baryon number associated with u and d quarks. Since no approximate parity doubling of light quark bound states is observed, the $U(2)_{\rm A}$ symmetry must be spontaneously broken (for example, no opposite parity analogues of protons and neutrons exist with a few tens of MeV separation in mass from the ordinary nucleons). The breaking of chiral symmetry is induced by the VEV of a quark condensate. For $N_{\rm f}=2$ this is $[\bar{u}_{\rm L}u_{\rm R}+\bar{d}_{\rm L}d_{\rm R}+{\rm h.c.}]$. A recent lattice calculation [208] has given for this condensate the value $[234\pm 18\,{\rm MeV}]^3$ (in $\overline{\rm MS}$, $N_{\rm f}=2+1$, with the physical m_s value, at the scale of 2 GeV). This scalar operator is an isospin singlet, so it preserves $U(2)_{\rm V}$, but breaks $U(2)_{\rm A}$. In fact, it transforms like (1/2,1/2) under $U(2)_{\rm L}\bigotimes U(2)_{\rm R}$, but is a singlet under the diagonal group $U(2)_{\rm V}$.

The pseudoscalar mesons are obvious candidates for the would-be Goldstone bosons associated with the breakdown of the axial group, in that they have the quantum number of the broken generators: the three pions are the approximately massless Goldstone bosons (exactly massless in the limit of vanishing u and d quark masses) associated with the breaking of three generators of $U(2)_{\rm L}\bigotimes U(2)_{\rm R}$ down to $SU(2)_{\rm V}\bigotimes U(1)_{\rm V}\bigotimes U(1)_{\rm A}$. The couplings of Goldstone bosons are very special: in particular only derivative couplings are allowed. The pions as pseudo-Goldstone bosons have couplings that satisfy strong constraints. An effective chiral Lagrangian formalism [362] allows one to systematically reproduce the low energy theorems implied by the approximate status of Goldstone particles for the pion, and successfully describes QCD for energies at scales below $\sim$1 GeV.

The breaking mechanism for the remaining $U(1)_{\rm A}$ arises from an even subtler mechanism. A state in the η–η' space cannot be the associated Goldstone particle because the masses are too large [361] and the η' mass does not vanish in the chiral limit [367]. Rather, the conservation of the singlet axial current $j_5^\mu = \sum \bar{q}_i \gamma^\mu \gamma_5 q_i$ is broken by the Adler–Bell–Jackiw anomaly [19]:

$$\partial_\mu j_5^\mu \equiv I(x) = N_{\rm f} \frac{\alpha_{\rm s}}{4\pi} \sum_{\rm A} F^A_{\mu\nu} \tilde{F}^{A\mu\nu} = N_{\rm f} \frac{\alpha_{\rm s}}{2\pi} {\rm Tr}(\mathbf{F}_{\boldsymbol{\mu\nu}} \tilde{\mathbf{F}}^{\boldsymbol{\mu\nu}}) \ , \tag{2.7}$$

recalling that $\mathbf{F}_{\boldsymbol{\mu\nu}} = \sum F^A_{\mu\nu} t^A$ and the normalization is ${\rm Tr}(t^A t^B) = 1/2\delta^{AB}$, with $F^A_{\mu\nu}$ given in (1.31) and j_5^μ the $u + d$ singlet axial current (the factor of $N_{\rm f}$, in this case $N_{\rm f} = 2$, in front of the right-hand side takes into account the fact that $N_{\rm f}$ flavours are involved), and

$$\tilde{F}^A_{\mu\nu} = \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} F^{A\rho\sigma} \ . \tag{2.8}$$

An important point is that the pseudoscalar quantity $I(x)$ is a four-divergence. More precisely, one can check that

$${\rm Tr}(\mathbf{F}_{\boldsymbol{\mu\nu}} \tilde{\mathbf{F}}^{\boldsymbol{\mu\nu}}) = \partial^\mu k_\mu \ , \tag{2.9}$$

with

$$k_\mu = \epsilon_{\mu\nu\lambda\sigma} {\rm Tr} \left[\mathbf{A}^\nu \left(\mathbf{F}^{\lambda\sigma} - \frac{2}{3} {\rm i} e_{\rm s} \mathbf{A}^\lambda \mathbf{A}^\sigma \right) \right] \ . \tag{2.10}$$

As a consequence the modified current $\tilde{j}_5^\mu$ and its associated charge $\tilde{Q}_5$ still appear to be conserved, viz.,

$$\partial_\mu \tilde{j}_5^\mu = \partial_\mu \left(j_5^\mu - N_{\rm f} \frac{\alpha_{\rm s}}{2\pi} k^\mu \right) = 0 \ , \tag{2.11}$$

and could act as a modified chiral current and charge with an additional gluonic component. But actually this charge is not conserved due to the topological structure of the QCD vacuum (instantons) as discussed in the following (for an introduction, see [308]).

The configuration where all gauge fields are zero $A^A_\mu = 0$ can be called "the vacuum". However, all configurations connected to $A^A_\mu = 0$ by a gauge transformation must also correspond to the same physical vacuum. For example, in an Abelian theory all gauge fields that can be written as the gradient of a scalar, i.e., $A^A_\mu = \partial_\mu \chi(x)$, are equivalent to $A^A_\mu = 0$. In non-Abelian gauge theories, there are some "large" gauge transformations that are topologically nontrivial and correspond to non-vanishing integer values of a topological charge, the "winding number". Taking $SU(2)$ for simplicity, although in QCD it could be any such

subgroup of colour $SU(3)$, we can consider the following time-independent gauge transformation:

$$\Omega_1(\mathbf{x}) = \frac{\mathbf{x}^2 - d^2 + 2\mathrm{i}d\boldsymbol{\tau}\cdot\mathbf{x}}{\mathbf{x}^2 + d^2} \,, \tag{2.12}$$

where d is a positive constant. Note that $\Omega_1^{-1} = \Omega_1^*$. Starting from $\mathbf{A}_\mu = (A_0, A_i) = (0,0)$ $(i = 1,2,3)$, with $\mathbf{A}_\mu = \sum A_\mu^a \tau^a/2$ and recalling the general expression of a gauge transformation in (1.15), the gauge transform of the potential by Ω_1 is

$$\mathbf{A}_\mathbf{j}^{(\mathbf{1})} = -\frac{\mathrm{i}}{e_\mathrm{s}}\big[\nabla_j \Omega_1(\mathbf{x})\big]\Omega_1^{-1}(\mathbf{x}) \,. \tag{2.13}$$

For the vector potential $\mathbf{A}^{(\mathbf{1})}$, which is a pure gauge and hence part of the "vacuum", the winding number n, defined in general by

$$n = \frac{\mathrm{i}e_\mathrm{s}^3}{24\pi^2}\int \mathrm{d}^3x\,\mathrm{Tr}\big[\mathbf{A_i}(\mathbf{x})\mathbf{A_j}(\mathbf{x})\mathbf{A_k}(\mathbf{x})\big]\epsilon^{ijk} \,, \tag{2.14}$$

is equal to 1, i.e., $n = 1$. Similarly, for $\mathbf{A}^{(\mathbf{m})}$ obtained from $\Omega_m = [\Omega_1]^m$, one has $n = m$. Given (2.9), we might expect the integrated four-divergence to vanish, but instead one finds

$$\frac{\alpha_\mathrm{s}}{4\pi}\int \mathrm{d}^4x\,\mathrm{Tr}(\mathbf{F}_{\boldsymbol{\mu\nu}}\tilde{\mathbf{F}}^{\boldsymbol{\mu\nu}}) = \frac{\alpha_\mathrm{s}}{4\pi}\int \mathrm{d}^4x\,\partial_\mu k^\mu = \frac{\alpha_\mathrm{s}}{4\pi}\left[\int \mathrm{d}^3x\,k_0\right]_{-\infty}^{+\infty} = n_+ - n_- \,, \tag{2.15}$$

for a configuration of gauge fields that vanish fast enough on the space sphere at infinity, and the winding numbers are $n_\mp$ at time $t = \mp\infty$ ("instantons").

From the above discussion it follows that in QCD all gauge fields can be classified in sectors with different n: there is a vacuum for each n, $|n\rangle$, and $\Omega_1|n\rangle = |n+1\rangle$ (not gauge invariant!). The true vacuum must be gauge invariant (up to a phase) and is obtained as a superposition of all $|n\rangle$:

$$|\theta\rangle = \sum_{-\infty}^{+\infty} \mathrm{e}^{-in\theta}|n\rangle \,. \tag{2.16}$$

In fact,

$$\Omega_1|\theta\rangle = \sum \mathrm{e}^{-\mathrm{i}n\theta}|n+1\rangle = \mathrm{e}^{\mathrm{i}\theta}|\theta\rangle \,. \tag{2.17}$$

If we compute the expectation value of any operator O in the θ vacuum, we find

$$\langle\theta|O|\theta\rangle = \sum_{m,n} \mathrm{e}^{\mathrm{i}(m-n)\theta}\langle m|O|n\rangle \,. \tag{2.18}$$

The path integral describing the O vacuum matrix element at $\theta = 0$ must be modified to reproduce the extra phase, taking (2.15) into account:

$$\langle\theta|O|\theta\rangle = \int \mathrm{d}A\mathrm{d}\bar{\psi}\mathrm{d}\psi O \exp\left[\mathrm{i}S_{\mathrm{QCD}} + \mathrm{i}\theta\frac{\alpha_{\mathrm{s}}}{4\pi}\int \mathrm{d}^4x\,\mathrm{Tr}(\mathbf{F}_{\mu\nu}\tilde{\mathbf{F}}^{\mu\nu})\right] . \tag{2.19}$$

This is equivalent to adding a θ term to the QCD Lagrangian:

$$\mathscr{L}_{\mathrm{QCD}} = \theta\frac{\alpha_{\mathrm{s}}}{4\pi}\int \mathrm{d}^4x\,\mathrm{Tr}(\mathbf{F}_{\mu\nu}\tilde{\mathbf{F}}^{\mu\nu}) . \tag{2.20}$$

The θ term is parity (P) odd and charge conjugation (C) even, so it introduces CP violation in the theory (and also time reversal (T) violation). A priori one would expect $\tilde{\theta}$ to be $O(1)$. But it would contribute to the neutron electric dipole moment, according to $d_n(e{\cdot}\mathrm{cm}) \sim 3\times 10^{-16}\tilde{\theta}$. The strong experimental bounds on d_n, viz., $d_n(e{\cdot}\mathrm{cm}) \leq 3\times 10^{-26}$ [307], imply that $\tilde{\theta}$ must be very small, viz., $\tilde{\theta} \leq 10^{-10}$. The so-called "strong CP problem" or "θ-problem" consists in finding an explanation for such a small value [263, 308]. An important point that is relevant for a possible solution is that a chiral transformation translates θ by a fixed amount. By recalling (2.11), we have

$$\mathrm{e}^{\mathrm{i}\delta\tilde{Q}_5}|\theta\rangle = |\theta - 2N_{\mathrm{f}}\delta\rangle . \tag{2.21}$$

To prove this relation we first observe that $\tilde{Q}_5$ is not gauge invariant under Ω_1, because it involves k_0 :

$$\Omega_1\tilde{Q}_5\Omega_1^{-1} = Q_5 - \Omega_1 2N_{\mathrm{f}}\frac{\alpha_{\mathrm{s}}}{4\pi}\left[\int \mathrm{d}^3x\,k_0\right]\Omega_1^{-1} = \tilde{Q}_5 - 2N_{\mathrm{f}} . \tag{2.22}$$

It then follows that

$$\Omega_1\mathrm{e}^{\mathrm{i}\delta\tilde{Q}_5}|\theta\rangle = \Omega_1\mathrm{e}^{\mathrm{i}\delta\tilde{Q}_5}\Omega_1^{-1}\Omega_1|\theta\rangle = \mathrm{e}^{\mathrm{i}(\theta-2N_{\mathrm{f}}\delta)}\mathrm{e}^{\mathrm{i}\delta\tilde{Q}_5}|\theta\rangle , \tag{2.23}$$

which implies (2.21). Thus in a chiral invariant theory, one could dispose of θ. For this it would be sufficient for a single quark mass to be zero, and the obvious candidate would be $m_u = 0$. But apparently this possibility has been excluded [263]. For non-vanishing quark masses, the transformation $m \to U_{\mathrm{L}}^{\dagger}mU_{\mathrm{R}}$ needed to make the mass matrix Hermitian (which implies γ_5-free) and diagonal involves a chiral transformation that affects θ. Considering that $U(N) = U(1)\bigotimes SU(N)$ and that for Hermitian m the argument of the determinant vanishes, i.e., $\arg\det m = 0$, the transformation from a generic m' to a real and diagonal m gives

$$\begin{aligned}\arg\det m = 0 &= \arg\det\ U_{\mathrm{L}}^{*} + \arg\det m' + \arg\det U_{\mathrm{R}}\\ &= -2N_{\mathrm{f}}(\delta_{\mathrm{L}} - \delta_{\mathrm{R}}) + \arg\det m' .\end{aligned} \tag{2.24}$$

From this equation one derives the phase $\delta_R - \delta_L$ of the chiral transformation and then, by (2.21), the important result for the effective θ value:

$$\theta_{\text{eff}} = \theta + \arg \det m' \,. \tag{2.25}$$

As we have seen the small empirical value of θ_{eff} poses a serious naturalness problem for the SM. Among the possible solutions, perhaps the most interesting option is a mechanism proposed by Peccei and Quinn [309]. One assumes that the SM or an enlarged theory is invariant under an additional chiral symmetry $U(1)_{PQ}$ acting on the fields of the theory. This symmetry is spontaneously broken by the vacuum expectation value v_{PQ} of a scalar field. The associated Goldstone boson, the axion, is actually not massless, because of the chiral anomaly. The parameter θ is canceled by the vacuum expectation value of the axion field due to the properties of the associated potential, also determined by the anomaly. Axions could contribute to the dark matter in the Universe, if their mass falls in a suitable narrow range (for a recent review, see, for example, [262]).

Alternative solutions to the θ-problem have also been suggested. Some of them can probably be discarded (for example, the idea that the up quark is exactly massless), while others are still possible: for example, in supersymmetric theories, if the smallness of θ could be guaranteed at the Planck scale by some feature of the more fundamental theory valid there, then the non-renormalization theorems of supersymmetry would preserve its small value throughout the transition down to low energy.

2.3 Massless QCD and Scale Invariance

As discussed in Chap. 2, the QCD Lagrangian in (1.28) only specifies the theory at the classical level. The procedure for quantizing gauge theories involves a number of complications that arise from the fact that not all degrees of freedom of gauge fields are physical because of the constraints from gauge invariance which can be used to eliminate the dependent variables. This is already true for Abelian theories and one is familiar with the QED case. One introduces a gauge fixing term (an additional term in the Lagrangian density that acts as a Lagrange multiplier in the action extremization). One can choose to preserve manifest Lorentz invariance. In this case, one adopts a covariant gauge, like the Lorentz gauge, and in QED one proceeds according to the formalism of Gupta and Bleuler [102]. Or one can give up explicit formal covariance and work in a non-covariant gauge, like the Coulomb or the axial gauges, and only quantize the physical degrees of freedom (in QED the transverse components of the photon field).

While this is all for an Abelian gauge theory, in the non-Abelian case some additional complications arise, in particular the need to introduce ghosts for the formulation of Feynman rules. As we have seen, there are in general as many ghost fields as gauge bosons, and they appear in the form of a transformation Jacobian

in the Feynman functional integral. Ghosts only propagate in closed loops and their vertices with gluons can be included as additional terms in the Lagrangian density, these being fixed once the gauge fixing terms and their infinitesimal gauge transformations are specified. Finally, the complete Feynman rules can be obtained in either the covariant or the axial gauges, and they appear in Fig. 2.1.

Once the Feynman rules are derived, we have a formal perturbative expansion, but loop diagrams generate infinities. First a regularization must be introduced, compatible with gauge symmetry and Lorentz invariance. This is possible in QCD. In principle, one can introduce a cutoff K (with dimensions of energy), for example, as done by Pauli and Villars [102]. But at present, the universally adopted regularization procedure is dimensional regularization, which we will describe briefly later on.

After regularization, the next step is renormalization. In a renormalizable theory (which is the case for all gauge theories in four spacetime dimensions and for QCD in particular), the dependence on the cutoff can be completely reabsorbed in a redefinition of particle masses, gauge coupling(s), and wave function normalizations. Once renormalization is achieved, the perturbative definition of the quantum theory that corresponds to a classical Lagrangian like (1.28) is completed.

In the QCD Lagrangian of (1.28), quark masses are the only parameters with physical dimensions (we work in the natural system of units $\hbar = c = 1$). Naively, we would expect massless QCD to be scale invariant. This is actually true at the classical level. Scale invariance implies that dimensionless observables should not depend on the absolute scale of energy, but only on ratios of energy-dimensional variables. The massless limit should be relevant for the large asymptotic energy limit of processes which are non-singular for $m \to 0$.

The naive expectation that massless QCD should be scale invariant is false in the quantum theory. The scale symmetry of the classical theory is unavoidably destroyed by the regularization and renormalization procedure, which introduce a dimensional parameter into the quantum version of the theory. When a symmetry of the classical theory is necessarily destroyed by quantization, regularization, and renormalization one talks of an "anomaly". So in this sense, scale invariance in massless QCD is anomalous.

While massless QCD is not in the end scale invariant, the departures from scaling are asymptotically small, logarithmic, and computable. In massive QCD, there are additional mass corrections suppressed by powers of m/E, where E is the energy scale (for processes that are non-singular in the limit $m \to 0$). At the parton level (q and g), we can consider applying the asymptotic predictions of massless QCD to processes and observables (we use the word "processes" for both) with the following properties ("hard processes"):

- All relevant energy variables must be large:

$$E_i = z_i Q \,, \quad Q \gg m_j \,, \quad z_i \text{ scaling variables } O(1) \,. \tag{2.26}$$

- There should be no infrared singularities (one talks of "infrared safe" processes).
- The processes concerned must be finite for $m \to 0$ (no mass singularities).

To have any chance of satisfying these criteria, processes must be as "inclusive" as possible: one should include all final states with massless gluon emission and add all mass degenerate final states (given that quarks are massless, $q\bar{q}$ pairs can also be massless if "collinear", that is moving together in the same direction at a common speed, the speed of light).

In perturbative QCD one computes inclusive rates for partons (the fields in the Lagrangian, that is, in QCD, quarks and gluons) and takes them as equal to rates for hadrons. Partons and hadrons are considered as two equivalent sets of complete states. This is called "global duality", and it is rather safe in the rare instance of a totally inclusive final state. It is less so for distributions, like distributions in the invariant mass M ("local duality"), where it can be reliable only if smeared over a sufficiently wide bin in M.

Let us discuss infrared and collinear safety in more detail. Consider, for example, a quark virtual line that ends up in a real quark plus a real gluon (Fig. 2.7). For the propagator we have

$$\text{propagator} = \frac{1}{(p+k)^2 - m^2} = \frac{1}{2(p\cdot k)} = \frac{1}{2E_k E_p}\cdot\frac{1}{1-\beta_p\cos\theta}\,. \tag{2.27}$$

Since the gluon is massless, E_k can vanish and this corresponds to an infrared singularity. Remember that we have to take the square of the amplitude and integrate it over the final state phase space, resulting in this case with $\mathrm{d}E_k/E_k$. Indeed, we get $1/E_k^2$ from the squared amplitude and $\mathrm{d}^3k/E_k \sim E_k \mathrm{d}E_k$ from the phase space. Further, for $m \to 0$, $\beta_p = \sqrt{1 - m^2/E_p^2} \to 1$ and $1-\beta_p\cos\theta$ vanishes at $\cos\theta = 1$, leading to a collinear mass singularity.

There are two very important theorems on infrared and mass singularities. The first one is the Bloch–Nordsieck theorem [103]: infrared singularities cancel between real and virtual diagrams (see Fig. 2.8) when all resolution-indistinguishable final states are added up. For example, for each real detector there is a minimum energy of gluon radiation that can be detected. For the cancellation of infrared divergences, one should add all possible gluon emission with a total energy below the detectable minimum.

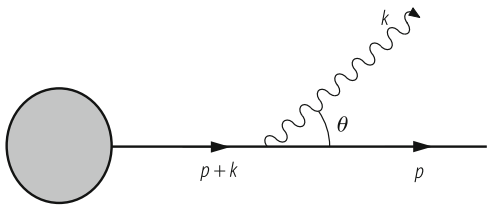

Fig. 2.7 The splitting of a virtual quark into a quark and a gluon

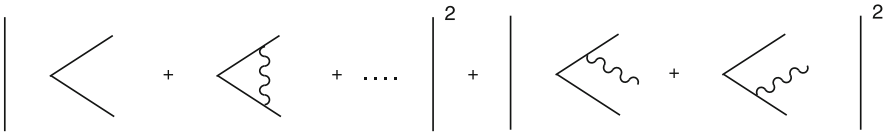

Fig. 2.8 Diagrams contributing to the total cross-section $e^+e^- \to$ hadrons at order α_s. For simplicity, only the final state quarks and (virtual or real) gluons are drawn

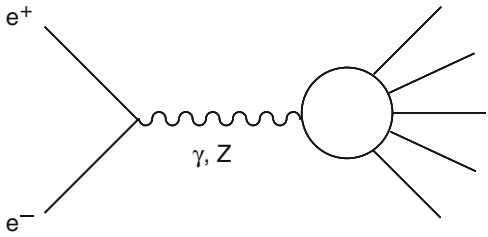

Fig. 2.9 Total cross-section $e^+e^- \to$ hadrons

The second one is the Kinoshita–Lee–Nauenberg theorem [265]: mass singularities connected with an external particle of mass m are canceled if all degenerate states (that is, with the same mass) are summed up. Hence, for a final state particle of mass m, we should add all final states that have the same mass in the limit $m \to 0$, including also gluons and massless pairs. If a completely inclusive final state is taken, only the mass singularities from the initial state particles remain (we shall see that they will be absorbed inside the non-perturbative parton densities, which are probability densities for finding the given parton in the initial hadron).

Hard processes to which the massless QCD asymptotics may possibly apply must be infrared and collinear safe, that is they must satisfy the requirements of the Bloch–Nordsieck and the Kinoshita–Lee–Nauenberg theorems. We now give some examples of important hard processes. One of the simplest hard processes is the totally inclusive cross-section for hadron production in e^+e^- annihilation (see Fig. 2.9), parameterized in terms of the already mentioned dimensionless observable $R = \sigma(e^+e^- \to \text{hadrons})/\sigma_{\text{point}}(e^+e^- \to \mu^+\mu^-)$. The pointlike cross-section in the denominator is given by $\sigma_{\text{point}} = 4\pi\alpha^2/3s$, where $s = Q^2 = 4E^2$ is the squared total center of mass energy and Q is the mass of the exchanged virtual gauge boson.

At parton level, the final state is $q\bar{q} + ng + n'q'\bar{q}'$, and n and n' are limited at each order of perturbation theory. It is assumed that the conversion of partons into hadrons does not affect the rate (it happens with probability 1). We have already mentioned that, in order for this to be true within a given accuracy, averaging over a sufficiently large bin of Q must be understood. The binning width is larger in the vicinity of thresholds: for example, when one goes across the charm $c\bar{c}$ threshold, the physical cross-section shows resonance bumps that are absent in the smooth partonic counterpart, which, however, gives an average of the cross-section.

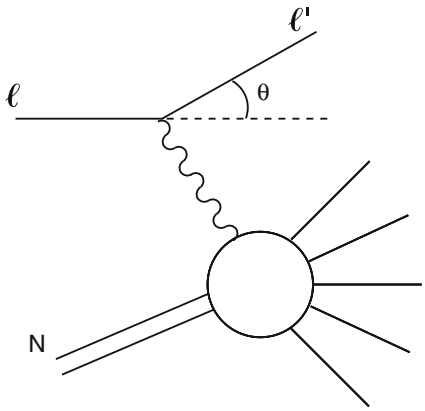

Fig. 2.10 Deep inelastic lepto-production

A very important class of hard processes is deep inelastic scattering (DIS):

$$l + N \to l' + X\,, \qquad l = e^{\pm},\, \mu^{\pm},\, \nu,\, \bar{\nu}\,. \tag{2.28}$$

This has played, and still plays, a very important role in our understanding of QCD and nucleon structure. For the processes in (2.28) (see Fig. 2.10), in the lab system where the nucleon of mass m is at rest, we have

$$Q^2 = -q^2 = -(k-k')^2 = 4EE' \sin^2\frac{\theta}{2}\,, \quad m\nu = (p.q)\,, \quad x = \frac{Q^2}{2m\nu}\,. \tag{2.29}$$

In this case the virtual momentum q of the gauge boson is spacelike. x is the familiar Bjorken variable. The DIS processes in QCD will be discussed extensively in Sect. 2.8.

2.4 The Renormalization Group and Asymptotic Freedom

In this section we aim to provide a reasonably detailed introduction to the renormalization group formalism and the concept of running coupling, which leads to the result that QCD has the property of asymptotic freedom. We start with a summary of how renormalization works.

In the simplest conceptual situation imagine that we implement regularization of divergent integrals by introducing a dimensional cutoff K that respects gauge and Lorentz invariance. The dependence of renormalized quantities on K is eliminated by absorbing it into a redefinition of m, the quark mass (for simplicity we assume a single flavour here), the gauge coupling e (which can be e in QED or e_s in QCD), and the wave function renormalization factors $Z_{q,g}^{1/2}$ for q and g, using suitable renormalization conditions (that is, precise definitions of m, g, and Z that can be

implemented order by order in perturbation theory). For example, we can define the renormalized mass m as the position of the pole in the quark propagator, and similarly, the normalization Z_q as the residue at the pole:

$$\text{propagator} = \frac{Z_q}{p^2 - m^2} + \text{no-pole terms}\,. \tag{2.30}$$

The renormalized coupling e can be defined in terms of a renormalized 3-point vertex at some specified values of the external momenta. More precisely, we consider a one-particle irreducible vertex (1PI). We recall that a connected Green function is the sum of all connected diagrams, while 1PI Green functions are the sum of all diagrams that cannot be separated into two disconnected parts by cutting only one line.

We now become more specific, by concentrating on the case of massless QCD. If we start from a vanishing mass at the classical (or "bare") level $m_0 = 0$, the mass is not renormalized because it is protected by a symmetry, namely, chiral symmetry. The conserved currents of chiral symmetry are axial currents: $\bar{q}\gamma_\mu\gamma_5 q$. Using the Dirac equation, divergence of the axial current gives $\partial^\mu(\bar{q}\gamma_\mu\gamma_5 q) = 2m\bar{q}\gamma_5 q$. So the axial current and the corresponding axial charge are conserved in the massless limit. Actually, the singlet axial current is not conserved due to the anomaly, but since QCD is a vector theory, we do not have to worry about chiral anomalies in the present context. As there are no γ_5 factors around, the chosen regularization preserves chiral symmetry as well as gauge and Lorentz symmetry, and the renormalized mass remains zero. The renormalized propagator has the form (2.30) with $m = 0$.

The renormalized coupling e_s can be defined from the renormalized 1PI 3-gluon vertex at a scale $-\mu^2$ (Fig. 2.11):

$$V_{\text{bare}}(p^2, q^2, r^2) = ZV_{\text{ren}}(p^2, q^2, r^2)\,, \quad Z = Z_g^{-3/2}\,, \quad V_{\text{ren}}(-\mu^2, -\mu^2, -\mu^2) \to e_s\,. \tag{2.31}$$

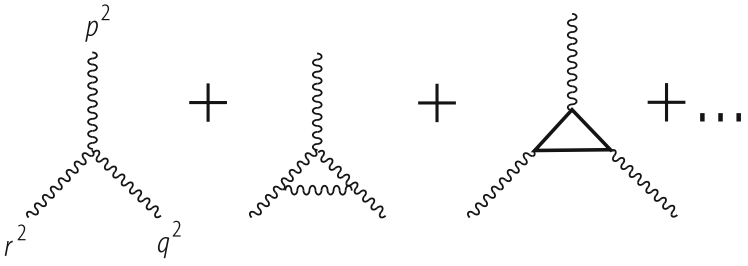

Fig. 2.11 Diagrams contributing to the 1PI 3-gluon vertex at the one-loop approximation level

We could just as well use the quark–gluon vertex or any other vertex which coincides with e_{s0} in lowest order (even the ghost–gluon vertex, if we want). With a regularization and renormalization that preserves gauge invariance, we can be sure that all these different definitions are equivalent.

Here V_{bare} is what is obtained from computing the Feynman diagrams including, for example, the 1-loop corrections at the lowest non-trivial order. V_{bare} is defined as the scalar function multiplying the 3-gluon vertex tensor (given in Fig. 2.1), normalized in such a way that it coincides with e_{s0} in lowest order. V_{bare} contains the cutoff K, but does not know about μ. Z is a factor that depends both on the cutoff and on μ, but not on momenta. Because of infrared singularities, the defining scale μ cannot vanish. The negative value $-\mu^2 < 0$ is chosen to stay away from physical cuts (a gluon with negative virtual mass cannot decay). Similarly, in the massless theory, we can define Z_g^{-1} as the inverse gluon propagator (the 1PI 2-point function) at the same scale $-\mu^2$ (the vanishing mass of the gluon is guaranteed by gauge invariance).

After computing all 1-loop diagrams indicated in Fig. 2.11, we have

$$\begin{aligned} V_{\text{bare}}(p^2,p^2,p^2) &= e_{s0}\left(1 + c\alpha_{s0}\log\frac{K^2}{p^2} + \cdots\right) \\ &= \left(1 + c\alpha_s \log\frac{K^2}{-\mu^2} + \ldots\right) e_{s0}\left(1 + c\alpha_{s0}\log\frac{-\mu^2}{p^2}\right) \\ &= Z_V^{-1} e_{s0}\left(1 + c\alpha_s \log\frac{-\mu^2}{p^2}\right) \\ &= \left(1 + d\alpha_s \log\frac{K^2}{-\mu^2} + \cdots\right) e_s\left(1 + c\alpha_s \log\frac{-\mu^2}{p^2}\right) \\ &= Z_g^{-3/2} V_{\text{ren}}\,. \end{aligned} \tag{2.32}$$

Note the replacement of α_{s0} with α_s in the second step, as we work at 1-loop accuracy. Then we change e_{s0} into e_s, given by $e_0 = Z_g^{-3/2} Z_V e$, and this implies changing c into d in the first bracket. The definition of e_s requires precise specification of what is included in Z. For this, in a given renormalization scheme, a prescription is fixed to specify the finite terms that go into Z, i.e., the terms of order α_s that accompany $\log K^2$. Then V_{ren} is specified and the renormalized coupling is defined from it according to (2.31). For example, in the momentum subtraction scheme we define $V_{\text{ren}}(p^2,p^2,p^2) = e_s + V_{\text{bare}}(p^2,p^2,p^2) - V_{\text{bare}}(-\mu^2,-\mu^2,-\mu^2)$, which is equivalent to saying that, at 1-loop, all finite terms that do not vanish at $p^2 = -\mu^2$ are included in Z.

A crucial observation is that V_{bare} depends on K, but not on μ, which is only introduced when Z, V_{ren}, and hence α_s are defined. (From here on, for simplicity, we write α to indicate either the QED coupling or the QCD coupling α_s.) Similarly, for

a generic Green function G, we have more generally

$$G_{\text{bare}}(K^2, \alpha_0, p_i^2) = Z_G G_{\text{ren}}(\mu^2, \alpha, p_i^2) \ , \tag{2.33}$$

whence

$$\frac{\mathrm{d}G_{\text{bare}}}{\mathrm{d}\log\mu^2} = \frac{\mathrm{d}}{\mathrm{d}\log\mu^2}(Z_G G_{\text{ren}}) = 0 \ , \tag{2.34}$$

or

$$Z_G \left(\frac{\partial}{\partial \log\mu^2} + \frac{\partial\alpha}{\partial\log\mu^2}\frac{\partial}{\partial\alpha} + \frac{1}{Z_G}\frac{\partial Z_G}{\partial\log\mu^2} \right) G_{\text{ren}} = 0 \ . \tag{2.35}$$

Finally, the renormalization group equation (RGE) can be written as

$$\left[\frac{\partial}{\partial \log\mu^2} + \beta(\alpha)\frac{\partial}{\partial\alpha} + \gamma_G(\alpha) \right] G_{\text{ren}} = 0 \ , \tag{2.36}$$

where

$$\beta(\alpha) = \frac{\partial\alpha}{\partial\log\mu^2} \tag{2.37}$$

and

$$\gamma_G(\alpha) = \frac{\partial \log Z_G}{\partial\log\mu^2} \ . \tag{2.38}$$

Note that $\beta(\alpha)$ does not depend on which Green function G we are considering. Actually, it is a property of the theory and of the renormalization scheme adopted, while $\gamma_G(\alpha)$ also depends on G. Strictly speaking the RGE as written above is only valid in the Landau gauge ($\lambda = 0$). In other gauges, an additional term that takes the variation of the gauge fixing parameter λ into account should also be included. We omit this term, for simplicity, as it is not relevant at the 1-loop level.

Suppose we want to apply the RGE to some hard process at a large scale Q, related to a Green function G that we can always take to be dimensionless (by multiplying by a suitable power of Q). Since the interesting dependence on Q will be logarithmic, we introduce the variable t as

$$t = \log\frac{Q^2}{\mu^2} \ . \tag{2.39}$$

Then we can write $G_{\text{ren}} \equiv F(t, \alpha, x_i)$, where x_i are scaling variables (we shall often omit them in the following). In the naive scaling limit, F should be independent of

t, according to the classical intuition that massless QCD is scale invariant. To find the actual dependence on t, we must solve the RGE

$$\left[-\frac{\partial}{\partial t}+\beta(\alpha)\frac{\partial}{\partial \alpha}+\gamma_G(\alpha)\right]G_{\text{ren}}=0\,, \tag{2.40}$$

with a given boundary condition at $t=0$ (or $Q^2=\mu^2$), viz., $F(0,\alpha)$.

We first solve the RGE in the simplest case, i.e., when $\gamma_G(\alpha)=0$. This is not an unphysical case. For example, it applies to

$$R=R_{e^+e^-}=\frac{\sigma(e^+e^-\to \text{hadrons})}{\sigma_{\text{point}}(e^+e^-\to \mu^+\mu^-)}\,,$$

where the vanishing of γ is related to the non-renormalization of the electric charge in QCD (otherwise the proton and the electron charge would not exactly balance, something we explain in Sect. 2.7). So we consider the equation

$$\left[-\frac{\partial}{\partial t}+\beta(\alpha)\frac{\partial}{\partial \alpha}\right]G_{\text{ren}}=0\,. \tag{2.41}$$

The solution is simply

$$F(t,\alpha)=F[0,\alpha(t)]\,, \tag{2.42}$$

where the "running coupling" $\alpha(t)$ is defined by

$$t=\int_{\alpha}^{\alpha(t)}\frac{1}{\beta(\alpha')}\mathrm{d}\alpha'\,. \tag{2.43}$$

Note that from this definition it follows that $\alpha(0)=\alpha$, so that the boundary condition is also satisfied. To prove that $F[0,\alpha(t)]$ is indeed the solution, we first take derivatives with respect of t and α (the two independent variables) of both sides of (2.43). By taking $\mathrm{d}/\mathrm{d}t$ we obtain

$$1=\frac{1}{\beta(\alpha(t))}\frac{\partial\alpha(t)}{\partial t}\,. \tag{2.44}$$

We then take $\mathrm{d}/\mathrm{d}\alpha$ and obtain

$$0=-\frac{1}{\beta(\alpha)}+\frac{1}{\beta(\alpha(t))}\frac{\partial\alpha(t)}{\partial\alpha}\,. \tag{2.45}$$

These two relations make explicit the dependence of the running coupling on t and α:

$$\frac{\partial \alpha(t)}{\partial t} = \beta(\alpha(t)) \ , \qquad \frac{\partial \alpha(t)}{\partial \alpha} = \frac{\beta(\alpha(t))}{\beta(\alpha)} \ . \tag{2.46}$$

Using these two equations, one immediately checks that $F[0, \alpha(t)]$ is indeed the solution.

Similarly, one finds that the solution of the more general equation (2.40) with $\gamma \neq 0$ is given by

$$F(t, \alpha) = F[0, \alpha(t)] \exp \int_{\alpha}^{\alpha(t)} \frac{\gamma(\alpha')}{\beta(\alpha')} \mathrm{d}\alpha' \ . \tag{2.47}$$

In fact the sum of the two derivatives acting on the factor $F[0, \alpha(t)]$ vanishes (as we have just seen), and the exponential is by itself a solution of the complete equation. Note that the boundary condition is also satisfied.

The important point is the appearance of the running coupling that determines the asymptotic departures from scaling. The next step is to study the functional form of the running coupling. From (2.46) we see that the rate of change of the running coupling with respect to t is determined by the function β. In turn, $\beta(\alpha)$ is determined by the μ dependence of the renormalized coupling through (2.37). Clearly, there is no dependence of the basic 3-gluon vertex on μ to lowest order (order e). The dependence starts at 1-loop, that is at order e^3 (one extra gluon has to be emitted and reabsorbed). Thus we find that, in perturbation theory,

$$\frac{\partial e}{\partial \log \mu^2} \propto e^3 \ . \tag{2.48}$$

Recalling that $\alpha = e^2/4\pi$, we have

$$\frac{\partial \alpha}{\partial \log \mu^2} \propto 2e \frac{\partial e}{\partial \log \mu^2} \propto e^4 \propto \alpha^2 \ . \tag{2.49}$$

Thus the behaviour of $\beta(\alpha)$ in perturbation theory is

$$\beta(\alpha) = \pm b \alpha^2 (1 + b' \alpha + \cdots) \ . \tag{2.50}$$

Since the sign of the leading term is crucial in the following discussion, we stipulate that $b > 0$ and we make the sign explicit in front.

Let us make the procedure more precise for computing the 1-loop beta function in QCD (or, similarly, in QED). The result of the 1-loop 1PI diagrams for $V_{\rm ren}$ can be written as

$$V_{\rm ren} = e\left(1 + \alpha B_{3g} \log \frac{\mu^2}{-p^2}\right) . \tag{2.51}$$

$V_{\rm ren}$ satisfies the RGE

$$\left[\frac{\partial}{\partial \log \mu^2} + \beta(\alpha)\frac{\partial e}{\partial \alpha}\frac{\partial}{\partial e} - \frac{3}{2}\gamma_g(\alpha)\right] V_{\rm ren} = 0 . \tag{2.52}$$

With respect to (2.36), the beta function term has been rewritten taking into account the fact that $V_{\rm ren}$ starts with e, and the anomalous dimension term arises from a factor $Z_g^{-1/2}$ for each gluon leg. In general, for an n-leg 1PI Green function $V_{n,\rm bare} = Z_g^{-n/2} V_{n,\rm ren}$, if all external legs are gluons. Note that, in the particular case of $V = V_3$ that is used to define e, other Z factors are absorbed in the replacement $Z_{\rm V}^{-1} Z_g^{3/2} e_0 = e$. At 1-loop accuracy, we replace $\beta(\alpha) = -b\alpha^2$ and $\gamma_g(\alpha) = \gamma_g^{(1)}\alpha$. One thus obtains

$$b = 2\left[B_{3g} - \frac{3}{2}\gamma_g^{(1)}\right] . \tag{2.53}$$

Similarly, we can write the diagrammatic expression and the RGE for the 1PI 2-gluon Green function, which is the inverse gluon propagator Π (a scalar function after removing the gauge invariant tensor):

$$\Pi_{\rm ren} = \left(1 + \alpha B_{2g} \log \frac{\mu^2}{-p^2} + \cdots\right) \tag{2.54}$$

and

$$\left[\frac{\partial}{\partial \log \mu^2} + \beta(\alpha)\frac{\partial}{\partial \alpha} - \gamma_g(\alpha)\right] \Pi_{\rm ren} = 0 . \tag{2.55}$$

Notice that the normalization and the phase of Π are specified by the lowest order term being 1. In this case the β function term is negligible, being of order α^2 (because Π is a function of e only through α) and we obtain

$$\gamma_g^{(1)} = B_{2g} . \tag{2.56}$$

Thus, finally,

$$b = 2\left(B_{3g} - \frac{3}{2}B_{2g}\right) . \tag{2.57}$$

By direct calculation at 1-loop level, one finds

$$\text{QED} \qquad \beta(\alpha) \sim +b\alpha^2 + \cdots , \quad b = \sum_i \frac{N_C Q_i^2}{3\pi} , \tag{2.58}$$

where $N_C = 3$ for quarks and $N_C = 1$ for leptons, and the sum runs over all fermions of charge $Q_i e$ that are coupled. One also finds

$$\text{QCD} \qquad \beta(\alpha) \sim -b\alpha^2 + \cdots , \quad b = \frac{11N_C - 2n_f}{12\pi} , \tag{2.59}$$

where, as usual, n_f is the number of coupled (see below) flavours of quarks (we assume here that $n_f \leq 16$, so that $b > 0$ in QCD).

If $\alpha(t)$ is small, we can compute $\beta(\alpha(t))$ in perturbation theory. The sign in front of b then decides the slope of the coupling: $\alpha(t)$ increases with t (or Q^2) if β is positive at small α (QED), or $\alpha(t)$ decreases with t (or Q^2) if β is negative at small α (QCD). A theory like QCD in which the running coupling vanishes asymptotically at large Q^2 is said to be (ultraviolet) "asymptotically free". An important result that has been proven [145] is that, in four spacetime dimensions, all and only non-Abelian gauge theories are asymptotically free.

Going back to (2.43), we replace $\beta(\alpha) \sim \pm b\alpha^2$, do the integral, and perform some simple algebra to find

$$\text{QED} \qquad \alpha(t) \sim \frac{\alpha}{1 - b\alpha t} \tag{2.60}$$

and

$$\text{QCD} \qquad \alpha(t) \sim \frac{\alpha}{1 + b\alpha t} . \tag{2.61}$$

A slightly different form is often used in QCD. Defining $1/\alpha = b \log \mu^2/\Lambda^2_{\text{QCD}}$, we can write

$$\alpha(t) \sim \frac{1}{\frac{1}{\alpha} + bt} = \frac{1}{b \log \frac{\mu^2}{\Lambda^2_{\text{QCD}}} + b \log \frac{Q^2}{\mu^2}} = \frac{1}{b \log \frac{Q^2}{\Lambda^2_{\text{QCD}}}} . \tag{2.62}$$

The parameter μ has been traded for the parameter Λ_{QCD}. We see that $\alpha(t)$ decreases logarithmically with Q^2 and that one can introduce a dimensional parameter Λ_{QCD} that replaces μ. In the following we will often simply write Λ for Λ_{QCD}. Note that it is clear that Λ depends on the particular definition of α, not only on the defining scale μ, but also on the renormalization scheme (see, for example, the discussion in the next section). Through the parameter b, and in general through the function β, it also depends on the number n_f of coupled flavours.

It is very important to note that QED and QCD are theories with "decoupling", i.e., up to the scale Q, only quarks with masses $m \ll Q$ contribute to the running of α. This is clearly very important, given that all applications of perturbative QCD so far apply to energies below the top quark mass m_t. For the validity of the decoupling theorem [60], the theory in which all the heavy particle internal lines are eliminated must still be renormalizable and the coupling constants must not vary with the mass. These requirements are satisfied for the masses of heavy quarks in QED and QCD, but they are not satisfied in the electroweak theory where the elimination of the top would violate $SU(2)$ symmetry (because the t and b left-handed quarks are in a doublet) and the quark couplings to the Higgs multiplet (hence to the longitudinal gauge bosons) are proportional to the mass.

In conclusion, in QED and QCD, quarks with $m \gg Q$ do not contribute to n_f in the coefficients of the relevant β function. The effects of heavy quarks are power suppressed and can be taken into account separately. For example, in e^+e^- annihilation for $2m_c < Q < 2m_b$, the relevant asymptotics is for $n_\mathrm{f} = 4$, while for $2m_b < Q < 2m_t$, it is for $n_\mathrm{f} = 5$. Going across the b threshold, the β function coefficients change, so the slope of $\alpha(t)$ changes. But $\alpha(t)$ is continuous, whence Λ changes so as to keep $\alpha(t)$ constant at the matching point at $Q \sim O(2m_b)$. The effect on Λ is large: approximately $\Lambda_5 \sim 0.65\Lambda_4$, where $\Lambda_{4,5}$ are for $n_\mathrm{f} = 4, 5$.

Note the presence of a pole at $\pm b\alpha t = 1$ in (2.60) and (2.61). This is called the Landau pole, since Landau had already realised its existence in QED in the 1950s. For $\mu \sim m_e$ (in QED), the pole occurs beyond the Planck mass. In QCD, the Landau pole is located for negative t or at $Q < \mu$ in the region of light hadron masses. Clearly the issue of the definition and the behaviour of the physical coupling (which is always finite, when defined in terms of some physical process) in the region around the perturbative Landau pole is a problem that lies outside the scope of perturbative QCD.

The non-leading terms in the asymptotic behaviour of the running coupling can in principle be evaluated by going back to (2.50) and computing b' at 2-loops and so on. But in general the perturbative coefficients of $\beta(\alpha)$ depend on the definition of the renormalized coupling α (the renormalization scheme), so one wonders whether it is worthwhile to do a complicated calculation to get b', if it must then be repeated for a different definition or scheme. In this respect it is interesting to note that both b and b' are actually independent of the definition of α, while higher order coefficients do depend on that. Here is the simple proof. Two different perturbative definitions of α are related by $\alpha' \sim \alpha(1 + c_1\alpha + \cdots)$. Then we have

$$\begin{aligned}
\beta(\alpha') = \frac{\mathrm{d}\alpha'}{\mathrm{d}\log\mu^2} &= \frac{\mathrm{d}\alpha}{\mathrm{d}\log\mu^2}(1 + 2c_1\alpha + \cdots) \\
&= \beta(\alpha)(1 + 2c_1\alpha + \ldots) \\
&= \pm b\alpha^2(1 + b'\alpha + \cdots)(1 + 2c_1\alpha + \cdots) \\
&= \pm b\alpha'^2(1 + b'\alpha' + \cdots)\,, \qquad (2.63)
\end{aligned}$$

which shows that, up to the first subleading order, $\beta(\alpha')$ has the same form as $\beta(\alpha)$.

In QCD ($N_C = 3$), it has been shown that [131]

$$b' = \frac{153 - 19n_f}{2\pi(33 - 2n_f)} . \tag{2.64}$$

By taking b' into account, one can write the expression for the running coupling at next to the leading order (NLO):

$$\alpha(Q^2) = \alpha_{LO}(Q^2)\left[1 - b'\alpha_{LO}(Q^2)\log\log\frac{Q^2}{\Lambda^2} + \cdots\right] , \tag{2.65}$$

where $\alpha_{LO}^{-1} = b\log Q^2/\Lambda^2$ is the LO result (actually at NLO, the definition of Λ is modified according to $b\log\mu^2/\Lambda^2 = 1/\alpha + b'\log b\alpha$).

Summarizing, we started from massless classical QCD which is scale invariant. But we have seen that the procedure of quantization, regularization, and renormalization necessarily breaks scale invariance. In the quantum QCD theory, there is a scale of energy Λ. From experiment, this is of the order of a few hundred MeV, its precise value depending on the definition, as we shall see in detail. Dimensionless quantities depend on the energy scale through the running coupling, which is a logarithmic function of Q^2/Λ^2. In QCD the running coupling decreases logarithmically at large Q^2 (asymptotic freedom), while in QED the coupling has the opposite behaviour.

2.5 More on the Running Coupling

In the last section we introduced the renormalized coupling α in terms of the 3-gluon vertex at $p^2 = -\mu^2$ (momentum subtraction). The Ward identities of QCD then ensure that the coupling defined from other vertices like the $\bar{q}qg$ vertex are renormalized in the same way and the finite radiative corrections are related. But at present the universally adopted definition of α_s is in terms of dimensional regularization [333], because of computational simplicity, which is essential given the great complexity of present day calculations. So we now briefly review the principles of dimensional regularization and the definition of minimal subtraction (MS) [335] and modified minimal subtraction ($\overline{\text{MS}}$) [82]. The $\overline{\text{MS}}$ definition of α_s is the one most commonly adopted in the literature, and values quoted for it normally refer to this definition.

Dimensional regularization (DR) is a gauge and Lorentz invariant regularization that consists in formulating the theory in $D < 4$ spacetime dimensions in order to make loop integrals ultraviolet finite. In DR one rewrites the theory in D dimensions (D is integer at the beginning, but then one realizes that the expression calculated from diagrams makes sense for all D, except for isolated singularities). The metric tensor is extended to a $D \times D$ matrix $g_{\mu\nu} = \text{diag}(1, -1, -1, \ldots, -1)$ and 4-vectors are given by $k^\mu = (k^0, k^1, \ldots, k^{D-1})$. The Dirac γ^μ are $f(D) \times f(D)$ matrices and

the precise form of the function $f(D)$ is not important. It is sufficient to extend the usual algebra in a straightforward way like $\{\gamma_\mu, \gamma_\nu\} = 2g_{\mu,\nu}I$, where I is the D-dimensional identity matrix, $\gamma^\mu\gamma^\nu\gamma_\mu = -(D-2)\gamma^\nu$, or $\mathrm{Tr}(\gamma^\mu\gamma^\nu) = f(D)g_{\mu\nu}$.

The physical dimensions of fields change in D dimensions, and as a consequence the gauge couplings become dimensional $e_D = \mu^\epsilon e$, where e is dimensionless, $D = 4-2\epsilon$, and μ is a mass scale (this is how a scale of mass is introduced in the DR of massless QCD). In fact, the dimension of the fields is determined by requiring the action $S = \int \mathrm{d}^D x \mathscr{L}$ to be dimensionless. By inserting terms like $m\bar{\Psi}\Psi$ or $m^2\phi^\dagger\phi$ or $e\bar{\Psi}\gamma^\mu\Psi A_\mu$ for $\mathscr{L}$, the dimensions of the fields and couplings m, Ψ, ϕ, A_μ, and e are determined as 1, $(D-1)/2$, $(D-2)/2$, $(D-2)/2$, and $(4-D)/2$, respectively. The formal expression of loop integrals can be written for any D. For example,

$$\int \frac{\mathrm{d}^D k}{(2\pi)^D} \frac{1}{(k^2-m^2)^2} = \frac{\Gamma(2-D/2)(-m^2)^{D/2-2}}{(4\pi)^{D/2}} \,. \tag{2.66}$$

For $D = 4-2\epsilon$, one can expand using

$$\Gamma(\epsilon) = \frac{1}{\epsilon} - \gamma_\mathrm{E} + O(\epsilon) \,, \quad \gamma_\mathrm{E} = 0.5772\ldots \,. \tag{2.67}$$

For some Green function G, normalized to 1 in lowest order (like V/e, with V the 3-gluon vertex function at the symmetric point $p^2 = q^2 = r^2$, considered in the previous section), we typically find, at the 1-loop level,

$$G_\mathrm{bare} = 1 + \alpha_0 \left(\frac{-\mu^2}{p^2}\right)^\epsilon \left[B\left(\frac{1}{\epsilon} + \log 4\pi - \gamma_\mathrm{E}\right) + A + O(\epsilon)\right] \,. \tag{2.68}$$

In $\overline{\mathrm{MS}}$, one rewrites this as (diagram by diagram, a virtue of the method)

$$\begin{aligned} G_\mathrm{bare} &= Z G_\mathrm{ren} \,, \\ Z &= 1 + \alpha\left[B\left(\frac{1}{\epsilon} + \log 4\pi - \gamma_\mathrm{E}\right)\right] \,, \\ G_\mathrm{ren} &= 1 + \alpha\left(B \log\frac{-\mu^2}{p^2} + A\right) \,. \end{aligned} \tag{2.69}$$

Here Z stands for the relevant product of renormalization factors. In the original MS prescription, only $1/\epsilon$ was subtracted (and this clearly plays the role of a cutoff), while $\log 4\pi$ and γ_E were not. Later, since these constants always appear in the expansion of Γ functions, it was decided to modify MS into $\overline{\mathrm{MS}}$. Note that the $\overline{\mathrm{MS}}$ definition of α is different than that in the momentum subtraction scheme, because the finite terms (those beyond logs) are different. In particular, the order α correction to G_ren does not vanish at $p^2 = -\mu^2$.

The third [337] and fourth [357] coefficients of the QCD β function are also known in the $\overline{\text{MS}}$ prescription (recall that only the first two coefficients are scheme-independent). The calculation of the last term involved the evaluation of some 50,000 four-loop diagrams. Translated in numbers, for $n_\text{f} = 5$, one obtains

$$\beta(\alpha) = -0.610\alpha^2\left[1 + 1.261\ldots\frac{\alpha}{\pi} + 1.475\ldots\left(\frac{\alpha}{\pi}\right)^2 + 9.836\ldots\left(\frac{\alpha}{\pi}\right)^3 + \cdots\right]. \tag{2.70}$$

It is interesting to remark that the expansion coefficients are of order 1 or 10 (only for the last one), so that the $\overline{\text{MS}}$ expansion looks reasonably well behaved.

2.6 On the Non-convergence of Perturbative Expansions

It is important to keep in mind that, after renormalization, all the coefficients in the QED and QCD perturbative series are finite, but the expansion does not converge. Actually, the perturbative series is not even Borel summable (for reviews see, for example, [31]). After the Borel resummation, for a given process, one is left with a result that is ambiguous up to terms typically going as $\exp(-n/b\alpha)$, where n is an integer and b the absolute value of the first β function coefficient. In QED, these corrective terms are extremely small and not very important in practice. However, in QCD, $\alpha = \alpha_\text{s}(Q^2) \sim 1/b\log(Q^2/\Lambda^2)$ and the ambiguous terms are of order $(1/Q^2)^n$, that is, they are power suppressed. It is interesting that, through this mechanism, the perturbative version of the theory is somehow able to take into account the power-suppressed corrections. A sequence of diagrams with factorial growth at large order n is constructed by dressing gluon propagators by any number of quark bubbles together with their gauge completions (renormalons). The problem of the precise relation between the ambiguities of the perturbative expansion and the power-suppressed corrections has been discussed in recent years, also for processes without light cone operator expansion [31, 324].

2.7 e^+e^- Annihilation and Related Processes

2.7.1 $R_{e^+e^-}$

The simplest hard process is

$$R = R_{e^+e^-} = \frac{\sigma(e^+e^- \to \text{hadrons})}{\sigma_\text{point}(e^+e^- \to \mu^+\mu^-)},$$

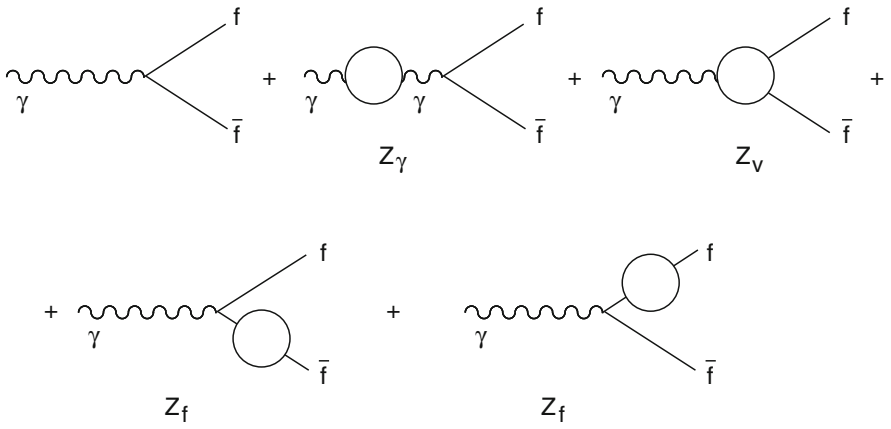

Fig. 2.12 Diagrams for charge renormalization in QED at 1-loop (the blob in each diagram represents the loop)

which we have already introduced. R is dimensionless and is given in perturbation theory by[1] $R = N_C \sum_i Q_i^2 F(t, \alpha_s)$, where $F = 1 + O(\alpha_s)$. We have already mentioned that for this process the "anomalous dimension" function vanishes, i.e., $\gamma(\alpha_s) = 0$, because of electric charge non-renormalization by strong interactions. Let us recall how this happens in detail.

The diagrams that are relevant for charge renormalization in QED at 1-loop are shown in Fig. 2.12. The Ward identity that follows from gauge invariance in QED requires the vertex (Z_V) and the self-energy (Z_f) renormalization factors to cancel, and the only divergence remains in Z_γ, the vacuum polarization of the photon. Hence, the charge is only renormalized by the photon vacuum polarization blob, and it is thus universal (the same factor for all fermions, independent of their charge) and not affected by QCD at 1-loop. It is true that at higher orders the photon vacuum polarization diagram is affected by QCD (for example, at 2-loops we can exchange a gluon between the quarks in the loop), but the renormalization induced by the divergent logs from the vacuum polarization diagram remain independent of the nature of the fermion to which the photon line is attached. The gluon contributions to the vertex (Z_V) and to the self-energy (Z_f) cancel, because they have exactly the same structure as in QED, and there is no gluon contribution to the photon blob at 1-loop, so that $\gamma(\alpha_s) = 0$.

At the 1-loop level, the diagrams relevant for the computation of R are shown in Fig. 2.13. There are virtual diagrams and also real diagrams with one additional gluon in the final state. Infrared divergences cancel between the interference term of the virtual diagrams and the absolute square of the real diagrams, according to the

[1] Actually, starting from order α_s^2, there are some "singlet" terms proportional to $(\sum_i Q_i)^2$. These small terms are included in F by dividing and multiplying by $\sum_i Q_i^2$.

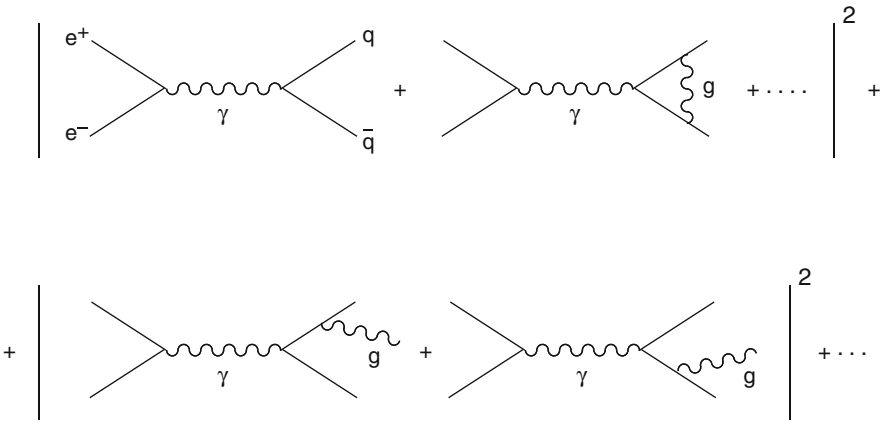

Fig. 2.13 Real and virtual diagrams relevant for the computation of R at 1-loop accuracy (the initial e^+e^- has been omitted to make the drawing simpler)

Bloch–Nordsieck theorem. Similarly, there are no mass singularities, in agreement with the Kinoshita–Lee–Nauenberg theorem, because the initial state is purely leptonic and all degenerate states that can appear at the given order are included in the final state. Given that $\gamma(\alpha_s) = 0$, the RGE prediction is simply given, as we have already seen, by $F(t, \alpha_s) = F[0, \alpha_s(t)]$. This means that, if we do, for example, a 2-loop calculation, we must obtain a result of the form

$$F(t, \alpha_s) = 1 + c_1\alpha_s(1 - b\alpha_s t) + c_2\alpha_s^2 + O(\alpha_s^3) \,. \tag{2.71}$$

In fact, taking into account the expression for the running coupling in (2.61), viz.,

$$\alpha_s(t) \sim \frac{\alpha_s}{1 + b\alpha_s t} \sim \alpha_s(1 - b\alpha_s t + \cdots) \,, \tag{2.72}$$

Eq. (2.71) can be rewritten as

$$F(t, \alpha_s) = 1 + c_1\alpha_s(t) + c_2\alpha_s^2(t) + O(\alpha_s^3(t)) = F[0, \alpha_s(t)] \,. \tag{2.73}$$

The content of the RGE prediction is, at this order, that there are no $\alpha_s t$ and $(\alpha_s t)^2$ terms (the leading log sequence must be absent), and the term of order $\alpha_s^2 t$ has the appropriate coefficient to be reabsorbed in the transformation of α_s into $\alpha_s(t)$.

At present the first four coefficients $c_1, \ldots, c_4$ have been computed in the $\overline{\text{MS}}$ scheme. The references are as follows: for c_2 [138], for c_3 [230], and for c_4 [74]. Clearly, $c_1 = 1/\pi$ does not depend on the definition of α_s, but the c_n with $n \geq 2$ do. The subleading coefficients also depend on the scale choice: if instead of expanding in $\alpha_s(Q)$, we decide to choose $\alpha_s(Q/2)$, the coefficients c_n $n \geq 2$ will change. In the $\overline{\text{MS}}$ scheme, for γ exchange and $n_f = 5$, which are good approximations for

$2m_b \ll Q \ll m_Z$, one has

$$F[0,\alpha_s(t)] = 1 + \frac{\alpha_s(t)}{\pi} + 1.409\ldots\left[\frac{\alpha_s(t)}{\pi}\right]^2 - 12.8\ldots\left[\frac{\alpha_s(t)}{\pi}\right]^3 - 80.0\ldots\left[\frac{\alpha_s(t)}{\pi}\right]^4 + \cdots . \tag{2.74}$$

Similar perturbative results at 3-loop accuracy also exist for

$$R_Z = \frac{\Gamma(Z \to \text{hadrons})}{\Gamma(Z \to \text{leptons})} , \qquad R_\tau = \frac{\Gamma(\tau \to \nu_\tau + \text{hadrons})}{\Gamma(\tau \to \nu_\tau + \text{leptons})} ,$$

and so on. We will discuss these results in Sect. 2.10, where we deal with measurements of α_s.

The perturbative expansion in powers of $\alpha_s(t)$ takes into account all contributions that are suppressed by powers of logarithms of the large scale Q^2 ("leading twist" terms). In addition, there are corrections suppressed by powers of the large scale Q^2 ("higher twist" terms). The pattern of power corrections is controlled by the light-cone operator product expansion (OPE) [112, 365], which leads (schematically) to

$$F = \text{pert.} + r_2\frac{m^2}{Q^2} + r_4\frac{\langle 0|\text{Tr}[\mathbf{F_{\mu\nu}F^{\mu\nu}}]|0\rangle}{Q^4} + \cdots + r_6\frac{\langle 0|O_6|0\rangle}{Q^6} + \cdots . \tag{2.75}$$

Here m^2 generically indicates mass corrections, for example from b quarks, beyond the b threshold, while top quark mass corrections only arise from loops, vanish in the limit $m_t \to \infty$, and are included in the coefficients like those in (2.74) and the analogous ones for higher twist terms; $\mathbf{F_{\mu\nu}} = \sum_A F^A_{\mu\nu} t^A$, O_6 is typically a 4-fermion operator, etc. For each possible gauge invariant operator, the corresponding negative power of Q^2 is fixed by dimensions.

We now consider the light-cone OPE in more detail. $R_{e^+e^-} \sim \Pi(Q^2)$, where $\Pi(Q^2)$ is the scalar spectral function related to the hadronic contribution to the imaginary part of the photon vacuum polarization $T_{\mu\nu}$:

$$T_{\mu\nu} = (-g_{\mu\nu}Q^2 + q_\mu q_\nu)\Pi(Q^2) = \int \mathrm{d}^4x \exp \mathrm{i}(q\cdot x)\langle 0|J^\dagger_\mu(x)J_\nu(0)|0\rangle$$
$$= \sum_n \langle 0|J^\dagger_\mu(0)|n\rangle\langle n|J_\nu(0)|0\rangle(2\pi)^4\delta^4(q-p_n) . \tag{2.76}$$

For $Q^2 \to \infty$, the $x^2 \to 0$ region is dominant. The light cone OPE is valid to all orders in perturbation theory. Schematically and dropping Lorentz indices for simplicity, near $x^2 \sim 0$, we have

$$J^\dagger(x)J(0) = I(x^2) + E(x^2)\sum_{n=0}^{\infty} c_n(x^2)x^{\mu_1}\ldots x^{\mu_n}\,O^n_{\mu_1\ldots\mu_n}(0) + \text{less sing. terms} . \tag{2.77}$$

Here $I(x^2)$, $E(x^2)$, ..., $c_n(x^2)$ are c-number singular functions and O^n is a string of local operators. $E(x^2)$ is the singularity of free field theory, while $I(x^2)$ and $c_n(x^2)$ in the interacting theory contain powers of $\log(\mu^2 x^2)$. Some O^n are already present in free field theory, while others appear when interactions are switched on. Given that $\Pi(Q^2)$ is related to the Fourier transform of the vacuum expectation value of the product of currents, less singular terms in x^2 lead to power-suppressed terms in $1/Q^2$. The perturbative terms, like those in (2.73), come from $I(x^2)$, which is the leading twist term, and the dominant logarithmic scaling violations induced by the running coupling are the logs in $I(x^2)$.

2.7.2 The Final State in e^+e^- Annihilation

Experiments on e^+e^- annihilation at high energy provide a remarkable opportunity for systematically testing the distinct signatures predicted by QCD for the structure of the final state averaged over a large number of events. Typical of asymptotic freedom is the hierarchy of configurations emerging as a consequence of the smallness of $\alpha_s(Q^2)$. When all corrections of order $\alpha_s(Q^2)$ are neglected, one recovers the naive parton model prediction for the final state: almost collinear events with two back-to-back jets with limited transverse momentum and an angular distribution $1 + \cos^2\theta$ with respect to the beam axis (typical of spin 1/2 parton quarks, while scalar quarks would lead to a $\sin^2\theta$ distribution). To order $\alpha_s(Q^2)$, a tail of events is predicted to appear with large transverse momentum $p_T \sim Q/2$ with respect to a suitably defined jet axis (for example, the thrust axis, see below). This small fraction of events with large p_T consists mainly of three-jet events with almost planar topology. The skeleton of a three-jet event, to leading order in $\alpha_s(Q^2)$, is formed by three hard partons $q\bar{q}g$, the third being a gluon emitted by a quark or antiquark line. To order $\alpha_s^2(Q^2)$, a hard perturbative non-planar component starts to build up, and a small fraction of four-jet events $q\bar{q}gg$ or $q\bar{q}q\bar{q}$ appear, and so on.

Event shape variables defined from the set of 4-momenta of final state particles are introduced to describe the topological structure of the final state energy flow in a quantitative manner [154]. The best known event shape variable is thrust (T) [192], defined as

$$T = \max \frac{\sum_i |\mathbf{p_i} \cdot \mathbf{n_T}|}{\sum_i |\mathbf{p_i}|} , \tag{2.78}$$

where the maximization is in terms of the axis defined by the unit vector n_T: the thrust axis is the axis that maximizes the sum of the absolute values of the longitudinal momenta of the final state particles. The thrust T varies between 1/2, for a spherical event, to 1 for a collinear (2-jet) event. Event shape variables are important for QCD tests and measurements of α_s, and also for more practical purposes, like a laboratory for assessing the reliability of event simulation programmes and a tool for the separation of signals and background.

A quantitatively specified definition of jets and of the number of jets in one event (jet counting) must be introduced for precise QCD tests and measurement of α_s, which must be infrared safe (i.e., not altered by soft particle emission or collinear splittings of massless particles) in order to be computable at the parton level and as insensitive as possible to the transformation of partons into hadrons (see, for example, [294]). For e^+e^- physics, one can use a jet algorithm based on a resolution parameter $y_{\rm cut}$ and a suitable pair variable. For example [172],

$$y_{ij} = \frac{2\min(E_i^2, E_j^2)(1-\cos\theta_{ij})}{s} \,. \tag{2.79}$$

Note that $1-\cos\theta_{ij} \sim \theta_{ij}^2/2$, so that the relative transverse momentum $k_{\rm T}^2$ is involved (hence, the name $k_{\rm T}$ algorithm). The particles i,j belong to different jets for $y_{ij} > y_{\rm cut}$. Clearly, the number of jets becomes a function of $y_{\rm cut}$, and in fact there are more jets for smaller $y_{\rm cut}$.

Recently, motivated by the LHC experiments, there has been a flurry of improved jet algorithm studies: it is essential that correct jet finding should be implemented by LHC experiments for optimal matching of theory and experiment [185, 317]. In particular, existing sequential recombination algorithms like $k_{\rm T}$ [132, 172] and Cambridge/Aachen [174] have been generalized. In these recursive definitions, one introduces distances d_{ij} between particles or clusters of particles i and j, and d_{iB} between i and the beam (B). The inclusive clustering proceeds by identifying the smallest of the distances and, if it is a d_{ij}, by recombining particles i and j, while if it is d_{iB}, calling i a jet and removing it from the list. The distances are recalculated and the procedure repeated until no i and j are left.

The extension relative to the $k_{\rm T}$ [132] and Cambridge/Aachen [174] algorithms lies in the definition of the distance measures:

$$d_{ij} = \min(k_{Ti}^{2p}, k_{Tj}^{2p})\frac{\Delta_{ij}^2}{R^2} \,, \tag{2.80}$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and k_{Ti}, y_i, and ϕ_i are the transverse momentum, rapidity, and azimuth of particle i, respectively. R is the radius of the jet, i.e., the radius of a cone which, by definition, contains the jet. The exponent p fixes the relative power of the energy versus geometrical (Δ_{ij}) scales.

For $p = 1$, one has the inclusive k_T algorithm. It can be shown in general that for $p \geq 0$ the behaviour of the jet algorithm with respect to soft radiation is rather similar to that observed for the k_T algorithm. The case $p = 0$ is special, and it corresponds to the inclusive Cambridge/Aachen algorithm [174]. Surprisingly (at first sight), taking p to be negative also yields an algorithm that is infrared and collinear safe and has sensible phenomenological behaviour. For $p = -1$, one obtains the recently introduced "anti-k_T" jet-clustering algorithm [126], which has particularly stable jet boundaries with respect to soft radiation and is suitable for practical use in experiments.

2.8 Deep Inelastic Scattering

Deep inelastic scattering (DIS) processes have played, and still play, a very important role in our understanding of QCD and of nucleon structure. This set of processes actually provides us with a rich laboratory for theory and experiment. There are several structure functions $F_i(x, Q^2)$ that can be studied, each a function of two variables. This is true separately for different beams and targets and different polarizations. Depending on the charges of ℓ and ℓ' [see (2.28)], we can have neutral currents (γ, Z) or charged currents in the ℓ–ℓ' channel (Fig. 2.10). In the past, DIS processes were crucial for establishing QCD as the theory of strong interactions and quarks and gluons as the QCD partons.

At present DIS remains very important for quantitative studies and tests of QCD. The theory of scaling violations for totally inclusive DIS structure functions, based on operator expansion or diagrammatic techniques and renormalization group methods, is crystal clear and the predicted Q^2 dependence can be tested at each value of x. The measurement of quark and gluon densities in the nucleon, as functions of x at some reference value of Q^2, which is an essential starting point for the calculation of all relevant hadronic hard processes, is performed in DIS processes. At the same time one measures $\alpha_s(Q^2)$, and the DIS values of the running coupling can be compared with those obtained from other processes. At all times new theoretical challenges arise from the study of DIS processes. Recent examples (see the following) are the so-called "spin crisis" in polarized DIS and the behaviour of singlet structure functions at small x, as revealed by HERA data. In the following we review the past successes and the present open problems in the physics of DIS.

The cross-section $\sigma \sim L^{\mu\nu} W_{\mu\nu}$ is given in terms of the product of a leptonic ($L^{\mu\nu}$) and a hadronic ($W_{\mu\nu}$) tensor. While $L^{\mu\nu}$ is simple and easily obtained from the lowest order electroweak (EW) vertex plus QED radiative corrections, the complicated strong interaction dynamics is contained in $W_{\mu\nu}$. The latter is proportional to the Fourier transform of the forward matrix element between the nucleon target states of the product of two EW currents:

$$W_{\mu\nu} = \int \mathrm{d}^4 y \, \exp \mathrm{i}(q \cdot y) \langle p | J_\mu^\dagger(y) J_\nu(0) | p \rangle \, . \tag{2.81}$$

Structure functions are defined starting from the general form of $W_{\mu\nu}$, given Lorentz invariance and current conservation. For example, for EW currents between unpolarized nucleons, we have

$$W_{\mu\nu} = \left(-g_{\mu\nu} + \frac{q_\mu q_\nu}{q^2} \right) W_1(\nu, Q^2) + \left(p_\mu - \frac{m\nu}{q^2} q_\mu \right) \left(p_\nu - \frac{m\nu}{q^2} q_\nu \right) \frac{W_2(\nu, Q^2)}{m^2}$$
$$- \frac{\mathrm{i}}{2m^2} \epsilon_{\mu\nu\lambda\rho} p^\lambda q^\rho W_3(\nu, Q^2) \, .$$

where variables are defined as in (2.28) and (2.29), and W_3 arises from VA interference and is absent for pure vector currents. In the limit $Q^2 \gg m^2$, with the Bjorken variable x fixed, the structure functions obey approximate Bjorken scaling, which is in fact broken by logarithmic corrections that can be computed in QCD:

$$mW_1(\nu, Q^2) \to F_1(x) \,, \qquad \nu W_{2,3}(\nu, Q^2) \to F_{2,3}(x) \,. \tag{2.82}$$

The γ–N cross-section is given by

$$\frac{\mathrm{d}\sigma^{\gamma}}{\mathrm{d}Q^2\mathrm{d}\nu} = \frac{4\pi\alpha^2 E'}{Q^4 E}\left(2\sin^2\frac{\theta}{2}W_1 + \cos^2\frac{\theta}{2}W_2\right) \,, \tag{2.83}$$

with $W_i = W_i(Q^2, \nu)$, while for the ν–N or $\bar{\nu}$–N cross-section one has

$$\frac{\mathrm{d}\sigma^{\nu,\bar{\nu}}}{\mathrm{d}Q^2\mathrm{d}\nu} = \frac{G_F^2 E'}{2\pi E}\left(\frac{m_W^2}{Q^2 + m_W^2}\right)^2\left(2\sin^2\frac{\theta}{2}W_1 + \cos^2\frac{\theta}{2}W_2 \pm \frac{E+E'}{m}\sin^2\frac{\theta}{2}W_3\right) \,, \tag{2.84}$$

with W_i for photons, and ν and $\bar{\nu}$ are all different, as we shall see in a moment.

In the scaling limit the longitudinal and transverse cross-sections are given by

$$\sigma_{\mathrm{L}} \sim \frac{1}{s}\left[\frac{F_2(x)}{2x} - F_1(x)\right] \,, \quad \sigma_{\mathrm{RH,LH}} \sim \frac{1}{s}\big[F_1(x) \pm F_3(x)\big] \,, \quad \sigma_{\mathrm{T}} = \sigma_{\mathrm{RH}} + \sigma_{\mathrm{LH}} \,, \tag{2.85}$$

where L, RH, LH refer to the helicity 0, 1, -1, respectively, of the exchanged gauge vector boson. For the photon case, $F_3 = 0$ and $\sigma_{\mathrm{RH}} = \sigma_{\mathrm{LH}}$.

In the 1960s the demise of hadrons from the status of fundamental particles to that of bound states of constituent quarks was the breakthrough that made possible the construction of a renormalizable field theory for strong interactions. The presence of an unlimited number of hadrons species, many of them with high spin values, presented an obvious dead-end for a manageable field theory. The evidence for constituent quarks emerged clearly from the systematics of hadron spectroscopy. The complications of the hadron spectrum could be explained in terms of the quantum numbers of spin 1/2, fractionally charged u, d, and s quarks. The notion of colour was introduced to reconcile the observed spectrum with Fermi statistics.

However, confinement, which forbids the observation of free quarks, was a clear obstacle towards the acceptance of quarks as real constituents and not just as fictitious entities describing some mathematical pattern (a doubt expressed even by Gell-Mann at the time). The early measurements of DIS at SLAC dissipated all doubts: the observation of Bjorken scaling and the success of Feynman's "naive" (not so much after all) parton model imposed quarks as the basic fields for describing the nucleon structure (parton quarks).

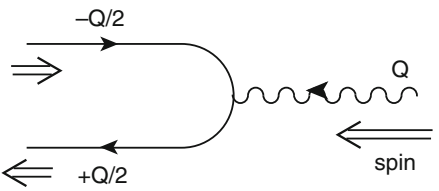

Fig. 2.14 Schematic diagram for the interaction of the virtual photon with a parton quark in the Breit frame

In the language of Bjorken and Feynman, the virtual γ (or, in general, any gauge boson) sees the quark partons inside the nucleon target as quasi-free, because their (Lorentz dilated) QCD interaction time is much longer than $\tau_\gamma \sim 1/Q$, the duration of the virtual photon interaction. Since the virtual photon 4-momentum is spacelike, we can go to a Lorentz frame where $E_\gamma = 0$ (Breit frame). In this frame $q = (E_\gamma = 0, 0, 0, Q)$ and the nucleon momentum, neglecting the mass $m \ll Q$, is $p = (Q/2x, 0, 0, -Q/2x)$. We note that this gives $q^2 = -Q^2$ and $x = Q^2/2(p \cdot q)$, as it should.

Consider the interaction of the photon with a quark (see Fig. 2.14) carrying a fraction y of the nucleon 4-momentum: $p_q = yp$ (we are neglecting the transverse components of p_q which are of order m). The incoming parton with $p_q = yp$ absorbs the photon and the final parton has 4-momentum p'_q. Since in the Breit frame the photon carries no energy, but only a longitudinal momentum Q, the photon can only be absorbed by those partons with $y = x$. Then the longitudinal component of $p_q = yp$ is $-yQ/2x = -Q/2$, and can be flipped into $+Q/2$ by the photon. As a result, the photon longitudinal momentum $+Q$ disappears, the parton quark momentum changes sign from $-Q/2$ to $+Q/2$ and the energy is not changed. So the structure functions are proportional to the density of partons with fraction x of the nucleon momentum, weighted by the squared charge.

Furthermore, recall that the helicity of a massless quark is conserved in a vector (or axial vector) interaction (see Sect. 1.5). So when the momentum is reversed, the spin must also flip. Since the process is collinear there is no orbital contribution, and only a photon with helicity ± 1 (transverse photon) can be absorbed. Alternatively, if partons were spin zero, only longitudinal photons would then contribute.

Using these results, which are maintained in QCD at leading order, the quantum numbers of the quarks were confirmed by early experiments. The observation that $R = \sigma_{\rm L}/\sigma_{\rm T} \to 0$ implies that the charged partons have spin 1/2. The quark charges were derived from the data on the electron and neutrino structure functions:

$$\begin{aligned} F_{ep} &= \frac{4}{9}u(x) + \frac{1}{9}d(x) + \cdots , \qquad & F_{en} &= \frac{4}{9}d(x) + \frac{1}{9}u(x) + \cdots , \\ F_{\nu p} &= F_{\bar{\nu} n} = 2d(x) + \cdots , \qquad & F_{\nu n} &= F_{\bar{\nu} p} = 2u(x) + \cdots , \end{aligned} \tag{2.86}$$

where $F \sim 2F_1 \sim F_2/x$ and $u(x)$, $d(x)$ are the parton number densities in the proton (with fraction x of the proton longitudinal momentum), which, in the scaling limit, do not depend on Q^2. The normalization of the structure functions and the parton densities are such that the charge relations hold:

$$\int_0^1 \big[u(x) - \bar{u}(x)\big]\mathrm{d}x = 2\ , \quad \int_0^1 \big[d(x) - \bar{d}(x)\big]\mathrm{d}x = 1\ , \quad \int_0^1 \big[s(x) - \bar{s}(x)\big]\mathrm{d}x = 0\ . \tag{2.87}$$

Furthermore, it was proven by experiment that, at values of Q^2 of a few GeV^2, in the scaling region, about half of the nucleon momentum, given by the momentum sum rule

$$\int_0^1 \Big[\sum_i \big[q_i(x) + \bar{q}_i(x)\big] + g(x)\Big]x\mathrm{d}x = 1\ , \tag{2.88}$$

is carried by neutral partons (gluons).

In QCD there are calculable log scaling violations induced by $\alpha_\mathrm{s}(t)$. The parton rules in (2.86) can be summarized in the schematic formula

$$F(x,t) = \int_x^1 \mathrm{d}y \frac{q_0(y)}{y} \sigma_\mathrm{point}\big(x/y, \alpha_\mathrm{s}(t)\big) + O(1/Q^2)\ . \tag{2.89}$$

Before QCD corrections $\sigma_\mathrm{point} = e^2\delta(x/y - 1)$ and $F = e^2 q_0(x)$ (here e denotes the charge of the quark in units of the positron charge, i.e., $e = 2/3$ for the u quark). QCD modifies σ_point at order α_s via the diagrams of Fig. 2.15. From a direct computation of the diagrams, one obtains a result of the following form:

$$\sigma_\mathrm{point}\big(z, \alpha_\mathrm{s}(t)\big) \simeq e^2\Big[\delta(z-1) + \frac{\alpha_\mathrm{s}}{2\pi}\big[tP(z) + f(z)\big]\Big]\ . \tag{2.90}$$

Note that the y integral in (2.89) is from x to 1, because the energy can only be lost by radiation before interacting with the photon (which eventually wants to find a fraction x, as we have explained). For $y > x$ the correction arises from diagrams with real gluon emission. Only the sum of the two real-gluon diagrams in Fig. 2.15 is gauge invariant, so the contribution of one given diagram will be gauge dependent. But in an axial gauge, which for this reason is sometimes also called the

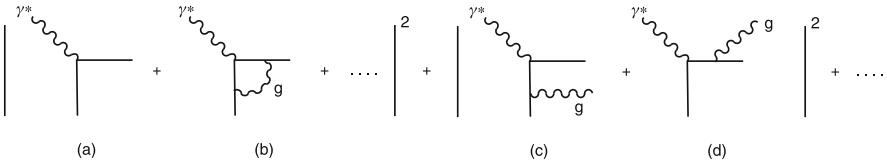

Fig. 2.15 First order QCD corrections to the virtual photon–quark cross-section. (**a**) Tree level, (**b**) vertex correction, (**c**) final-state radiation of one leg, (**d**) final-state radiation off the other leg

"physical gauge", the diagram of Fig. 2.15c, among real diagrams, gives the whole t-proportional term at $0 < x < 1$. It is obviously not essential to go to this gauge, but this diagram has a direct physical interpretation: a quark in the proton has a fraction $y > x$ of the parent 4-momentum; it then radiates a gluon and loses energy down to a fraction x, before interacting with the photon. The log arises from the virtual quark propagator, according to the discussion of collinear mass singularities in (2.27). In fact, in the massless limit, one has (k and h are the 4-momenta of the initial quark and the emitted gluon, respectively):

$$\begin{aligned} \text{propagator} &= \frac{1}{r^2} = \frac{1}{(k-h)^2} = \frac{-1}{2E_k E_h} \frac{1}{1-\cos\theta} \\ &= \frac{-1}{4E_k E_h} \frac{1}{\sin^2\theta/2} \propto \frac{-1}{p_{\mathrm{T}}^2} , \end{aligned} \tag{2.91}$$

where p_{T} is the transverse momentum of the virtual quark. So the square of the propagator goes like $1/p_{\mathrm{T}}^4$. But there is a p_{T}^2 factor in the numerator, because in the collinear limit, when $\theta = 0$ and the initial and final quarks and the emitted gluon are all aligned, the quark helicity cannot flip (vector interaction), so that the gluon should carry zero helicity, while a real gluon can only have ± 1 helicity. Thus the numerator vanishes as p_{T}^2 in the forward direction and the cross-section behaves as

$$\sigma \sim \int^{Q^2} \frac{1}{p_{\mathrm{T}}^2} \mathrm{d}p_{\mathrm{T}}^2 \sim \log Q^2 . \tag{2.92}$$

Actually, the log should be read as $\log Q^2/m^2$, because in the massless limit a genuine mass singularity appears. In fact, the mass singularity connected with the initial quark line is not cancelled, because we do not have the sum of all degenerate initial states [265], but only a single quark. But in correspondence with the initial quark, we have the (bare) quark density $q_0(y)$ which appears in the convolution integral. This is a non-perturbative quantity determined by the nucleon wave function. So we can factorize the mass singularity in a redefinition of the quark density: we replace $q_0(y) \to q(y,t) = q_0(y) + \Delta q(y,t)$ with

$$\Delta q(x,t) = \frac{\alpha_{\mathrm{s}}}{2\pi} t \int_x^1 \mathrm{d}y \frac{q_0(y)}{y} P(x/y) . \tag{2.93}$$

Here the factor of t is a bit symbolic: it stands for $\log Q^2/m^2$, but what exactly we put under Q^2 depends on the definition of the renormalized quark density, which also fixes the exact form of the finite term $f(z)$ in (2.90).

The effective parton density $q(y,t)$ that we have defined is now scale dependent. In terms of this scale dependent density, we have the following relations, where we have also replaced the fixed coupling with the running coupling according to the

prescription derived from the RGE:

$$F(x,t) = \int_x^1 \mathrm{d}y \frac{q(y,t)}{y} e^2 \left[\delta\left(\frac{x}{y}-1\right) + \frac{\alpha_\mathrm{s}(t)}{2\pi} f\left(\frac{x}{y}\right)\right] = e^2 q(x,t) + O\big(\alpha_\mathrm{s}(t)\big),$$
$$\frac{\mathrm{d}}{\mathrm{d}t} q(x,t) = \frac{\alpha_\mathrm{s}(t)}{2\pi} \int_x^1 \mathrm{d}y \frac{q(y,t)}{y} P\left(\frac{x}{y}\right) + O\big(\alpha_\mathrm{s}(t)^2\big). \tag{2.94}$$

We see that at lowest order we reproduce the naive parton model formulae for the structure functions in terms of effective parton densities that are scale dependent. The evolution equations for the parton densities are written down in terms of kernels (the "splitting functions" [40]), which can be expanded in powers of the running coupling. At leading order, we can interpret the evolution equation by saying that the variation of the quark density at x is given by the convolution of the quark density at y and the probability of emitting a gluon with fraction x/y of the quark momentum.

It is interesting that the integro-differential QCD evolution equation for densities can be transformed into an infinite set of ordinary differential equations for Mellin moments [234]. The Mellin moment f_n of a density $f(x)$ is defined by

$$f_n = \int_0^1 \mathrm{d}x\, x^{n-1} f(x)\ . \tag{2.95}$$

By taking moments of both sides of the second equation in (2.94), and changing the order of integration, one finds the simpler equation for the n th moment:

$$\frac{\mathrm{d}}{\mathrm{d}t} q_n(t) = \frac{\alpha_\mathrm{s}(t)}{2\pi} P_n q_n(t)\ . \tag{2.96}$$

To solve this equation we observe that it is equivalent to

$$\log \frac{q_n(t)}{q_n(0)} = \frac{P_n}{2\pi} \int_0^t \alpha_\mathrm{s}(t)\mathrm{d}t = \frac{P_n}{2\pi} \int_{\alpha_\mathrm{s}}^{\alpha_\mathrm{s}(t)} \frac{\mathrm{d}\alpha'}{-b\alpha'}\ . \tag{2.97}$$

To see the equivalence just take the t derivative of both sides. Here we used (2.46) to change the integration variable from $\mathrm{d}t$ to $\mathrm{d}\alpha(t)$ (denoted $\mathrm{d}\alpha'$) and

$$\beta(\alpha) \simeq -b\alpha^2 + \cdots\ .$$

Finally, the solution is

$$q_n(t) = \left[\frac{\alpha_\mathrm{s}}{\alpha_\mathrm{s}(t)}\right]^{P_n/2\pi b} q_n(0)\ . \tag{2.98}$$

The connection between these results and the RGE general formalism occurs via the light cone OPE [recall (2.81) for $W_{\mu\nu}$ and (2.77) for the OPE of two currents]. In the case of DIS, the c-number term $I(x^2)$ does not contribute, because we are interested in the connected part of the matrix element $\langle p|\ldots|p\rangle - \langle 0|\ldots|0\rangle$. The relevant terms are

$$J^\dagger(x)J(0) = E(x^2)\sum_{n=0}^{\infty} c_n(x^2)x^{\mu_1}\ldots x^{\mu_n}O^n_{\mu_1\ldots\mu_n}(0) + \text{less singular terms}\,. \tag{2.99}$$

A formally intricate but conceptually simple argument based on the analyticity properties of the forward virtual Compton amplitude shows that the Mellin moments M_n of structure functions are related to the individual terms in the OPE, in fact, precisely to the Fourier transform $c_n(Q^2)$, which we will write as $c_n(t,\alpha)$, of the coefficient $c_n(x^2)$ times a reduced matrix element h_n from the operators O^n: $\langle p|O^n_{\mu_1\ldots\mu_n}(0)|p\rangle = h_n p_{\mu_1}\ldots p_{\mu_n}$:

$$c_n\langle p|O^n|p\rangle \longrightarrow M_n = \int_0^1 \mathrm{d}x\, x^{n-1}F(x)\,. \tag{2.100}$$

Since the matrix element of the products of currents satisfy the RGE, so do the moments M_n. Hence, the general form of the Q^2 dependence is given by the RGE solution [see (2.47)]:

$$M_n(t,\alpha) = c_n[0,\alpha(t)]\exp\int_\alpha^{\alpha(t)}\frac{\gamma_n(\alpha')}{\beta(\alpha')}\mathrm{d}\alpha' h_n(\alpha)\,. \tag{2.101}$$

At lowest order, in the simplest case, identifying M_n with q_n, we have

$$\gamma_n(\alpha) = \frac{P_n}{2\pi}\alpha + \cdots\,, \quad \beta(\alpha) = -b\alpha^2 + \cdots\,, \tag{2.102}$$

and

$$q_n(t) = q_n(0)\exp\int_\alpha^{\alpha(t)}\frac{\gamma_n(\alpha')}{\beta(\alpha')}\mathrm{d}\alpha' = \left[\frac{\alpha_s}{\alpha_s(t)}\right]^{P_n/2\pi b} q_n(0)\,, \tag{2.103}$$

which exactly coincides with (2.98).

Up to this point we have implicitly restricted our attention to non-singlet (under the flavour group) structure functions. The Q^2 evolution equations become non-diagonal as soon as we take into account the presence of gluons in the target. In fact, the quark which is seen by the photon can be generated by a gluon in the target (Fig. 2.16). The quark evolution equation becomes:

$$\frac{\mathrm{d}}{\mathrm{d}t}q_i(x,t) = \frac{\alpha_s(t)}{2\pi}[q_i\otimes P_{qq}] + \frac{\alpha_s(t)}{2\pi}[g\otimes P_{qg}]\,, \tag{2.104}$$

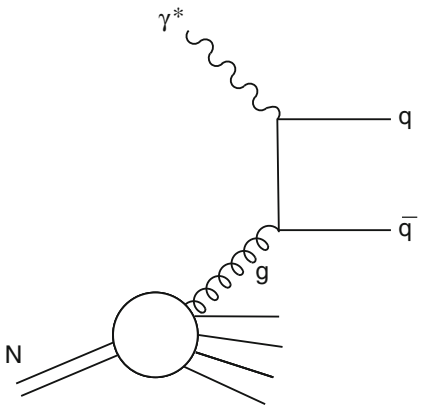

Fig. 2.16 Lowest order diagram for the interaction of the virtual photon with a parton gluon

where we have introduced the shorthand notation

$$[q \otimes P] = [P \otimes q] = \int_x^1 \mathrm{d}y \frac{q(y,t)}{y} P(x/y) \; . \tag{2.105}$$

It is easy to check that the convolution defined in this way is commutative, like an ordinary product. At leading order, the interpretation of (2.104) is simply that the variation of the quark density is due to the convolution of the quark density at a higher energy times the probability of finding a quark in a quark (with the right energy fraction) plus the gluon density at a higher energy times the probability of finding a quark (of the given flavour i) in a gluon. The evolution equation for the gluon density, needed to close the system,[2] can be obtained by suitably extending the same line of reasoning to a gedanken probe sensitive to colour charges, for example,

[2]The evolution equations are now often called the DGLAP equations (after Dokshitzer, Gribov, Lipatov, Altarelli, and Parisi). The first article by Gribov and Lipatov was published in 1972 [233] (even before the works by Gross and Wilczek and by Politzer!), and was followed in 1974 by a paper by Lipatov [283] (these dates correspond to the publication in Russian). All these articles refer to an Abelian vector theory (treated in parallel with a pseudoscalar theory). Seen from the point of view of the evolution equations, these papers, in the context of the Abelian theory, ask the right question and extract the relevant logarithmic terms from the dominant class of diagrams. But from their formal presentation, the relation to real physics is somewhat hidden (in this respect the 1974 paper by Lipatov makes some progress and explicitly refers to the parton model). The article by Dokshitser [171] was exactly contemporary to that by Altarelli and Parisi [40]. It now refers to the non-Abelian theory (with running coupling), and the discussion is more complete and explicit than in the Gribov–Lipatov articles. But, for example, the connection to the parton model, the notion of the evolution as a branching process, and the independence of the kernels from the process are not emphasized.

a virtual gluon. The resulting equation is of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}g(x,t) = \frac{\alpha_\mathrm{s}(t)}{2\pi}\left[\sum_i (q_i + \bar{q}_i) \otimes P_{gq}\right] + \frac{\alpha_\mathrm{s}(t)}{2\pi}[g \otimes P_{\mathrm{gg}}] \,. \tag{2.106}$$

The explicit form of the splitting functions in lowest order [40, 171, 233] can be directly derived from the QCD vertices [40]. They are a property of the theory and do not depend on the particular process the parton density is taking part in. The results are as follows:

$$\begin{aligned}
P_{qq} &= \frac{4}{3}\left[\frac{1+x^2}{(1-x)_+} + \frac{3}{2}\delta(1-x)\right] + O(\alpha_\mathrm{s}),\\
P_{gq} &= \frac{4}{3}\frac{1+(1-x)^2}{x} + O(\alpha_\mathrm{s}),\\
P_{qg} &= \frac{1}{2}\left[x^2 + (1-x)^2\right] + O(\alpha_\mathrm{s}),\\
P_{\mathrm{gg}} &= 6\left[\frac{x}{(1-x)_+} + \frac{1-x}{x} + x(1-x)\right] + \frac{33-2n_\mathrm{f}}{6}\delta(1-x) + O(\alpha_\mathrm{s}).
\end{aligned} \tag{2.107}$$

For a generic non-singular weight function $f(x)$, the "+" distribution is defined as

$$\int_0^1 \frac{f(x)}{(1-x)_+}\mathrm{d}x = \int_0^1 \frac{f(x)-f(1)}{1-x}\mathrm{d}x \,. \tag{2.108}$$

The $\delta(1-x)$ terms arise from the virtual corrections to the lowest order tree diagrams. Their coefficient can be simply obtained by imposing the validity of charge and momentum sum rules. In fact, from the request that the charge sum rules in (2.87) are not affected by the Q^2 dependence, one derives

$$\int_0^1 P_{qq}(x)\mathrm{d}x = 0 \,, \tag{2.109}$$

which can be used to fix the coefficient of the $\delta(1-x)$ terms of P_{qq}. Similarly, by taking the t derivative of the momentum sum rule in (2.88) and requiring it to vanish for generic q_i and g, one obtains

$$\int_0^1 \left[P_{qq}(x) + P_{gq}(x)\right]x\mathrm{d}x = 0 \,, \quad \int_0^1 \left[2n_\mathrm{f}P_{qg}(x) + P_{\mathrm{gg}}(x)\right]x\mathrm{d}x = 0 \,. \tag{2.110}$$

At higher orders, the evolution equations are easily generalized, but the calculation of the splitting functions rapidly becomes very complicated. For many years the splitting functions were only completely known at NLO accuracy [198], that is, $\alpha_s P \sim \alpha_s P_1 + \alpha_s^2 P_2 + \cdots$. But in recent years, the NNLO results P_3 were first derived in analytic form for the first few moments, and then the full NNLO analytic calculation, a really monumental work, was completed in 2004 by Moch et al. [292].

Beyond leading order, a precise definition of parton densities should be specified. One can take a physical definition: for example, quark densities can be defined so as to keep the LO expression for the structure function F_2 valid at all orders, the so-called DIS definition [42], and the gluon density can be defined starting from F_L, the longitudinal structure function. Alternatively, one can adopt a more abstract specification, for example, in terms of the $\overline{\text{MS}}$ prescription. Once the definition of parton densities is fixed, the coefficients that relate the different structure functions to the parton densities at each fixed order can be computed. Similarly, the higher order splitting functions also depend, to some extent, on the definition of parton densities, and a consistent set of coefficients and splitting functions must be used at each order.

The scaling violations are clearly observed by experiment (Fig. 2.17), and their pattern is well reproduced by QCD fits at NLO (Figs. 2.18 and 2.19) [349]. These fits provide an impressive confirmation of a quantitative QCD prediction, a measurement of $q_i(x, Q_0^2)$ and $g(x, Q_0^2)$, at some reference value Q_0^2 of Q^2, and a precise measurement of $\alpha_s(Q^2)$.

2.8.1 The Longitudinal Structure Function

After SLAC established the dominance of the transverse cross-section it took about 40 years to get meaningful data on the longitudinal structure function F_L [see (2.85)]! These data are an experimental highlight of recent years. They were obtained by H1 at HERA [237]. The data are shown in Fig. 2.20. For spin 1/2 charged partons, F_L vanishes asymptotically. In QCD F_L starts at order $\alpha_s(Q^2)$. At LO the simple 30-year-old formula is valid (for $N_f = 4$) [39]:

$$F_L(x, Q^2) = \frac{\alpha_s(Q^2)}{2\pi} x^2 \int_x^1 \frac{dy}{y^3} \left[\frac{8}{3} F_2(y, Q^2) + \frac{40}{9} y g(y, Q^2) \left(1 - \frac{x}{y}\right) \right] . \tag{2.111}$$

The $O(\alpha_s^2)$ [372] and $O(\alpha_s^3)$ [293] corrections are now also known. One would not have expected it to take such a long time to have a meaningful test of this simple prediction! And in fact better data would be highly desirable. But how and when they will be obtained is at present not clear at all.

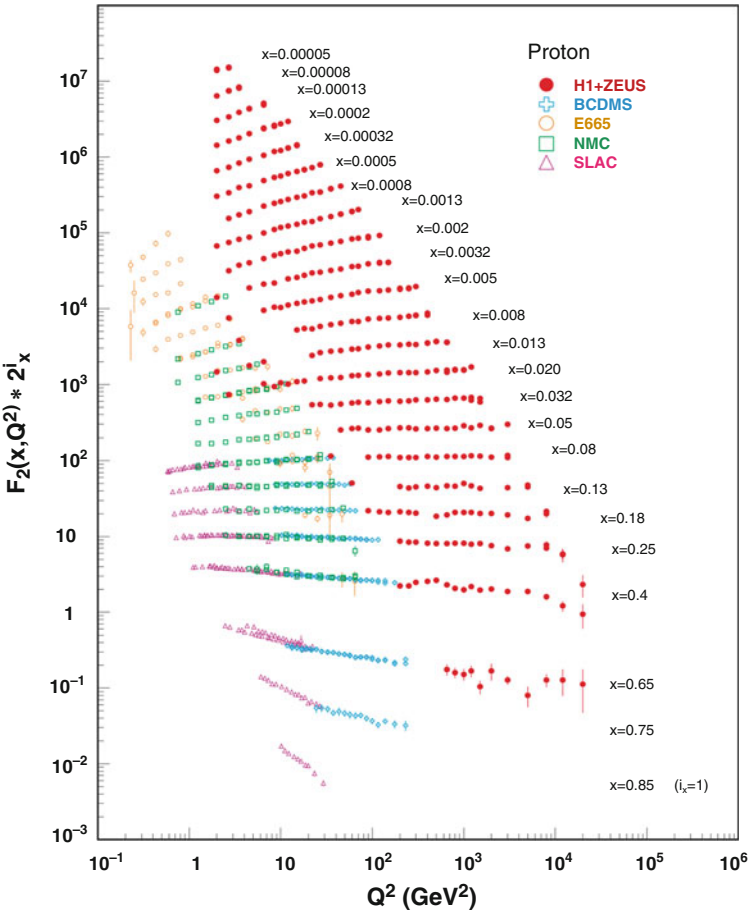

Fig. 2.17 A representative selection of data on the proton electromagnetic structure function F_2^p, from collider (HERA) and fixed target experiments [307], clearly showing the pattern of scaling violations. Figure reproduced with permission. Copyright (c) 2012 by *American Physical Society*

2.8.2 *Large and Small x Resummations for Structure Functions*

At values of x either near 0 or near 1 (with Q^2 large), those terms of higher order in α_s, in both the coefficients or the splitting functions, which are multiplied by powers of $\log 1/x$ or $\log(1-x)$ eventually become important and should be taken into account. Fortunately, the sequences of leading and subleading logs can be evaluated at all orders by special techniques, and resummed to all orders.

For $x \sim 1$ resummation [329], I refer to the recent papers [202, 211] (the latter also involving higher twist corrections, which are important at large x), where a

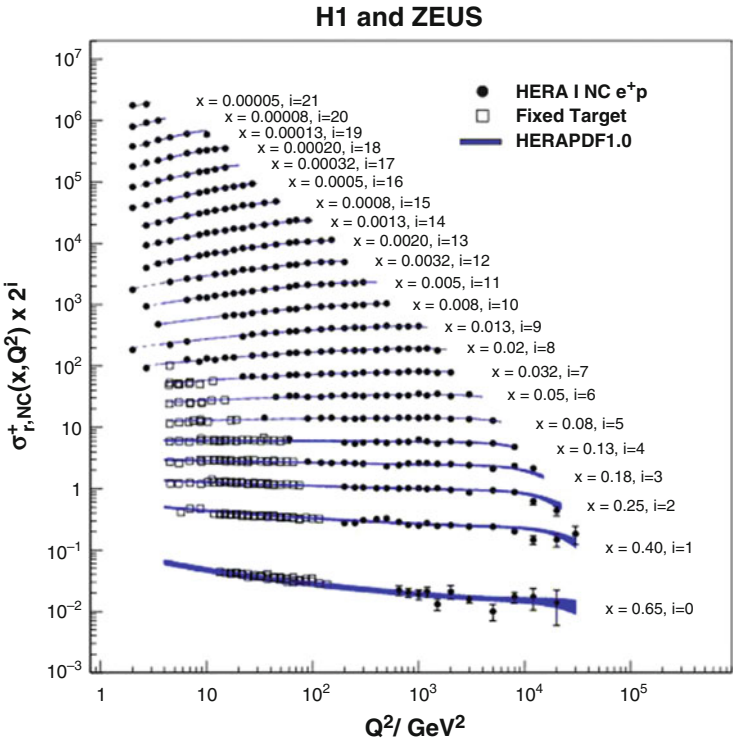

Fig. 2.18 NLO QCD fit to the combined HERA data with $Q^2 \geq 3.5\,\mathrm{GeV}^2$: $\chi^2/\mathrm{dof} = 574/582$ [349]

list of references to previous work can be found. More important is the small x resummation because, the singlet structure functions are large in this domain of x (while all structure functions vanish near $x = 1$). Here we briefly summarize the small-x case for the singlet structure function, which is the dominant channel at HERA, dominated by the sharp rise of the gluon and sea parton densities at small x.

The small x data collected by HERA can be fitted reasonably well, even at the smallest measured values of x, by the NLO QCD evolution equations, so that there is no dramatic evidence in the data for departures. This is surprising also in view of the fact that the NNLO effects in the evolution have recently become available and are quite large [292]. Resummation effects have been shown to resolve this apparent paradox. For the singlet splitting function, the coefficients of all LO and NLO corrections of order $[\alpha_s(Q^2)\log 1/x]^n$ and $\alpha_s(Q^2)[\alpha_s(Q^2)\log 1/x]^n$, respectively, are explicitly known from the Balitski, Fadin, Kuraev, Lipatov (BFKL) analysis of virtual gluon–virtual gluon scattering [191, 284]. But the simple addition of these higher order terms to the perturbative result (with subtraction of all double counting) does not lead to a converging expansion (the NLO logs completely override the LO logs in the relevant domain of x and Q^2).

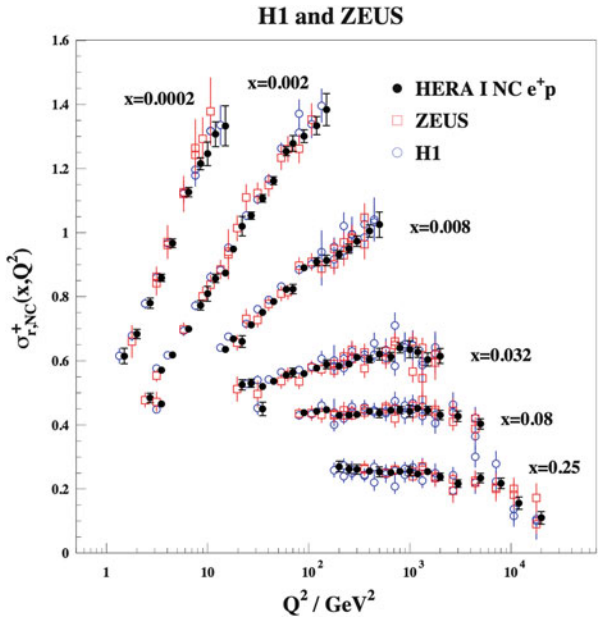

Fig. 2.19 More detailed view of the NLO QCD fit to a selection of the HERA data [349]

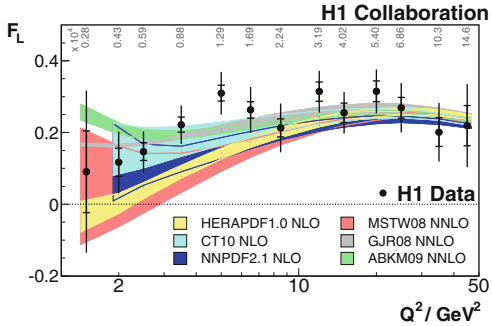

Fig. 2.20 Longitudinal structure function F_L measured by H1 at HERA, as a function of Q^2 for different values of x. The theoretical curves are obtained from different sets of parton densities as indicated

A sensible expansion is only obtained by a proper treatment of momentum conservation constraints, also using the underlying symmetry of the BFKL kernel under exchange of the two external gluons, and especially, of the running coupling effects (see the analysis in [49, 141] and references therein). In Fig. 2.21, we present the results for the dominant singlet splitting function $xP_{gg}(x, \alpha_s(Q^2))$ for $\alpha_s(Q^2) \sim 0.2$. We see that, while the NNLO perturbative splitting function deviates sharply from the NLO approximation at small x, the resummed result only shows a moderate dip with respect to the NLO perturbative splitting function in the region

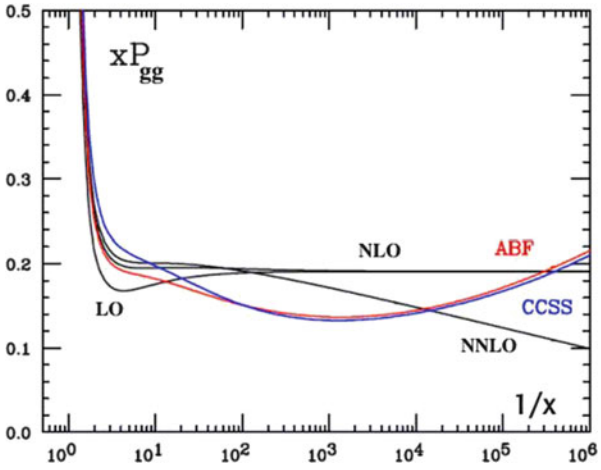

Fig. 2.21 Dominant singlet splitting function $xP_{gg}(x, \alpha_s(Q^2))$ for $\alpha_s(Q^2) \sim 0.2$. The resummed results from [49] (labeled ABF) and from [141] (CCSS), which are in good mutual agreement, are compared with the LO, NLO, and NNLO perturbative results (adapted from [49] and [141])

of HERA data, and the full effect of the true small x asymptotics is only felt at much smaller values of x. The related effects are not very important for most processes at the LHC, but could become relevant for the next generation of hadron colliders.

2.8.3 Polarized Deep Inelastic Scattering

Polarized DIS is a subject where our knowledge is still far from satisfactory, in spite of a great experimental effort (for recent reviews, see, for example, [24]). One major question is how the proton helicity is distributed among quarks, gluons, and orbital angular momentum:

$$\frac{1}{2}\Delta\Sigma + \Delta g + L_z = \frac{1}{2}\,. \tag{2.112}$$

Experiments with polarized leptons on polarized nucleons are sensitive to the polarized parton densities $\Delta q = q_+ - q_-$, the difference of quark densities with helicity plus and minus, in a proton with helicity plus. These differences are related to the quark matrix elements of the axial current. The polarized densities satisfy evolution equations analogous to (2.104) and (2.106), but with modified splitting functions that were derived in [40] (the corresponding anomalous dimensions were obtained in [22]).

Measurements have shown that the quark moment $\Delta\Sigma$ is small. This is the "spin crisis" started by [65]: values from recent fits [104, 159, 244, 277, 303, 326] lie in

the range $\Delta\Sigma \sim 0.2$–0.3. In any case, it is a less pronounced crisis than it used to be in the past. From the spin sum rule, one finds that either $\Delta g + L_z$ is relatively large or there are contributions to $\Delta\Sigma$ at very small x, outside of the measured region. Denoting the first moment of the net helicity carried by the sum $q + \bar{q}$ by Δq, we have the relations [104, 159]

$$a_3 = \Delta u - \Delta d = (F + D)(1 + \epsilon_2) = 1.269 \pm 0.003 \,, \tag{2.113}$$

$$a_8 = \Delta u + \Delta d - 2\Delta s = (3F - D)(1 + \epsilon_3) = 0.586 \pm 0.031 \,, \tag{2.114}$$

where the F and D couplings are defined in the $SU(3)$ flavour symmetry limit, and ϵ_2 and ϵ_3 describe the $SU(2)$ and $SU(3)$ breakings, respectively. From the measured first moment of the structure function g_1, one obtains the value of $a_0 = \Delta\Sigma$:

$$\Gamma_1 = \int \mathrm{d}x\, g_1(x) = \frac{1}{12}\left[a_3 + \frac{1}{3}(a_8 + 4a_0)\right] \,, \tag{2.115}$$

with the result, at $Q^2 \sim 4\,\mathrm{GeV}^2$,

$$a_0 = \Delta\Sigma = \Delta u + \Delta d + \Delta s = a_8 + 3\Delta s \sim 0.25 \,. \tag{2.116}$$

In turn, in the $SU(3)$ limit $\epsilon_2 = \epsilon_3 = 0$, one then obtains

$$\Delta u \sim 0.82 \,, \qquad \Delta d \sim -0.45 \,, \qquad \Delta s \sim -0.11 \,. \tag{2.117}$$

This is an important result! Given F, D, and Γ_1, we know Δu, Δd, Δs, and $\Delta\Sigma$ in the $SU(3)$ limit, which should be reasonably accurate. The x distribution of g_1 is known down to $x \sim 10^{-4}$ on proton and deuterium, and the first moment of g_1 does not seem to get much from the unmeasured range at small x (also, theoretically, g_1 should be smooth at small x [190]).

The value of $\Delta s \sim -0.11$ from totally inclusive data and $SU(3)$ appears to be at variance with the value extracted from single-particle inclusive DIS (SIDIS), where one obtains a nearly vanishing result for Δs in a fit to all data [159, 326] that leads to puzzling results. There is, in fact, an apparent tension between the first moments as determined by using the approximate $SU(3)$ symmetry and from fitting the data on SIDIS ($x \geq 0.001$) (in particular for the strange density). But the adequacy of the SIDIS data is questionable (in particular the kaon data which fix Δs) and so is their theoretical treatment (for example, the application of parton results at too low an energy and the ambiguities in the kaon fragmentation function).

$\Delta\Sigma$ is conserved in perturbation theory at LO (i.e., it does not evolve with Q^2). Regarding conserved quantities, we would expect them to be the same for constituent and for parton quarks. But actually, the conservation of $\Delta\Sigma$ is broken by the axial anomaly and, in fact, in perturbation theory beyond LO, the conserved density is actually $\Delta\Sigma' = \Delta\Sigma + \Delta g(n_\mathrm{f}/2\pi\alpha_\mathrm{s})$ [41]. Note also that $\alpha_\mathrm{s}\Delta g$ is conserved in LO, that is $\Delta g \sim \log Q^2$. This behaviour is not controversial, but it will

be a long time before the log growth of Δg is confirmed by experiment! However, by establishing this behaviour, one would show that the extraction of Δg from the data is correct and that the QCD evolution works as expected.

If Δg were large enough, it could account for the difference between partons ($\Delta\Sigma$) and constituents ($\Delta\Sigma'$). From the spin sum rule it is clear that the log increase should cancel between Δg and L_z. This cancelation is automatic, as a consequence of helicity conservation in the basic QCD vertices. Δg can be measured indirectly by scaling violations and directly from asymmetries, e.g., in SIDIS. Existing measurements by HERMES, COMPASS, and at RHIC are still crude, but show no hint of a large Δg at accessible values of x and Q^2. Present data, affected by large errors (see, in particular, [303] for a discussion of this point) are consistent [104, 159, 244, 277, 303, 326] with a sizable contribution of Δg to the spin sum rule in (2.112), but there is no indication that $\alpha_s \Delta g$ effects can explain the difference between constituents and parton quarks.

2.9 Hadron Collider Processes and Factorization

There are three classes of hard processes: those with no hadronic particles in the initial state, like e^+e^- annihilation, those initiated by a lepton and a hadron, like DIS, and those with two incoming hadrons. The parton densities, defined and measured in DIS, are instrumental to compute hard processes initiated by collisions of two hadrons, like $p\bar{p}$ (Tevatron) or pp (LHC). Suppose we have a hadronic process of the form $h_1 + h_2 \to X + \text{all}$, where h_i are hadrons and X is some triggering particle, or pair of particles, or one or more jets which specify the large scale Q^2 relevant for the process, in general somewhat, but not much, smaller than s, the total centre-of-mass squared mass. For example, X can be a $W^\pm$, or a Z, or a virtual photon with large Q^2 (Drell–Yan processes), or a jet with large transverse momentum p_T, or a quark–antiquark pair with heavy components (of mass M). By "all" we mean a totally inclusive collection of hadronic particles. The factorization theorem (FT) states that, for the total cross-section or some other sufficiently inclusive distribution, we can write, apart from power-suppressed corrections, the expression (see also Fig. 2.22)

$$\sigma(s,\tau) = \sum_{AB} \int \mathrm{d}x_1 \mathrm{d}x_2 p_{1A}(x_1, Q^2) p_{2B}(x_2, Q^2) \sigma_{AB}(x_1 x_2 s, \tau) \,. \tag{2.118}$$

Here $\tau = Q^2/s$ is a scaling variable, p_{iA} are the densities for a parton of type A inside the hadron h_i, and σ_{AB} is the partonic cross-section for

$$\text{parton } A + \text{parton } B \to X + \text{ all}' \,.$$

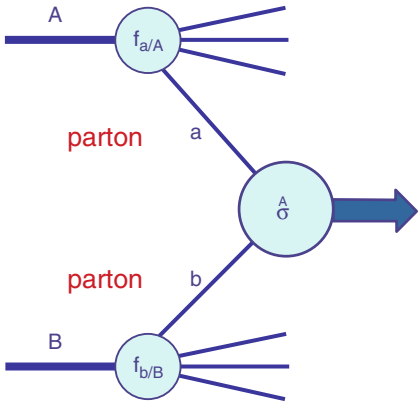

Fig. 2.22 Diagram for the factorization theorem

Here all′ is the partonic version of "all", i.e., a totally inclusive collection of quarks, antiquarks, and gluons. This result is based on the fact that the mass singularities associated with the initial legs are of universal nature, so that one can reproduce the same modified parton densities by absorbing these singularities into the bare parton densities, as in DIS. Once the parton densities and α_s are known from other measurements, the prediction of the rate for a given hard process is obtained without much ambiguity (e.g., from scale dependence or hadronization effects).

At least an NLO calculation of the reduced partonic cross-section σ_{AB} is needed in order to correctly specify the scale, and in general also the definition of the parton densities and of the running coupling in the leading term. The residual scale and scheme dependence is often the most important source of theoretical error. It is important to ask to what extent the FT has been proven? In perturbation theory up to NNLO, it has been explicitly checked to hold for many processes: if corrections exist we already know that they must be small (we stress that we are only considering totally inclusive processes). At all orders, the most in-depth discussions have been carried out in [146], in particular for Drell–Yan processes. The LHC experiments offer a wonderful opportunity for testing the FT by comparing precise theoretical predictions with accurate data on a wide variety of processes (for a recent review, see, for example, [119]).

A great effort has been and is being devoted to the theoretical preparation and interpretation of the LHC experiments. For this purpose very, difficult calculations are needed at NLO and beyond because the strong coupling, even at the large Q^2 values involved, is not that small. Further powerful techniques for amplitude calculations have been devised.

An interesting development at the interface between string theory and QCD is twistor calculus. A precursor was the Parke–Taylor result in 1986 [305] on the amplitudes for n incoming gluons with given $\pm$ helicities [91]. Inspired by dual models, they derived a compact formula for the maximum non-vanishing helicity-violating amplitude (with $n-2$ plus and 2 minus helicities) in terms of spinor products. In 2003, using the relation between strings and gauge theories in

twistor space, Witten developed [368] a formalism in terms of effective vertices and propagators that allows one to compute all helicity amplitudes. The method, alternative to other modern techniques for the evaluation of Feynman diagrams [163], leads to very compact results.

Since then, there has been rapid progress (for reviews, see [128]). The method was extended to include massless external fermions [217] and also external EW vector bosons [96] and Higgs particles [167]. The level attained is already important for multijet events at the LHC. The study of loop diagrams came next. The basic idea is that loops can be fully reconstructed from their unitarity cuts. First proposed by Bern et al. [95], the technique was revived by Britto et al. [114] and then perfected by Ossola et al. [304] and further extended to massive particles in [186]. For a recent review of these new methods see [188].

In parallel with this, activity on event simulation has received a big boost from preparations at the LHC (see, for example, the review [130]). Powerful techniques have been developed to generate numerical results at NLO for processes with complicated final states: matrix element calculation has been matched with modeling of parton showers in packages like Black Hat [92] (on-shell methods for loops), used in association with Sherpa [227] (for real emission), or POWHEG BOX [299] or aMC@NLO [203], the automated version of the general framework MC@NLO [206]. In a complete simulation, the matrix element calculation, improved by resummation of large logs, provides the hard skeleton (with large p_{T} branchings), while the parton shower is constructed by a sequence of factorized collinear emissions fixed by the QCD splitting functions. In addition, at low scales, a model of hadronization completes the simulation. The importance of all the components, matrix element, parton shower, and hadronization, can be appreciated in simulations of hard events compared with Tevatron and LHC data. One can say that the computation of NLO corrections in perturbative QCD has now been completely automated.

A partial list of examples of recent NLO calculations in pp collisions, obtained with these techniques is: $W + 3$ jets [187], Z, $\gamma^* + 3$ jets [93], W, $Z + 4$ jets [94], $W + 5$ jets [97], $t\bar{t}b\bar{b}$ [113], $t\bar{t} + 2$ jets [100], $t\bar{t}\,W$ [129], $WW + 2$ jets [289], $WWb\bar{b}$ [161], $b\bar{b}b\bar{b}$ [232], etc. In the following we shall detail a number of the most important and simplest examples, without any pretension to completeness.

2.9.1 Vector Boson Production

Drell–Yan processes which include lepton pair production via virtual γ, W, or Z exchange, offer a particularly good opportunity to test QCD. This process, among those quadratic in parton densities with a totally inclusive final state, is perhaps the simplest one from a theoretical point of view. The large scale is specified and measured by the invariant mass squared Q^2 of the lepton pair, which is not itself strongly interacting (so there are no dangerous hadronization effects). The improved QCD parton model leads directly to a prediction for the total rate as a function of

s and $\tau = Q^2/s$. The value of the LO cross-section is inversely proportional to the number of colours N_C, because a quark of given colour can only annihilate with an antiquark of the same colour to produce a colourless lepton pair. The order $\alpha_s(Q^2)$ NLO corrections to the total rate were computed long ago [42, 273] and found to be particularly large, when the quark densities are defined from the structure function F_2 measured in DIS at $q^2 = -Q^2$. The ratio $\sigma_{corr}/\sigma_{LO}$ of the corrected and the Born cross-sections was called the K-factor [28], because it is almost a constant in rapidity. More recently, the NNLO full calculation of the K-factor was completed in a truly remarkable calculation [240].

Over the years the QCD predictions for W and Z production, a better testing ground than the older fixed-target Drell–Yan experiments, have been compared with experiments at CERN $Sp\bar{p}S$ and Tevatron energies and now at the LHC. $Q \sim m_{W,Z}$ is large enough to make the prediction reliable (with a not too large K-factor) and the ratio $\sqrt{\tau} = Q/\sqrt{s}$ is not too small. Recall that, in lowest order, one has $x_1 x_2 s = Q^2$, so that the parton densities are probed at x values around $\sqrt{\tau}$. We have $\sqrt{\tau} = 0.13$–0.15 (for W and Z production, respectively) at $\sqrt{s} = 630\,\text{GeV}$ (CERN $Sp\bar{p}S$ collider) and $\sqrt{\tau} = 0.04$–0.05 at the Tevatron. At the LHC at 8 TeV or at 14 TeV, one has $\sqrt{\tau} \sim 10^{-2}$ or $\sim 6 \times 10^{-3}$, respectively (for both W and Z production). A comparison of the experimental total rates for W and Z with the QCD predictions at hadron colliders [327] is shown in Fig. 2.23. It is also important to mention that the cross-sections for di-boson production (i.e., WW, WZ, ZZ, $W\gamma$, $Z\gamma$) have been measured at the Tevatron and the LHC and are in fair agreement with the SM prediction (see, for example, the summary in [285] and references therein). The typical precision is comparable to or better than the size of NLO corrections.

The calculation of the W/Z p_T distribution is a classic challenge in QCD. For large p_T, for example $p_T \sim O(m_W)$, the p_T distribution can be reliably computed in perturbation theory, and this was done up to NLO in the late 1970s and early 1980s [183]. A problem arises in the intermediate range $\Lambda_{QCD} \ll p_T \ll m_W$, where the bulk of the data is concentrated, because terms of order $\alpha_s(p_T^2) \log m_W^2/p_T^2$ become

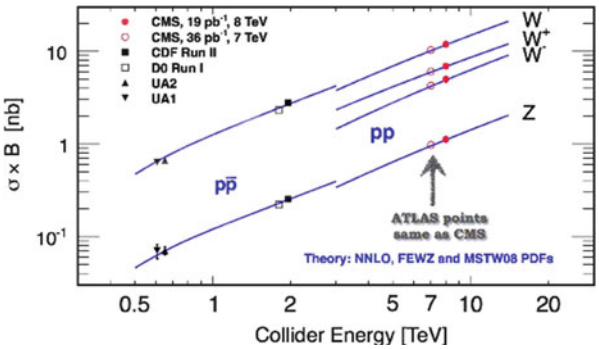

Fig. 2.23 Data vs. theory for W and Z production at hadron colliders [327] (included with permission)

of order 1 and should be included to all orders [330]. At order α_s, we have

$$\frac{1}{\sigma_0}\frac{d\sigma_0}{dp_T^2} = (1+A)\delta(p_T^2) + \frac{B}{p_T^2}\log\frac{m_W^2}{(p_T^2)_+} + \frac{C}{(p_T^2)_+} + D(p_T^2)\ , \tag{2.119}$$

where A, B, C, D are coefficients of order α_s. The "+" distribution is defined in complete analogy with (2.108):

$$\int_0^{p_{T\,MAX}^2} g(z)f(z)_+ dz = \int_0^{p_{T\,MAX}^2} \big[g(z)-g(0)\big]f(z)dz\ . \tag{2.120}$$

The content of this, at first sight mysterious, definition is that the singular "+" terms do not contribute to the total cross-section. In fact, for the cross-section, the weight function is $g(z) = 1$ and we obtain

$$\sigma = \sigma_0\left[(1+A) + \int_0^{p_{T\,MAX}^2} D(z)dz\right]\ . \tag{2.121}$$

The singular terms, of infrared origin, are present at the not completely inclusive level, but disappear in the total cross-section. Solid arguments have been given [330] to suggest that these singularities exponentiate. Explicit calculations in low order support the exponentiation, and this leads to the following expression:

$$\frac{1}{\sigma_0}\frac{d\sigma_0}{dp_T^2} = \int \frac{d^2b}{4\pi}\exp(-ib\cdot p_T)(1+A)\exp S(b)\ , \tag{2.122}$$

with

$$S(b) = \int_0^{p_{T\,MAX}} \frac{d^2k_T}{2\pi}\big[\exp(ik_T\cdot b) - 1\big]\left(\frac{B}{k_T^2}\log\frac{m_W^2}{k_T^2} + \frac{C}{k_T^2}\right)\ . \tag{2.123}$$

At large p_T the perturbative expansion is recovered. At intermediate p_T the infrared p_T singularities are resummed (the Sudakov log terms, which are typical of vector gluons, are related to the fact that for a charged particle in acceleration, it is impossible not to radiate, so that the amplitude for no soft gluon emission is exponentially suppressed). A delicate procedure for matching perturbative and resummed terms is needed [43]. However, this formula has problems at small p_T, for example, because of the presence of α_s under the integral for $S(b)$. Presumably, the relevant scale is of order k_T^2. So it must be completed by some non-perturbative ansatz or an extrapolation into the soft region [330].

All the formalism has been extended to NLO accuracy [64], where one starts from the perturbative expansion at order α_s^2, and generalises the resummation to include also NLO terms of order $\alpha_s(p_T^2)^2\log m_W^2/p_T^2$. The comparison with the data is very impressive. Figure 2.24 shows the p_T distribution as predicted in QCD (with

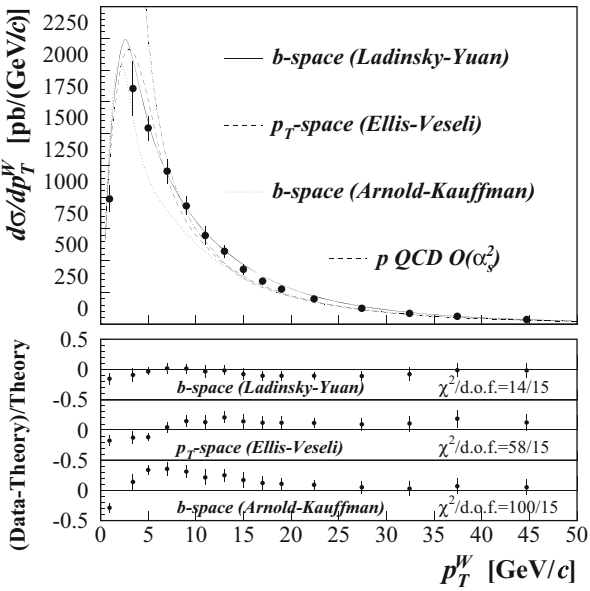

Fig. 2.24 QCD predictions for the W p_T distribution compared with recent D0 data at the Tevatron ($\sqrt{s} = 1.8$ TeV) (adapted from [64, 347])

a number of variants that differ mainly in the approach to the soft region) compared with some recent data at the Tevatron [347]. The W and Z p_T distributions have also been measured at the LHC and are in fair agreement with the theoretical expectation [343].

The rapidity distributions of the produced W and Z have also been measured with fair accuracy at the Tevatron and at the LHC, and predicted at NLO [55]. A representative example of great significance is provided by the combined LHC results for the W charge asymmetry, defined as $A \sim (W^+ - W^-)/(W^+ + W^-)$, as a function of the pseudo-rapidity η [340]. These data combine the ATLAS and CMS results at smaller values of η with those of the LHCb experiments at larger η (in the forward direction). This is very important input for the disentangling of the different quark parton densities.

2.9.2 *Jets at Large Transverse Momentum*

Another simple and important process at hadron colliders is the inclusive production of jets at high energy $\sqrt{s}$ and transverse momentum p_T. A comparison of the data with the QCD NLO predictions [147, 180] in pp or $p\bar{p}$ collisions is shown in Fig. 2.25 [369]. This is a particularly significant test because the rates at different centre-of-mass energies and, for each energy, at different values of p_T, span many orders of magnitude. This steep behaviour is determined by the sharp drop in the parton densities with increasing x. Moreover, the corresponding values of $\sqrt{s}$ and

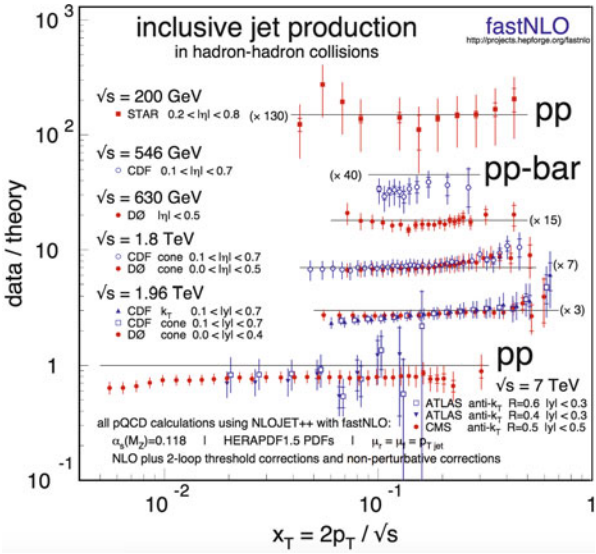

Fig. 2.25 Jet production cross-section at pp or $p\bar{p}$ colliders, as a function of p_T [369]. Theoretical predictions are from NLO perturbative calculations with state-of-the-art parton densities with the corresponding value of α_s plus a non-perturbative correction factor due to hadronization and the underlying event, obtained using Monte Carlo event generators (included with permission)

p_T are large enough to be well inside the perturbative region. The overall agreement of the data from ISR, UA1,2, STAR (at RHIC), CDF/D0, and now ATLAS/CMS, is indeed spectacular. In fact, the uncertainties in the resulting experiment/theory ratio, due to systematics and to ambiguities in the parton densities, the value of α_s, the scale choice, and so on, which can reach a factor of 2–3, are much smaller than the spread of the cross-section values over many orders of magnitude.

Similar results also hold for the production of photons at large p_T. The ATLAS data [342], shown in Fig. 2.26, are in fair agreement with the theoretical predictions. For the same process, a less clear situation was found with fixed target data. Here, first of all, the experimental results show some internal discrepancies. Moreover, the accessible values of p_T being smaller, the theoretical uncertainties are greater.

2.9.3 Heavy Quark Production

We now discuss heavy quark production at colliders. The totally inclusive cross-sections have been known at NLO for a long time [300]. The resummation of leading and next-to-leading logarithmically enhanced effects in the vicinity of the threshold region have also been studied [108]. The bottom production at the Tevatron has represented a problem for some time: the total rate and the

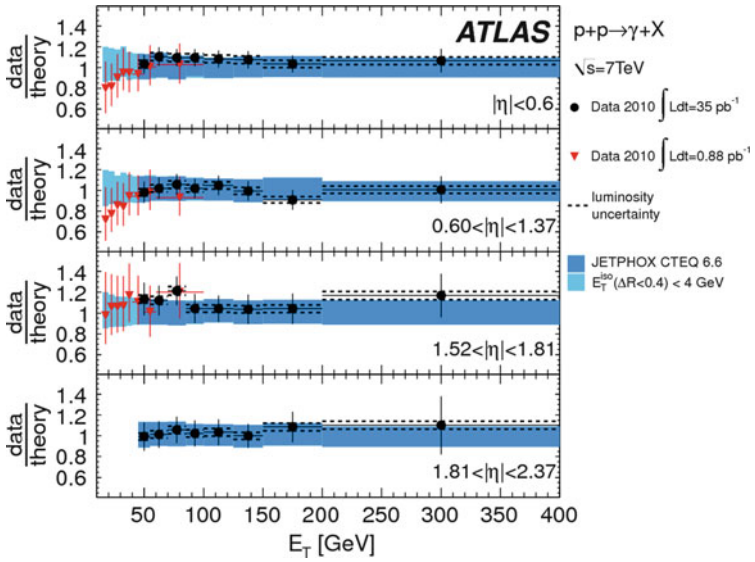

Fig. 2.26 Single-photon production in $p\bar{p}$ colliders as a function of p_T [342] (included with permission)

p_T distribution of b quarks observed at CDF and D0 appeared in excess of the prediction, up to the highest measured values of p_T [83, 124]. But this is a complicated problem, with different scales present at the same time: $\sqrt{s}, p_T, m_b$. The discrepancy was finally explained by more carefully taking into account a number of small effects from resummation of large logarithms, the difference between b hadrons and b partons, the inclusion of better fragmentation functions, etc. [125]. At present the LHC data on b production are in satisfactory agreement with the theoretical predictions (Fig. 2.27 [67]).

The top quark is really special: its mass is of the order of the Higgs VEV or its Yukawa coupling is of order 1, and in this sense, it is the only "normal" case among all quarks and charged leptons. Due to its heavy mass, it decays so fast that it has no time to be bound in a hadron: thus it can be studied as a quark. It is very important to determine its mass and couplings for different precision predictions of the SM. The top quark may be particularly sensitive to new heavy states or have a connection to the Higgs sector when we go beyond the SM theories.

Top quark physics has thus attracted much attention, both from the experimental side, at hadron colliders, and from the theoretical point of view. In particular, the top–antitop inclusive cross-section has been measured in $p\bar{p}$ collisions at the Tevatron [15], and now in pp collisions at the LHC [339, 346]. The QCD prediction is at present completely known at NNLO [150]. Soft gluon resummation has also been performed at NNLL [127]. The agreement between theory and experiment is good for the best available parton density functions together with the values of α_s

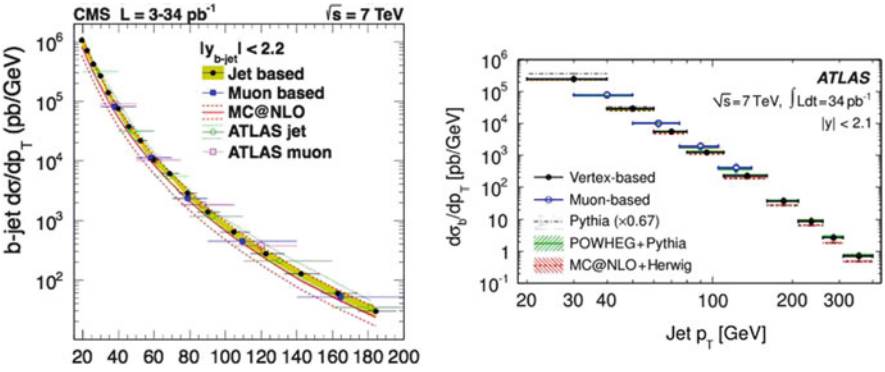

Fig. 2.27 The b production p_{T} distribution at the LHC [67]

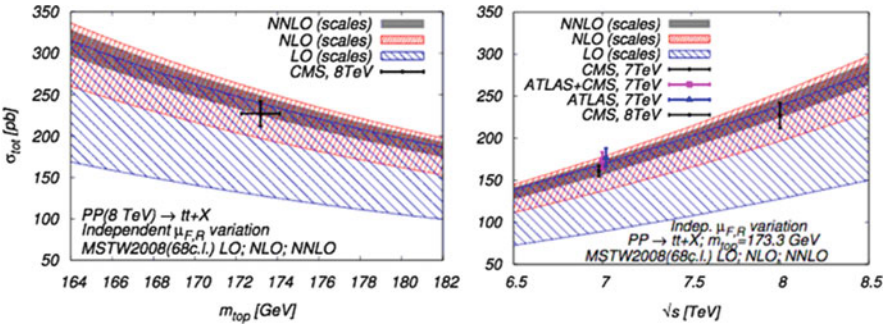

Fig. 2.28 The $t\bar{t}$ production cross-section at the LHC collider. Scale dependence of the total cross-section at LO (*blue*), NLO (*red*), and NNLO (*black*) as a function of m_{top} (*left*) or $\sqrt{s}$ (*right*) at the LHC 8 TeV [150] (included with permission)

and m_t measured separately (the top mass is measured from the invariant mass of the decay products), as can be seen from Fig. 2.28 [150].

The mass of the top (and the value of α_{s}) can be determined from the cross-section, assuming that QCD is correct, and compared with the more precise value from the final decay state. The value of the top pole mass derived in [27] from the cross-section data, using the best available parton densities with the correlated value of α_{s}, is $m_t^{\mathrm{pole}} = 173.3 \pm 2.8\,\mathrm{GeV}$. This is to be compared with the value measured at the Tevatron by the CDF and D0 collaborations, viz., $m_t^{\mathrm{exp}} = 173.2 \pm 0.9\,\mathrm{GeV}$. This quoted error is clearly too optimistic, especially if one identifies this value with the pole mass which it resembles. This error is only adequate within the specific procedure used by the experimental collaborations to define their mass (including Montecarlo, with assumptions about higher order terms, non-perturbative effects, etc.). The problem is how to export this value to other processes. Leaving aside the thorny issue of the precise relation between m_t^{exp} with m_t^{pole}, it is clear that there is good overall consistency.

The inclusive forward–backward asymmetry, $A_{\rm FB}$, in the $t\bar{t}$ rest frame has been measured by both the CDF [6] and D0 [9] collaborations, and found to be in excess of the SM prediction, by about 2σ [247]. For CDF the discrepancy increases at large $t\bar{t}$ invariant mass, and reaches about 2.5σ for $M_{t\bar{t}} \geq 450\,\text{GeV}$. Recently, CDF has obtained [7] the first measurement of the top quark pair production differential cross-section as a function of $\cos\theta$, with θ the production angle of the top quark. The coefficient of the $\cos\theta$ term in the differential cross-section, viz., $a_1 = 0.40 \pm 0.12$, is found to be in excess of the NLO SM prediction, viz., $0.15^{+0.07}_{-0.03}$, while all other terms are in good agreement with the NLO SM prediction, and $A_{\rm FB}$ is dominated by this excess linear term. Is this a real discrepancy? The evidence is far from compelling, but this effect has received much attention from theorists [321]. A related observable at the LHC is the charge asymmetry $A_{\rm C}$ in $t\bar{t}$ production. In contrast to $A_{\rm FB}$, the combined value of $A_{\rm C}$ reported by ATLAS [1] and CMS [144] agrees with the SM, within the still limited accuracy of the data.

2.9.4 Higgs Boson Production

We now turn to the discussion of the SM Higgs inclusive production cross-section (for a review and a list of references, see [165]). The most important Higgs production modes are gluon fusion, vector boson fusion, Higgs strahlung, and associated production with top quark pairs. Some typical Feynman diagrams for these different modes are depicted in Fig. 2.29. The predicted rates are shown in Fig. 2.30 [168].

The most important channel at the LHC is Higgs production via $g + g \to H$. The amplitude is dominated by the top quark loop [216]. The NLO corrections turn out to be particularly large [156], as can be seen in Fig. 2.31. Higher order corrections can be computed either in the effective Lagrangian approach, where the heavy top is integrated away and the loop is shrunk down to a point [182] (the coefficient of the effective vertex is known to $\alpha_{\rm s}^4$ accuracy [139]), or in the full theory. At the NLO, the two approaches agree very well for the rate as a function of $m_{\rm H}$ [270]. The NNLO corrections have been computed in the effective vertex approximation [133] (see Fig. 2.31). Beyond fixed order, resummation of large logs has been carried out [134]. Further, the NLO EW contributions have been computed [20]. Rapidity (at NNLO) [56] and $p_{\rm T}$ distributions (at NLO) [158] have also been evaluated. At smaller $p_{\rm T}$, the large logarithms $[\text{Log}(p_{\rm T}/m_{\rm H})]^n$ have been resummed in analogy with what was done long ago for W and Z production [110]. For additional recent works on Higgs physics at colliders, see, for example, [184].

So far we have seen examples of resummation of large logs. This is a very important chapter of modern QCD. The resummation of soft gluon logs enter into different problems, and the related theory is subtle. The reader is referred here to some recent papers where additional references can be found [77]. A particularly interesting related development has to do with the so-called non-global logs (see, for example, [153]). If in the measurement of an observable some experimental cuts

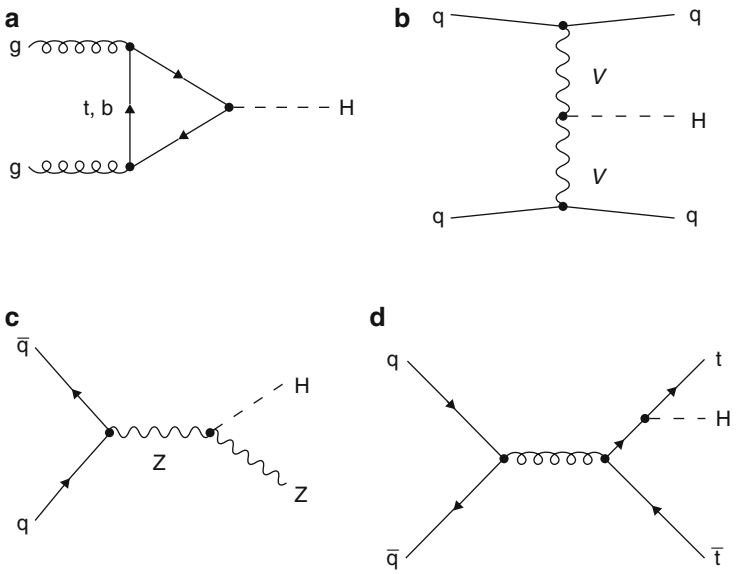

Fig. 2.29 Representative Feynman diagrams for the Higgs production cross-section mechanisms. (**a**) Gluon fusion. (**b**) Vector boson fusion ($V = W, Z$). (**c**) Higgs strahlung from a Z boson (an analogous diagram can be drawn for the W boson). (**d**) $t\bar{t}$ associated production

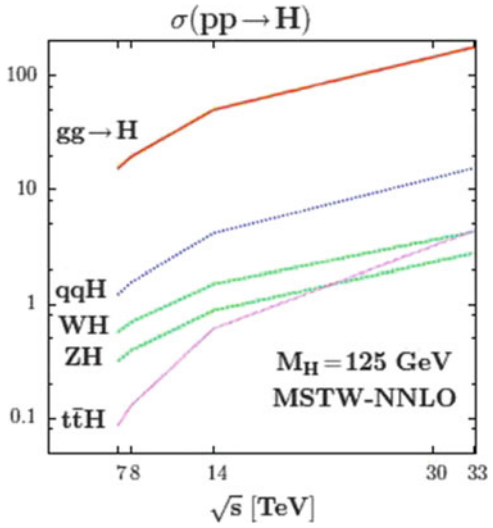

Fig. 2.30 Production cross-sections at the LHC for a Higgs with mass $M_{\mathrm{H}} \sim 125\,\mathrm{GeV}$ for different centre-of-mass energies [168]

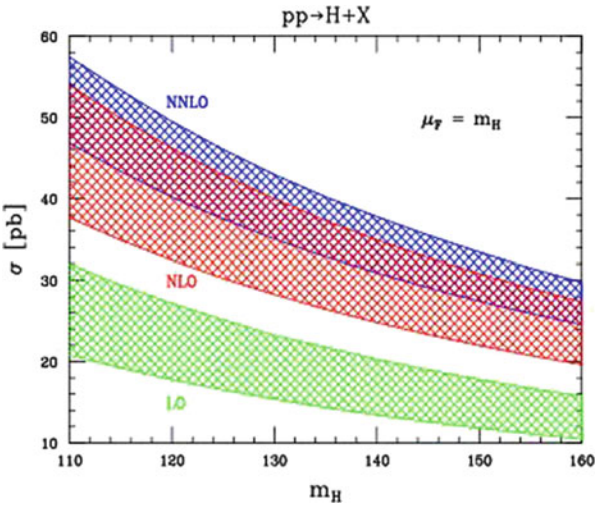

Fig. 2.31 Higgs gluon fusion cross-section in LO, NLO, and NLLO [57]. Figure reproduced with permission. Copyright (c) 2005 by *American Physical Society*

are introduced, which is very often the case, then a number of large logs can arise from the corresponding breaking of inclusiveness. It is also important to mention the development of software for the automated implementation of resummation (see, for example, [78]).

2.10 Measurements of α_s

Very precise and reliable measurements of $\alpha_s(m_Z)$ are obtained from e^+e^- colliders (in particular LEP), from deep inelastic scattering, and from the hadron colliders (Tevatron and LHC). The "official" compilation due to Bethke [99, 311], included in the 2012 edition of the PDG [307], is reproduced here in Fig. 2.32. The agreement among so many different ways of measuring α_s is a strong quantitative test of QCD. However, for some entries the stated error is taken directly from the original works and is not transparent enough when viewed from the outside (e.g., the lattice determination). In my opinion one should select a few of the theoretically cleanest processes for measuring α_s and consider all other ways as tests of the theory. Note that, in QED, α is measured from a single very precise and theoretically clean observable (one possible calibration process is at present the electron $g-2$ [242]). The cleanest processes for measuring α_s are the totally inclusive ones (no hadronic corrections) with light cone dominance, like Z decay, scaling violations in DIS, and perhaps τ decay (but for τ the energy scale is dangerously low). We will review these cleanest methods for measuring α_s in the following.

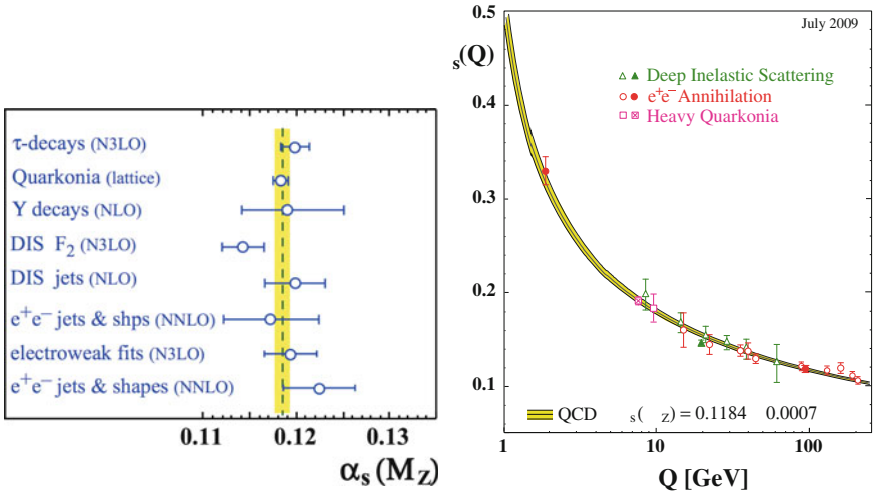

Fig. 2.32 *Left*: Summary of measurements of $\alpha_s(m_Z)$. The *yellow band* is the proposed average: $\alpha_s(m_Z) = 0.1184 \pm 0.0007$. *Right*: Summary of measurements of α_s as a function of the respective energy scale Q. Figures from [99]

2.10.1 α_s from e^+e^- Colliders

The totally inclusive processes for measuring α_s at e^+e^- colliders are hadronic Z decays (R_l, σ_h, σ_l, Γ_Z) and hadronic τ decays. As we have seen in Sect. 2.7.1, for a quantity like R_l we can write a general expression of the form

$$R_l = \frac{\Gamma(Z, \tau \to \text{hadrons})}{\Gamma(Z, \tau \to \text{leptons})} \sim R^{\text{EW}}(1 + \delta_{\text{QCD}} + \delta_{\text{NP}}) \,, \tag{2.124}$$

where R^{EW} is the electroweak-corrected Born approximation, and δ_{QCD}, δ_{NP} are the perturbative (logarithmic) and non-perturbative (power suppressed) QCD corrections. For a measurement of α_s (in the following we always refer to the $\overline{\text{MS}}$ definition of α_s) at the Z resonance peak, one can use all the information from R_l, $\Gamma_Z = 3\Gamma_l + \Gamma_h + \Gamma_{\text{inv}}$, and $\sigma_F = 12\pi\Gamma_l\Gamma_F/(m_Z^2\Gamma_Z^2)$, where F stands for h or l.

In the past, the measurement from R_l was preferred (taken by itself it leads to $\alpha_s(m_Z) = 0.1226 \pm 0.0038$, a bit on the large side), but after LEP there is no reason for this preference. In all these quantities α_s enters through Γ_h, but the measurements of, say, Γ_Z, R_l, and σ_l are really independent, as they are affected by entirely different systematics: Γ_Z is extracted from the line shape, and R_l and σ_l are measured at the peak, but R_l does not depend on the absolute luminosity, while σ_l does. The most sensitive single quantity is σ_l. It gives $\alpha_s(m_Z) = 0.1183 \pm 0.0030$. The combined value from the measurements at the Z (assuming the validity of the SM and the

observed Higgs mass) is [268]

$$\alpha_s(m_Z) = 0.1187 \pm 0.0027 \,. \tag{2.125}$$

Similarly, by adding all other electroweak precision tests (in particular m_W), one finds [350]

$$\alpha_s(m_Z) = 0.1186 \pm 0.0026 \,. \tag{2.126}$$

These results have been obtained from the $\delta_{\rm QCD}$ expansion up to and including the c_3 term of order α_s^3. But by now the c_4 term (NNNLO!) has also been computed [74] for inclusive hadronic Z and τ decay. For $n_f = 5$ and $a_s = \alpha_s(m_Z)/\pi$, this remarkable calculation of about 20,000 diagrams for the inclusive hadronic Z width leads to the result

$$\delta_{\rm QCD} = 1 + a_s + 0.76264 a_s^2 - 15.49 a_s^3 - 68.2 a_s^4 + \cdots \,. \tag{2.127}$$

This result can be used to improve the value of $\alpha_s(m_Z)$ from the EW fit given in (2.126), which becomes

$$\alpha_s(m_Z) = 0.1190 \pm 0.0026 \,. \tag{2.128}$$

Note that the error shown is dominated by the experimental errors. Ambiguities from higher perturbative orders [328], from power corrections, and also from uncertainties on the Bhabha luminometer (which affect $\sigma_{\rm h,l}$) [157] are very small. In particular, the fact of having now fixed m_H does not decrease the error significantly [73] (Grunewald, M., for the LEP EW Group, private communication). The main source of error is the assumption of no new physics, for example, in the $Zb\bar{b}$ vertex, which may affect the Γ_h prediction.

We now consider the measurement of $\alpha_s(m_Z)$ from τ decay. R_τ has a number of advantages which, at least in part, tend to compensate for the smallness of $m_\tau = 1.777\,\text{GeV}$. First, R_τ is maximally inclusive, more so than $R_{e^+e^-}(s)$, because one also integrates over all values of the invariant hadronic squared mass:

$$R_\tau = \frac{1}{\pi} \int_0^{m_\tau^2} \frac{{\rm d}s}{m_\tau^2} \left(1 - \frac{s}{m_\tau^2}\right)^2 {\rm Im}\, \Pi_\tau(s) \,. \tag{2.129}$$

As we have seen, the perturbative contribution is now known at NNNLO [74]. Analyticity can be used to transform the integral into one on the circle at $|s| = m_\tau^2$:

$$R_\tau = \frac{1}{2\pi {\rm i}} \oint_{|s|=m_\tau^2} \frac{{\rm d}s}{m_\tau^2} \left(1 - \frac{s}{m_\tau^2}\right)^2 \Pi_\tau(s) \,. \tag{2.130}$$

Furthermore, the factor $(1 - s/m_\tau^2)^2$ is important to kill the sensitivity in the region $\mathrm{Re}[s] = m_\tau^2$ where the physical cut and the associated thresholds are located. However, the sensitivity to hadronic effects in the vicinity of the cut is still a non-negligible source of theoretical error which the formulation of duality violation models tries to decrease. But the main feature that has attracted attention to τ decays for the measurement of $\alpha_s(m_Z)$ is that even a rough determination of $\Lambda_{\rm QCD}$ at a low scale $Q \sim m_\tau$ leads to a very precise prediction of α_s at the scale m_Z, just because in $\log Q/\Lambda_{\rm QCD}$ the value of $\Lambda_{\rm QCD}$ counts less and less as Q increases. The absolute error in α_s shrinks by a factor of about one order of magnitude in going from $\alpha_s(m_\tau)$ to $\alpha_s(m_Z)$.

Still it seems a little suspicious that, in order to obtain a better measurement of $\alpha_s(m_Z)$, we have to go down to lower and lower energy scales. And in fact, in general, one finds that the decreased control of higher order perturbative and non-perturbative corrections makes the apparent advantage totally illusory. For α_s from R_τ, the quoted amazing precision is obtained by taking for granted that corrections suppressed by $1/m_\tau^2$ are negligible. The argument is that, in the massless theory, the light cone expansion is given by

$$\delta_{\rm NP} = \frac{\rm ZERO}{m_\tau^2} + c_4 \frac{\langle O_4 \rangle}{m_\tau^4} + c_6 \frac{\langle O_6 \rangle}{m_\tau^6} + \cdots . \tag{2.131}$$

In fact there are no 2D Lorentz and gauge invariant operators. For example, $\mathrm{Tr}[\mathbf{g}_\mu \mathbf{g}^\mu]$ [recall (1.12)] is not gauge invariant. In the massive theory, ZERO here is replaced by the light quark mass-squared m^2. This is still negligible if m is taken as a Lagrangian mass of a few MeV. If on the other hand the mass were taken to be the constituent mass of order $\Lambda_{\rm QCD}$, this term would not be negligible at all, and would substantially affect the result [note that $\alpha_s(m_\tau)/\pi \sim 0.1 \sim (0.6\,\mathrm{GeV}/m_\tau)^2$ and that $\Lambda_{\rm QCD}$ for three flavours is large]. The principle that coefficients in the operator expansion can be computed from the perturbative theory in terms of parton masses has never really been tested (due to ambiguities in the determination of condensates) and this particular case with a ZERO there is unique in making the issue crucial. Many distinguished theorists believe the optimistic version. I am not convinced that the gap is not filled up by ambiguities in $O(\Lambda_{\rm QCD}^2/m_\tau^2)$ from $\delta_{\rm pert}$ [45].

There is a vast and sophisticated literature on α_s from τ decay. Unbelievably small errors are obtained in one or the other of several different procedures and assumptions that have been adopted to end up with a specified result. With time there has been an increasing awareness of the problem of controlling higher orders and non-perturbative effects. In particular, fixed order perturbation theory (FOPT) has been compared with resummation of leading beta function effects in the so-called contour-improved perturbation theory (CIPT). The results are sizeably different in the two cases, and there have been many arguments in the literature about which method is best.

One important piece of progress comes from the experimental measurement of moments of the τ decay mass distributions, defined by modifying the weight function in the integral in (2.129). In principle, one can measure α_s from the

sum rules obtained from different weight functions that emphasize different mass intervals and different operator dimensions in the light cone operator expansion. A thorough study of the dependence of the measured value of α_s on the choice of the weight function, and in general of higher order and non-perturbative corrections, has appeared in [89], and the interested reader is advised to look at that paper and the references therein.

We consider here the recent evaluations of α_s from τ decay based on the NNNLO perturbative calculations [74] and different procedures for estimating the different kinds of corrections. From the papers given in [90], we obtain an average value and error that agrees with the Erler and Langacker's values as given in PDG 12 [307]:

$$\alpha_s(m_\tau) = 0.3285 \pm 0.018 \,, \tag{2.132}$$

or

$$\alpha_s(m_Z) = 0.1194 \pm 0.0021 \,. \tag{2.133}$$

In any case, one can discuss the error, but what is true and remarkable is that the central value of α_s from τ decay, obtained at very small Q^2, is in good agreement with all other precise determinations of α_s at more typical LEP values of Q^2.

2.10.2 α_s from Deep Inelastic Scattering

In principle, DIS is expected to be an ideal laboratory for the determination of α_s, but in practice the outcome is still to some extent unsatisfactory. QCD predicts the Q^2 dependence of $F(x, Q^2)$ at each fixed x, not the x shape. But the Q^2 dependence is related to the x shape by the QCD evolution equations. For each x bin, the data can be used to extract the slope of an approximately straight line in $\mathrm{d}\log F(x, Q^2)/\mathrm{d}\log Q^2$, i.e., the log slope. The Q^2 span and the precision of the data are not very sensitive to the curvature, for most x values. A single value of Λ_{QCD} must be fitted to reproduce the collection of the log slopes. For the determination of α_s, the scaling violations of non-singlet structure functions would be ideal, because of the minimal impact of the choice of input parton densities. We can write the non-singlet evolution equations in the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\log F(x,t) = \frac{\alpha_s(t)}{2\pi}\int_x^1 \frac{\mathrm{d}y}{y}\frac{F(y,t)}{F(x,t)}P_{qq}\left(\frac{x}{y}, \alpha_s(t)\right) \,, \tag{2.134}$$

where P_{qq} is the splitting function. At present, NLO and NNLO corrections are known. It is clear from this form that, for example, the normalization error on the input density drops out, and the dependence on the input is reduced to a minimum (indeed, only a single density appears here, while in general there are quark and gluon densities).

Unfortunately, the data on non-singlet structure functions are not very accurate. If we take the difference $F_{\rm p} - F_p$ in the data on protons and neutrons, experimental errors add up and become large in the end. The $F_{3\nu{\rm N}}$ data are directly non-singlet, but are not very precise. Another possibility is to neglect sea and glue in F_2 at sufficiently large x. But by only taking data at $x > x_0$, one decreases the sample and introduces a dependence on x_0 and an error from residual singlet terms. A recent fit to non singlet structure functions in electron or muon production extracted from proton and deuterium data, neglecting sea and gluons at $x > 0.3$ (error to be evaluated), has led to the results [105]:

$$\alpha_{\rm s}(m_Z) = 0.1148 \pm 0.0019({\rm exp}) + ? \quad ({\rm NLO}) \,, \tag{2.135}$$

$$\alpha_{\rm s}(m_Z) = 0.1134 \pm 0.0020({\rm exp}) + ? \quad ({\rm NNLO}) \,. \tag{2.136}$$

The central values are rather low and there is not much difference between NLO and NNLO. The question marks refer to the uncertainties from the residual singlet component at $x > 0.3$, and also to the fact that the old BCDMS data, whose systematics has been questioned, are very important at $x > 0.3$ and push the fit towards small values of $\alpha_{\rm s}$.

When one measures $\alpha_{\rm s}$ from scaling violations in F_2, measured with e or μ beams, the data are abundant, the statistical errors are small, the ambiguities from the treatment of heavy quarks and the effects of the longitudinal structure function $F_{\rm L}$ can be controlled, but there is an increased dependence on input parton densities, and most importantly a strong correlation between the result on $\alpha_{\rm s}$ and the adopted parametrization of the gluon density. In the following we restrict our attention to recent determinations of $\alpha_{\rm s}$ from scaling violations at NNLO accuracy, such as those in [26, 254] which report the results:

$$\alpha_{\rm s}(m_Z) = 0.1134 \pm 0.0011({\rm exp}) + ? \,, \tag{2.137}$$

$$\alpha_{\rm s}(m_Z) = 0.1158 \pm 0.0035 \,. \tag{2.138}$$

In the first line the question mark refers to the issue of the $\alpha_{\rm s}$–gluon correlation. In fact, $\alpha_{\rm s}$ tends to slide towards low values ($\alpha_{\rm s} \sim 0.113$–$0.116$) if the gluon input problem is not fixed. Indeed, in the second line, taken from [254], the large error also includes an estimate of the ambiguity from the gluon density parametrization. One way to restrict the gluon density is to use the Tevatron and LHC high $p_{\rm T}$ jet data to fix the gluon parton density at large x. Via the momentum conservation sum rule, this also constrains the small x values of the same density. Of course, in this way one has to go outside the pure domain of DIS. Further, the jet rates have been computed at NLO only. In a simultaneous fit of $\alpha_{\rm s}$ and the parton densities from a set of data which, although dominated by DIS data, also contains Tevatron jets and Drell–Yan production, the result was [287]

$$\alpha_{\rm s}(m_Z) = 0.1171 \pm 0.0014 + ? \,. \tag{2.139}$$

The authors of [287] attribute their higher value of α_s to a more flexible parametrization of the gluon and the inclusion of Tevatron jet data, which are important to fix the gluon at large x.

An alternative way to cope with the gluon problem is to drastically suppress the gluon parametrization rigidity by adopting the neural network approach. With this method, the following value was obtained, in [76], from DIS data alone, treated at NNLO accuracy:

$$\alpha_s(m_Z) = 0.1166 \pm 0.0008(\text{exp}) \pm 0.0009(\text{th}) + ? \,, \tag{2.140}$$

where the stated theoretical error is that quoted by the authors within their framework, while the question mark has to do with possible additional systematics from the method adopted. Interestingly, in the same approach, not much difference is found by also including the Tevatron jets and the Drell–Yan data:

$$\alpha_s(m_Z) = 0.1173 \pm 0.0007(\text{exp}) \pm 0.0009(\text{th}) + ? \,. \tag{2.141}$$

We see that, when the gluon input problem is suitably addressed, the fitted value of α_s is increased.

As we have seen there is some spread of results, even among the most recent determinations based on NNLO splitting functions. We tend to favour determinations from the whole DIS set of data (i.e., beyond the pure non-singlet case) and with attention paid to the gluon ambiguity problem (even if some non DIS data from Tevatron jets at NLO have to be included). A conservative proposal for the resulting value of α_s from DIS which emerges from the above discussion would be something like

$$\alpha_s(m_Z) = 0.1165 \pm 0.0020 \,. \tag{2.142}$$

The central value is below those obtained from Z and τ decays, but perfectly compatible with those results.

2.10.3 Recommended Value of $\alpha_s(m_Z)$

According to my proposal to calibrate $\alpha_s(m_Z)$ from the theoretically cleanest and most transparent methods, identified as the totally inclusive, light cone operator expansion dominated processes, I collect here my understanding of the results:

- From Z decays and EW precision tests, i.e., (2.126):

$$\alpha_s(m_Z) = 0.1190 \pm 0.0026 \,. \tag{2.143}$$

- From scaling violations in DIS, i.e., (2.142):

$$\alpha_s(m_Z) = 0.1165 \pm 0.0020 \,. \tag{2.144}$$

- From R_τ (2.133):

$$\alpha_s(m_Z) = 0.1194 \pm 0.0021. \tag{2.145}$$

If one wants to be on the safe side, one can take the average of Z decay and DIS, i.e.,

$$\alpha_s(m_Z) = 0.1174 \pm 0.0016\ . \tag{2.146}$$

This is my recommended value. If one adds to the average the rather conservative R_τ value and error given above in (2.145), which takes into account the dangerously low energy scale of the process, one obtains

$$\alpha_s(m_Z) = 0.1184 \pm 0.0011\ . \tag{2.147}$$

Note that this essentially coincides with the "official" average, with a moderate increase in the error.

2.10.4 Other $\alpha_s(m_Z)$ Measurements as QCD Tests

There are a number of other determinations of α_s that are important because they arise from qualitatively different observables and methods. Here I will give a few examples of the most interesting measurements.

A classic set of measurements comes from a number of infrared-safe observables related to event rates and jet shapes in e^+e^- annihilation. One important feature of these measurements is that they can be repeated at different energies in the same detector, like the JADE detector in the energy range of PETRA (most of the intermediate energy points in the right-hand panel of Fig. 2.32 are from this class of measurements) or the LEP detectors from LEP1 to LEP2 energies. As a result, one obtains a striking direct confirmation of the running of the coupling according to the renormalization group prediction. The perturbative part is known at NNLO [213], and resummations of leading logs arising from the vicinity of cuts and/or boundaries have been performed in many cases using effective field theory methods. The main problem with these measurements is the possibly large impact of non-perturbative hadronization effects on the result, and therefore on the theoretical error.

According to [99], a summarizing result that takes into account the central values and the spread from the JADE measurements at PETRA, in the range 14–46 GeV, is

$$\alpha_s(m_Z) = 0.1172 \pm 0.0051\ ,$$

while from the ALEPH data at LEP, in the range 90–206 GeV, the reported value [164] is

$$\alpha_s(m_Z) = 0.1224 \pm 0.0039\ .$$

It is amazing to note that among the related works there are a couple of papers by Abbate et al. [10, 11] where an extremely sophisticated formalism is developed for the thrust distribution, based on NNLO perturbation theory with resummations at NNNLL plus a data/theory-based estimate of non-perturbative corrections. The final quoted results are unbelievably precise:

$$\alpha_s(m_Z) = 0.1135 \pm 0.0011 \,,$$

from the tail of the thrust distribution [10], and

$$\alpha_s(m_Z) = 0.1140 \pm 0.0015 \,,$$

from the first moment of the thrust distribution [11]. I think that this is a good example of an underestimated error which is obtained within a given machinery without considering the limits of the method itself.

Another allegedly very precise determination of $\alpha_s(m_Z)$ is obtained from lattice QCD by several groups [288] with different methods and compatible results. A value that summarizes these different results is [307]

$$\alpha_s(m_Z) = 0.1185 \pm 0.0007 \,.$$

With all due respect to the lattice community, I think this small error is totally unrealistic. But we have shown that a sufficiently precise measurement of $\alpha_s(m_Z)$ can be obtained, viz., (2.146) and (2.147), by using only the simplest processes, where the control of theoretical errors is maximal. One is left free to judge whether a further restriction of theoretical errors is really on solid ground.

The value of Λ (for $n_f = 5$) which corresponds to (2.146) is

$$\Lambda_5 = 202 \pm 18\,\text{MeV} \,, \tag{2.148}$$

while the value from (2.147) is

$$\Lambda_5 = 213 \pm 13\,\text{MeV} \,. \tag{2.149}$$

Λ is the scale of mass that finally appears in massless QCD. It is the scale where $\alpha_s(\Lambda)$ is of order 1. Hadron masses are determined by Λ. Actually, the ρ mass or the nucleon mass receive little contribution from the quark masses (the case of pseudoscalar mesons is special, as they are the pseudo-Goldstone bosons of broken chiral invariance). Hadron masses would be almost the same in massless QCD.

2.11 Conclusion

We have seen that perturbative QCD based on asymptotic freedom offers a rich variety of tests, and we have described some examples in detail. QCD tests are not as precise as for the electroweak sector. But the number and diversity of such tests has established a very firm experimental foundation for QCD as a theory of strong interactions. The physics content of QCD is very large and our knowledge, especially in the non-perturbative domain, is still very limited, but progress both from experiment (Tevatron, RHIC, LHC, etc.) and from theory is continuing at a healthy rate. And all the QCD predictions that we have been able to formulate and to test appear to be in very good agreement with experiment.

The field of QCD appears to be one of great maturity, but also of robust vitality, with many rich branches and plenty of new blossoms. I may mention the very exciting explorations of supersymmetric extensions of QCD and the connections with string theory (for a recent review and a list of references, see [166]). In particular, $N = 4$ SUSY QCD (that is, with four spinor charge generators) has a vanishing beta function and is loop-finite. In the limit $N_C \to \infty$ with $\lambda = e_s^2 N_C$ fixed, planar diagrams are dominant. There is progress towards a solution of planar $N = 4$ SUSY QCD. The large λ limit corresponds by the AdS/CFT duality (anti-de Sitter/conformal field theory), a string theory concept, to the weakly coupled string (gravity) theory on $AdS_5 \times S_5$ (the 10 dimensions are compactified in a 5-dimensional anti-de Sitter space times a 5-dimensional sphere). By moving along this very tentative route, one can transfer some results (assumed to be of sufficiently universal nature) from the computable weak limit of the associated string theory to the non-perturbative ordinary QCD domain. Further along this line of investigation, there are studies of $N = 8$ supergravity, related to $N = 4$ SUSY Yang–Mills, which has been proven finite up to four loops. It could possibly lead to a finite field theory of gravity in four dimensions.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

http://www.springer.com/978-3-319-51919-7

Collider Physics within the Standard Model
A Primer
Altarelli, G. - Wells, J.D. (Ed.)
2017, XIV, 173 p. 60 illus., 34 illus. in color., Softcover
ISBN: 978-3-319-51919-7