written using Hamiltonian and stationary states, or using Lagrangian and periodic Euclidean paths. Further elucidation of this duality regarding QCD monopoles [1425] shed light on their density and the long-known absence of classical solutions for them. All of these hint that the different faces of "gauge topology" we discussed will "asymptotically" converge into a single semiclassical theory.

6 Effective field theories

Conveners:
Franz Gross and Mike Strickland

In this second section on approximate methods, we discuss effective field theories (EFTs), a powerful technique that can be used when there are widely separated energy scales appearing in a problem. A classic example of this is nonrelativistic QCD (NRQCD), which emerges in the limit of a large quark mass M (Sect. 6.1). For $v = p/M \ll 1$, there is a large separation between 'hard' modes, with energy on the order of M; soft modes, with energy on the order of Mv; and ulrasoft (potential) modes, with energy/momentum on the order Mv^2. Using EFT methods, one can write an effective NRQCD Lagrangian that includes all terms allowed by QCD symmetries. The coefficients in this effective Lagrangian can be computed systematically by a matching procedure, which ensures that quantities calculated in the EFT are the same as those computed in QCD itself up to a given order in v. The NRQCD EFT can be extended by further integrating out the soft scale to obtain an effective theory called potential NRQCD (pNRQCD), which is written entirely in terms of singlet and octet quark–antiquark pairs. As a result, at leading order in pNRQCD, the physics of heavy quarkonium reduces to solving a Schrödinger equation for bound state wave functions.

This is but one example. The use of EFTs applied to QCD has allowed systematic progress on many fronts in the last decades. These include a systematically extendable model of low-energy hadronic physics called chiral perturbation theory (Sect. 6.2), which can be used as a foundation for nuclear physics (Sect. 6.3) giving both a successful description of the NN interaction up to 200 MeV, and the properties of light nuclei up to $A \leq 12$. In the realm of jets, soft-collinear effective theory implements power counting in the transverse momentum of gluon radiation (Sect. 6.4).

EFT methods have also been used to understand high-temperature QCD thermodynamics, in which case the hard, soft, and ultrasoft scales are T, gT, and g^2T, respectively (Sect. 6.5). The resulting EFTs, called electrostatic QCD (EQCD) and magnetostatic QCD (MQCD) allow one to systematically calculate the equation of state of a high-temperature quark–gluon plasma. Together with other finite-temperature resummation schemes such as hard-thermal-loop perturbation theory these methods have provided a way to calculate the QCD equation state that agrees well with lattice calculations down to temperatures just above the quark–gluon plasma phase transition temperature. Finally, Sect. 6.6 describes how EFTs have recently been applied to non-equilibrium QCD physics such as the quantum transport of bottomonium through the quark–gluon plasma.

6.1 Nonrelativistic effective theory

Antonio Vairo

In QCD, quarks may be divided into two fundamental sets: heavy quarks (charm, bottom, top) whose masses m_h are much larger than the typical hadronic scale $\Lambda_{\rm QCD}$ and light quarks (up, down, strange) whose masses m_ℓ are much smaller than $\Lambda_{\rm QCD}$. Both the hierarchies, $m_h \gg \Lambda_{\rm QCD}$ and $m_\ell \ll \Lambda_{\rm QCD}$, allow for an effective field theory (EFT) treatment of hadrons that exploits the symmetries that the hadrons manifest in the large and small mass limits. Because these symmetries are not manifest in QCD, the EFT is typically simpler and more predictive than the full QCD treatment, at least at the lowest orders in the effective expansion. At higher orders in the effective expansion the original symmetries of QCD are restored. We discuss EFTs for heavy quarks in this section, while EFTs for light quarks, i.e., chiral EFTs, are reviewed in the following sections.

In general, an effective field theory of QCD is constructed as an expansion in the ratio Λ_ℓ/Λ_h of a low energy scale Λ_ℓ, e.g. $\Lambda_{\rm QCD}$, and a high energy scale Λ_h, e.g. m_h. Each term in the expansion is made of the fields describing the system at the low-energy scale; these terms may have any form consistent with the symmetries of QCD. The low-energy fields are the effective degrees of freedom. The resulting scattering matrix is the most general one consistent with analyticity, perturbative unitarity, cluster decomposition and the symmetry principles [1426].

It is said that the large energy scale "has been integrated out" from QCD. Analytic terms in the expansion parameter Λ_ℓ/Λ_h are accounted for by the operators of the EFT. Non-analytic terms, carrying the contributions of the high-energy modes in QCD, which are no longer dynamical in the EFT, are encoded in the parameters multiplying the EFT operators. These parameters are the Wilson coefficients of the EFT, also called matching coefficients, or low-energy constants in the chiral EFT. Hence, EFTs automatically factorize, for any observable, high-energy from low-energy contributions. The Wilson coefficients of the EFT Lagrangian are fixed by matching to the fundamental theory, i.e., by requiring that the EFT and the fundamental theory describe the same physics (observables, Green functions, scattering matrices,…) at any given order in the expansion parameter Λ_ℓ/Λ_h.

The advantage of dealing with heavy quarks is that the matching coefficients associated with the heavy quark mass scale are guaranteed to be computable in perturbative QCD, i.e., order by order in $\alpha_s(m_h)$, as a consequence of asymptotic freedom. This is not the case for matching coefficients associated with lower energy scales or for the low-energy constants that need to be computed either numerically in lattice QCD or fixed on experimental data.

To allow for controlled calculations based on the effective Lagrangian, operators, as well as the quantum corrections, are organized according to their expected importance. Operators in the Lagrangian are counted in powers of the small expansion parameter Λ_ℓ/Λ_h, whereas quantum corrections are either computed exactly or counted in powers of the coupling constant. For example, a strict expansion in terms of the coupling is possible, as remarked above, when integrating out the heavy quark mass.

EFTs are renormalizable at each order in the expansion parameter. Hence, the EFT produces finite and controlled expansions for any observable of the effective degrees of freedom that may be computed respecting the energy scale hierarchy upon which the EFT is based. The power counting of the EFT, i.e., the way to assess the size of the different terms in the effective expansion, may or may not be obvious. The power counting turns out to be obvious if the system is characterized by just one dynamical energy scale. Reducing the description of a system to that one of an effective one scale system is the ultimate goal of any effective field theory.

In this section, we restrict ourself to EFTs for heavy quarks, where the heavy quark mass is the largest scale. These EFTs are called nonrelativistic EFTs, because requiring the heavy quark mass to be the largest scale implies that it is also larger than the momentum p of the heavy quark in the hadron reference frame: the condition $m_h \gg p$ qualifies the quark as nonrelativistic. The presentation of this section follows the one of Ref. [1427].

For hadrons made of one heavy quark, like heavy-light mesons and baryons, the proper nonrelativistic EFT is called Heavy Quark Effective Theory (HQET). Heavy-light hadrons are systems characterized by just two relevant energy scales, m_h and $\Lambda_{\rm QCD}$. HQET follows from QCD by integrating out modes associated with the heavy quark mass and exploiting the hierarchy $m_h \gg \Lambda_{\rm QCD}$. In the context of HQET one deals with heavy-light hadrons made of either a charm or a bottom quark (the top quark has no time to form a bound state before decaying weakly into a b quark). HQET is discussed in Sect. 6.1.1.

Systems made of more than one heavy quark, like quarkonia (e.g. charmonium and bottomonium) or quarkonium-like states or doubly-heavy baryons are characterized by more energy scales. The typical distance between the heavy quarks is of the order of $1/(m_h v)$, $v \ll 1$ being the relative velocity of the heavy quark, which implies that the typical momentum transfer between the heavy quarks is of order $m_h v$, and the typical binding and kinetic energy is of order $m_h v^2$. The inverse of $m_h v^2$ sets the time scale of the bound state. These systems are to some extent the QCD equivalent of positronium in QED. In a positronium, an electron and a positron move with a relative velocity $v \sim \alpha$, where α is the fine structure constant, at a typical distance given by the Bohr radius, which is proportional to $1/(m\alpha)$, and are bound with the energy given by the Bohr levels, which are proportional to $m\alpha^2$.

At each of the energy scales one can construct an EFT, specifically, nonrelativistic QCD (NRQCD) at the scale $m_h v$, which is discussed in Sect. 6.1.2, and potential NRQCD (pNRQCD) at the scale $m_h v^2$, which is discussed in Sect. 6.1.3.

6.1.1 Heavy quark effective theory

Heavy Quark Effective Theory was the first nonrelativistic EFT of QCD with a fully developed nonrelativistic expansion, computation of higher-order radiative corrections, renormalization group equations, and a wide range of physical applications [704,1047,1250,1428] (for an early review see, for instance, Ref. [1429], for a textbook see Ref. [711]). This despite the fact that nonrelativistic QCD and QED, the EFTs for two nonrelativistic particles that we discuss in Sect. 6.1.2, were suggested before [1430].

In a sense, HQET describes QCD in the opposite limit of the chiral EFT, however, it is important to realize that HQET is not the large mass limit of QCD, but the EFT suited to describe heavy-light hadrons, i.e., hadrons made of one heavy particle and light degrees of freedom. The heavy particle may be taken to be a heavy quark, but also a composite particle made by more than one heavy quark when the internal modes of the composite heavy particle can be ignored. The light degrees of freedom are made by light quarks and gluons. Among the light quarks we may distinguish between valence quarks and sea quarks, where the first ones are those that establish, together with the heavy degrees of freedom, the quantum numbers of the heavy-light hadron.

The HQET Lagrangian is made of low-energy degrees of freedom living at the low-energy scale $\Lambda_{\rm QCD}$. These are the low-energy modes of the heavy quark (antiquark), described by a Pauli spinor ψ (χ) that annihilates (creates) the heavy quark (antiquark), and low-energy gluons and light quarks. The HQET is constructed as an expansion in $1/m_h$: the heavy quark expansion. Matrix elements of operators of dimension d are of order $\Lambda_{\rm QCD}^d$, hence the higher the dimension of the operator the higher the suppression in $\Lambda_{\rm QCD}/m_h$. In the rest frame of the heavy-light hadron, the HQET Lagrangian density for a heavy quark reads up to order $1/m_h^2$ and including the $1/m_h^3$ kinetic operator (HQET up to order $1/m_h^4$ can be found in Refs. [1431,1432])

$$\mathcal{L}_{\rm HQET} = \mathcal{L}^{\psi} + \mathcal{L}^{\ell}, \tag{6.1}$$

with

$$\mathcal{L}^{\psi} = \psi^{\dagger}\Bigg\{ i D_0 + \frac{\boldsymbol{D}^2}{2m_h} + \frac{\boldsymbol{D}^4}{8m_h^3} - c_F \frac{\boldsymbol{\sigma}\cdot g\boldsymbol{B}}{2m_h} - c_D \frac{[\boldsymbol{D}\cdot, g\boldsymbol{E}]}{8m_h^2} - i c_S \frac{\boldsymbol{\sigma}\cdot[\boldsymbol{D}\times, g\boldsymbol{E}]}{8m_h^2}\Bigg\}\psi, \tag{6.2}$$

$$\mathcal{L}^{\ell} = -\frac{1}{4}F^A_{\mu\nu}F^{A\,\mu\nu} + \frac{d_2}{m_h^2}F^A_{\mu\nu}D^2F^{A\,\mu\nu} - \frac{d_3}{m_h^2}g f_{ABC}F^A_{\mu\nu}F^B_{\mu\alpha}F^C_{\nu\alpha} + \sum_{\ell=1}^{n_\ell}\bar{q}_\ell\left(i\not{D} - m_\ell\right)q_\ell, \tag{6.3}$$

where $i\boldsymbol{D} = i\boldsymbol{\nabla} - g\boldsymbol{A}$, $iD_0 = i\partial_0 + gA_0$, $[\boldsymbol{D}\cdot, g\boldsymbol{E}] = \boldsymbol{D}\cdot g\boldsymbol{E} - g\boldsymbol{E}\cdot\boldsymbol{D}$ and $[\boldsymbol{D}\times, g\boldsymbol{E}] = \boldsymbol{D}\times g\boldsymbol{E} - g\boldsymbol{E}\times\boldsymbol{D}$, $\boldsymbol{E}^i = F^{i0}$ is the chromoelectric field, $\boldsymbol{B}^i = -\epsilon_{ijk}F^{jk}/2$ the chromomagnetic field, and $\boldsymbol{\sigma}$ are the Pauli matrices. The fields q_ℓ are n_ℓ light-quark fields. The heavy-quark mass, m_h, has to be understood as the heavy quark pole mass, hence not the mass that is in the QCD Lagrangian. The coefficients c_F, c_D, c_S, d_2, and d_3 are Wilson coefficients of the EFT. They encode the contributions of the high-energy modes that have been integrated out from QCD. Since the high-energy scale, m_h, is larger than $\Lambda_{\rm QCD}$, the Wilson coefficients may be computed in perturbation theory and organized as an expansion in α_s at a typical scale of order m_h. The coefficients c_F, c_D, and c_S are 1 at leading order, while the perturbative series of the coefficients d_2 and d_3 starts at order α_s. The one-loop expression of the coefficients may be found in Ref. [1433]. Some of the coefficients are known far beyond one loop. For instance, the Fermi coefficient c_F, which plays a crucial role in the spin splittings, is known up to three loops [1434]. In Eq. (6.3) we have not considered $1/m_h^2$ suppressed operators involving light quarks [1435,1436] since their impact is negligible in most hadronic observables. The HQET Lagrangian for a heavy antiquark may be obtained from Eqs. (6.1) and (6.2) by charge conjugation.

The HQET Lagrangian provides a description of heavy-light hadrons that is the same as QCD order by order in $\Lambda_{\rm QCD}/m_h$. Because QCD is a Lorentz invariant theory, this symmetry must be somehow maintained in HQET, although HQET breaks manifest Lorentz invariance by the nonrelativistic expansion. Indeed, Lorentz invariance is realized in HQET by constraining the Wilson coefficients [1433,1437–1439]. For instance, Lorentz invariance relates c_F and the spin–orbit coefficient c_S: $c_S = 2c_F - 1$. This relation is exact, which means that it holds to any order in α_s.

The impact of HQET on the physics involving heavy-light hadrons and, in particular, their weak decays has been enormous. The reason is that the leading-order HQET Lagrangian,

$$\mathcal{L}^{(0)}_{\rm HQET} = \psi^{\dagger} i D_0 \psi - \frac{1}{4}F^A_{\mu\nu}F^{A\,\mu\nu} + \sum_{\ell=1}^{n_\ell}\bar{q}_\ell\left(i\not{D} - m_\ell\right)q_\ell, \tag{6.4}$$

makes manifest a hidden symmetry of heavy-light hadrons. This symmetry is the heavy-quark symmetry and stands for invariance with respect to the heavy-quark flavor and spin. A consequence of the heavy-quark symmetry is that electroweak transitions in the heavy-light meson sector depend on only one form factor, the Isgur–Wise function $\xi(w)$, whose absolute normalization is $\xi(0) = 1$ [1250,1428]. Moreover, the leading-order HQET Lagrangian is exactly renormalizable.

Higher-order operators in Eq. (6.1) break the heavy-quark symmetry (and exact renormalizability), however, they do it in a perturbative way controlled by powers of $\Lambda_{\rm QCD}/m_h$. Hence, observables computed up to some order in the HQET expansion depend on fewer and more universal nonperturbative matrix elements than they would in a full QCD calculation. This makes the heavy quark expansion more predictive than a full QCD calculation.

As an application, let us consider the heavy-light meson masses. Expressed in the HQET as an expansion up to order $1/m_h$, they read [1440]

$$M_{H^{(*)}} = m_h + \bar{\Lambda} + \frac{\mu_\pi^2}{2m_h} - d_{H^{(*)}}\frac{\mu_G^2(m_h)}{2m_h} + \mathcal{O}\left(\frac{1}{m_h^2}\right), \tag{6.5}$$

where $M_{H^{(*)}}$ is the spin singlet (triplet) meson mass, m_h the heavy quark pole mass, $\bar{\Lambda}$ the binding energy in the static limit, of order $\Lambda_{\rm QCD}$, $\mu_\pi^2/2m_h$ the kinetic energy of the heavy quark (μ_π^2 is the matrix element of $\psi^\dagger\boldsymbol{D}^2\psi$), of order $\Lambda^2_{\rm QCD}/m_h$, $d_{H^{(*)}}$ is 1 for H and $-1/3$ for H^*, and $d_{H^{(*)}}\mu_G^2(m_h)/2m_h$ is the matrix element of $c_F\,\psi^\dagger\boldsymbol{\sigma}\cdot g\boldsymbol{B}/(2m_h)\psi$, of order $\Lambda^2_{\rm QCD}/m_h$. The heavy quark symmetry manifests itself through the universality of the leading term $M_{H^{(*)}} - m_h \approx \bar{\Lambda}$, and of the matrix elements μ_π^2 and $\mu_G^2(m_h)/c_F(m_h)$, which depend neither on the heavy quark flavor nor on the heavy quark spin. The flavor dependence of $\mu_G^2(m_h)$ comes entirely from the Wilson coefficient c_F, which depends on m_h through the running of the strong coupling. Equation (6.5) can be used to extract the heavy quark masses from the measured meson masses. One can also use lattice QCD data to determine meson masses for fictitious heavy quarks of any mass m_h, so to reconstruct $M_{H^{(*)}}$ as a function of m_h. One general difficulty in this kind of study is that the relation between the $\overline{\rm MS}$ mass, which is the short distance quantity that appears in the renormalized QCD Lagrangian, and the pole mass, which is the quantity that

appears in the HQET Lagrangian, is plagued by a poorly convergent perturbative series (at present, the relation between the $\overline{\rm MS}$ mass and the pole mass is known up to four loops [1441,1442]). The large terms in the perturbative series trace back to a renormalon singularity in the Borel plane of order $\Lambda_{\rm QCD}$. This singularity may be subtracted from the pole mass and reabsorbed into a redefinition of the other nonperturbative parameters appearing in Eq. (6.5). There are many possible subtraction schemes [1443–1448]. For illustration, we present the heavy quark masses and matrix elements appearing in (6.5) obtained from lattice QCD data set to reproduce the physical D_s and B_s masses in Ref. [1449]:

$$\begin{aligned}
\overline{m}_c &= 1273(10)\ \text{MeV},\\
\overline{m}_b &= 4195(14)\ \text{MeV},\\
\bar{\Lambda} &= 555(31)\ \text{MeV},\\
\mu_\pi^2 &= 0.05(22)\ \text{GeV}^2,\\
\mu_G^2(m_b) &= 0.38(2)\ \text{GeV}^2,
\end{aligned}$$

where $\overline{m}_h$ is the $\overline{\rm MS}$ mass of the quark h at the scale of its $\overline{\rm MS}$ mass, $\bar{\Lambda}$ is in the renormalon subtraction scheme adopted in Ref. [1448,1449] and the quantity μ_G^2 has been evaluated for the b quark. Note the approximate scaling of the nonperturbative parameters according to the power counting of HQET (with a somewhat smaller μ_π^2).

Equation (6.5) can be immediately extended to heavy-light baryons. What changes is the explicit value of the nonperturbative matrix elements, as the light degrees of freedom are different from the mesonic case. Also doubly-heavy baryons may be described by the same mass formula if the typical distance between the two heavy quarks is much smaller than the typical size of a heavy-light meson, which is of order $1/\Lambda_{\rm QCD}$. In this case, at a distance of order $1/\Lambda_{\rm QCD}$ one cannot resolve the inner structure of the heavy diquark system, which effectively behaves as a pointlike particle in an antitriplet color configuration, i.e., as a heavy antiquark of mass $2m_h$; under some conditions, effects due to the heavy quark–quark interaction my be added perturbatively in the framework of the nonrelativistic EFTs developed in the following sections [796,1450–1459]. Finally, the heavy quark symmetry may be also applied to link doubly-heavy tetraquarks (tetraquarks made of two heavy and two light quarks) with heavy-light baryons sharing the same light-quark content [1455,1460,1461]. Many of the new charmonium- and bottomonium-like states observed at colliders in the last decades have a doubly-heavy tetraquark content [1427].

6.1.2 Nonrelativistic QCD

Hadrons made of two or more nonrelativistic particles, like two heavy quarks or a heavy quark and a heavy antiquark, or more generally just heavy quark–antiquark pairs near threshold, are multiscale systems characterized by a hierarchy of dynamically generated scales:

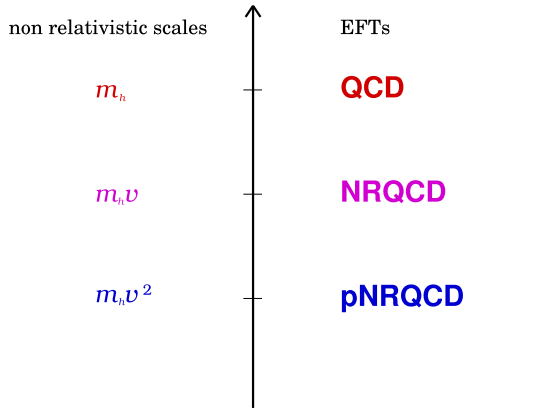


Fig. 132 Hierarchy of energy scales and EFTs for systems made of a heavy quark and (anti)quark pair near threshold

$$m_h \gg m_h v \gg m_h v^2. \tag{6.6}$$

We discussed the origin of these energy scales at the beginning of the section. The nonrelativistic energy scales are correlated. To reach a situation like in HQET, i.e. an EFT with just one dynamical low-energy scale, we need to construct at least two nonrelativistic EFTs: one following from integrating out from QCD modes associated with the energy scale m_h and one following from integrating out modes associated with the energy scale $m_h v$, ending up with an ultimate EFT at the energy scale $m_h v^2$ [1462]. An illustration of the tower of energy scales and corresponding EFTs is in Fig. 132. In the last 20 years, the development of such nonrelativistic EFTs of QCD has been the major theoretical breakthrough in the description of quarkonium and quarkonium-like systems [1463–1465]. For a more historical perspective, see Ref. [1466].

NRQCD is the EFT suited to describe systems made of a heavy quark and (anti)quark pair near threshold that follows from QCD by integrating out the energy scale associated with the heavy quark mass, m_h [1430]. In a heavy quark–antiquark bound state, the virial theorem constrains the kinetic energy of the heavy particles to be of the same order as the binding energy. As a consequence, the power counting of NRQCD must be such that the leading-order NRQCD Lagrangian includes the kinetic energy operators, $\psi^\dagger\ \nabla^2/(2m_h)\ \psi\ -\ \chi^\dagger\nabla^2/(2m_h)\chi$, making the NRQCD Lagrangian, even at leading order, non renormalizable. This is different from HQET.

NRQCD posed initially also some difficulties in finding a computational scheme to integrate consistently over the different momentum and energy regions in dimensional regularization. NRQCD or its QED equivalent were therefore used for a long time either for analytical calculations in QED

with a hard cut off [1467,1468] or for lattice QCD calculations involving heavy quarks [289,1469]. The advantage for lattice NRQCD calculations is that, once the heavy quark mass has been integrated out, the lattice spacing, a, is not constrained, as in full lattice QCD calculations, to be smaller than $1/m_h$, which would amount to requiring a very fine lattice if the quark is heavy. In lattice NRQCD the constraint is relaxed to $a < 1/(m_h v)$. Since at the same time the lattice size has to be large enough to include distances of the order of $1/\Lambda_{\rm QCD}$ for quenched calculations and $1/M_\pi$ for full calculations, simulations with heavy quarks in full QCD are computationally quite demanding. Lattice NRQCD has been, for a long time, the sole way to compute nonperturbatively observables involving bottom quarks in full QCD (see, for instance, Refs. [281,1470–1473]). Only recently the first full lattice QCD calculations of bottomonium-like systems have become available [1474].

After the development of HQET, NRQCD became a systematic tool for analytical calculations of quarkonium observables. NRQCD is well suited to the description of heavy quark–antiquark annihilation, because this happens at the scale m_h, which is the energy scale that has been integrated out from QCD to construct NRQCD. In NRQCD, the information about annihilation goes into the (imaginary part) of the Wilson coefficients associated with the four-fermion operators. These four fermion operators, a novel feature of NRQCD with respect to HQET, are not only essential to the description of the annihilation processes, but also to the correct description of all short-distance interactions between the heavy particles. In NRQCD, annihilation processes factorize therefore into a short-distance part, which may be computed in perturbative QCD and is encoded in the Wilson coefficients, and into matrix elements of the NRQCD operators that encode the low-energy dynamics of the heavy quark–antiquark bound state. Processes involving heavy quark–antiquark annihilations are quarkonium inclusive and electromagnetic decay [1475,1476] and quarkonium production [1476]. The large amount of data on quarkonium production in hadron and lepton colliders, together with the predictive power of NRQCD and its success in most of the predictions, has established NRQCD as a standard tool for studying quarkonium annihilation [1463–1465,1477–1479].

Because four-fermion operators projecting onto color octet quark–antiquark states are possible, NRQCD naturally allows for production and decay of quark–antiquark states in a color octet configuration. These states constitute a suppressed, in v, component of the Fock state describing a physical quarkonium, but are necessary in the quarkonium phenomenology [1463–1465]. They are also necessary for the consistency of the theory, as they cancel infrared divergences in quarkonium decay and production observables and eventually provide finite, physical results [1475,1476]. It should be noted, however, that the NRQCD factorization has been rigorously proved only for quarkonium decay but not for quarkonium production [1275,1313,1480–1482].

A last crucial progress in establishing NRQCD as a valuable tool for analytical calculations came when it was shown that the computation of the Wilson coefficients of NRQCD in dimensional regularization requires expanding in the heavy quark mass to avoid integrating over the high momentum region. This means that, even if the power countings of NRQCD and HQET are different, the matching to QCD proceeds in the same way, leading to the same operators and Wilson coefficients in the two-fermion and gauge sectors [1433].

The NRQCD Lagrangian density for systems made of a heavy quark and a heavy antiquark of equal masses m_h up to order $1/m_h^2$, and including the $1/m_h^3$ kinetic operator, is given by

$$\mathcal{L}_{\rm NRQCD} = \mathcal{L}^\psi + \mathcal{L}^\chi + \mathcal{L}^{\psi\chi} + \mathcal{L}^\ell, \tag{6.7}$$

where $\mathcal{L}^\psi$ and $\mathcal{L}^\chi$ are the HQET Lagrangian densities for the quark (see Eq. 6.2) and antiquark, respectively, $\mathcal{L}^\ell$ is the Lagrangian density (6.3) for the light degrees of freedom, and $\mathcal{L}^{\psi\chi}$ is the four-fermion sector, which up to order $1/m_h^2$ reads

$$\begin{aligned}\mathcal{L}^{\psi\chi} = {} & \frac{f_1(^1S_0)}{m_h^2}\psi^\dagger\chi\chi^\dagger\psi + \frac{f_1(^3S_1)}{m_h^2}\psi^\dagger\boldsymbol{\sigma}\chi\cdot\chi^\dagger\boldsymbol{\sigma}\psi \\ & + \frac{f_8(^1S_0)}{m_h^2}\psi^\dagger \mathrm{T}^A\chi\chi^\dagger \mathrm{T}^A\psi + \frac{f_8(^3S_1)}{m_h^2}\psi^\dagger \mathrm{T}^A\boldsymbol{\sigma}\chi\cdot\chi^\dagger \mathrm{T}^A\boldsymbol{\sigma}\psi.\end{aligned} \tag{6.8}$$

As in the HQET case, m_h is the pole mass. The four-fermion Lagrangian in Eq. (6.8) is made of all possible four-fermion operators of dimension 6. The corresponding Wilson coefficients are $f_1(^1S_0)$, $f_1(^3S_1)$, $f_8(^1S_0)$, and $f_8(^3S_1)$. The the first (second) four-fermion operator projects on a heavy quark–antiquark pair in a color singlet configuration with quantum numbers $^1S_0$ ($^3S_1$), whereas the third (fourth) four-fermion operator projects on a heavy quark–antiquark pair in a color octet configuration with quantum numbers $^1S_0$ ($^3S_1$). The matrices T^A are the SU(3) generators $\lambda^A/2$. The four-fermion Wilson coefficients have been computed in Refs. [1476,1483]. They have a real part that starts at order α_s for $f_8(^3S_1)$ and at order α_s^2 for the other coefficients, and they have an imaginary part, coming from one loop or higher annihilation diagrams, which is of order α_s^2 for Im $f_1(^1S_0)$, Im $f_8(^1S_0)$, and Im $f_8(^3S_1)$, and of order α_s^3 for Im $f_1(^3S_1)$. A list of imaginary parts of four-fermion Wilson coefficients in NRQCD and related bibliography can be found in Ref. [1484].

The four-fermion sector of the NRQCD Lagrangian has been derived up to order $1/m_h^4$ (complete) and orders $1/m_h^5$ and $1/m_h^6$ (partial) in Refs. [1485–1487]. Like for the Wilson coefficients in the two-fermion sector, also the

coefficients in the four-fermion sector are not all independent: some are related by Poincaré invariance [1486, 1487].

Sometimes it is useful to isolate the electromagnetic component of the four-fermion operator and its corresponding Wilson coefficient. This is the case when computing electromagnetic decay widths and photoproduction cross sections in NRQCD. The electromagnetic operators are obtained by projecting on an intermediate QCD vacuum state, $|0\rangle$, e.g., $\psi^\dagger \chi \chi^\dagger \psi \to \psi^\dagger \chi \, |0\rangle\langle 0| \, \chi^\dagger \psi$.

Unlike in HQET, the power counting of NRQCD is not unique. The reason is that, while HQET is a one-scale EFT, its only dynamical scale being $\Lambda_{\rm QCD}$, NRQCD is still a multiscale EFT. The dynamical scales of NRQCD are, at least, $m_h v$, $m_h v^2$, and $\Lambda_{\rm QCD}$. In more complicated settings even more scales may be relevant. Hence, one can imagine different power countings: some more conservative, like assuming that the matrix elements all scale according to the largest dynamical scale, i.e., $m_h v$, [1488], and some less conservative or closer to a perturbative counting [1476]. All the power countings have in common that the kinetic energy scales like the binding energy and that therefore $\psi^\dagger i\partial_0 \psi$ is of the same order as $\psi^\dagger \boldsymbol{\nabla}^2/(2m_h)\psi$, and analogously for the antiquark. As we have mentioned above, this reflects the virial theorem, an unavoidable consequence of the dynamics of a nonrelativistic bound state.

The leading-order NRQCD Lagrangian reads in Coulomb gauge [1476]

$$\mathcal{L}^{(0)}_{\rm NRQCD} = \psi^\dagger \left(i D_0 + \frac{\boldsymbol{\nabla}^2}{2m_h} \right) \psi + \chi^\dagger \left(i D_0 - \frac{\boldsymbol{\nabla}^2}{2m_h} \right) \chi - \frac{1}{4} F^A_{\mu\nu} F^{A\,\mu\nu} + \sum_{\ell=1}^{n_\ell} \bar{q}_\ell \left(i \not{D} - m_\ell \right) q_\ell . \tag{6.9}$$

Note that this Lagrangian contains the heavy quark mass, and therefore violates the heavy-quark flavor symmetry; hence the bottomonium binding energy is different, even at leading order, from the charmonium one. In the power counting of Ref. [1476] one further assumes: $D_0 \sim m_h v^2$ (when acting on ψ or χ), $\boldsymbol{D} \sim m_h v$ (when acting on ψ or χ), $g\boldsymbol{E} \sim m_h^2 v^3$, and $g\boldsymbol{B} \sim m_h^2 v^4$. A consequence is that the heavy-quark spin symmetry is a symmetry of the leading-order NRQCD Lagrangian. Another consequence is that the order $1/m_h^3$ kinetic energy operator $\psi^\dagger \boldsymbol{D}^4/(8m_h^3)\psi$ and its charge conjugated are of the same order as the $1/m_h$ and $1/m_h^2$ operators in $\mathcal{L}^\psi$ and $\mathcal{L}^\chi$. Matrix elements of octet operators on quarkonium states are further suppressed by the fact that they project on subleading components of the quarkonium Fock state, the ones made of a heavy quark–antiquark pair in a color octet configuration and gluons. The amount of suppression depends on the adopted power counting.

6.1.3 Potential nonrelativistic QCD

Nonrelativistic bound states involve energy scales, $m_h v$, $m_h v^2$, and $\Lambda_{\rm QCD}$, that are still dynamical and entangled in NRQCD. A consequence of this is that, although the equations of motion that follow from the NRQCD Lagrangian (6.9) resemble a Schrödinger equation for nonrelativistic bound states, they are not quite that. They involve gauge fields and do not supply a field theoretical definition and derivation of the potential that would appear in a Schrödinger equation. Nevertheless, we expect that, in some nonrelativistic limit, a Schrödinger equation describing the quantum mechanics of the nonrelativistic bound state should emerge from field theory, since field theory may be understood as an extension of quantum mechanics that includes relativistic and radiative corrections. Another consequence already remarked in the previous section is that the power counting of NRQCD is not unique.

Since the scales $m_h v$ and $m_h v^2$ are hierarchically ordered, they may be disentangled by systematically integrating out modes associated with scales larger than the smallest scale, $m_h v^2$, and matching to a lower energy EFT, where only degrees of freedom resolved at distances of order $1/(m_h v^2)$ are left dynamical [1462]. This EFT is pNRQCD [1489,1490]. Because the scale $m_h v$ has been integrated out, the power counting of pNRQCD is less ambiguous than the one of NRQCD. In situations where we can neglect the hadronic scale $\Lambda_{\rm QCD}$, the power counting of pNRQCD is indeed unique, as its only dynamical scale is $m_h v^2$.

Having integrated out the scale $m_h v$ associated with the inverse of the distance r between the heavy quark and antiquark, implies that pNRQCD is constructed as an expansion in r, with Wilson coefficients encoding non-analytic contributions in r. This is analogous to how HQET and NRQCD are constructed; there the heavy quark mass, m_h, is integrated out and the EFTs are organized as expansions in $1/m_h$, with Wilson coefficients encoding the non-analytic contributions in the form of logarithms of m_h. Some of the Wilson coefficients of pNRQCD may be identified with the potentials in the Schrödinger equation of quarkonium.

The specific form of pNRQCD depends on the scale $\Lambda_{\rm QCD}$. If $\Lambda_{\rm QCD} \lesssim m_h v^2$, then one deals with weakly-coupled bound states and the EFT is called weakly-coupled pNRQCD. At distances of the order of or smaller than $1/(m_h v^2)$, one may still resolve colored degrees of freedom (gluons, quarks, and antiquarks), as color confinement has not yet set in. Hence gluons, quarks, and antiquarks are the degrees of freedom of weakly-coupled pNRQCD. Weakly-coupled pNRQCD is well suited to describe tightly bound quarkonia, like bottomonium and (to a less extent) charmonium ground states, the B_c ground state, and threshold effects in $t\bar{t}$ production. If $\Lambda_{\rm QCD} \gtrsim m_h v^2$, then one deals with strongly-coupled bound states and the EFT is

called strongly-coupled pNRQCD. At distances of the order of $1/(m_h v^2)$, confinement has set in and the only available degrees of freedom are color singlets. These are, in principle, all, ordinary and exotic, heavy, heavy-light and light hadrons that we might have in the spectrum. Strongly-coupled pNRQCD is suited to describe higher states in the bottomonium and charmonium spectra, as well as quarkonium exotica. If $m_h v \gg \Lambda_{\rm QCD} \gg m_h v^2$, the matching to pNRQCD may be done in two steps, first integrating out (perturbatively) $m_h v$ then (nonperturbatively) $\Lambda_{\rm QCD}$. Contributions coming from these two scales will be automatically factorized in pNRQCD observables.

Weakly-coupled pNRQCD

The degrees of freedom of weakly-coupled pNRQCD are heavy quarks and antiquarks of momentum $m_h v$ and energy $m_h v^2$, gluons of momentum and energy $m_h v^2$ (sometimes called ultrasoft gluons), and light quarks of momentum and energy $m_h v^2$; these remain after gluons (sometimes called soft gluons) and light quarks of energy or momentum $m_h v$ have been integrated out from NRQCD. Because a single heavy quark and antiquark cannot be resolved at the scale $m_h v^2$, it is useful to cast heavy quark and antiquark fields into bilocal fields that depend on time, t, the center of mass coordinate, $\boldsymbol{R}$, and the relative coordinate, $\boldsymbol{r}$. We call the color singlet component of the quark and antiquark field S, and its color octet component O, normalized to $\mathrm{S} = \mathbb{1}_{3\times 3} S/\sqrt{3}$ and $\mathrm{O} = \sqrt{2} O^A T^A$. The distance r scales like $1/(m_h v)$, while the center of mass coordinate, R, and the time, t, scale like $1/(m_h v^2)$, because the quark–antiquark pair may only recoil against ultrasoft gluons. To ensure that gluons are ultrasoft in the pNRQCD Lagrangian, gauge fields are multipole expanded in $\boldsymbol{r}$. Hence gauge fields in the pNRQCD Lagrangian only depend on time and the center of mass coordinate. The pNRQCD Lagrangian is organized as a double expansion in $1/m_h$ and r. At order r in the multipole expansion, the weakly-coupled pNRQCD Lagrangian density has the form [1489,1490]

$$\mathcal{L}^{\rm weak}_{\rm pNRQCD} = \mathcal{L}^{S,O} + \mathcal{L}^{\ell}, \tag{6.10}$$

with

$$\begin{aligned}\mathcal{L}^{S,O} = &\int d^3r\, \mathrm{Tr}\left\{\mathrm{S}^\dagger (i\partial_0 - h_s)\mathrm{S} + \mathrm{O}^\dagger (iD_0 - h_o)\mathrm{O}\right\} \\ &- V_A \mathrm{Tr}\left\{\mathrm{O}^\dagger \boldsymbol{r}\cdot g\boldsymbol{E}\,\mathrm{S} + \mathrm{S}^\dagger \boldsymbol{r}\cdot g\boldsymbol{E}\,\mathrm{O}\right\} \\ &- \frac{V_B}{2}\mathrm{Tr}\left\{\mathrm{O}^\dagger \boldsymbol{r}\cdot g\boldsymbol{E}\,\mathrm{O} + \mathrm{O}^\dagger \mathrm{O}\boldsymbol{r}\cdot g\boldsymbol{E}\right\},\end{aligned} \tag{6.11}$$

where, up to order $1/m_h^2$, and including the $1/m_h^3$ terms in the kinetic energies,

$$h_s = \frac{\boldsymbol{p}^2}{m_h} + \frac{\boldsymbol{P}^2}{4m_h} - \frac{\boldsymbol{p}^4}{4m_h^3} + \cdots + V_s, \tag{6.12}$$

$$h_o = \frac{\boldsymbol{p}^2}{m_h} + \frac{\boldsymbol{P}^2}{4m_h} - \frac{\boldsymbol{p}^4}{4m_h^3} + \cdots + V_o. \tag{6.13}$$

The covariant derivative acting on the octet field is defined as $iD_0\mathrm{O} = i\partial_0\mathrm{O} + g[A_0(\boldsymbol{R},t),\mathrm{O}]$, $\boldsymbol{P} = -i\boldsymbol{D}_{\boldsymbol{R}}$ is the (gauge covariant) center of mass momentum, $\boldsymbol{p} = -i\boldsymbol{\nabla}_{\boldsymbol{r}}$ is the relative momentum, and h_s and h_o may be interpreted as the Hamiltonians of the color singlet and color octet heavy quark–antiquark fields. The dots in Eqs. (6.12) and (6.13) stand for higher-order terms in the nonrelativistic expansion of the kinetic energy. The trace in Eq. (6.11) is in spin and in color space. The Lagrangian $\mathcal{L}^{\ell}$ is the Lagrangian of the light degrees of freedom (6.3) inherited from NRQCD.

The quantities V_s, V_o, V_A, and V_B are Wilson coefficients of pNRQCD. They encode contributions from the soft degrees of freedom that have been integrated out from NRQCD. Because (under the hierarchy of weakly-coupled pNRQCD) the soft scale, $m_h v$, is larger than $\Lambda_{\rm QCD}$, the Wilson coefficients may be computed in perturbation theory, order by order in α_s. They are, in general, functions of $\boldsymbol{r}$, as well as of the spin and momentum. At leading order, V_A and V_B are 1; they get possible corrections at order α_s^2 [1491]. The coefficients V_s and V_o may be identified with the color singlet and octet potentials, respectively. To leading order $V_s^{(0)} = -4\alpha_s/(3r)$ and $V_o^{(0)} = \alpha_s/(6r)$, which are the Coulomb potentials in the color SU(3) fundamental and adjoint representation, respectively. The potentials V_s and V_o contain, however, also momentum- and spin-dependent corrections. For the singlet case (the octet case is analogous) we can write, up to order $1/m_h^2$:

$$V_s = V_s^{(0)}(r) + \frac{V_s^{(1)}(r)}{m_h} + \frac{V_{\rm SI}^{(2)}}{m_h^2} + \frac{V_{\rm SD}^{(2)}}{m_h^2}, \tag{6.14}$$

where, at order $1/m_h^2$ we have distinguished between spin-independent (SI) and spin-dependent (SD) terms. In turn, they can be organized as

$$\begin{aligned}V_{\rm SI}^{(2)} = &V_r^{(2)}(r) + \frac{1}{4}V_{p^2,\rm CM}^{(2)}(r)\boldsymbol{P}^2 + \frac{1}{4}\frac{V_{L^2,\rm CM}^{(2)}(r)}{r^2}(\boldsymbol{r}\times\boldsymbol{P})^2 \\ &+ \frac{1}{2}\left\{V_{p^2}^{(2)}(r), \boldsymbol{p}^2\right\} + \frac{V_{L^2}^{(2)}(r)}{r^2}\boldsymbol{L}^2,\end{aligned} \tag{6.15}$$

$$\begin{aligned}V_{\rm SD}^{(2)} = &\frac{1}{2}V_{LS,\rm CM}^{(2)}(r)\,(\boldsymbol{r}\times\boldsymbol{P})\cdot(\boldsymbol{S}_1 - \boldsymbol{S}_2) + V_{LS}^{(2)}(r)\,\boldsymbol{L}\cdot\boldsymbol{S} \\ &+ V_{S^2}^{(2)}(r)\,\boldsymbol{S}^2 + V_{S_{12}}^{(2)}(r)\,S_{12},\end{aligned} \tag{6.16}$$

where $\{\ ,\ \}$ stands for the anticommutator,

$$\boldsymbol{S} = \boldsymbol{S}_1 + \boldsymbol{S}_2 = (\boldsymbol{\sigma}_1 + \boldsymbol{\sigma}_2)/2$$

is the total spin ($\boldsymbol{S}_i = \boldsymbol{\sigma}_i/2$ is the spin of the particle i), $\boldsymbol{L} = \boldsymbol{r}\times\boldsymbol{p}$ is the relative orbital angular momentum, and

$$S_{12} = 3(\hat{\boldsymbol{r}}\cdot\boldsymbol{\sigma}_1)(\hat{\boldsymbol{r}}\cdot\boldsymbol{\sigma}_2) - \boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2.$$

The potential $V_s^{(0)}$ is the static potential, the potential proportional to $V_{LS}^{(2)}$ may be identified with the spin–orbit potential, the potential proportional to $V_{S^2}^{(2)}$ with the spin–spin potential and the potential proportional to $V_{S_{12}}^{(2)}$ with the spin tensor potential. The potentials that contribute in the center of mass reference frame are, at leading (non-vanishing) order in perturbation theory (see, e.g., Ref. [1462]):

$$V_s^{(1)}(r) = -\frac{2\alpha_s^2}{r^2}, \tag{6.17}$$

$$V_r^{(2)}(r) = \frac{4}{3}\pi\alpha_s\delta^{(3)}(\boldsymbol{r}), \quad V_{p^2}^{(2)}(r) = -\frac{4\alpha_s}{3r}, \tag{6.18}$$

$$V_{L^2}^{(2)}(r) = \frac{2\alpha_s}{3r}, \quad V_{LS}^{(2)}(r) = \frac{2\alpha_s}{r^3}, \tag{6.19}$$

$$V_{S^2}^{(2)}(r) = \frac{16\pi\alpha_s}{9}\delta^{(3)}(\boldsymbol{r}), \quad V_{S_{12}}^{(2)}(r) = \frac{\alpha_s}{3r^3}. \tag{6.20}$$

Potentials that depend on the center of mass momentum are relevant only if the quarkonium is recoiling, like in hadronic and electromagnetic transitions.

Beyond leading order, the static potential is known up to three-loop accuracy [1492–1494], and also subleading logarithms showing up at four loops have been computed [1495]; the $1/m_h$ potential is known up to order α_s^3 [1496], and $1/m_h^2$ potentials up to order α_s^2 (these potentials have a long history, see Ref. [1497] and references therein). We have assumed that the heavy quark and antiquark have equal masses; for the case of a quark and an antiquark of different masses, we refer, for instance, to Refs. [1438,1462,1498–1500].

The Wilson coefficients of pNRQCD inherit the Wilson coefficients of NRQCD. Hence, some of the couplings appearing in the expansion of the Wilson coefficients are naturally computed at the scale of NRQCD, m_h, while others, encoding the soft contributions, are naturally computed at the soft scale, $m_h v$. In weakly-coupled pNRQCD, because the leading potential is the Coulomb potential, the Bohr radius is proportional to $1/(m_h\alpha_s)$ and $v \sim \alpha_s$. Finally, like in any non relativistic EFT, also the Wilson coefficients of pNRQCD are related through constraints imposed by the Poincaré invariance of QCD, as we have seen at work in HQET and NRQCD. These constraints set the coefficients of the kinetic terms appearing in Eqs. (6.12) and (6.13) to be the ones coming from expanding the relativistic kinetic energies of a free quark and antiquark. Furthermore they fix the potentials depending on the center of mass momentum by expressing them in terms of the static potential,

$$V_{LS,\mathrm{CM}} = -\frac{1}{2r}\frac{dV_s^{(0)}}{dr}, \quad V_{L^2,\mathrm{CM}} = -\frac{r}{2}\frac{dV_s^{(0)}}{dr},$$

$$V_{p^2,\mathrm{CM}} = \frac{r}{2}\frac{dV_s^{(0)}}{dr} - \frac{1}{2}V_s^{(0)}.$$

These and other constraints have been derived in Refs. [1438, 1487,1499,1501,1502]. These relations are exact, i.e., valid at any order in perturbation theory and, when applicable, also nonperturbatively.

In the pNRQCD Lagrangian the relative coordinate $\boldsymbol{r}$ plays the role of a continuous parameter labeling different fields. The dynamical spacetime coordinates of the Lagrangian density are the time t and the coordinate $\boldsymbol{R}$, which, in the case of the fields S and O, is the center of mass coordinate. Having written the Lagrangian in terms of singlet and octet fields has made each term in Eq. (6.11) explicitly gauge invariant.

The power counting of weakly-coupled pNRQCD is straightforward. We have already found that $r \sim 1/(m_h v)$ and t, $R \sim 1/(m_h v^2)$. Momenta scale like $p \sim m_h v$ and $P \sim m_h v^2$. Gluon fields and light quark fields are ultrasoft and scale like $m_h v^2$ or Λ_{QCD} to their dimension. The leading-order singlet Hamiltonian, $\boldsymbol{p}^2/m_h + V_s^{(0)}$, scales like $m_h v^2$ (and analogously in the octet case), which is the order of the Bohr levels. The potentials listed in Eqs. (6.17)–(6.20) contribute to V_s at order $m_h v^4$, as $\alpha_s \sim v$.

The first correction to a pure potential picture of the quarkonium interaction comes from the chromoelectric dipole interaction terms in the second line of Eq. (6.11). These operators are of order $g(m_h v^2)^2/(m_h v) \sim g m_h v^3$. In order to project on color singlet states, the chromoelectric dipole interaction may enter only in loop diagrams, which at leading order is a self-energy diagram with two chromoelectric dipole vertices. Such a self-energy diagram is of order $g^2(m_h v^2)^3/(m_h v)^2 \sim g^2 m_h v^4$. The coupling g^2 is computed at the ultrasoft scale. Hence, if $\Lambda_{\mathrm{QCD}} \ll m_h v^2$, the coupling is perturbative and the self-energy diagram with two chromoelectric dipole vertices is suppressed with respect to the contributions coming from the potentials in Eqs. (6.17)–(6.20). Elsewhere, if $\Lambda_{\mathrm{QCD}} \sim m_h v^2$, it is of the same order.

At leading order in the multipole expansion, the equation of motion for the singlet field is

$$i\partial_0 \mathrm{S} = h_s \mathrm{S}, \tag{6.21}$$

which is the Schrödinger equation that in quantum mechanics describes the evolution of a nonrelativistic bound state. Potential NRQCD provides therefore a field theoretical foundation of the Schrödinger equation, a rigorous QCD definition and derivation of its potential, and the range of validity of the quantum mechanical picture. Ultrasoft gluons start contributing, and therefore correcting the potential picture, at order $m_h v^4$ (for $\Lambda_{\mathrm{QCD}} \sim m_h v^2$) or $m_h v^5$ (for $\Lambda_{\mathrm{QCD}} \ll m_h v^2$) in the spectrum.

Not only the static potential is derived from first principles in pNRQCD, but all higher-order corrections in the nonrelativistic expansion, including the spin–orbit, spin–spin and Darwin term as well. In general, the potentials factorize soft contributions from radiative corrections at the scale m_h. These last ones are encoded in the matching coefficients inherited from NRQCD, e.g., all $1/m_h^2$ spin-dependent

potentials contain the Fermi coefficient c_F. Since the potentials are Wilson coefficients of an EFT, they are regularized, undergo renormalization and satisfy renormalization group equations that allow to resum potentially large logarithms in their expressions [1491,1492,1503–1509]. The scale dependence of the Wilson coefficients cancels in physical observables. For instance, the QCD static potential is infrared sensitive at three loops [1510], a sensitivity that stems from the fact that a static quark–antiquark pair may change its color status by emitting an ultrasoft gluon. The infrared sensitivity of the static potential cancels in the computation of the static energy against the self-energy diagram with two chromoelectric dipole vertices considered above, in a sort of non-Abelian Lamb shift mechanism [1492].

Weakly-coupled pNRQCD requires the fulfillment of the condition $\Lambda_{\rm QCD} \lesssim m_h v^2$. The condition $\Lambda_{\rm QCD} \ll m_t v^2$ is certainly fulfilled by top–antitop quark pairs near threshold. The production of $t\bar{t}$ pairs near threshold is expected to be measured with precision at future linear colliders, providing, among others, a determination of the top mass with an uncertainty well below 100 MeV, which is a crucial input to test the Standard Model. This requires a comparable theoretical accuracy, which has led in the last decades to several high-order calculations of the near threshold cross section in the framework of nonrelativistic EFTs of QCD [1511–1517]. The condition $\Lambda_{\rm QCD} \lesssim m_h v^2$ is also fulfilled by compact and Coulombic quarkonia. Examples are the bottomonium ground state, the ground state of the B_c system, and, to a somewhat lesser extent, the charmonium ground state, and the first bottomonium excited states. We recall that in a Coulombic system the size is proportional to the inverse of the mass and to the principal quantum number. A review on applications of weakly-coupled pNRQCD to several tightly bound quarkonia can be found in Ref. [1518].

Weak-coupling determinations of the bottomonium ground state masses are typically used to extract the charm and bottom masses [1446,1519–1527]. Hence, they provide alternative observables for the extraction of the heavy quark masses to the heavy-light meson masses discussed in Sect. 6.1.1. The results are consistent with the ones presented in Sect. 6.1.1. The present precision is N^{3}LO; the determination of the bottom mass includes the effects due to the charm mass at two loops. Once the heavy quark masses have been established for one set of spectroscopy observables, they can be used for others like the B_c mass or the B_c spectrum (see Ref. [1498] for an early work and Ref. [1500] for a state of the art calculation at N^{3}LO). Fine and hyperfine splittings of charmonium and bottomonium have been computed perturbatively in Refs. [1528,1529] and to NLL accuracy in Ref. [1530], similarly for the B_c^*-B_c hyperfine splitting in Ref. [1531]. After more than one decade of work the whole perturbative heavy quarkonium spectrum has been computed at N^{3}LO [1504,1532–1537]. Recently, this result has been further improved reaching N^{3}LL accuracy up to a missing contribution of the two-loop soft running [1508,1509]. The N^{3}LL order represents the presently achievable precision of these calculations. Going beyond this precision will require a major computational effort, like the four-loop determination of the static potential. Electromagnetic decays of the bottomonium lowest levels have been computed including N^{2}LL corrections in Refs. [1513,1538]. A different power counting that includes at leading order the exact static potential has been used for these quantities in Ref. [1539]. Corrections to the wave function and leptonic decay width of the $\Upsilon(1S)$ at N^{3}LO have been computed in Refs. [1540,1541]. Nonperturbative corrections in the form of condensates have been included in Refs. [1542,1543]. Radiative quarkonium decays have been analyzed in Refs. [1544–1549]. Radiative and inclusive decays of the $\Upsilon(1S)$ may also serve as a determination of α_s at the bottom mass scale [1550]. Radiative transitions, M1 and E1, at relative order v^2 in the velocity expansion have been computed in Refs. [1551–1554]; noteworthily, pNRQCD may explain the observed tiny $\Upsilon(2S) \to \gamma\, \eta_b(1S)$ branching fraction. Finally, the photon line shape in the radiative transition $J/\psi \to \gamma\, \eta_c(1S)$ has been studied in Ref. [1555].

Strongly-coupled pNRQCD

When the hierarchy of scales is $\Lambda_{\rm QCD} \gg m_h v^2$, pNRQCD is a strongly-coupled theory. This condition may be appropriate to describe higher quarkonium states, and quarkonium exotica. Strongly-coupled pNRQCD is obtained by integrating out the hadronic scale $\Lambda_{\rm QCD}$, which means that all colored degrees of freedom are absent [1462,1488,1556–1559].

Let us consider the case of strongly-coupled pNRQCD for ordinary quarkonia. Lattice QCD shows evidence that the quarkonium static energy is separated by a gap of order $\Lambda_{\rm QCD}$ from the energies of the gluonic excitations between the static quark–antiquark pair [1560]. If, in addition, the binding energies of the states that can be constructed out of the quarkonium static energy are also separated by a gap of order $\Lambda_{\rm QCD}$ from the binding energies of the states that can be constructed out of the static energies of the gluonic excitations, and from open-flavor states, then one can integrate out all these latter states. The resulting EFT is made of a quark–antiquark color singlet field, whose modes are the quarkonium states, and light hadrons. The coupling of quarkonia with light hadrons has been considered in the framework of pNRQCD in Ref. [1561]. It impacts very mildly spectral properties (masses, widths) of quarkonia that lie well below the open-flavor threshold. For such quarkonia we may neglect their couplings with light hadrons and the pNRQCD Lagrangian density assumes the particularly simple form:

$$\mathcal{L}^{\rm strong}_{\rm pNRQCD} = \int d^3r\, {\rm Tr}\left\{ {\rm S}^\dagger \left(i\partial_0 - h_s \right) {\rm S} \right\}. \tag{6.22}$$

The Hamiltonian, h_s, has the same form as in Eqs. (6.12) and (6.14)–(6.16). The equation of motion is the Schrödinger equation (6.21).

The nonperturbative dynamics is encoded in the potentials, which at order $1/m_h$ is $V_s^{(1)}$ and at order $1/m_h^2$ are the spin-independent and spin-dependent terms identified in Eqs. (6.15) and (6.16). What distinguishes the EFT from phenomenological potential models can be summarized as follows:

(i) The potentials are products of Wilson coefficients, factorizing contributions from the high-energy scale, m_h, and low-energy matrix elements, encoding contributions coming from the scales $m_h v$ and $\Lambda_{\rm QCD}$. The exact expressions follow from matching pNRQCD with its high-energy completion, which is NRQCD.

(ii) The high-energy Wilson coefficients of pNRQCD are inherited from NRQCD. These are the Wilson coefficients in the NRQCD Lagrangian (6.7). Because the NRQCD Wilson coefficients have a real and an imaginary part, also the pNRQCD potentials develop a real part, responsible for the quarkonium binding, and an imaginary part, responsible for the quarkonium annihilation. At higher orders, also contributions coming from the scale $\sqrt{m_h \Lambda_{\rm QCD}}$ may become relevant [1559].

(iii) The low-energy matrix elements are nonperturbative. Their field-theoretical expressions, relevant for potentials up to order $1/m_h^2$, are known.

The static potential is equal to $\lim_{T\to\infty} i \ln W/T$, where W is the expectation value of a rectangular Wilson loop of spatial extension r and temporal extension T [97,1562–1564]. Similarly, the low-energy real parts of the other potentials can be expressed in terms of Wilson loops and field insertions on them [803,1556,1557]. These Wilson loops may be computed in weakly-coupled QCD giving back the weak-coupling potentials listed at leading order in Eqs. (6.17)–(6.20) [1565] or nonperturbatively via lattice QCD. Indeed, the computation of these potentials has a long history that begins with the inception of lattice QCD. Their most recent determinations can be found in Refs. [804,1566,1567], see also Ref. [1568]. Noteworthily the long-distance behaviour of the potentials agrees with the expectations of the effective string theory [1565,1569–1571].

The low-energy contributions to the imaginary parts of the potential are matrix elements of the NRQCD four-fermion operators. Hence they are local terms proportional to $\delta^3(\boldsymbol{r})$ or derivatives of it. Nonperturbative contributions are encoded into constants that may be expressed in terms of momenta of correlators of chromoelectric and/or chromomagnetic fields [1488,1558], and eventually fixed on data or computed with lattice QCD.

(iv) Finally, pNRQCD is renormalizable order by order in the expansion parameters in both its weak-coupling and strong-coupling versions. In particular, quantum-mechanical perturbation theory can be implemented at any order without incurring uncanceled divergences like in purely phenomenological potential models.

Strongly-coupled pNRQCD has been used to compute quarkonium inclusive and electromagnetic decay widths [1463,1488,1558,1559], and hadronic and electromagnetic production cross sections [1572–1576]. The advantage with respect to the NRQCD approach is that, while the NRQCD four-fermion matrix elements depend on the quarkonium state, their pNRQCD expression factorizes all the quarkonium dependence into the wave function at the origin (or its derivatives) squared. The wave function at the origin squared gets multiplied by momenta of correlators of field-strength tensors, F, that are universal, quarkonium independent, constants. Schematically, one obtains for the expression of a generic NRQCD four-fermion matrix element in pNRQCD, entering either quarkonium production or decay, that

$$\langle \text{4-fermion} \rangle \sim |\text{wave-function}(0)|^2 \times \int dt \cdots \langle F(t) \cdots F(0) \rangle . \tag{6.23}$$

This leads to a significant reduction in the number of nonperturbative parameters and allows to use information gained in the charmonium sector to make predictions in the bottomonium sector. Finally, pNRQCD combined with the multipole expansion has been used to compute quarkonium hadronic transitions in Ref. [1577].

pNRQCD for systems other than quarkonia

Potential NRQCD can be used to describe systems with three valence quarks, two of them heavy [1451,1452,1458,1459,1578,1579]. This is because the nonrelativistic hierarchy of scales given in Eq. (6.6) is preserved, which allows to systematically integrate out these scales to describe eventually the baryon with a suitable EFT. If the heavy quark–quark distance is of the order of $1/\Lambda_{\rm QCD}$, then the valence light-quark affects the quark–quark potential. Elsewhere, if the heavy quark–quark distance is smaller than $1/\Lambda_{\rm QCD}$, then we may disentangle the quark–quark dynamics, described by a perturbative quark–quark potential, from the coupling of the heavy-quark pair with the light quark. Since in this last case, the light quark sees the heavy-quark pair as a pointlike particle, its interaction with the heavy-quark pair is described by HQET.

One can devise EFTs for describing low-energy modes of baryons made of three heavy quarks. These states have not been discovered yet in experiments, but they offer a unique tool to study confinement and the transition region from the Coulombic regime to the confined one in non-trivial geometrical settings [1578], using, for instance the three-quark static potential computed in lattice QCD [1580–1582].

Possible bound states made of two quarkonia or of a quarkonium and a nucleon (hadroquarkonium) may be characterized by even lower energy scales than those characterizing the binding in quarkonia or baryons made of at least two heavy quarks. These lower energy scales are those associated with pion exchanges responsible for the long-range interaction. One can treat these systems in an EFT framework by starting from the pNRQCD description of the quarkonium and the heavy-baryon chiral effective theory description of the nucleon. The long-range pion exchange interaction sets the scale of the typical size of the system to be of the order of $1/M_\pi$, i.e., much larger than the size of the quarkonium and even larger than its typical time scale, which is of the order of the inverse of the binding energy.

Once modes associated with the quarkonium binding energy and M_π have been integrated out, the quarkonium–quarkonium or the quarkonium–nucleon interaction is described by a potential that, in this way, has been systematically computed from QCD. The coupling of quarkonium with the pions is encoded in a Wilson coefficient that may be identified with the quarkonium chromoelectric polarizability [1583]. In the quarkonium–quarkonium system, the lowest energy EFT describing modes of energy and momentum of order $M_\pi^2/(2m_h)$ is called van der Waals EFT (WEFT) [1561,1584]. The resulting potential is, in fact, the van der Waals potential. In the quarkonium–nucleon system, the lowest energy EFT describing modes of energy and momentum of order $M_\pi^2/(2\Lambda_\chi)$ has been dubbed potential quarkonium–nucleon EFT (pQNEFT) [1585]. Such frameworks may be relevant to describe heavy pentaquarks.

Quarkonium-like multiparticle systems, where the light degrees of freedom remain adiabatically in a stationary state with respect to the heavy quark motion, can be studied within the Born–Oppenheimer approximation that may be implemented in a suitable version of pNRQCD called Born–Oppenheimer effective field theory (BOEFT) [414,1586,1587]. This framework has been applied to quarkonium hybrids, quarkonium tetraquarks and to heavy-meson threshold effects [1461,1588–1591]. Finally, nonrelativistic EFTs like pNRQCD are also advantageous in describing the propagation of quarkonium in a thermal medium either in equilibrium [1592–1597] or out of equilibrium [1598–1602]; see also Sect. 6.6.

6.2 Chiral perturbation theory

Stefan Scherer and Matthias Schindler

Chiral perturbation theory (ChPT) is an effective field theory that describes the properties of strongly-interacting systems at energies far below typical hadron masses. The degrees of freedom are hadrons instead of the underlying quarks and gluons. ChPT is a systematic and model-independent approximation method based on an expansion of amplitudes in terms of light-quark masses and momenta. The following is a brief overview of ChPT that is largely based on Ref. [1603], which can be referred to for a more detailed introduction.

6.2.1 QCD and chiral symmetry

The QCD Lagrangian, obtained by applying the gauge principle with respect to the SU(3) color group to the free Lagrangians of six quark flavors with masses m_f, reads

$$\mathcal{L}_{\rm QCD} = \sum_{f=u,\ldots,t} \bar{q}_f \left(i\not{\mathbf{D}} - m_f\right) q_f - \frac{1}{2}{\rm Tr}_c\left(\mathbf{F}_{\mu\nu}\mathbf{F}^{\mu\nu}\right). \tag{6.24}$$

For each quark flavor f, the quark field q_f is a color triplet, transforming in the triplet representation,

$$q_f(x) \mapsto U(x) q_f(x), \tag{6.25}$$

where $U(x)$ denotes a smooth space-time-dependent SU(3) matrix. Using the Gell-Mann matrices [1604], the eight gluon fields $\mathcal{A}_\mu^A$ are collected in a traceless, Hermitian, 3×3 matrix $\mathbf{A}_\mu = \lambda^A \mathcal{A}_\mu^A/2$ (summation over repeated indices implied), transforming inhomogeneously under a gauge transformation,

$$\mathbf{A}_\mu(x) \mapsto U(x)\mathbf{A}_\mu(x)U^\dagger(x) + \frac{i}{g_s}\partial_\mu U(x) U^\dagger(x), \tag{6.26}$$

where g_s denotes the SU(3) gauge coupling constant. In terms of $\mathbf{A}_\mu$, the covariant derivative of the quark fields is defined as

$$\mathbf{D}_\mu q_f = \left(\partial_\mu + i g_s \mathbf{A}_\mu\right) q_f. \tag{6.27}$$

Finally, the field strength tensor is given by

$$\mathbf{F}_{\mu\nu} = \partial_\mu \mathbf{A}_\nu - \partial_\nu \mathbf{A}_\mu + i g_s [\mathbf{A}_\mu, \mathbf{A}_\nu]. \tag{6.28}$$

By construction, the Lagrangian of Eq. (6.24) is invariant under the combined transformations of Eqs. (6.25) and (6.26). From the point of view of gauge invariance, the strong-interaction Lagrangian could also involve a term of the type (c.f. Eq. (1.6) from Sect. 1.1)

$$\mathcal{L}_\theta = \frac{g_s^2\bar{\theta}}{32\pi^2}\epsilon_{\mu\nu\rho\sigma}{\rm Tr}_c\left(\mathbf{F}^{\mu\nu}\mathbf{F}^{\rho\sigma}\right), \quad \epsilon_{0123} = 1, \tag{6.29}$$

where $\epsilon_{\mu\nu\rho\sigma}$ denotes the totally antisymmetric Levi-Civita tensor. The so-called θ term of Eq. (6.29) implies an explicit P and CP violation of the strong interactions. The present empirical information on the neutron electric dipole moment [1605] indicates that the θ term is small and, in the following, we will omit Eq. (6.29) from our discussion.

Since the covariant derivative of the quark fields is flavor independent, the Lagrangian of Eq. (6.24) has additional, accidental, and in this case global, symmetries aside from the gauge symmetry. Both the dynamics of the theory (via spontaneous symmetry breaking) and the values of the quark

masses impact how these symmetries are (approximately) realized in nature. Dynamical chiral symmetry breaking introduces the scale $\Lambda_\chi = 4\pi F_0$ (see below) of the order of 1 GeV [1606]. In this context it is common to divide the six quark flavors into the three light quarks u, d, and s with $m_l < \Lambda_\chi$ and the three heavy flavors c, b, and t with $m_h > \Lambda_\chi$ (as discussed in the previous subsection). As a theoretical starting point, one may consider two limits, namely, sending the light-quark masses to zero (chiral limit) and the heavy-quark masses to infinity. In Ref. [1607], this situation is referred to as a "theoretician's paradise." In the following, we exclusively concentrate on the chiral limit for either two (u, d) or three (u, d, s) light quarks and omit the heavy quarks from our discussion. Introducing left-handed and right-handed quark fields (color and flavor indices omitted) as

$$q_L = \frac{1}{2}(\mathbb{1} - \gamma_5)\, q, \quad q_R = \frac{1}{2}(\mathbb{1} + \gamma_5)\, q, \quad \gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3, \tag{6.30}$$

the QCD Lagrangian in the chiral limit decomposes into

$$\mathcal{L}^0_{\text{QCD}} = \sum_{l=u,d,s} \left(\bar{q}_{L,l} i \not{D} q_{L,l} + \bar{q}_{R,l} i \not{D} q_{L,R}\right) - \frac{1}{2}\text{Tr}_c\left(\mathbf{F}_{\mu\nu}\mathbf{F}^{\mu\nu}\right). \tag{6.31}$$

In the massless limit, the helicity of a quark is a good quantum number which is conserved in the interaction with gluons. Moreover, the classical Lagrangian in the chiral limit has a global $\text{U}(3)_L \times \text{U}(3)_R$ symmetry, i.e., it is invariant under independent unitary flavor transformations of the left-handed and the right-handed quark fields,

$$q_L \mapsto U_L q_L, \quad q_R \mapsto U_R q_R.$$

At the classical level, this chiral symmetry results in $2 \times (8 + 1) = 18$ conserved currents:

$$L^\mu_a = \bar{q}_L \gamma^\mu \frac{\lambda_a}{2} q_L, \quad R^\mu_a = \bar{q}_R \gamma^\mu \frac{\lambda_a}{2} q_R, \quad a = 1, \ldots, 8,$$
$$V^\mu = \bar{q}_R \gamma^\mu q_R + \bar{q}_L \gamma^\mu q_L, \quad A^\mu = \bar{q}_R \gamma^\mu q_R - \bar{q}_L \gamma^\mu q_L.$$

Here, the Gell-Mann matrices act in flavor space, since q_R and q_L are flavor triplets.[64] Because of quantum effects the singlet axial-vector current $A^\mu = \bar{q}\gamma^\mu\gamma_5 q$ develops a so-called anomaly, resulting in the divergence equation

$$\partial_\mu A^\mu = \frac{3g_s^2}{16\pi^2}\epsilon_{\mu\nu\rho\sigma}\text{Tr}_c\left(\mathbf{F}^{\mu\nu}\mathbf{F}^{\rho\sigma}\right). \tag{6.32}$$

The factor of three originates from the number of flavors. In the large N_c (number of colors) limit of Ref. [1162] the singlet axial-vector current is conserved, because the strong coupling constant behaves as $g_s^2 \sim N_c^{-1}$.

In the quantized theory, the spatial integrals over the charge densities of the symmetry currents give rise to the charge operators Q_{La}, Q_{Ra} $(a = 1, \ldots, 8)$, and Q_V. They are generators of the group $\text{SU}(3)_L \times \text{SU}(3)_R \times \text{U}(1)_V$, acting on the Hilbert space of QCD, and satisfy the commutation relations

$$[Q_{La}, Q_{Lb}] = i f_{abc} Q_{Lc}, \tag{6.33a}$$
$$[Q_{Ra}, Q_{Rb}] = i f_{abc} Q_{Rc}, \tag{6.33b}$$
$$[Q_{La}, Q_{Rb}] = 0, \tag{6.33c}$$
$$[Q_{La}, Q_V] = [Q_{Ra}, Q_V] = 0, \tag{6.33d}$$

where the f_{abc} are the totally antisymmetric structure constants of the Lie algebra of SU(3) [1604]. In the chiral limit, these operators are time independent, i.e., they commute with the Hamiltonian in the chiral limit,

$$[Q_{La}, H^0_{\text{QCD}}] = [Q_{Ra}, H^0_{\text{QCD}}] = [Q_V, H^0_{\text{QCD}}] = 0. \tag{6.34}$$

It is convenient to consider the linear combinations $Q_{Aa} \equiv Q_{Ra} - Q_{La}$ and $Q_{Va} \equiv Q_{Ra} + Q_{La}$, which transform as $Q_{Aa} \mapsto -Q_{Aa}$ and $Q_{Va} \mapsto Q_{Va}$ under parity. The hadron spectrum can be organized in multiplets belonging to irreducible representations of $\text{SU}(3)_V$ with a given baryon number. If not only the vector subgroup, but the full group were realized linearly by the spectrum of the hadrons, one would expect a so-called parity doubling of mass-degenerate states. The absence of such a doubling in the low-energy spectrum is an indication that the $\text{SU}(3)_L \times \text{SU}(3)_R$ chiral symmetry is dynamically broken in the ground state. One then assumes that the axial generators Q_{Aa} do not annihilate the ground state of QCD,

$$Q_{Aa}|0\rangle \neq 0. \tag{6.35}$$

As a consequence of the Goldstone theorem [12], each axial generator Q_{Aa} not annihilating the ground state corresponds to a *massless* Goldstone-boson field ϕ_a with spin 0, whose symmetry properties are tightly connected to the generator in question. The Goldstone bosons have the same transformation behavior under parity as the axial generators,

$$\phi_a(t, \vec{x}) \overset{P}{\mapsto} -\phi_a(t, -\vec{x}), \tag{6.36}$$

i.e., they are pseudoscalars. From Eqs. (6.33a) and (6.33b) one obtains $[Q_{Va}, Q_{Ab}] = i f_{abc} Q_{Ac}$ and thus the Goldstone bosons transform under the subgroup $\text{SU}(3)_V$, which leaves the vacuum invariant, as an octet:

$$[Q_{Va}, \phi_b(x)] = i f_{abc} \phi_c(x). \tag{6.37}$$

The members of the pseudoscalar octet (π, K, η) of the real world are identified as the Goldstone bosons of QCD and would be massless for massless quarks.

[64] Lower case Roman letters denote SU(3) flavor indices.

After turning on the quark masses in terms of the mass term

$$\mathcal{L}_{\mathcal{M}} = -\bar{q}\mathcal{M}q = -\left(\bar{q}_R\mathcal{M}q_L + \bar{q}_L\mathcal{M}^\dagger q_R\right),$$
$$\mathcal{M} = \mathrm{diag}(m_u, m_d, m_s),$$

the Goldstone bosons will no longer be massless (see below). Moreover, the symmetry currents are no longer conserved. In terms of the vector currents $V_a^\mu = R_a^\mu - L_a^\mu$ and the axial-vector currents $A_a^\mu = R_a^\mu - L_a^\mu$, the corresponding divergences read

$$\partial_\mu V_a^\mu = i\bar{q}\left[\mathcal{M}, \frac{\lambda_a}{2}\right]q, \quad \partial_\mu A_a^\mu = i\bar{q}\gamma_5\left\{\frac{\lambda_a}{2}, \mathcal{M}\right\}q. \tag{6.38}$$

The properties of the currents corresponding to the approximate chiral symmetry of QCD can be summarized as follows:

1. In the limit of massless quarks, the sixteen currents L_a^μ and R_a^μ or, alternatively, $V_a^\mu = R_a^\mu + L_a^\mu$ and $A_a^\mu = R_a^\mu - L_a^\mu$ are conserved. The same is true for the singlet vector current V^μ, whereas the singlet axial-vector current A^μ has an anomaly (see Eq. (6.32)).
2. For any values of quark masses, the individual flavor currents $\bar{u}\gamma^\mu u$, $\bar{d}\gamma^\mu d$, and $\bar{s}\gamma^\mu s$ are always conserved in the strong interactions reflecting the flavor independence of the strong coupling and the diagonal form of the quark–mass matrix. Of course, the singlet vector current V^μ, being the sum of the three flavor currents, is always conserved.
3. In addition to the anomaly, the singlet axial-vector current has an explicit divergence due to the quark masses:

$$\partial_\mu A^\mu = 2i\bar{q}\gamma_5\mathcal{M}q + \frac{3g_s^2}{16\pi^2}\epsilon_{\mu\nu\rho\sigma}\mathrm{Tr}_c\left(\mathbf{F}^{\mu\nu}\mathbf{F}^{\rho\sigma}\right).$$

4. For equal quark masses, $m_u = m_d = m_s$, the eight vector currents V_a^μ are conserved, because $[\lambda_a, \mathbb{1}] = 0$. Such a scenario is the origin of the SU(3) symmetry originally proposed by Gell-Mann and Ne'eman [1608]. The eight axial-vector currents A_a^μ are not conserved. The divergences of the octet axial-vector currents of Eq. (6.38) are proportional to pseudoscalar quadratic forms. This can be interpreted as the microscopic origin of the PCAC relation (partially conserved axial-vector current) [19,1609] which states that the divergences of the axial-vector currents are proportional to renormalized field operators representing the lowest-lying pseudoscalar octet.
5. Taking $m_u = m_d \neq m_s$ reduces SU(3) flavor symmetry to SU(2) isospin symmetry.
6. Taking $m_u \neq m_d$ leads to isospin-symmetry breaking.

Besides the conservation properties of the currents, one may also calculate their commutators (current algebra), which may then be used to derive certain relations among QCD Green functions analogous to the Ward identities of Quantum Electrodynamics. The set of all QCD Green functions involving color-neutral quark bilinears is very efficiently collected in a generating functional,

$$\exp\left(i\,Z_{\mathrm{QCD}}[v, a, s, p]\right) = \langle 0|T\exp\left[i\int d^4x\,\mathcal{L}_{\mathrm{ext}}(x)\right]|0\rangle_0, \tag{6.39}$$

where [69,1610]:

$$\begin{aligned}\mathcal{L}_{\mathrm{ext}} &= \sum_{a=1}^{8} v_a^\mu\,\bar{q}\gamma_\mu\frac{\lambda_a}{2}q + v_{(s)}^\mu\,\frac{1}{3}\bar{q}\gamma_\mu q + \sum_{a=1}^{8} a_a^\mu\,\bar{q}\gamma_\mu\gamma_5\frac{\lambda_a}{2}q \\ &\quad - \sum_{a=0}^{8} s_a\,\bar{q}\lambda_a q + \sum_{a=0}^{8} p_a\,i\bar{q}\gamma_5\lambda_a q \\ &= \bar{q}\gamma_\mu\left(v^\mu + \frac{1}{3}v_{(s)}^\mu + \gamma_5 a^\mu\right)q - \bar{q}(s - i\gamma_5 p)q,\end{aligned} \tag{6.40}$$

where $\lambda_0 = \sqrt{\frac{2}{3}}\mathbb{1}$. A particular Green function is then obtained through a partial functional derivative with respect to the corresponding external fields. Note that both the quark field operators q in $\mathcal{L}_{\mathrm{ext}}$ and the ground state $|0\rangle$ refer to the chiral limit, indicated by the subscript 0 in Eq. (6.39). The quark fields are operators in the Heisenberg picture and have to satisfy the equation of motion and the canonical anticommutation relations. From the generating functional, we can even obtain Green functions of the "real world," where the quark fields and the ground state are those with finite quark masses. To that end one needs to evaluate the functional derivative of Eq. (6.39) at $s = \mathrm{diag}(m_u, m_d, m_s)$. The chiral Ward identities result from an invariance of the generating functional of Eq. (6.39) under a *local* transformation of the quark fields and a simultaneous transformation of the external fields [69,1610],

$$q_L \mapsto \exp\left(-i\frac{\Theta(x)}{3}\right)V_L(x)q_L, \tag{6.41a}$$
$$q_R \mapsto \exp\left(-i\frac{\Theta(x)}{3}\right)V_R(x)q_R, \tag{6.41b}$$

where $V_L(x)$ and $V_R(x)$ are independent space-time-dependent SU(3) matrices, provided the external fields are subject to the transformations

$$l_\mu \mapsto V_L l_\mu V_L^\dagger + i\,V_L\partial_\mu V_L^\dagger, \tag{6.42a}$$
$$r_\mu \mapsto V_R r_\mu V_R^\dagger + i\,V_R\partial_\mu V_R^\dagger, \tag{6.42b}$$
$$v_\mu^{(s)} \mapsto v_\mu^{(s)} - \partial_\mu\Theta, \tag{6.42c}$$

$$s + ip \mapsto V_R(s + ip)V_L^\dagger, \tag{6.42d}$$

$$s - ip \mapsto V_L(s - ip)V_R^\dagger. \tag{6.42e}$$

The derivative terms in Eqs. (6.42a)–(6.42c) serve the same purpose as in the construction of gauge theories, i.e., they cancel analogous terms originating from the kinetic part of the quark Lagrangian.

6.2.2 Chiral perturbation theory for mesons

Effective field theory (EFT) is a powerful tool for describing the strong interactions at low energies. The essential idea behind EFT was formulated by Weinberg in Ref. [1426] as follows:

> "…if one writes down the most general possible Lagrangian, including all terms consistent with assumed symmetry principles, and then calculates matrix elements with this Lagrangian to any given order of perturbation theory, the result will simply be the most general possible S–matrix consistent with analyticity, perturbative unitarity, cluster decomposition and the assumed symmetry principles."

In the present context, we want to describe the low-energy dynamics of QCD in terms of its Goldstone bosons as effective degrees of freedom rather than in terms of quarks and gluons. The resulting low-energy approximation is called (mesonic) chiral perturbation theory (ChPT). Its foundations are discussed in Ref. [1611]. Since the interaction strength of the Goldstone bosons vanishes in the zero-energy limit and the quark masses are regarded as small perturbations around the chiral limit, the mesonic Lagrangian is organized in a simultaneous derivative and a quark–mass expansion. This Lagrangian is expected to have exactly eight pseudoscalar degrees of freedom transforming as an octet under flavor $\mathrm{SU}(3)_V$. Moreover, taking account of spontaneous symmetry breaking, the ground state should only be invariant under $\mathrm{SU}(3)_V \times \mathrm{U}(1)_V$. Finally, in the chiral limit, we want the effective Lagrangian to be invariant under $\mathrm{SU}(3)_L \times \mathrm{SU}(3)_R \times \mathrm{U}(1)_V$.

Our goal is to approximate the "true" generating functional $Z_{\mathrm{QCD}}[v, a, s, p]$ of Eq. (6.39) by a sequence

$$Z_{\mathrm{eff}}^{(2)}[v, a, s, p] + Z_{\mathrm{eff}}^{(4)}[v, a, s, p] + \cdots,$$

where the effective generating functionals are obtained using the effective field theory. The rationale underlying this approach is the assumption that including all of the infinite number of effective functionals $Z_{\mathrm{eff}}^{(2n)}[v, a, s, p]$ will, at least in the low-energy region, generate a result which is equivalent to that obtained from $Z_{\mathrm{QCD}}[v, a, s, p]$. Because of spontaneous symmetry breaking, the chiral group $\mathrm{SU}(3)_L \times \mathrm{SU}(3)_R$ is realized nonlinearly on the Goldstone-boson fields [1426,1612]. We define the SU(3) matrix

$$U(x) = \exp\left(i\frac{\phi(x)}{F_0}\right), \tag{6.43}$$

where the field matrix ϕ is a Hermitian, traceless 3×3 matrix,

$$\phi(x) = \sum_{a=1}^{8} \phi_a \lambda_a \equiv \begin{pmatrix} \pi^0 + \frac{1}{\sqrt{3}}\eta & \sqrt{2}\pi^+ & \sqrt{2}K^+ \\ \sqrt{2}\pi^- & -\pi^0 + \frac{1}{\sqrt{3}}\eta & \sqrt{2}K^0 \\ \sqrt{2}K^- & \sqrt{2}\bar{K}^0 & -\frac{2}{\sqrt{3}}\eta \end{pmatrix}, \tag{6.44}$$

and the parameter F_0 is the chiral limit of the pion-decay constant. Under *local* chiral transformations, $U(x)$ transforms as [69]

$$U(x) \mapsto V_R(x)U(x)V_L^\dagger(x). \tag{6.45}$$

In particular, Eq. (6.45) implies for the field matrix ϕ the transformation behavior $\phi(x) \mapsto V\phi(x)V^\dagger$ under global flavor $\mathrm{SU}(3)_V$, i.e., the Goldstone bosons indeed form an SU(3) octet [see Eq. (6.37)]. The most general Lagrangian with the smallest (nonzero) number of external fields is given by [69]

$$\mathcal{L}_2 = \frac{F_0^2}{4}\mathrm{Tr}[D_\mu U (D^\mu U)^\dagger] + \frac{F_0^2}{4}\mathrm{Tr}(\chi U^\dagger + U\chi^\dagger), \tag{6.46}$$

where

$$D_\mu U \equiv \partial_\mu U - i r_\mu U + i U l_\mu \mapsto V_R D_\mu U V_L^\dagger, \tag{6.47a}$$

$$\chi \equiv 2B_0(s + ip) \mapsto V_R \chi V_L^\dagger. \tag{6.47b}$$

If we denote a small four momentum as of $\mathcal{O}(q)$, the covariant derivative counts as $\mathcal{O}(q)$ and χ as $\mathcal{O}(q^2)$ (see below), such that the lowest-order Lagrangian is of $\mathcal{O}(q^2)$, indicated by the subscript 2. Using the cyclic property of the trace, $\mathcal{L}_2$ is easily seen to be invariant under the transformations of Eqs. (6.42a)–(6.42e) and (6.45). Moreover, $\mathcal{L}_2$ is invariant under the simultaneous replacements $U \leftrightarrow U^\dagger$, $l_\mu \leftrightarrow r_\mu$, and $\chi \leftrightarrow \chi^\dagger$. It is said to be of even intrinsic parity.

At lowest order, the effective field theory contains two parameters F_0 and B_0. In order to pin down the meaning of F_0, we consider the axial-vector current J_{Aa}^μ associated with $\mathcal{L}_2$:

$$J_{Aa}^\mu = -i\frac{F_0^2}{4}\mathrm{Tr}\left(\lambda_a\{U, \partial^\mu U^\dagger\}\right). \tag{6.48}$$

Expanding U in terms of the field matrix ϕ, and using $\mathrm{Tr}(\lambda_a\lambda_b) = 2\delta_{ab}$ results in

$$J_{Aa}^\mu = -F_0\partial^\mu\phi_a + \mathcal{O}(\phi^3), \tag{6.49}$$

from which we conclude that the axial-vector current has a nonvanishing matrix element when evaluated between the vacuum and a one-Goldstone-boson state:

$$\langle 0|J_{Aa}^\mu(x)|\phi_b(p)\rangle = ip^\mu F_0 \exp(-ip\cdot x)\delta_{ab}. \tag{6.50}$$

Equation (6.50) holds at leading order (LO) in ChPT. It is the current–density analog of Eq. (6.35), i.e., a nonvanishing value of F_0 is a necessary and sufficient criterion for spontaneous symmetry breaking in QCD.

The expansion of the first term of Eq. (6.46) in the field matrix ϕ yields

$$\frac{1}{4}\mathrm{Tr}\left(\partial_\mu\phi\partial^\mu\phi\right)+\frac{1}{48F^2}\mathrm{Tr}\left([\phi,\partial_\mu\phi][\phi,\partial^\mu\phi]\right)+\cdots. \tag{6.51}$$

The first term of Eq. (6.51) describes the kinetic term of the eight Goldstone bosons and the second term contributes to the scattering of Goldstone bosons. The second term of Eq. (6.46) is an example how the explicit symmetry breaking by the quark masses is transferred from the QCD level to the EFT level. Both, $\mathcal{L}^0_{\mathrm{QCD}}+\mathcal{L}_{\mathrm{ext}}$ and $\mathcal{L}_2$ are invariant under *local* chiral transformations. Inserting $\mathcal{L}_{\mathrm{ext}}=\mathcal{L}_{\mathcal{M}}$ corresponds to $s=\mathrm{diag}(m_u,m_d,m_s)$ and it is the same s that is to be used in the effective Lagrangian. The expansion of the χ term gives rise to

$$F_0^2B_0(m_u+m_d+m_s)-\frac{B_0}{2}\mathrm{Tr}\left(\phi^2\mathcal{M}\right)+2B_0\mathrm{Tr}\left(\mathcal{M}\phi^4\right)+\cdots. \tag{6.52}$$

Even though the first term of Eq. (6.52) is of no dynamical significance for the interaction among the Goldstone bosons, it represents an interesting effect. Its negative is the energy density of the vacuum, $\langle\mathcal{H}_{\mathrm{eff}}\rangle_{\mathrm{min}}$, which is shifted relative to the chiral limit because of the nonzero quark masses. We compare the partial derivative of $\langle\mathcal{H}_{\mathrm{eff}}\rangle_{\mathrm{min}}$ with respect to (any of) the light-quark masses m_l with the corresponding quantity in QCD,

$$\left.\frac{\partial\langle 0|\mathcal{H}_{\mathrm{QCD}}|0\rangle}{\partial m_l}\right|_{m_u=m_d=m_s=0}=\frac{1}{3}\langle 0|\bar{q}q|0\rangle_0=\frac{1}{3}\langle\bar{q}q\rangle_0, \tag{6.53}$$

where $\langle\bar{q}q\rangle_0$ is the scalar singlet quark condensate. Within the framework of the lowest-order effective Lagrangian, the constant B_0 is thus related to the scalar singlet quark condensate by

$$3F_0^2B_0=-\langle\bar{q}q\rangle_0. \tag{6.54}$$

For an overview of recent lattice QCD determinations of $\langle\bar{q}q\rangle_0$ see Ref. [1613]. Because of the second term of Eq. (6.52), the Goldstone bosons are no longer massless. If, for the sake of simplicity, we consider the isospin-symmetric limit $m_u=m_d=\hat{m}$ (so that there is no π^0-η mixing), we obtain for the masses of the Goldstone bosons, to lowest order in the quark masses ($\mathcal{O}(q^2)$, denoted by the subscript 2),

$$M^2_{\pi,2}=2B_0\hat{m}, \tag{6.55a}$$

$$M^2_{K,2}=B_0(\hat{m}+m_s), \tag{6.55b}$$

$$M^2_{\eta,2}=\frac{2}{3}B_0\left(\hat{m}+2m_s\right). \tag{6.55c}$$

These results, in combination with Eq. (6.54), correspond to relations obtained in Ref. [1614] and are referred to as the Gell-Mann, Oakes, and Renner relations. Because of the on-shell condition $p^2=M^2$, Eqs. (6.55a)–(6.55c) justify the assignment $\chi=\mathcal{O}(q^2)$. Inserting the empirical values $M_\pi=135$ MeV, $M_K=496$ MeV, and $M_\eta=548$ MeV for the lowest-order predictions provides a first estimate for the ratio of the quark masses,

$$\frac{M_K^2}{M_\pi^2}=\frac{\hat{m}+m_s}{2\hat{m}}\Rightarrow\frac{m_s}{\hat{m}}=25.9, \tag{6.56a}$$

$$\frac{M_\eta^2}{M_\pi^2}=\frac{2m_s+\hat{m}}{3\hat{m}}\Rightarrow\frac{m_s}{\hat{m}}=24.3. \tag{6.56b}$$

A remarkable feature of Eq. (6.46) is the fact that, once F_0 is known (from pion decay), chiral symmetry allows us to make absolute predictions about other processes. For example, the lowest-order results for the scattering of Goldstone bosons can be derived straight-forwardly from the $\mathcal{O}(\phi^4)$ contributions of Eqs. (6.51) and (6.52). In particular, the s-wave $\pi\pi$-scattering lengths for the isospin channels $I=0$ and $I=2$ are obtained as [1610]

$$a_0^0=\frac{7M_\pi^2}{32\pi F_\pi^2}=0.160,\quad a_0^2=-\frac{M_\pi^2}{16\pi F_\pi^2}=-0.0456, \tag{6.57}$$

where we replaced F_0 by the physical pion-decay constant and made use of the numerical values $F_\pi=92.2$ MeV and $M_\pi=M_{\pi^+}=139.57$ MeV. These results are identical with the current–algebra predictions of Ref. [22]. Actually, they serve as an illustration of the fact that the results of current algebra can (more easily) be reproduced from lowest-order perturbation theory in terms of an effective Lagrangian [1615] – in the present case the lowest-order mesonic ChPT Lagrangian.

However, ChPT is much more powerful than the effective Lagrangians of the 1960s, which, by definition, were meant to be applied only in lowest-order perturbation theory (see, e.g., the second footnote in Ref. [1616]). In ChPT, a systematic improvement beyond the tree-level of the lowest-order Lagrangian of Eq. (6.46) is accomplished by calculating loop corrections in combination with tree-level contributions from Lagrangians of higher order. For a long time it was believed that performing loop calculations using the Lagrangian of Eq. (6.46) would make no sense, because it is not renormalizable (in the traditional sense [813]). However, as emphasized by Weinberg [1426,1617], the cancellation of ultraviolet divergences does not really depend on renormalizability; as long as one includes every one of the infinite number of interactions allowed by symmetries, the so-called

non-renormalizable theories are actually just as renormalizable as renormalizable theories [1617]. This still leaves open the question of how to organize a perturbative description of observables. For that purpose, one needs a power-counting scheme to assess the importance of various diagrams calculated from the most general effective Lagrangian. Using Weinberg's power counting scheme [1426], one may analyze the behavior of a given diagram of mesonic ChPT under a linear re-scaling of all *external* momenta, $p_i \mapsto t p_i$, and a quadratic re-scaling of the light-quark masses, $m_l \mapsto t^2 m_l$, which, in terms of the Goldstone-boson masses, corresponds to $M^2 \mapsto t^2 M^2$. The chiral dimension D of a given diagram with amplitude $\mathcal{M}(p_i, m_l)$ is defined by

$$\mathcal{M}(t p_i, t^2 m_l) = t^D \mathcal{M}(p_i, m_l), \tag{6.58}$$

where, in n dimensions,

$$D = n N_L - 2 N_I + \sum_{k=1}^{\infty} 2k N_{2k} \tag{6.59}$$

$$= 2 + (n-2) N_L + \sum_{k=1}^{\infty} 2(k-1) N_{2k} \tag{6.60}$$

$$\geq 2 \text{ in 4 dimensions.}$$

Here, N_L is the number of independent loops, N_I the number of internal Goldstone-boson lines, and N_{2k} the number of vertices originating from $\mathcal{L}_{2k}$. A diagram with chiral dimension D is said to be of order $\mathcal{O}(q^D)$. Clearly, for small enough momenta and masses, diagrams with small D, such as $D = 2$ or $D = 4$, should dominate. Of course, the re-scaling of Eq. (6.58) must be viewed as a mathematical tool. While external three-momenta can, to a certain extent, be made arbitrarily small, the re-scaling of the quark masses is a theoretical instrument only. Note that, for $n = 4$, loop diagrams are always suppressed due to the term $2N_L$ in Eq. (6.60). In other words, we have a perturbative scheme in terms of external momenta and masses which are small compared to some scale (here $4\pi F_0 \approx 1$ GeV).

The most general Lagrangian at $\mathcal{O}(q^4)$ was constructed by Gasser and Leutwyler [69] and contains twelve low-energy constants (LECs) $(L_1, \ldots, L_{10}, H_1 H_2)$,

$$\mathcal{L}_4 = L_1 \left\{ \mathrm{Tr}[D_\mu U (D^\mu U)^\dagger] \right\}^2 + \cdots + H_2 \mathrm{Tr}\left(\chi \chi^\dagger\right). \tag{6.61}$$

The numerical values of the low-energy constants L_i are not determined by chiral symmetry. In analogy to F_0 and B_0 of $\mathcal{L}_2$ they are parameters containing information on the underlying dynamics. For an extensive review of the status of these coupling constants, see Ref. [1618] as well as [1613].

As an example of a one-loop calculation let us consider the $\mathcal{O}(q^4)$ corrections to the masses of the Goldstone bosons. For that purpose one needs to evaluate the self-energy diagrams shown in Fig. 133.

Fig. 133 Self-energy diagrams at $\mathcal{O}(q^4)$. Vertices derived from $\mathcal{L}_{2n}$ are denoted by $2n$ in the interaction blobs

The corresponding expressions for the masses were first given in Ref. [69], of which we show the squared pion mass as a representative example:

$$M_{\pi,4}^2 = M_{\pi,2}^2 \left\{ 1 + \frac{M_{\pi,2}^2}{32\pi^2 F_0^2} \ln\left(\frac{M_{\pi,2}^2}{\mu^2}\right) - \frac{M_{\eta,2}^2}{96\pi^2 F_0^2} \ln\left(\frac{M_{\eta,2}^2}{\mu^2}\right) + \frac{16}{F_0^2}\left[(2\hat{m} + m_s) B_0 (2L_6^r - L_4^r) + \hat{m} B_0 (2L_8^r - L_5^r)\right] \right\}. \tag{6.62}$$

Because of the overall factor $M_{\pi,2}^2$, the pion stays massless as $m_l \to 0$. This is, of course, what we expected from QCD in the chiral limit, but it is comforting to see that the self interaction in $\mathcal{L}_2$ (in the absence of quark masses) does not generate Goldstone-boson masses at higher order. The ultraviolet divergences generated by the loop diagram of Fig. 133 are cancelled by a suitable adjustment of the parameters of $\mathcal{L}_4$. This is Weinberg's argument on renormalizability at work; as long as one works with the most general Lagrangian all ultraviolet divergences can be absorbed in the parameters of the theory. At $\mathcal{O}(q^4)$, the squared Goldstone-boson masses contain terms which are analytic in the quark masses, namely, of the form m_l^2 multiplied by the renormalized low-energy constants L_i^r. However, there are also nonanalytic terms of the type $m_l^2 \ln(m_l)$ – so-called chiral logarithms – which do not involve new parameters. Such a behavior is an illustration of the mechanism found by Li and Pagels [1619], who noticed that a perturbation theory around a symmetry, which is realized in the Nambu–Goldstone mode, results in both analytic as well as nonanalytic expressions in the perturbation. Finally, by construction, the scale dependence of the renormalized coefficients L_i^r entering Eq. (6.62) is such that it cancels the scale dependence of the chiral logarithms [69]. Thus, physical observables do not depend on the scale μ.

In terms of Fig. 133 and the result of Eq. (6.62), we can also comment on the so-called chiral-symmetry-breaking scale Λ_χ to be $\Lambda_\chi = 4\pi F_0$ [1606]. In a loop correction, every endpoint of an internal Goldstone-boson line is multiplied by a factor of $1/F_0$, since the SU(3) matrix of Eq. (6.43) contains the Goldstone-boson fields in the combination ϕ/F_0. On the other hand, external momenta q or Goldstone-boson masses produce factors of q^2 or M^2 (see Eqs. (6.51) and (6.52)).

Together with a factor $1/(16\pi^2)$ remaining after integration in four dimensions they combine to corrections of the order of $[q/(4\pi F_0)]^2$ for each independent loop. Strictly speaking, this particular integral generates an additional factor of 2, and the factor of $1/(16\pi^2)$ should be considered an estimate.

The Lagrangians discussed so far are of even intrinsic parity. At $\mathcal{O}(q^4)$, they are incomplete, because they do not describe processes such as $K^+K^- \to \pi^+\pi^-\pi^0$ or $\pi^0 \to \gamma\gamma$. The missing piece is the effective Wess–Zumino–Witten (WZW) action [1620,1621], which accounts for the chiral anomaly. The chiral anomaly results in the so-called anomalous Ward identities that give a particular form to the *variation* of the generating functional [1610,1620]. At leading order, $\mathcal{O}(q^4)$, and in the absence of external fields, the WZW action reads [1620,1621],

$$S^0_{\rm ano} = N_c\, S^0_{\rm WZW},$$
$$S^0_{\rm WZW} = -\frac{i}{240\pi^2}\int_0^1 d\alpha \int d^4x \epsilon^{ijklm}\mathrm{Tr}(\mathcal{U}_i^L\mathcal{U}_j^L\mathcal{U}_k^L\mathcal{U}_l^L\mathcal{U}_m^L). \tag{6.63}$$

For the construction of the WZW action, the domain of definition of U needs to be extended to a (hypothetical) fifth dimension,

$$U(y) = \exp\left(i\alpha\frac{\phi(x)}{F_0}\right), \tag{6.64}$$

where $y^i = (x^\mu, \alpha)$, $i = 0,\ldots,4$, and $0 \le \alpha \le 1$. Minkowski space is defined as the surface of the five-dimensional space for $\alpha = 1$. The indices $i,\ldots,m$ in Eq. (6.63) run from 0 to 4, $y_4 = y^4 = \alpha$, ϵ_{ijklm} is the completely antisymmetric (five-dimensional) tensor with $\epsilon_{01234} = -\epsilon^{01234} = 1$, and $\mathcal{U}_i^L = U^\dagger\partial U/\partial y^i$.

In contrast to $\mathcal{L}_2$ and $\mathcal{L}_4$, $S^0_{\rm ano}$ is of odd intrinsic parity, i.e., it changes sign under $\phi \to -\phi$. Expanding the SU(3) matrix $U(y)$ in terms of the Goldstone-boson fields, $U(y) = \mathbb{1} + i\alpha\phi(x)/F_0 + \mathcal{O}(\phi^2)$, one obtains an infinite series of terms, each involving an odd number of Goldstone bosons. For example, after some rearrangements, the term with the smallest number of Goldstone bosons reads

$$S^{5\phi}_{\rm WZW} = \frac{1}{240\pi^2F_0^5}\int d^4x\,\epsilon^{\mu\nu\rho\sigma}\mathrm{Tr}(\phi\partial_\mu\phi\partial_\nu\phi\partial_\rho\phi\partial_\sigma\phi). \tag{6.65}$$

In particular, the WZW action without external fields involves at least five Goldstone bosons [1620]. Again, once F_0 is known, after inserting $N_c = 3$ one obtains a parameter-free prediction for, e.g., the process $K^+K^- \to \pi^+\pi^-\pi^0$.

In the presence of external fields, the anomalous action receives an additional term [1621–1623]

$$S_{\rm ano} = N_c(S^0_{\rm WZW} + S^{\rm ext}_{\rm WZW}) \tag{6.66}$$

given by

$$S^{\rm ext}_{\rm WZW} = -\frac{i}{48\pi^2}\int d^4x\epsilon^{\mu\nu\rho\sigma}\mathrm{Tr}\big[Z_{\mu\nu\rho\sigma}(U,l,r) - Z_{\mu\nu\rho\sigma}(\mathbb{1},l,r)\big]. \tag{6.67}$$

where the explicit form of $Z_{\mu\nu\rho\sigma}(U,l,r)$ can be found in [1622,1623]. At leading order, the action of Eq. (6.67) is responsible for the two-photon decays of the π^0 or the η. Quantum corrections to the WZW classical action do not renormalize the coefficient of the WZW term. The counter terms needed to renormalize the one-loop singularities at $\mathcal{O}(q^6)$ are of a conventional chirally invariant structure. In the three-flavor sector, the most general odd-intrinsic-parity Lagrangian at $\mathcal{O}(q^6)$ contains 23 independent terms [1624,1625]. For an overview of applications in the odd-intrinsic-parity sector, we refer to Ref. [1623].

6.2.3 ChPT for baryons

ChPT was first extended to the baryon sector in Ref. [1626], which considered a variety of matrix elements with single-nucleon incoming and outgoing states. While the general approach is analogous to that in the mesonic sector, i.e., one considers the most general Lagrangian consistent with the symmetries of QCD and expands observables in a quark-mass and low-momentum expansion, the baryon sector exhibits some new features. In particular, unlike the Goldstone-boson masses, the baryon masses do not vanish in the chiral limit. This has important consequences for obtaining a proper power counting of diagrams containing baryon lines and for the regularization and renormalization of loop diagrams. In the following we restrict the discussion to $\mathrm{SU}(2)_L \times \mathrm{SU}(2)_R$ chiral symmetry; for the extension to $\mathrm{SU}(3)_L \times \mathrm{SU}(3)_R$ see, e.g., the reviews of Refs. [1627,1628] and references therein. To construct the pion-nucleon Lagrangian, the proton (p) and neutron (n) fields are combined into an SU(2) doublet Ψ,

$$\Psi = \begin{pmatrix} p \\ n \end{pmatrix}. \tag{6.68}$$

The nucleon fields are chosen to transform under local $\mathrm{SU}(2)_L \times \mathrm{SU}(2)_R$ transformations as

$$\Psi \to K(V_L, V_R, U)\Psi, \tag{6.69}$$

where the SU(2) matrix K depends on the left- and right-handed transformations as well as on the pion fields collected in U,

$$K(V_L, V_R, U) = \sqrt{V_R U V_L^\dagger}^{-1}\, V_R\sqrt{U}. \tag{6.70}$$

The baryon Lagrangian also contains the covariant derivative of the nucleon field given by

$$D_\mu \Psi = (\partial_\mu + \Gamma_\mu - i v_\mu^{(s)})\Psi, \tag{6.71}$$

with the connection [1626,1629]

$$\Gamma_\mu = \frac{1}{2}\left[u^\dagger(\partial_\mu - i r_\mu)u + u(\partial_\mu - i l_\mu)u^\dagger\right], \tag{6.72}$$

where $u^2 = U$, and the isoscalar vector field $v_\mu^{(s)}$. Further, it is convenient to define

$$u_\mu = i\left[u^\dagger(\partial_\mu - i r_\mu)u - u(\partial_\mu - i l_\mu)u^\dagger\right]. \tag{6.73}$$

The LO Lagrangian can be written as [1626]

$$\mathcal{L}_{\pi N}^{(1)} = \bar{\Psi}\left(i \not{D} - \mathtt{m} + \frac{\mathtt{g}_A}{2}\gamma^\mu \gamma_5 u_\mu\right)\Psi. \tag{6.74}$$

It contains two LECs: $\mathtt{m}$ and $\mathtt{g}_A$. These correspond to the nucleon mass ($\mathtt{m}$) and the nucleon axial-vector coupling constant ($\mathtt{g}_A$), both taken in the chiral limit. The corresponding physical values will be denoted as m_N and g_A in the following. The superscript (1) in Eq. (6.74) denotes that the Lagrangian is of first order in the power counting. While neither the nucleon energy nor the chiral-limit nucleon mass are small parameters, the combination $i \not{D} - \mathtt{m}$ can be assumed to be a small quantity as long as the nucleon three-momentum is $\mathcal{O}(q)$.

This Lagrangian can be used to calculate the first loop contribution to the nucleon mass. The power counting predicts this contribution to be of $\mathcal{O}(q^3)$. However, the application of dimensional regularization and the minimal subtraction scheme of ChPT ($\widetilde{\text{MS}}$) as used in the meson sector results in terms that are of lower order than predicted by the power counting. Analogous issues also arise for other observables and higher-order contributions. The authors of Ref. [1626] pointed out that the failure of the power counting is related to the regularization and renormalization schemes and that the "same phenomenon would occur in the meson sector, if one did not make use of dimensional regularization." Several methods to address the power counting issue have been proposed [1630–1635].

One commonly used method is Heavy Baryon ChPT (HBChPT) [1630], which was inspired by Heavy Quark Effective Theory [704,1252] (see the discussion in Sect. 6.1). Because the nucleon mass is large compared to the pion mass, an additional expansion of the pion-nucleon Lagrangian is performed in inverse powers of the nucleon mass. In this formalism, application of dimensional regularization in combination with $\widetilde{\text{MS}}$ to loop diagrams, as in the meson sector, leads to a consistent power counting, connecting the chiral to the loop expansion. The heavy-baryon Lagrangian up to and including order q^4 is given in Ref. [1636]. For an introduction to, and applications of, this method see, e.g., Refs. [1627,1637].

While the heavy-baryon formalism makes it possible to use techniques from the meson sector, the additional expansion in powers of the inverse nucleon mass results in a large number of terms in the higher-order Lagrangians. Some of the higher-order terms are related to those at lower orders by Lorentz invariance [1437]. Calculated amplitudes can be expressed in Lorentz-invariant forms, but Lorentz invariance is not manifest throughout intermediate steps of the calculations. Further, issues with analyticity arise in some specific cases because the heavy-baryon expansion results in a shift of the poles in the nucleon propagator [1632].

A manifestly Lorentz-invariant approach to baryon ChPT that addresses these issues was formulated in Ref. [1632], referred to as infrared regularization. While infrared regularization also uses dimensional regularization, the renormalization procedure is different from minimal subtraction. Loop integrals are separated into infrared-singular and infrared-regular parts. The infrared-singular parts contain the same infrared singularities as the original integral and they satisfy the power counting. The infrared-regular parts are analytic in small parameters for arbitrary spacetime dimensions and contain the power-counting-violating terms. Since the infrared-regular parts are analytic, they can be absorbed in the LECs of the baryon Lagrangian. Infrared regularization in its original formulation was applicable to one-loop diagrams. It has been widely used in the calculation of baryon properties, see, e.g., Ref. [1638] for a review.

The expansion of the infrared-regular parts in small parameters contains not only the terms violating the power counting, but also an infinite set of terms that satisfy the power counting. The extended on-mass-shell (EOMS) scheme [1635] provides a method to isolate the terms that violate the power counting and to absorb only these terms in the LECs of the Lagrangian. The EOMS scheme was also shown to be applicable to multi-loop diagrams [1639] and diagrams containing particles other than pions and nucleons [1640]. By reformulating infrared regularization analogously to the EOMS scheme [1641], it can be applied beyond one-loop pion-nucleon diagrams [1639]; see also Ref. [1642] for a different extension of infrared regularization.

The nucleon mass presents an example of the application of baryon ChPT. It has been determined to one-loop order in several approaches, including HBChPT [1643], infrared regularization [1632], and the EOMS scheme [1635]. Up to and including order q^3, the chiral expansion of the nucleon mass is given by

$$m_N = \mathtt{m} - 4c_1 M^2 - \frac{3\mathtt{g}_A^2}{32\pi F^2} M^3 + \cdots, \tag{6.75}$$

where F denotes the pion-decay constant in the two-flavor chiral limit, $F_\pi = F[1 + \mathcal{O}(\hat{m})] = 92.2$ MeV and $M^2 = 2B\hat{m}$ is the lowest-order expression for the squared pion mass.

The result of Eq. (6.75) exhibits some general features of baryon ChPT: The expansion contains not just even powers in the small parameter q like the meson sector, but also odd powers. As a result, the convergence of chiral expansions is expected to be slower in the baryon sector. The second-order contribution is proportional to the LEC c_1 from the second-order Lagrangian. On the other hand, the coefficient of the nonanalytic term proportional to M^3 is given entirely in terms of the LO LEC g_A and F. Similar features also appear at higher orders. The general form of the chiral expansion of the nucleon mass to higher orders is given by

$$\begin{aligned} m_N = {} & \mathrm{m} + k_1 M^2 + k_2 M^3 + k_3 M^4 \ln\left(\frac{M}{\mu}\right) + k_4 M^4 \\ & + k_5 M^5 \ln\left(\frac{M}{\mu}\right) + k_6 M^5 \\ & + k_7 M^6 \ln^2\left(\frac{M}{\mu}\right) + k_8 M^6 \ln\left(\frac{M}{\mu}\right) + k_9 M^6 + \cdots , \end{aligned} \tag{6.76}$$

where μ is the renormalization scale and the ellipsis denotes higher-order terms. The coefficients k_i are linear combinations of various LECs. k_1 through k_4 can be determined by considering at most one-loop diagrams, while k_5 through k_9 receive contributions from two-loop diagrams. Using estimates of the LECs entering the k_i, Ref. [1644] estimated the nucleon mass in the chiral limit from an EOMS calculation to order q^4 to be

$$\begin{aligned} \mathrm{m} &= [938.3 - 74.8 + 15.3 + 4.7 - 0.7]\,\mathrm{MeV} \\ &= 882.8\,\mathrm{MeV}. \end{aligned} \tag{6.77}$$

Two-loop contributions to order q^5 were considered in Ref. [1645], while Refs. [1646,1647] determined m_N to order q^6. Because several currently undetermined LECs enter the expressions for several of the higher-order k_i, no reliable estimate of the complete two-loop contributions is possible. However, the coefficient k_5 of the leading nonanalytic contribution at order q^5 only depends on g_A and the pion-decay constant F and can therefore be compared to lower-order terms. At the physical pion mass and with $\mu = m_N$, $k_5 M^5 \ln(M/m_N) = -4.8\,\mathrm{MeV}$.

Chiral expansions like that of Eq. (6.76) are also important at nonphysical pion masses in the extrapolation of lattice QCD results (for an introduction see, e.g., Ref. [1648]). The fifth-order term $k_5 M^5 \ln(M/m_N)$ becomes as large as the third-order term $k_2 M^3$, where k_2 also only depends on g_A and F, for a pion mass of about 360 MeV. While this comparison includes only one part of the two-loop contributions, it indicates a limit to the applicability of the power counting. This estimate agrees with others found using different methods in Refs. [1649,1650].

Even though the nucleon mass is a static quantity, it is not entirely surprising that a combined chiral and momentum expansion in the baryon sector does not converge well for energies beyond about 300 MeV. This roughly corresponds to the mass gap between the nucleon and the $\Delta(1232)$ resonance. At the physical point, treating the Δ as an explicit degree of freedom has limited impact on the nucleon mass [1651,1652]. However, the $\Delta(1232)$ also couples strongly to the πN channel and has relatively large photon decay amplitudes, resulting in important contributions to processes such as pion-nucleon scattering, Compton scattering, and electromagnetic pion production. These issues were already pointed out in Ref. [1630], which advocated for treating Δ degrees of freedom as dynamic. In baryon ChPT with only pions and nucleons as degrees of freedom, effects of the $\Delta(1232)$ enter implicitly through the values of the LECs. However, these contributions can be proportional to powers of M/δ, where $\delta = (m_\Delta - m)$. This ratio is small as the quark masses approach the chiral limit, but it is a rather large expansion parameter at the physical values, especially when combined with the strong coupling of the Δ. By formulating a theory that also includes the Δ as an active degree of freedom, one hopes to improve the convergence of the perturbative expansion and potentially to increase the kinematic range of applicability.

The inclusion of Δ degrees of freedom poses additional challenges to the construction of the most general Lagrangian and to the power counting. The covariant description of spin-$\frac{3}{2}$, isospin-$\frac{3}{2}$ fields introduces unphysical degrees of freedom [1653,1654]. For the free Lagrangian, these can be eliminated by subsidiary equations and projection operators. The correct number of degrees of freedom also has to be preserved when including interactions with pions, nucleons, and external fields. Various approaches addressing this issue have been considered, see, e.g., Refs. [1655–1660].

The main issue for the power counting is how to count the Δ-nucleon mass difference δ. In one version of the power counting [1657], it is a small quantity of the same order as the pion mass, $\delta \sim \mathcal{O}(q)$. In a different approach [1661], it is argued that (for physical quark masses) $M_\pi < \delta$ and that $M_\pi/\delta \sim \delta/\Lambda$, where $\Lambda \sim 1\,\mathrm{GeV}$ is the breakdown scale of the EFT. Denoting $\bar{\delta} \equiv \delta/\Lambda$ implies that $M_\pi/\Lambda \sim \bar{\delta}^2$, i.e., the pion mass is of higher order than the Δ-nucleon mass difference in this power counting.

6.2.4 *Conclusions*

Over the last few decades, ChPT has developed into a mature and comprehensive approach to the low-energy interactions between Goldstone bosons, nucleons, and external fields, with numerous successful applications. ChPT has played an important role in interpreting lattice QCD calculations performed at unphysical pion masses. It has also served as a prototype for semi-phenomenological approaches in other systems. The application of ChPT methods to the interac-

tions between two and more nucleons is discussed in the contribution by Epelbaum and Pastore.

6.3 Chiral EFT and nuclear physics

Evgeny Epelbaum and Saori Pastore

As explained in the previous section, ChPT allows one to describe the low-energy interactions between hadrons in the Goldstone-boson and single-baryon sectors by means of a perturbative expansion in light-quark masses and particle momenta in line with the symmetries of QCD. In this section we briefly review the extension and application of this systematic and model-independent method to systems with several baryons, focusing on the non-strange sector. This extension goes beyond strict perturbation theory and is commonly referred to as chiral effective field theory, or ChEFT, in order to make the distinction with ChPT clear.

6.3.1 The foundations of ChEFT

EFT methods enjoy increasing popularity in nuclear physics.[65] A unified description of few-nucleon systems, medium mass and heavy nuclei as well as nuclear matter up to the saturation density calls for an EFT applicable at nucleon momenta $p \sim M_\pi$, which must include pions as dynamical DoF. The corresponding framework, commonly referred to as chiral EFT (ChEFT), was pioneered by Weinberg [1666,1667] and represents the most widely used EFT approach in nuclear physics applications. The method relies on the spontaneously broken approximate chiral symmetry of QCD and makes use of the effective Lagrangian for pions and nucleons already introduced in the previous section. Specifically, the $\mathcal{O}(q^2)$ and $\mathcal{O}(q^4)$ mesonic Lagrangians are given in Eqs. (6.46) and (6.61), respectively, while the LO pion–nucleon (πN) Lagrangian is written in Eq. (6.74). Most of the applications to few-nucleon systems are carried out using the heavy-baryon (HB) Lagrangian for the velocity-dependent nucleon field $N(x) = e^{imv\cdot x} P_v^+ \Psi(x)$, with $P_v^+ = (1 + v\cdot\gamma)/2$ being the velocity projection operator [1630]. The LO HB πN Lagrangian obtained from the covariant expression in Eq. (6.74) takes the form

$$\mathcal{L}_{\pi N}^{(1)} = N^\dagger (iv\cdot D + g_A S\cdot u) N, \tag{6.78}$$

where $S_\mu = -\gamma_5[\gamma_\mu, \gamma_\nu]v^\nu/4$ is the covariant spin-operator that is given by the usual Pauli matrices $S^\mu = (0, \vec{\sigma}/2)$ in the rest-frame system of the nucleon with $v_\mu = (1, \vec{0})$. Higher-order terms in the HB πN Lagrangian can be found in Refs. [1627,1636]. Finally, one also needs to include in the effective Lagrangian terms with more than two nucleon fields. The corresponding LO Lagrangian has the form [1666, 1667]

$$\mathcal{L}_{NN}^{(0)} = -\frac{1}{2} C_S (N^\dagger N)^2 + 2 C_T N^\dagger S_\mu N\, N^\dagger S^\mu N, \tag{6.79}$$

with C_S, C_T being low-energy constants (LECs).

While both ChPT and ChEFT rely on the same effective Lagrangian, the two frameworks are applied to describe rather different phenomenological situations. Contrary to the meson and single-baryon sectors, the scattering amplitudes for few-nucleon systems exhibit low-lying poles corresponding to bound (and virtual) states, which signal the breakdown of perturbation theory at very low momenta. For example, in the ${}^3S_1$ and ${}^1S_0$ channels of neutron–proton scattering, the poles are located at $p_{\rm cms} \sim 45i$ MeV and $p_{\rm cms} \sim -8i$ MeV, respectively, which is well within the validity domain of chiral (and even pion-less) EFT. This is in strong contrast to pion–pion scattering, where the lowest-lying resonances reside at momenta of the order of the breakdown scale of ChPT, and the scattering amplitude admits a perturbative expansion in powers of momenta for $p \sim M_\pi$. It is worth emphasizing that while the spontaneously broken chiral symmetry of QCD leads to a strong suppression of the interactions between Goldstone bosons (pions) at low energy, which is at the heart of ChPT, it does not constrain the strength of the interaction between the nucleons for $|\vec{p}\,| \to 0$, see Eq. (6.79).

So how can ChEFT be reconciled with the nonperturbative nature of the two-nucleon interaction? To answer this question one needs a power-counting scheme that determines the importance of *renormalized* contributions to the scattering amplitude. The power counting of mesonic ChPT was already given in Eqs. (6.58)–(6.60). Using the HB framework to avoid the appearance of positive powers of the nucleon mass in renormalized expressions as explained in the previous section, the power counting can be straightforwardly extended to single- and few-nucleon scattering amplitudes. A connected contribution to the scattering amplitude for N nucleons with generic momenta $|\vec{p}\,| \sim M_\pi$ involving N_L independent loop integrals is found to scale as $\mathcal{M} \sim q^D$, where $q \in \{|\vec{p}\,|/\Lambda_b,\ M_\pi/\Lambda_b\}$ with Λ_b being the breakdown scale of ChEFT. In four space-time dimensions, the power D

[65] In the past decades, a variety of EFTs utilizing different degrees of freedom (DoF) have been developed to describe phenomena characterized by specific energy scales. For example, an EFT description of rotational bands of deformed heavy nuclei with excitation energies $E \ll 1$ MeV can be efficiently achieved in terms of collective coordinates with no need to resolve the internal structure of a nucleus under consideration [1662]. Low-energy properties of nuclei consisting of a dense core, surrounded by weakly bound nucleons, have been studied in halo-EFT [1663]. This framework treats the core nucleus as a point-like particle and utilizes the expansion in powers of $p/p_{\rm core}$, with p and $p_{\rm core}$ denoting the binding momenta of the nucleons and of the core nucleus, respectively. Another EFT approach, the so-called pion-less EFT, is formulated in terms of nucleons as the only dynamical DoF and is well suited to describe the dynamics of few-nucleon systems at momenta $p \ll M_\pi$. This framework has proven to be particularly efficient for uncovering universal features of few-body systems around the unitary limit [1664,1665].

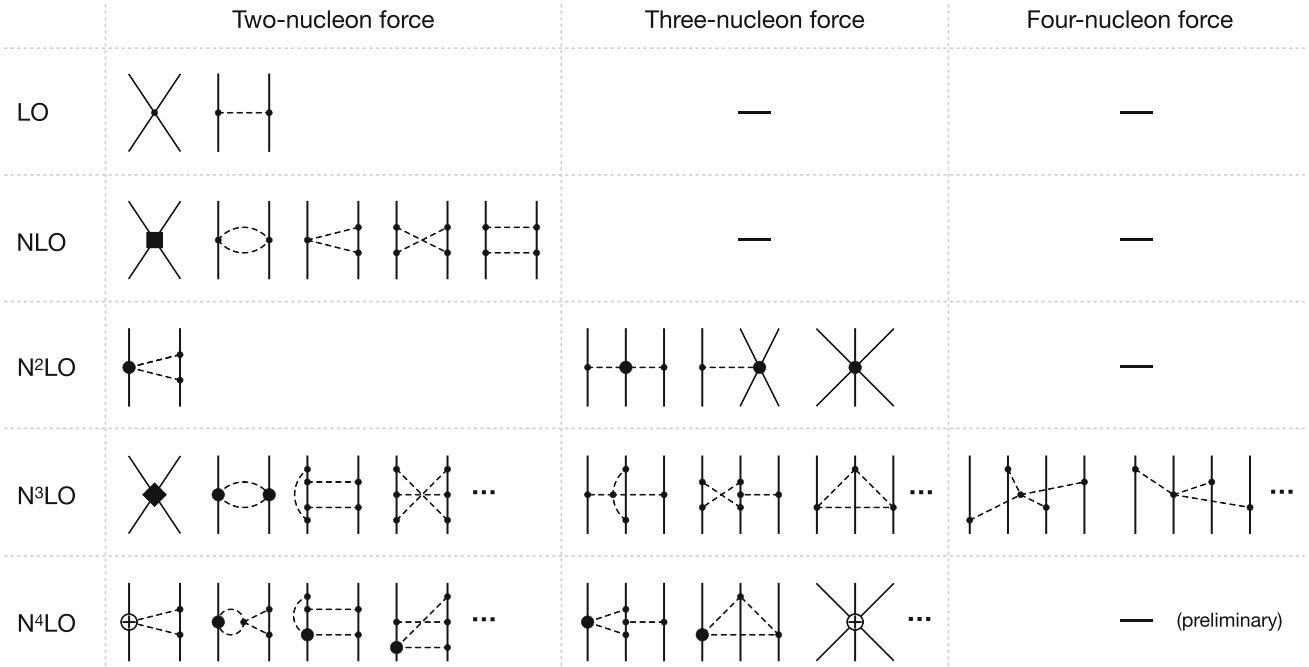


Fig. 134 Diagrams contributing to the two-, three- and four-nucleon forces up to fifth order $\mathcal{O}(q^5)$ in ChEFT. Solid and dashed lines denote nucleons and pions, respectively. Solid dots, filled circles, filled squares, crossed circles and filled diamonds denote vertices with $\Delta = 0, 1, 2, 3$ and 4, respectively

is given by [1666,1667]

$$D = 2 - N + 2N_L + \sum_i V_i \Delta_i , \tag{6.80}$$

where V_i denotes the number of vertices of type i, whose dimension Δ_i is given by

$$\Delta_i = -2 + \frac{1}{2} n_i + d_i . \tag{6.81}$$

Here, n_i is the number of nucleon fields while d_i refers to the number of derivatives and/or insertions of M_π. Using Eq. (6.80), one can draw the relevant Feynman diagrams contributing to the multi-nucleon scattering amplitude at increasing orders in chiral EFT, see Fig. 134. The terms LO, NLO, N^{2}LO, N^{3}LO and N^{4}LO refer to the ChEFT orders q^0, q^2, q^3, q^4 and q^5, respectively. Notice that contributions at order q^1 are forbidden by parity conservation. However, the above classification of Feynman diagrams implies a perturbative nature of multi-nucleon scattering amplitudes, which is in contradiction with the empirical evidence. The key insight of Weinberg was the observation that certain contributions to the amplitude are enhanced beyond what is expected based on Eq. (6.80) [1666,1667]. Consider, for example, the two-pion exchange planar box diagram (the last diagram in the second line of Fig. 134):

$$\mathcal{M} = i \int \frac{d^4 l_1}{(2\pi)^4} l_1^i l_1^j l_2^k l_2^l \, \hat{O}_{ijkl} \, \frac{i}{l_1^2 - M_\pi^2 + i\epsilon} \frac{i}{l_2^2 - M_\pi^2 + i\epsilon} \times \frac{2im}{(p_1 - l_1)^2 - m^2 + i\epsilon} \frac{2im}{(p_2 + l_1)^2 - m^2 + i\epsilon} , \tag{6.82}$$

where $p_1^\mu = (\sqrt{\vec{p}^{\,2} + m^2}, \, \vec{p})$ and $p_2^\mu = (\sqrt{\vec{p}^{\,2} + m^2}, \, -\vec{p})$ are the initial four-momenta of the nucleons, l_1 and $l_2 = p_1' - p_1 + l_1$ are pion momenta and we have used the relativistic rather than the strict HB expressions for the nucleon propagators for reasons to be given below. The spin-isospin operator $\hat{O}_{ijkl}$ with $i, j, k, l = 1, \ldots, 3$ emerges from four πN vertices $\propto g_A$ with $\Delta = 0$. Assuming $|\vec{p}\,|, l_1, l_2 \sim M_\pi \ll m$ and applying naive dimensional analysis (NDA) to the integrand in Eq. (6.82), the renormalized amplitude for the box diagram is expected to be of the order of $\mathcal{M} \sim M_\pi^2$ in agreement with the power counting formula in Eq. (6.80). On the other hand, performing the integration over l_1^0 using the residue theorem, one obtains

$$\mathcal{M} = \int \frac{d^3 l_1}{(2\pi)^3} \hat{O}_{ijkl} \left[\frac{l_1^i l_1^j}{\omega_1^2} \frac{m}{\vec{p}^{\,2} - (\vec{p} - \vec{l}_1)^2 + i\epsilon} \frac{l_2^k l_2^l}{\omega_2^2} + \frac{\omega_1^2 + \omega_1 \omega_2 + \omega_2^2}{2\omega_1^3 \omega_2^3 (\omega_1 + \omega_2)} l_1^i l_1^j l_2^k l_2^l + \mathcal{O}\left(\frac{1}{m}\right) \right] , \tag{6.83}$$

where $\omega_i = \sqrt{\vec{l}_i^{\,2} + M_\pi^2}$ are the energies of the exchanged pions. Remarkably, the first term in the square brackets is enhanced by the factor m/M_π compared to the power

counting estimation. The origin of this enhancement can be traced back to the pinch singularity in the $m \to \infty$ limit [1666,1667], which is why we used the relativistic expressions for the nucleon propagators.[66] Notice that infrared divergences of this kind do not appear in the single-baryon sector of ChPT.

To identify all enhanced types of contributions to the amplitude it is useful to recall that performing the integration over l_0 leads to a decomposition of Feynman diagrams into a sum of diagrams emerging in old-fashioned time-ordered perturbation theory (TOPT). Indeed, the first (enhanced) term in the square brackets in Eq. (6.83) stems from two-nucleon-reducible TOPT diagrams which have an intermediate state involving two nucleons and no pions. Energy denominators associated with such purely nucleonic intermediate states of TOPT diagrams involve only nucleon kinetic energies $\sim M_\pi^2/m \ll M_\pi$ and are smaller than what is expected from NDA. This leads to the enhancement of reducible-type diagrams beyond the power counting estimation in Eq. (6.80).[67] In contrast, the second term in the square brackets of Eq. (6.83) emerges from irreducible two-pion exchange diagrams with intermediate states involving at least one pion and results in the contribution $\mathcal{M} \sim M_\pi^2$ in agreement with Eq. (6.80).

In his seminal work [1666,1667], Weinberg has argued that the breakdown of perturbation theory for the scattering amplitude in the few-nucleon sector of ChEFT can be traced back to the enhancement of reducible diagrams, which need to be resummed to all orders. He also noticed that ladder-type reducible TOPT diagrams automatically get resummed by solving the Lippmann–Schwinger-type integral equations for the amplitude

$$\mathcal{M} = V + VG_0\mathcal{M} = V + VG_0V + VG_0VG_0V + \cdots .$$

Indeed, the terms on the right-hand side of Eq. (6.83) can be easily identified with the iterated one-pion exchange potential (OPEP) and the leading two-pion exchange potential (TPEP), $\mathcal{M} = V_{1\pi}G_0V_{1\pi} + V_{2\pi} + \cdots$. Thus, low-energy processes involving several nucleons can be calculated in a systematically improvable way by applying ChPT to the kernel of the dynamical equation, defined as a sum of all possible few-nucleon-irreducible time-ordered diagrams, rather than to the scattering amplitude. The contributions to the nuclear forces depicted in Fig. 134 are to be understood as (sums of) the corresponding few-nucleon-irreducible time-ordered-like graphs rather than Feynman diagrams. Switching on external classical sources in the effective Lagrangian as explained in the previous section, the same framework can be used to derive nuclear current operators and to analyze low-energy electroweak processes (see the discussion below).

It is worth emphasizing that the enhancement of reducible diagrams mentioned above is insufficient to justify the need for a non-perturbative resummation of the amplitude if one counts $m \sim \Lambda_b$ as done in ChPT. For example, the iterated OPEP contributes at order $V_{1\pi}G_0V_{1\pi} \sim mM_\pi/\Lambda_b^2$ (assuming that all intermediate momenta are $\sim M_\pi$ after renormalization) and is thus suppressed relative to the tree-level term $V_{1\pi} = \mathcal{O}(1)$. To have a self-consistent non-perturbative approach, Weinberg proposed an alternative counting scheme for the nucleon mass by assigning $m \sim \Lambda_b^2/M_\pi \gg \Lambda_b$, which is supported by the large-N_c arguments given that $\Lambda_b \sim M_\rho = \mathcal{O}(1)$ while $m = \mathcal{O}(N_c)$. On the other hand, it is shown in Ref. [1668] that Weinberg's power counting can be realized via a suitable choice of renormalization conditions with no need to depart from the standard ChPT counting for the nucleon mass, see also Ref. [1669] for a related discussion.

Weinberg's power counting suggests that the LO potential stemming from the derivativeless contact interactions $\propto C_{S,T}$, see Eq. (6.79), and the OPEP as shown in Fig. 134 has to be iterated to all orders. For the contact interactions alone, the scattering amplitude resulting from solving the Lippmann–Schwinger (LS) equation can be calculated analytically and is renormalizable in the usual sense.[68] In contrast, iterations of the OPEP in spin-triplet channels lead to ultraviolet divergences whose cancellation requires counterterms with an increasing power of momenta. This feature, along with the numerical nature of the calculations in the presence of the OPEP, make renormalization of chiral EFT a complicated matter; see Ref. [1670] for a collection of perspectives.

Notice that the existence of shallow bound states alone does not necessarily imply a nonperturbative nature of the OPEP, but merely indicates a fine tuning of the LECs $C_{S,T}$ [1667,1671]. An alternative approach based on a perturbative treatment of the OPEP was proposed by Kaplan, Savage and Wise (KSW) in the late nineties of the last century [1672,1673]. This framework allows one to compute the NN scattering amplitude analytically and to implement the renormalization program in a straightforward way with no need to introduce a finite cutoff. However, extensive calculations performed in the KSW approach have revealed poor convergence (at least) in certain spin-triplet channels [1674,1675], see also [1676–1678] for a related discussion, indicating that the OPEP should indeed be treated nonperturbatively in low partial waves.

[66] This singularity is the basis of the covariant spectator theory discussed in Sect. 5.2.

[67] Reducible and irreducible diagrams also play a central role in the derivation of the Bethe–Salpeter equation; see Sect. 5.2.

[68] That is, all ultraviolet divergences emerging from the iterations of the LS equation can be absorbed into a redefinition of $C_{S,T}$.

The most advanced applications of chiral EFT to nuclear systems are carried out utilizing the finite-cutoff formulation of Ref. [1679]. In essence, it amounts to solving the quantum-mechanical A-body problem using the nuclear potentials calculated in ChPT and regularized with some finite cutoff Λ taken of the order of $\Lambda \sim \Lambda_b$. The calculated scattering amplitudes are implicitly renormalized by tuning the bare LECs $C_S(\Lambda)$, $C_T(\Lambda)$, etc., of multi-nucleon vertices to low-energy observables. The resulting (renormalized) scattering amplitudes depend on the physical parameters and the cutoff Λ. The residual Λ-dependence of the calculated observables is expected to introduce an uncertainty beyond the order one is working at and offers a nontrivial a posteriori consistency check. For more details on the foundations and applications of the finite-cutoff formulation of chiral EFT see Refs. [1679,1680]. Finally, a first step towards a formal renormalizability proof of the finite-cutoff scheme to all orders in the iterated OPEP using the Bogoliubov–Parasiuk–Hepp–Zimmermann (BPHZ) subtraction technique can be found in Ref. [1681].

6.3.2 Nuclear interactions from ChEFT

In ChPT, the S-matrix is usually obtained by applying the Feynman graph technique to the effective chiral Lagrangian. To derive nuclear forces, it is more natural and convenient to employ non-covariant old-fashioned perturbation theory as already mentioned above. This approach is based on the Hamiltonian rather than Lagrangian, so the first step amounts to using the canonical formalism for constructing the Hamiltonian $H = H_0 + H_I$ for interacting pions and nucleons from the effective chiral Lagrangian [1666,1667]. The NN scattering amplitude between the initial and final states $|i\rangle$ and $|f\rangle$, respectively, can be written as

$$\langle f|\mathcal{M}|i\rangle = \langle f|H_I \sum_{n=0}^{\infty} \left(\frac{1}{E_i - H_0 + i\epsilon} H_I\right)^n |i\rangle, \tag{6.84}$$

where E_i is the energy of the nucleons in the state $|i\rangle$. Notice that the intermediate states in the above equation include both pions and nucleons. Let η and λ denote the projection operators on the purely nucleonic subspace and the rest of the Fock space, respectively. Eq. (6.84) can be cast into the form of the LS equation

$$\langle f|\mathcal{M}|i\rangle = \langle f|V \sum_{n=0}^{\infty} \left(\frac{\eta}{E_i - H_0 + i\epsilon} V\right)^n |i\rangle, \tag{6.85}$$

where the potential V can e.g. be chosen in the energy-dependent form as done in Refs. [1666,1667,1682,1683]:

$$V(E_i) = \eta H_I \sum_{n=0}^{\infty} \left(\frac{\lambda}{E_i - H_0 + i\epsilon} H_I\right)^n \eta. \tag{6.86}$$

The explicit energy dependence of V is a higher-order effect, see e.g. Eq. (6.83), and can be eliminated yielding an energy independent hermitian NN potential. The method can be applied to many-body forces and has also been used to derive nuclear currents starting from the effective Lagrangian with external sources.

It is important to keep in mind that nuclear potentials, in contrast to the on-shell amplitude $\langle f|\mathcal{M}|i\rangle$, are not directly observable and represent scheme-dependent quantities. This intrinsic ambiguity reflects the arbitrariness in making off-shell extensions of the scattering amplitude. Clearly, such off-shell ambiguities cannot lead to measurable effects. Being a quantum-field-theory-based method, chiral EFT by construction maintains consistency between many-body interactions and current operators and ensures that calculated observables are independent of the off-shell ambiguities (up to higher-order corrections).

The method of deriving nuclear forces and currents by matching to the scattering amplitude as outlined above was used e.g. in Refs. [1684–1688] and is usually referred to as TOPT. Another closely related approach amounts to block-diagonalizing the pion–nucleon Hamiltonian via a suitable unitary transformation [1689]

$$H \to H' = U^\dagger H U = \begin{pmatrix} \eta H' \eta & 0 \\ 0 & \lambda H' \lambda \end{pmatrix}. \tag{6.87}$$

Both the unitary operator U and the nuclear potential $V = \eta(H' - H_0)\eta$ are calculated perturbatively using the standard power counting of ChPT as explained in Ref. [1690]. The method of unitary transformation (MUT) to derive nuclear forces and currents was applied e.g. in Refs [1691–1699]. A pedagogical discussion of methods outlined above can be found in Ref. [1700].

So far, we have left out renormalization of nuclear potentials. In contrast to the scattering amplitude, renormalizability of nuclear forces and currents derived in ChPT is not guaranteed by construction and was shown to impose severe constraints on their off-shell behavior starting from N^{3}LO [1690,1695,1697–1699,1701].

Having introduced various methods to derive nuclear potentials from the effective chiral Lagrangian, we are now in the position to discuss the ChEFT expansion of the long-range NN force. The one- and two-pion exchange contributions up to N^{2}LO depend solely on the momentum transfer $\vec{q}$ and are, therefore, local. The resulting potentials have a clear and intuitive interpretation in coordinate space. Using the decomposition

$$\begin{aligned} V(\vec{r}\,) = {} & V_C(r) + V_S(r)\vec{\sigma}_1 \cdot \vec{\sigma}_2 + V_T(r) S_{12} \\ & + \big[W_C(r) + W_S(r)\vec{\sigma}_1 \cdot \vec{\sigma}_2 + W_T(r) S_{12}\big]\vec{\tau}_1 \cdot \vec{\tau}_2, \end{aligned} \tag{6.88}$$

where $S_{12} = 3\vec{\sigma}_1 \cdot \hat{r}\vec{\sigma}_2 \cdot \hat{r} - \vec{\sigma}_1 \cdot \vec{\sigma}_2$ is the tensor operator while $\vec{\tau}_i$ refer to the isospin Pauli matrices of the nucleon i, the LO contribution due to the OPEP is given by

$$W^{(0)}_{T,\,1\pi}(r) = \frac{g_A^2}{48\pi F_\pi^2}\frac{e^{-x}}{r^3}(3+3x+x^2),$$
$$W^{(0)}_{S,\,1\pi}(r) = \frac{g_A^2 M_\pi^2}{48\pi F_\pi^2}\frac{e^{-x}}{r}, \tag{6.89}$$

where the superscript of the potentials gives the ChEFT order. Further, $x \equiv M_\pi r$ while g_A and F_π denote the physical values of the nucleon axial-vector coupling and pion decay constant, respectively. Notice that only the $W^{(0)}_{T,\,1\pi}(r) \propto r^{-3}$ part of the tensor potential survives in the chiral limit of $M_\pi \to 0$. It is precisely this singular interaction that leads to the already mentioned non-renormalizability of the OPEP in all spin-triplet channels of NN scattering. The NLO contributions to the long-range NN interaction stem from the TPEP and are given by [1682,1684,1689,1702]:

$$\begin{aligned}
W^{(2)}_{C,\,2\pi}(r) &= \frac{M_\pi}{128\pi^3 F_\pi^4}\frac{1}{r^4}\big\{K_1(2x)\\
&\quad\times\big[1+2g_A^2(5+2x^2)-g_A^4(23+12x^2)\big]\\
&\quad+xK_0(2x)\big[1+10g_A^2-g_A^4(23+4x^2)\big]\big\},\\
V^{(2)}_{T,\,2\pi}(r) &= -\frac{g_A^4 M_\pi}{128\pi^3 F_\pi^4}\frac{1}{r^4}\\
&\quad\times\big\{12xK_0(2x)+(15+4x^2)K_1(2x)\big\},\\
V^{(2)}_{S,\,2\pi}(r) &= \frac{g_A^4 M_\pi}{32\pi^3 F_\pi^4}\frac{1}{r^4}\\
&\quad\times\big\{3xK_0(2x)+(3+2x^2)K_1(2x)\big\},
\end{aligned} \tag{6.90}$$

where $K_{0,1}(x)$ denote the modified Bessel functions. To arrive at these expressions, one first needs to evaluate the three-dimensional loop integrals for the corresponding TOPT diagrams[69] using e.g. dimensional regularization. The resulting p-space potentials cannot be Fourier transformed to r-space directly since the Fourier integrals diverge at high momenta. Eq. (6.90) is obtained by Fourier transforming the regularized momentum-space potentials and subsequently removing the regulator.

Similarly, at N^{2}LO, the TPEP receives contributions given by [1684,1689]

$$\begin{aligned}
V^{(3)}_{C,\,2\pi}(r) &= \frac{3g_A^2}{32\pi^2 F_\pi^4}\frac{e^{-2x}}{r^6}\big\{2c_1x^2(1+x)^2\\
&\quad+c_3(6+12x+10x^2+4x^3+x^4)\big\},
\end{aligned}$$

$$W^{(3)}_{T,\,2\pi}(r) = -\frac{g_A^2 c_4}{48\pi^2 F_\pi^4}\frac{e^{-2x}}{r^6}(1+x)(3+3x+x^2),$$
$$W^{(3)}_{S,\,2\pi}(r) = \frac{g_A^2 c_4}{48\pi^2 F_\pi^4}\frac{e^{-2x}}{r^6}(1+x)(3+3x+2x^2), \tag{6.91}$$

where c_i are LECs accompanying the subleading $\pi\pi$NN vertices with $\Delta = 1$.

The expressions for the OPEP and TPEP, Eqs. (6.89) to (6.91), illustrate the general features of the chiral expansion of the long-range nuclear interactions:

- The chiral expansion of the N-pion exchange potential generally corresponds to the expansion in powers of M_π/Λ_χ, where the chiral symmetry breaking scale Λ_χ is given by $4\pi F_\pi$ and/or the scale that governs the πN LECs starting from the subleading ones. The expansion pattern is the same as for ChPT in the meson and single-baryon sectors. The chiral expansion for $V(\vec{r})$ is expected to converge at distances $r \gtrsim 1/M_\pi$ and larger. In contrast, at short distances $r \ll 1/M_\pi$, the expansion diverges yielding highly singular van der Waals-like behaviour $V^{(s)}_{N\pi}(\vec{r}) \sim 1/r^{3+s}$; see also Ref. [1703] for further insights and examples. In the finite-cutoff formulation of chiral EFT, this unphysical short-distance behavior is removed by the regulator.
- Since all relevant πN LECs can nowadays be reliably determined from the pion–nucleon scattering amplitude in the subthreshold region, obtained from the dispersive Roy–Steiner-equation analysis [1704–1706], ChEFT yields parameter-free predictions for the long-range behavior of the nuclear forces and currents. These predictions are model-independent and represent non-trivial manifestations of the spontaneously broken chiral symmetry of QCD.
- Eqs. (6.89)–(6.91) also point towards some limitations of ChPT, which relies on NDA and cannot capture possible enhancements due to large dimensionless prefactors. In the NN sector, this especially affects the N^{2}LO contributions to the TPEP. The corresponding triangle diagram, see Fig. 134, leads to the contribution enhanced by a factor of 4π relative to what is expected based on the power counting, so that Λ_χ is in this case better estimated as $\Lambda_\chi \sim \sqrt{4\pi}F_\pi$ than $\Lambda_\chi \sim 4\pi F_\pi$. Enhancements of this kind are also not uncommon in the single-nucleon sector of ChPT. For the subleading central potential $V^{(3)}_{C,\,2\pi}(r)$, this enhancement combines with the large numerical coefficients and a large value of the LEC c_3 driven by the intermediate $\Delta(1232)$ excitation [1707]. Altogether, this results in $V^{(3)}_{C,\,2\pi}(r)$ being by far the dominant TPE component, whose strength is comparable to that of the OPEP even at $r \sim 2$ fm. The strongly attractive nature of the isoscalar central potential at intermediate distances is supported by phenomenology and often

[69] E.g., the second term in the square brackets in Eq. (6.83) gives the TPEP $\propto g_A^4$ stemming from the last diagram in the second row of Fig. 134 (planar box diagram).

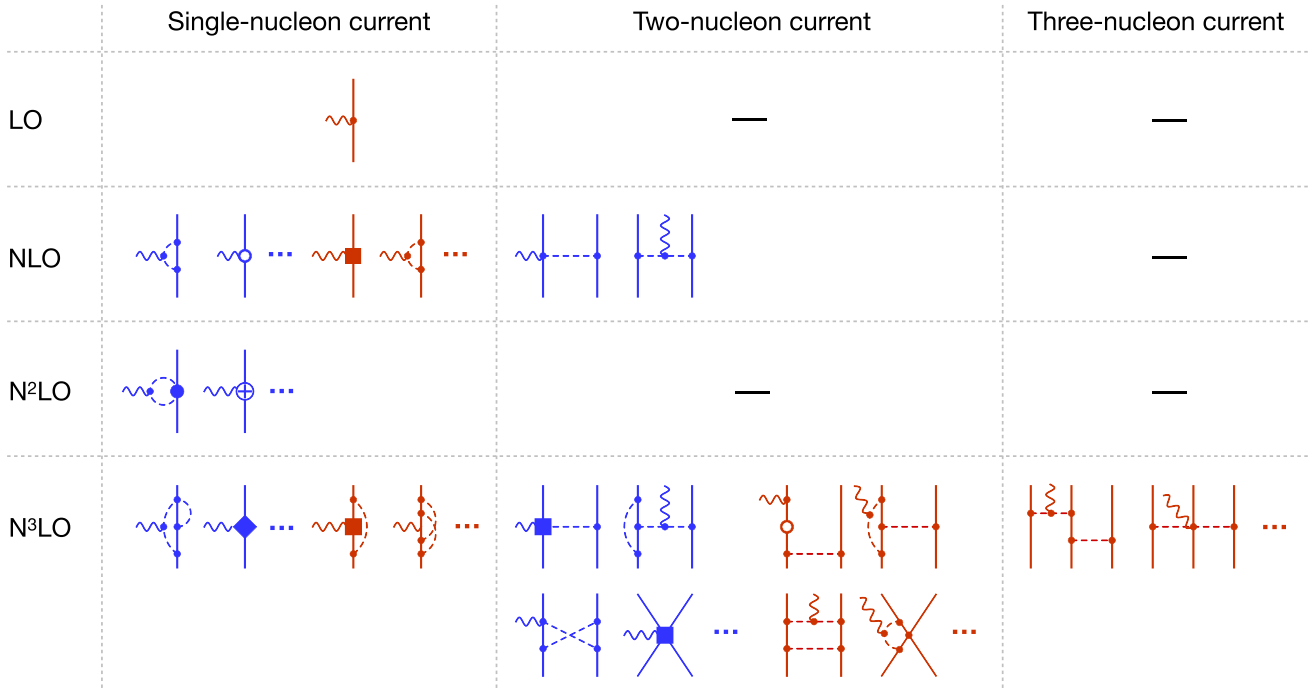


Fig. 135 Diagrams contributing to the single-, two- and three-nucleon electromagnetic current operators at lowest orders of chiral EFT using the counting scheme with $m \sim \Lambda_b^2/M_\pi$. Wiggly lines denote photons. Blue and red diagrams depict the contributions to the current and charge densities, respectively. An open circle shows an insertion of the kinetic energy term with $\Delta = 2$. For remaining notations see Fig. 134

attributed to the σ-meson exchange in traditional nuclear physics jargon. The chiral expansion of the TPEP has been extended to N^{4}LO [1708–1710] and even beyond and was shown to yield converged results [1711,1712].

The current status of the derivation of nuclear potentials in ChEFT is visualized in Fig. 134,[70] see Refs. [1713,1714] for comprehensive review articles. On the qualitative level, ChEFT provides a justification of the observed hierarchy of nuclear forces with $V_{2N} \gg V_{3N} \gg V_{4N} \gg \ldots$ [1666,1667].

The leading contributions to the three-nucleon force (3NF) at N^{2}LO have been known for a long time [1715,1716]. The expressions for the N^{3}LO and (most of the) N^{4}LO corrections have been worked out in Refs. [1691–1694,1717–1719]. The four-nucleon force is further suppressed relative to the 3NF and appears first at N^{3}LO [1690,1701]. Isospin-breaking as well as parity- and time-reversal-violating nuclear potentials have also been worked out, see Refs. [1713,1720] and references therein.

The first application of ChEFT to study nuclear current operators goes back to the pioneering papers by Park et al. [1721,1722]. In the past decade, the vector [1685–1687, 1695,1696,1698], axial-vector [1688,1697], pseudoscalar [1697] and scalar [1699,1723] current operators have been worked out to the leading one-loop-order accuracy for the two-body contributions (i.e., to N^{3}LO using the counting scheme with $m \sim \Lambda_b^2/M_\pi$). As an example, the ChEFT expansion of the electromagnetic nuclear currents is shown in Fig. 135. Similarly to the case of the nuclear forces, the chiral power counting leads, in general, to a suppression of many-body operators. On the other hand, the leading contributions to the single- and two-nucleon current density both appear at NLO. In contrast, the exchange charge density contributions are strongly suppressed relative to the LO term (the charge operator of the nucleon), with both two- and three-nucleon contributions appearing at N^{3}LO. A comprehensive review of nuclear currents in ChEFT, including a detailed comparison of results obtained by different groups and a thorough discussion of the differences between them, can be found in Ref. [1724].

All results described above are based on the effective chiral Lagrangian involving pions and nucleons as the only explicit DoF. As already emphasized in the previous section, given the low excitation energy of the Δ-resonance and its strong coupling to the πN system, it might be advantageous to also treat the Δ DoF as dynamic. This formulation of ChEFT was already applied to derive the NN force and most of the 3NF contributions up through N^{3}LO [1683,1725–1728]. The explicit treatment of the Δ leads to a reshuffling of certain contributions to lower orders in the EFT expansion. In par-

[70] In some approaches, NN contact interactions are promoted to orders different than those derived by NDA.

ticular, a part of the unnaturally strong N^{2}LO TPEP is shifted to NLO, and the LECs $c_{3,4}$ take more natural numerical values [1706]. These results indeed support the expected better convergence pattern of ChEFT with explicit Δ DoF.

Last but not least, ChEFT has also been extended to the SU(3) sector and applied to study the interactions between nucleons and hyperons, see e.g. Refs. [1729–1731] and Ref. [1732] for a recent review article.

6.3.3 Applications

As already pointed out, nuclear interactions derived in ChEFT are singular at short distances and need to be regularized prior to solving the dynamical equation. A broad range of regulators featuring different functional dependence on momenta and relative distances have been proposed in the literature, see Refs. [1712,1733–1737] for some examples and Ref. [1738] for a related discussion. For the long-range OPEP and TPEP, it is advantageous to use a local regularization in order to preserve the analytic structure of the amplitude [1712,1736]. For short-range terms, angle-independent nonlocal regulators maintain a one-to-one correspondence between the plane-wave and partial-wave bases, which simplifies the determination of the corresponding LECs. This choice is utilized in both available N^{4}LO implementations of the NN potentials [1712,1739] which, however, differ in their way of regularizing the long-range terms. In both cases, the LECs accompanying the NN short-range interactions were determined solely from the neutron–proton and proton–proton data. Alternative fitting strategies, which include information about light and medium-mass nuclei and even nuclear matter, are also being explored [1740].

The very accurate and precise NN potentials of [1712, 1741], derived in chiral EFT with pions and nucleons as the only active DoF, provide an outstanding description of NN data up to the pion production threshold.[71] In fact, the results of Ref. [1741] comprise a full-fledged partial wave analysis of NN scattering data based solely on chiral EFT. For more details and comparison between different NN potentials see Ref. [1742].

To give an impression about the convergence pattern of ChEFT consider the total cross section for neutron–proton scattering at $E_{\rm lab} = 100$ MeV as a representative example. Using the potentials from Ref. [1741] one obtains for the cutoff $\Lambda = 450$ MeV (in mb)

$$\sigma_{\rm tot} = 84.0_{[q^0]} - 10.2_{[q^2]} + 0.4_{[q^3]} \\ -0.4_{[q^4]} + 0.6_{[q^5]} - 0.0_{[q^6]},$$

where the last term gives the contribution of the order-q^6 F-wave contact interactions. Given that the expansion parameter is $q = p_{\rm cms}/\Lambda_b \sim 1/3$, where we have used $\Lambda_b = 650$ MeV [1736,1743,1744], one observes that the order-q^3 and q^4 contributions appear to be smaller, while the order-q^5 correction is somewhat larger than naively expected. The truncation error of the calculated value can be estimated using a Bayesian approach by inferring the information about the convergence pattern of the ChEFT from the results at all available orders [1743]; see also Ref. [1736] for a related earlier work. Using the Bayesian model from Ref. [1745], the N^{4}LO truncation error for the case at hand is estimated to be $\delta\sigma_{\rm tot} = 0.14$ mb at 68% confidence level. The final result then reads $\sigma_{\rm tot} = 74.35(14)(17)(1)$ mb, where the last two errors refer to the statistical error and uncertainty in the πN LECs.

The sub-percent accuracy level of ChEFT has also been reached for other low-energy observables in the NN sector [1742]. In particular, the charge and quadrupole form factors of the deuteron were analyzed to N^{4}LO in Refs. [1746,1747]. The predicted value for the deuteron structure radius, $r_{str} = 1.9729^{+0.0015}_{-0.0012}$ fm, was used, in combination with the very precise measurement of the charge radius difference between ^{2}H and the proton [1748], to determine the neutron radius. The obtained value of the quadrupole moment $Q_d = 0.2854^{+0.0038}_{-0.0017}$ fm^2 [1747] is in a very good agreement with the spectroscopy determination $Q_d = 0.285699(15)(18)$ fm^2 [1749].

The spontaneously broken approximate chiral symmetry of QCD, together with the experimental information about the πN system, allow one to predict the long-range behavior of the nuclear forces. In the NN sector, these predictions have been verified from experimental data. For example, the only order-q^3 contribution to the NN force comes from the TPEP in Eqs. (6.91) (since the contact interactions contribute at orders q^{2i}, $i = 0, 1, 2, \ldots$). Adding these parameter-free contributions to the potential was demonstrated to very significantly improve the description of the data [1736,1750,1751]. A similar improvement is observed by adding the order-q^5 TPEP [1710,1712,1752]. It is also worth mentioning that the potentials of [1712] achieve a comparable precision to that of the available high-precision phenomenological potentials while having a much smaller number of adjustable parameters[72] This is yet another evidence of the important role played by chiral symmetry. Finally, the convergence of the chiral EFT expansion can be further improved by the inclusion of Δ's as explicit DoF of the theory. This is supported by the recently developed Norfolk chiral many-body interactions [1753]; see also Ref. [1754] for a related discussion.

[71] This requires the inclusion of four order-q^6 contact interactions that contribute to F-waves [1712,1739].

[72] The N^{4}LO potentials of [1712] depend on 27 LECs fitted to NN data, while the realistic potentials typically involve 40-50 adjustable parameters.

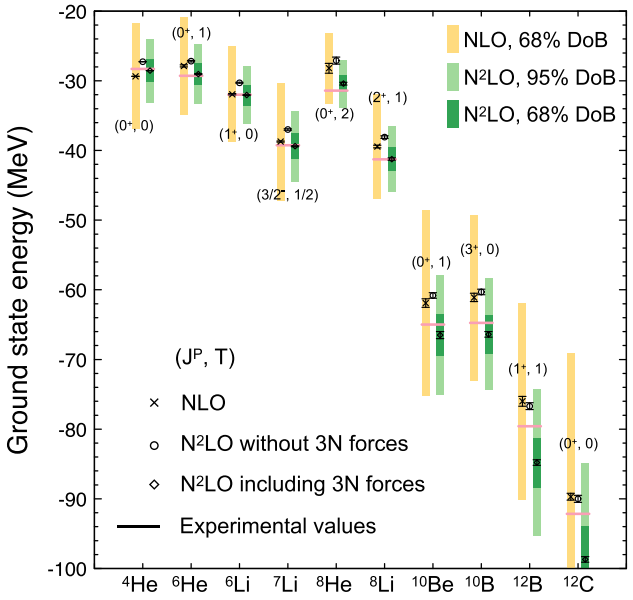


Fig. 136 Predictions for ground state energies of selected p-shell nuclei at NLO and N^{2}LO using the chiral EFT NN potentials from Ref. [1712] together with the consistently regularized 3NF for $\Lambda = 450$ MeV. Black error bars indicate the uncertainties from the employed many-body method, while shaded bars refer to the EFT truncation errors (not shown for incomplete N^{2}LO calculations based on the NN force only). Figure adapted from Ref. [1755]

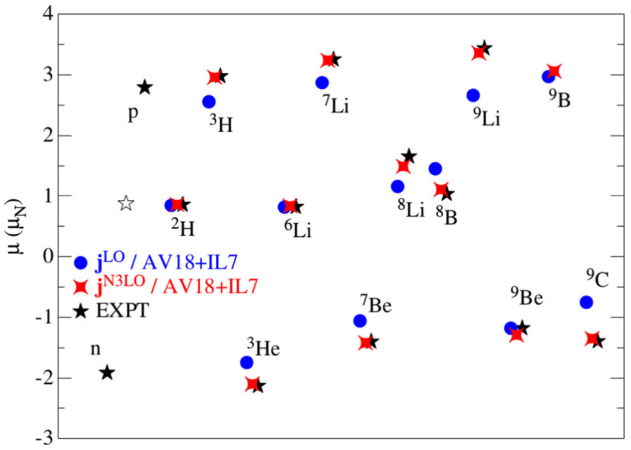


Fig. 137 Magnetic moments in nuclear magnetons for $A \leq 9$ nuclei from Ref. [1767]. Black stars indicate the experimental values while blue dots (red diamonds) represent Green's Function Monte Carlo calculations which include the LO one-body currents (one-body plus two-body currents at N3LO) from ChEFT. For more details and references to the experimental data see [1767]

Beyond the two-nucleon system, the results are presently limited to the N^{2}LO accuracy level due to the lack of *consistently regularized* many-body interactions and exchange currents starting from N^{3}LO. As discussed in Refs. [1680, 1724, 1742], using dimensional regularization in the derivation of nuclear interactions in combination with a cutoff regularization of the Schrödinger equation leads, in general, to violations of chiral symmetry. This issue affects all loop contributions to the 3NF and exchange current operators, which therefore need to be re-derived using symmetry-preserving cutoff regularization.

At the N^{2}LO level, the results for three-nucleon scattering observables [1745, 1755–1757] and the spectra of light- and medium-mass nuclei [1755, 1757–1764] are mostly consistent with experimental data within errors; see also Refs. [1765, 1766] for review articles. As a representative example, we show in Fig. 136 the calculated ground state energies of p-shell nuclei from Ref. [1755].

ChEFT interactions and associated currents have been vigorously utilized in the past 10 years to study both static and dynamical electroweak properties of nuclei, including electromagnetic form factors [852, 1747, 1768], electromagnetic moments [1768–1770], electroweak decays [1771, 1772], and low-energy reactions such as electroweak captures[1773, 1774]. ChEFT currents were first used in calculations of nuclei with $A > 3$ in Ref. [1775] where they are used to study magnetic moments and electromagnetic transitions in $A \leq 10$ systems. Two-body currents were found to improve the agreement between experimental data and theoretical calculations. For example, a long standing under-prediction [1776] of the measured ^{9}C magnetic moment by less sophisticated theoretical calculations is explained by the $\sim 40\%$ correction generated by two-body electromagnetic currents in Ref. [1775]. This enhancement can be appreciated in Fig. 137 by comparing blue dots (representing calculations based on the single nucleon paradigm) and red diamonds (representing calculations with two-body electromagnetic currents).

Axial currents are tested primarily in beta decays and electron capture processes for which data are readily available and known for the most part with great accuracy. The long-standing problem of the systematic over-prediction of Gamow–Teller beta decay matrix elements [1778] in simplified nuclear calculations, also known as the 'g_A problem', has been recently addressed by several groups [1772, 1777, 1779]. The authors of Refs. [1772, 1777] calculated the Gamow–Teller matrix elements in A = 6–10 nuclei accounting systematically for many-body effects in nuclear interactions and coupling to the axial current, both derived in ChEFT. The agreement of the calculations with the data is excellent for $A = 3, 6$ and 7 systems, with two-body currents providing a small ($\sim 2\%$) contribution to the matrix elements. Decays in the $A = 8$ and 10 systems, instead, require further developments of the nuclear wave functions [1777, 1779]. The 'g_A-problem' can be resolved in light nuclei largely by correlation effects in the nuclear wave functions. A summary of these calculations is reported in Fig. 138. Similar results for these light nuclei obtained using the No-core shell model are reported in Ref. [1779].

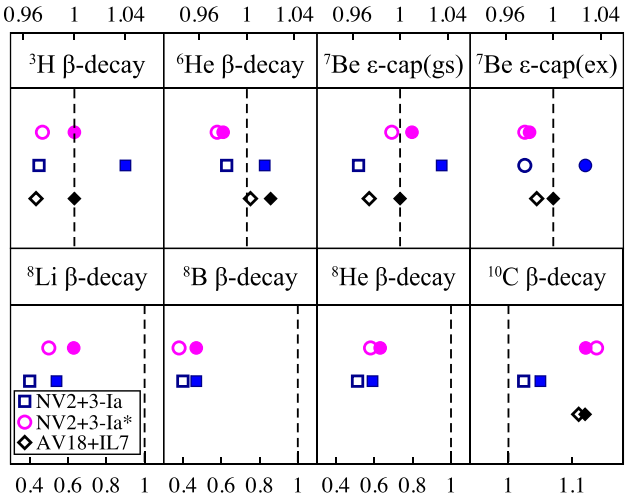


Fig. 138 Ratios of Green's function Monte Carlo calculations to experimental values of the Gamow–Teller reduced matrix elements in the ^{3}H, ^{6}He, ^{7}Be, ^{8}B, ^{8}Be, ^{8}He and ^{10}C weak transitions from Refs. [1772,1777]. Theory predictions correspond to the ChEFT axial current at LO (empty symbols) and up to N3LO (filled symbols)

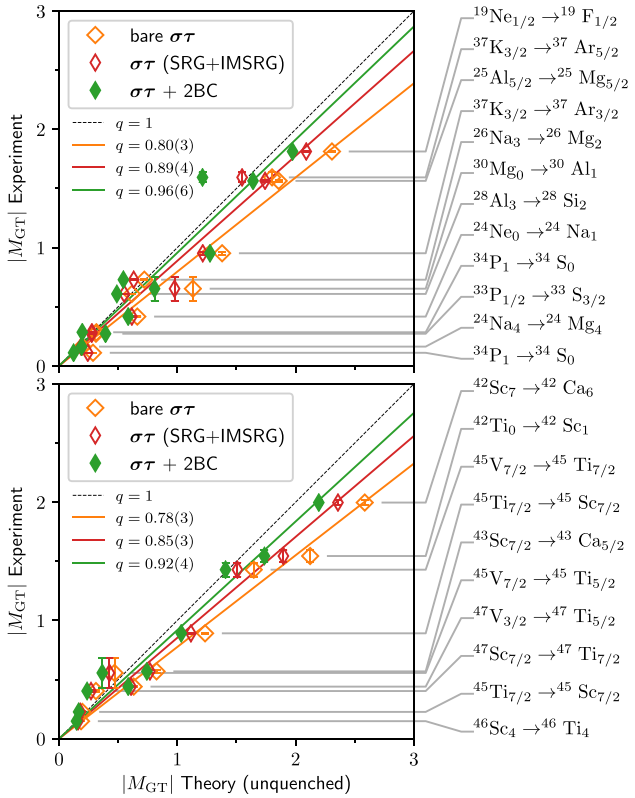


Fig. 139 Comparison of experimental (y-axis) and theoretical (x-axis) Gamow–Teller matrix elements for medium-mass nuclei. The theoretical results were obtained using (i) a bare Gamow–Teller one-body operator, (ii) Gamow–Teller one-body operator consistently evolved with the Hamiltonian [1779], and (iii) a consistently-evolved Gamow–Teller operator that includes both one- and two-body currents. See Ref. [1779] for details

The ChEFT approach is also being implemented in studies of medium-mass nuclei [1779]. As a representative of this class of electroweak calculations we show the results of Ref. [1779] on beta decay matrix elements visualized in Fig. 139. Here, the authors demonstrate that the quenching in the nuclear matrix elements arises primarily from ChEFT axial two-body currents and strong correlations in the nucleus. Nuclei from $A = 3$ to ^{100}Sn are calculated based on ChEFT in agreement with experimental data.

To summarize, there has been exceptional progress in studying nuclear physics using ChEFT. In the last two decades this framework, rooted in the symmetries of QCD and their breaking pattern, has allowed for the calculation of many low-energy nuclear processes, such as electromagnetic reactions and β decays in both light and medium-mass nuclei, has reached a remarkable agreement with experiment, and has contributed to solving long-standing anomalies in nuclear theory. As chiral interactions and currents are being refined and pushed to higher orders, we have entered the precision era of this powerful framework.

6.3.4 Connections to lattice QCD

Lattice QCD (LQCD) offers a first-principles approach to study hadronic and nuclear systems. Several LQCD groups have studied baryon-baryon systems as well as light (hyper-) nuclei at unphysically heavy pion masses using different methods. For non-strange nuclear systems, the current status of LQCD remains controversial, see [1780] for a review. On the EFT side, efforts concentrated on extrapolating lattice QCD results as follows:

- *Chiral extrapolations* of few-nucleon observables have been studied using a variety of ChEFT formulations, see e.g. Refs. [1781–1786]. Currently, the main limiting factor for constraining the quark mass dependence of the nuclear interactions is the lack of reliable LQCD results for not-too-heavy quark masses within the applicability domain of ChEFT.
- *Extrapolations of the NN scattering amplitude in energy* at fixed values of the quark masses were performed [1787, 1788] by exploiting the knowledge of the longest-range interaction due to the OPEP.
- *Infinite-volume extrapolations* of LQCD results for heavy pion masses were carried out in both pion-less [1789–1791] and chiral [1792] EFT.
- Finally, *extrapolations of LQCD results to heavier systems* were considered in Ref. [1793] using the framework of pion-less EFT and in Ref. [1794] utilizing a discretized formulation of ChEFT.

These studies demonstrate remarkable synergy between LQCD and EFT. In the future, LQCD is expected to provide

valuable input for EFT calculations of systems and processes where scarce experimental data exist such as e.g. strange multi-baryon systems and nuclear matrix elements for BSM searches [1780].

6.3.5 Challenges and outlook

To summarize, ChEFT has revolutionized the field of nuclear physics over the past three decades by providing a systematically improvable and theoretically well founded approach to low-energy nuclear interactions, which relies on the symmetries of QCD (and their breaking pattern). The method has proven to be phenomenologically successful and has led to new research directions such as e.g. nuclear lattice simulations [1795–1797]. In the two-nucleon sector, ChEFT has already reached maturity to become a precision tool.

One of the most pressing remaining challenges is the development of accurate and precise three-nucleon interactions needed to shed light onto the long-standing discrepancies in the three-nucleon continuum [1765]. Pushing the ChEFT expansion for many-body forces and exchange currents to $\mathrm{N^3LO}$ and beyond calls for a symmetry preserving regularization [1742], and it will also require new ideas to overcome computational challenges related to the determination of LECs; see Refs. [1798–1800] for recent steps along these lines. Other frontiers include the derivation of consistently regularized electroweak currents, better understanding of renormalization in ChEFT, precision studies of nuclear structure, reactions and the equation of state of nuclear matter as well as applications to searches for BSM physics in processes involving nuclear systems.

6.4 Soft collinear effective theory

Iain Stewart

6.4.1 Introduction

Effective field theory is a powerful tool which enables the organization of QCD dynamics at different momentum scales. The most well known examples of EFTs involve the dynamics of massive particles, like integrating out the heavy electroweak W and Z bosons to obtain the Electroweak Hamiltonian, or systematically treating the mass scale of heavy quarks like the t, b, and c in HQET or NRQCD. On the other hand, much of our knowledge about strong interactions comes from hard scattering interactions of light quarks and gluons, which are the most important processes in pp, e^-p, or e^+e^- colliders. Such processes are the way we search for new particles or fundamental interactions at short distances, and indeed were key to the discovery of the c, b, and t quarks, the W and Z bosons, and the Higgs H. In these processes we must simultaneously deal with perturbative QCD dynamics at the hard interaction scale Q governing the dynamics of the high energy collision, as well as nonperturbative physics at the scale $\Lambda \ll Q$, which is responsible for the confinement and hadronization of partons. Many processes studied at colliders also have additional important intermediate scales Δ, with $\Lambda \ll \Delta \ll Q$. Examples of Δ include the transverse momentum of particles inside an energetic jet produced from the collimated shower of a high energy quark or gluon, or the measurement of differential distributions of a kinematic variable Δ, where the largest cross section contributions typically arise from the $\Lambda \ll \Delta \ll Q$ kinematic situation. The appropriate effective field theory for these processes is the Soft Collinear Effective Theory (SCET) [1801–1804]. Traditional QCD methods, outside the framework of EFT, have a long tradition for describing the physics of hard processes, including the Brodsky–Lepage/Efremov–Radyushkin formalism [225,226,1805] for exclusive hadronic processes, and the Collins–Soper–Sterman formalism [242,1280,1347,1806] for inclusive cross sections. SCET builds naturally on this foundation.

SCET is an effective theory which systematically describes the infrared QCD dynamics in hard collisions, including the associated dynamics of soft and collinear degrees of freedom. Its popularity stems in part from the fact that it enables the description of a huge variety of collider processes [1807]. This includes processes that involve energetic hadrons such as large Q^2 form factors $\gamma^*\gamma \to \pi^0$, $\gamma^*\pi^+ \to \pi^+$, or fragmentation to one or more hadrons h_i in processes like $e^+e- \to h_1h_2X$ and $pp \to h_1X$. Other examples include energetic hadronic collisions like at the Large Hadron Collider, including Higgs production $pp \to HX$ and Drell–Yan $pp \to X\ell^+\ell^-$, Deep Inelastic Scattering (DIS) $e^-p \to e^-X$ or e^--ion $\to e^-X$, and Semi-Inclusive DIS $e^-p \to e^-hX$ (for the latter see Ref. [1808]). SCET also describes processes that produce energetic jets instead of (or in addition to) energetic hadrons, such as $e^+e^- \to$ 2-jets [302,1809–1812], $pp \to H +$ 1-jet [1813,1814], or $pp \to$ 2-jets [1815,1816]. In addition it can be used to describe jet-substructure, the dynamics of particles and sub-jets inside an identified jet [1817–1830]. Finally, it can also be used to describe the dynamics of heavy particle production and decay. Indeed some of the original applications of SCET were to processes like $B \to \pi\ell\nu$ [1802,1831–1833], $B \to D\pi$ [1834,1835], $B \to \pi\pi$ [1832,1836], and $B \to X_s\gamma$ [1801,1803,1837–1841] (where SCET is combined with HQET), as well as $e^+e^- \to J/\Psi X$ [1842–1845] and $\Upsilon \to X\gamma$ [1545,1546,1846–1849] (where SCET is combined with NRQCD). Recent applications of SCET include its extension to forward scattering and Regge phenomena [1850–1853], heavy-ion collisions [1854–1859], gravitational effects [1860–1865], the resummation of large electroweak logarithms [1866–1872], large logs in dark mat-

ter annihilation cross sections [1873–1877], and radiative corrections in neutrino-nucleon scattering [1878,1879].

Features of SCET that people find useful include: the universal steps in deriving *factorization*, whereby observables split themselves into independent functions governing the hard, collinear and soft dynamics of a process, the transparency in carrying out higher order resummation of large logarithms, the ability to generalize factorization to more complicated processes and multiscale observables, and the capability to systematically study power corrections.

6.4.2 Degrees of freedom

SCET describes collinear particles that are constituents of energetic hadrons or jets and have a large momentum along a particular light-like direction n_i^μ. For each collinear direction we have two reference vectors n_i^μ and $\bar{n}_i^\mu$ such that $n_i^2 = \bar{n}_i^2 = 0$ and $n_i \cdot \bar{n}_i = 2$. A common choice is $n_i^\mu = (1, \hat{n}_i)$ and $\bar{n}_i^\mu = (1, -\hat{n}_i)$, with $\hat{n}_i$ a unit three-vector in the collinear direction. Any four-momentum p can be decomposed in terms of these as

$$p^\mu = \bar{n}_i \cdot p \, \frac{n_i^\mu}{2} + n_i \cdot p \, \frac{\bar{n}_i^\mu}{2} + p_{n_i\perp}^\mu . \tag{6.92}$$

Particles with p^μ close to n_i^μ are referred to as n_i-collinear and have $(n_i \cdot p, \bar{n}_i \cdot p, p_{n_i\perp}) = (p^+, p^-, p_\perp) \sim Q\,(\lambda^2, 1, \lambda)$, where $\lambda \ll 1$ is the small SCET power counting parameter, determined by scales and kinematics or by measurements restricting QCD radiation. SCET also describes particles with soft momenta

$$p^\mu \sim Q(\lambda, \lambda, \lambda)$$

and with ultrasoft (usoft) $p^\mu \sim Q(\lambda^2, \lambda^2, \lambda^2)$.

Examples are shown in Fig. 140. In the $B \to D\pi$ process, $Q = \{m_b, m_c, m_b - m_c\}$ and $\lambda = \Lambda/Q$, with the B and D composed of a heavy quark, light soft quarks, and soft gluons. The pion has $E_\pi = 2.3\,\text{GeV} = Q \gg \Lambda$, and has collinear quark and gluon constituents. In the $e^+e^- \to$ 2-jets process, we have back-to-back jets with energy Q, where Q^2 is the invariant mass of the e^+e^- pair, and $\lambda = \Delta/Q$ with $\Lambda \ll \Delta \ll Q$. Here Δ is a scale that characterizes the transverse size of the jet, and associated to measurements made on the jets. For example, if a hemisphere jet mass m_J is measured, then $\Delta = m_J$, while if thrust $1-\tau$ is measured, $\Delta^2 = Q^2\tau$.

To ensure that collinear directions n_i and n_j are distinct, we must have $n_i \cdot n_j \gg \lambda^2$ for $i \neq j$. Since distinct reference vectors, n_i and n_i', with $n_i \cdot n_i' \sim \lambda^2$ both describe the same collinear physics, one can label a collinear sector by any member of an equivalence class of vectors, $\{n_i\}$. This freedom manifests as a symmetry of the effective theory known as reparametrization invariance (RPI) [1880,1881]. Three classes of RPI transformations are

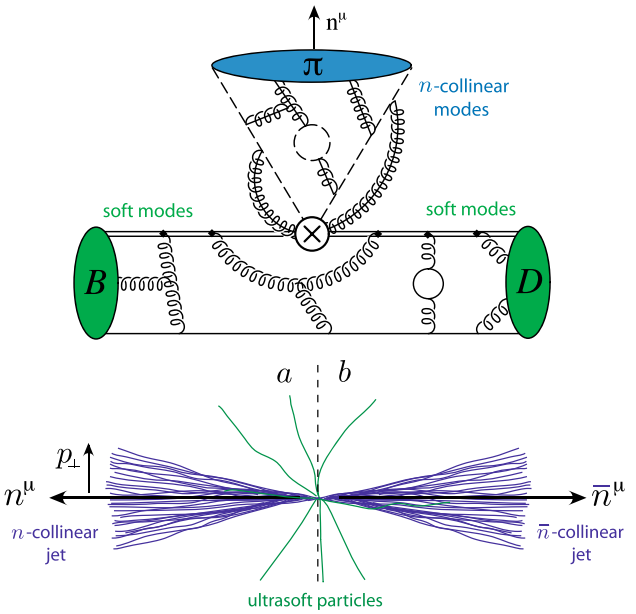


Fig. 140 Example processes $B \to D\pi$ and $e^+e^- \to$ 2-jets

$$\begin{array}{lll} \text{RPI-I} & \text{RPI-II} & \text{RPI-III} \\ n_{i\mu} \to n_{i\mu} + \Delta_\mu^\perp & n_{i\mu} \to n_{i\mu} & n_{i\mu} \to e^\alpha n_{i\mu} \\ \bar{n}_{i\mu} \to \bar{n}_{i\mu} & \bar{n}_{i\mu} \to \bar{n}_{i\mu} + \epsilon_\mu^\perp & \bar{n}_{i\mu} \to e^{-\alpha} \bar{n}_{i\mu}, \end{array} \tag{6.93}$$

where $\alpha \sim \lambda^0$ and infinitesimal parameters $\Delta^\perp \sim \lambda$ and $\epsilon^\perp \sim \lambda^0$. These parameters satisfy $n_i \cdot \Delta^\perp = \bar{n}_i \cdot \Delta^\perp = n_i \cdot \epsilon^\perp = \bar{n}_i \cdot \epsilon^\perp = 0$.

The effective theory is constructed by separating collinear momenta into large (label) $\tilde{p}$ and small (residual) p_r components

$$p^\mu = \tilde{p}^\mu + p_r^\mu = \bar{n}_i \cdot \tilde{p} \, \frac{n_i^\mu}{2} + \tilde{p}_{n_i\perp}^\mu + p_r^\mu , \tag{6.94}$$

with $\bar{n}_i \cdot \tilde{p} \sim Q$, $\tilde{p}_{n_i\perp} \sim \lambda Q$. The small $p_r^\mu \sim \lambda^2 Q$ describes fluctuations about the label momentum. To simultaneously describe different regions of momentum space with operators that have manifest power counting, it is necessary to have multiple fields for the same fundamental particle. Namely, for each collinear direction we have collinear quark fields $\xi_{n_i} \sim \lambda$ and collinear gluon fields $A_{n_i}^\mu \sim (\lambda^2, \lambda^0, \lambda)$, as well as soft quark $q_s \sim \lambda^{3/2}$ and soft gluon $A_s^\mu \sim \lambda$ fields, and/or usoft quark $q_{us} \sim \lambda^3$ and usoft gluon $A_{us}^\mu \sim \lambda^2$ fields. These power counting assignments ensure that the corresponding kinetic terms in the action are $\mathcal{O}(\lambda^0)$.

The precise degrees of freedom depend on the process. Often only usoft or soft fields are present, in which case the theories are referred to as SCET_I and SCET_II respectively [1832]. SCET_I is relevant for measurements sensi-

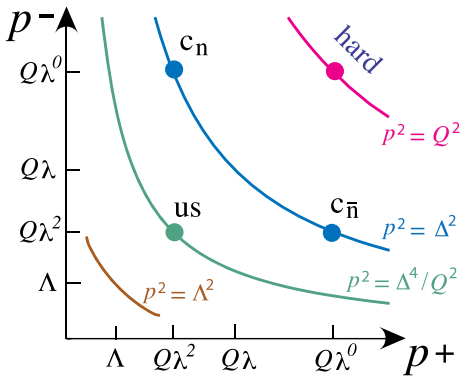


Fig. 141 Degrees of freedom for jet mass in $e^+e^- \to$ 2-jets

tive to the small $n_i \cdot p \sim Q\lambda^2$ momentum, such as jet mass in $e^+e^- \to$ 2-jets, see Fig. 141; while SCET_{II} is relevant for measurements that involve transverse momenta or collinear and soft modes with the same invariant mass. Examples also exist that require mixed soft-collinear modes with $p^\mu \sim Q'(\lambda^2, 1, \lambda)$ where $Q' \ll Q$, in which case the theory is referred to as SCET_+; see Ref. [1827]. Independent collinear, soft, and usoft gauge symmetries are also enforced for each set of fields [1804]. A general SCET λ power counting formula can be used to determine the order of any diagram entirely from operators inserted at its vertices plus topological factors [1850,1882].

To fully expand in λ one must carry out a multipole expansion for the fields in SCET. There are two equivalent ways that this expansion has been constructed in the literature, either in a combination with momentum space for large label momenta and position space for the residuals, with fields written as $\xi_{n_i,\tilde{p}}(x)$ [1803], or with the multipole expansion carried out entirely in position space [1831]. We will use the former, and facilitate the expansion by defining two derivative operators, a label momentum operator $\mathcal{P}^\mu_{n_i}$ giving large momentum components, such as $\mathcal{P}^\mu_{n_i}\, \xi_{n_i,\tilde{p}} = \tilde{p}^\mu\, \xi_{n_i,\tilde{p}}$, and a residual momentum operator giving residual small components, such as $i\partial^\mu \xi_{n_i,\tilde{p}}(x) \sim Q\lambda^2 \xi_{n_i,\tilde{p}}(x)$. The shorthand $\bar{\mathcal{P}} = \bar{n}_i \cdot \mathcal{P}_{n_i}$ is used for the largest $\mathcal{O}(\lambda^0)$ label momentum. Useful covariant derivatives include

$$\begin{aligned} i\bar{n}\cdot\mathbf{D}_n &= \bar{\mathcal{P}} + g\bar{n}\cdot\mathbf{A}_n, \quad i\mathbf{D}^\mu_{n\perp} = \mathcal{P}^\mu_\perp + g\mathbf{A}^\mu_{n\perp} \\ in\cdot\mathbf{D}_n &= in\cdot\partial + gn\cdot\mathbf{A}_n, \quad i\mathbf{D}^\mu_{us} = i\partial^\mu + g\mathbf{A}^\mu_{us} \\ in\cdot\mathbf{D} &= in\cdot\partial + gn\cdot\mathbf{A}_{us} + gn\cdot\mathbf{A}_n, \end{aligned} \tag{6.95}$$

where $\mathbf{A}^\mu_n \equiv A^{A\mu}_n T^A$ and $igF^{A\mu\nu}_n T^A = [i\mathbf{D}^\mu_n, i\mathbf{D}^\nu_n] = ig\mathbf{F}^{\mu\nu}_n$. This is the standard sign convention for g used in the SCET literature. It differs from the QCD summary above ($g \to -g$).

6.4.3 SCET Lagrangian and factorization

The SCET Lagrangian is

$$\mathcal{L}_{\text{SCET}} = \mathcal{L}_{\text{hard}} + \mathcal{L}_{\text{dyn}} = \sum_{i\geq 0}\Big(\mathcal{L}^{(i)}_{\text{hard}} + \mathcal{L}^{(i)}_{\text{dyn}}\Big) + \mathcal{L}^{(0)}_G, \tag{6.96}$$

where the superscript (i) indicates terms suppressed by $\mathcal{O}(\lambda^i)$ relative to the leading power Lagrangian. Here the hard short distance interactions are encoded in $\mathcal{L}^{(i)}_{\text{hard}}$ with only one of these appearing in each amplitude (unless we study multiple hard scatterings). They contain multiple types of collinear (and soft) fields. The dynamic Lagrangians $\mathcal{L}^{(i)}_{\text{dyn}}$ describe the evolution and interactions of collinear and (u)soft particles. We have singled out the so-called Glauber Lagrangian $\mathcal{L}^{(0)}_G$ for special treatment since it is the only term that violates factorization of collinear and (u)soft modes [1850].

At leading power the dynamic SCET_{I} and SCET_{II} Lagrangians are [1804]

$$\begin{aligned} \mathcal{L}^{\text{I}(0)}_{\text{dyn}} &= \sum_n \mathcal{L}^{(0)}_n + \mathcal{L}^{(0)}_{us}, \\ \mathcal{L}^{\text{II}(0)}_{\text{dyn}} &= \sum_n \mathcal{L}^{(0)}_n + \mathcal{L}^{(0)}_s, \end{aligned} \tag{6.97}$$

where the first terms sum over all needed independent collinear sectors. In SCET_{II} each of $\mathcal{L}^{(0)}_n$ and $\mathcal{L}^{(0)}_s$ only involves collinear or soft fields, so the sectors are immediately factorized by the power expansion. In SCET_{I} the $n\cdot\mathbf{A}_{us}$ fields still interact with collinear fields since they are $\mathcal{O}(\lambda^2)$ just like $n\cdot\partial$ and $n\cdot\mathbf{A}_n$, and do not knock the collinear particles offshell (meaning that initial and final particles have momenta satisfying the collinear power counting). These $n\cdot\mathbf{A}_{us}$ interactions can be decoupled by the BPS field redefinition [1804]

$$\xi_n(x) \to Y_n(x)\xi_n(x), \quad \mathbf{A}^\mu_n(x) \to Y_n(x)\mathbf{A}^\mu_n(x)Y^\dagger_n(x), \tag{6.98}$$

where Y_n is an ultrasoft Wilson line

$$Y_n(x;-\infty,0) = P\exp\Big(ig\int_{-\infty}^0 \mathrm{d}s\; n\cdot\mathbf{A}_{us}(x+ns)\Big), \tag{6.99}$$

and P is path ordering of color matrices with s. This transformation moves usoft interactions into the hard scattering operators, and leaves factorized Lagrangians $\mathcal{L}^{(0)}_n$ and $\mathcal{L}^{(0)}_{us}$, which only depend on collinear or usoft fields respectively. For example, for collinear quarks in SCET_{I} we have

$$\mathcal{L}^{\text{I}(0)}_{n\xi} = e^{-ix\cdot\mathcal{P}}\,\bar{\xi}_n\Big(in\cdot\mathbf{D} + i\not{\mathbf{D}}_{n\perp}\frac{1}{i\bar{n}\cdot\mathbf{D}_n}i\not{\mathbf{D}}_{n\perp}\Big)\frac{\not{\bar{n}}}{2}\xi_n$$

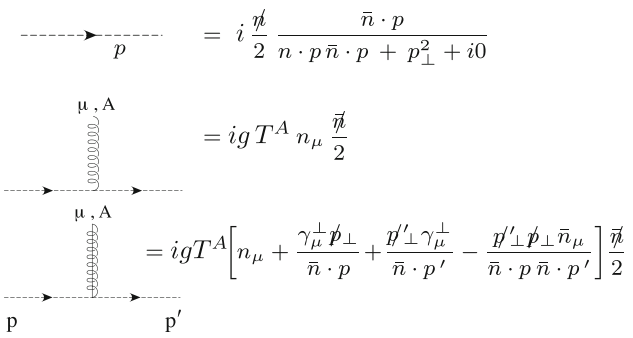


Fig. 142 $\mathcal{O}(\lambda^0)$ Feynman rules for collinear quarks (dashed) interacting with a soft gluon (spring) or collinear gluon (spring with a line through it). Rules with more collinear gluons are not shown

$$\to e^{-ix\cdot\mathcal{P}}\,\bar{\xi}_n\Big(in\cdot\mathbf{D}_n + i\not{\mathbf{D}}_{n\perp}\frac{1}{i\bar{n}\cdot\mathbf{D}_n}i\not{\mathbf{D}}_{n\perp}\Big)\frac{\not{\bar{n}}}{2}\,\xi_n. \tag{6.100}$$

The first few Feynman rules prior to the field redefinition are shown in Fig. 142, and the one in the second line is removed from $\mathcal{L}^{\mathrm{I}(0)}_{n\xi}$ after implementing Eq. (6.98). After the transformation the $\mathcal{L}^{(0)}_n$ Lagrangian has the same form in $\mathrm{SCET_I}$ and $\mathrm{SCET_{II}}$.

The construction of SCET hard scattering Lagrangians $\mathcal{L}^{(i)}_{\mathrm{hard}}$ requires integrating out offshell fields, which have larger p^2 than the collinear and (u)soft fields (see Fig. 141 for example). When collinear particles in two different sectors interact, the resulting particles are hard and offshell with $p^2 \sim Q^2$. Likewise when collinear and soft particles interact this results in offshell hard-collinear particles with $p^2 \sim Q^2\lambda$. Systematically integrating out the corresponding offshell fields results in collinear and soft Wilson lines appearing in operators [1802–1804]. This involves an infinite number of gluon attachments and can be carried out analytically with background field techniques [1804,1807]. In label momentum space the resulting collinear Wilson lines are defined as

$$W_{n_i}(x) = \Big[\sum_{\text{perms}} \exp\Big(-\frac{g}{\bar{\mathcal{P}}}\,\bar{n}\cdot\mathbf{A}_{n_i}(x)\Big)\Big]. \tag{6.101}$$

Note that it is the $\bar{n}_i\cdot\mathbf{A}_n \sim \lambda^0$ component of the gluon field that appears in these Wilson lines. In general all $\mathcal{O}(\lambda^0)$ gluon components can be traded for Wilson lines using $i\bar{n}_i\cdot\mathbf{D}_{n_i} = W_{n_i}\bar{\mathcal{P}}W^\dagger_{n_i}$. Unlike Y_n, the subscript on W_{n_i} refers to the collinear fields it is built out of, not the Wilson line direction (which is $\bar{n}_i$). For zero residual momentum $x = (0, x^-, x_\perp)$, the $W_n(x)$ is simply the Fourier transform ($b^+ \leftrightarrow p^-$) of a standard position-space Wilson line ending at $b = (b^+, x^-, x_\perp)$:

$$W_n(b;-\infty,0) = P\,\exp\Big(ig\int_{-\infty}^{0} ds\;\bar{n}\cdot\mathbf{A}_n(b+\bar{n}s)\Big). \tag{6.102}$$

Since the construction of hard-collinear interactions in $\mathrm{SCET_{II}}$ can be facilitated by matching QCD→$\mathrm{SCET_I}$ → $\mathrm{SCET_{II}}$ [1832], it suffices to primarily focus on matching for $\mathrm{SCET_I}$. The definition for the soft Wilson line $S_n(x;-\infty,0)$ appearing in $\mathrm{SCET_{II}}$ is identical to Eq. (6.99) with $\mathbf{A}_{us} \to \mathbf{A}_s$.

Hard interactions involving collinear fermions provide a frame of reference that allows us to simplify the Dirac structures that appear, since so-called good fermion components dominate over bad components in the λ expansion. In SCET this is encoded by the projection relations $(\not{n}_i\not{\bar{n}}_i/4)\xi_{n_i} = \xi_{n_i}$, which also implies $\not{n}_i\xi_{n_i} = 0$. The same formulae also hold for χ_{n_i}. Only the good components are needed to construct operators in SCET at any order in the power expansion, and indeed we have already written $\mathcal{L}^{(0)}_{n\xi}$ in Eq. (6.100) using them. Note that on its own, Eq. (6.100) is equivalent to a QCD Lagrangian for collinear quarks (indeed it has the same form as the light-cone QCD Lagrangian [1883]), with a distinction made only by which fermion components are sourced in the path integral.

Integrating out offshell fluctuations also results in Wilson coefficients that depend on the large $\mathcal{O}(\lambda^0)$ momenta of collinear fields. It is straightforward to see why this is the case, since if we annihilate or produce two collinear particles with $p^\mu_n = \omega_1 n^\mu/2$ and $p^\mu_{\bar{n}} = \omega_2\bar{n}^\mu/2$, then $q = p_n + p_{\bar{n}}$ has $q^2 = \omega_1\omega_2 \sim Q^2$. Thus offshell fluctuations that depend on Q^2 also depend on the large momenta $\omega_i \sim \lambda^0$ of collinear fields. Two other constraints on the form of hard operators are SCET gauge invariance and the ability to use the equations of motion to reduce the operators basis to a minimal set. This is summarized by the fact that all operators can be constructed out of a minimal set of building blocks, formed from combinations of fields and Wilson lines [1802,1803,1884]. The collinearly gauge-invariant quark and gluon building block fields are defined as

$$\begin{aligned}\chi_{n_i,\omega}(x) &= \Big[\delta(\omega-\bar{\mathcal{P}}_{n_i})\,W^\dagger_{n_i}(x)\,\xi_{n_i}(x)\Big],\\ \mathcal{B}^\mu_{n_i\perp,\omega}(x) &= \frac{1}{g}\Big[\delta(\omega+\bar{\mathcal{P}}_{n_i})\,W^\dagger_{n_i}(x)\,i\mathbf{D}^\mu_{n_i\perp}W_{n_i}(x)\Big].\end{aligned} \tag{6.103}$$

The Wilson lines $W_{n_i}(x)$ are localized with respect to the position x, and we can therefore treat $\chi_{n_i,\omega}(x)$ and $\mathcal{B}^\mu_{n_i,\omega}(x)$ as local quark and gluon fields from the perspective of ultrasoft derivatives ∂^μ that act on x. Our conventions for $\chi_{n_i,\omega}$ have $\omega > 0$ for an incoming quark and $\omega < 0$ for an outgoing antiquark at lowest order. For $\mathcal{B}_{n_i\perp,\omega}$, $\omega > 0$ ($\omega < 0$) corresponds to outgoing (incoming) gluons at lowest order.

For $\mathrm{SCET_I}$ the complete set of building blocks and their power counting is summarized in Table 7.

Both the χ_n and $\mathcal{B}_{n\perp}$ building block fields scale as $\mathcal{O}(\lambda)$. For the majority of jet processes there is a single collinear

Table 7 Power counting for building block operators in $\mathrm{SCET_I}$

Operator	$\mathcal{B}^{\mu}_{n_i\perp}$	χ_{n_i}	$\mathcal{P}^{\mu}_{\perp}$	q_{us}	$\mathbf{D}^{\mu}_{us}$
Power counting	λ	λ	λ	λ^3	λ^2

field operator for each collinear sector at leading power. For exclusive processes that directly produce energetic hadrons at the hard interaction (rather than by fragmentation) there are multiple building blocks from the same sector in the leading power operators, since we must form a color singlet in each sector in order to directly produce a color singlet hadron. The $\mathcal{P}_\perp \sim \lambda$ is not typically present at leading power. At subleading power, operators for all processes can involve multiple collinear fields in the same collinear sector, as well as $\mathcal{P}_\perp$ operator insertions. The power counting for an operator is obtained by simply adding up the powers for the building blocks it contains. To ensure consistency under renormalization group evolution the operator basis in SCET must be complete, namely all operators consistent with the symmetries of the problem must be included. The counting of subleading power operators is greatly facilitated by spinor-helicity SCET techniques [1885–1889].

A few examples of hard scattering operators can help clarify the above points. For $\mathrm{SCET_I}$ processes like thrust, jet mass, or other dijet event shapes in e^+e^- collisions, or for threshold resummation in Drell–Yan or DIS, the leading power Lagrangian from the electromagnetic current is

$$\mathcal{L}^{\mathrm{I}(0)}_{\mathrm{hard}}(0) = \frac{ie^2}{Q^2} J^{\mu}_{\bar\ell\ell} \int \mathrm{d}\omega_1 \mathrm{d}\omega_2\, C_f^{(0)}(\omega_1\omega_2,\mu) \times \big[(\bar\chi^f_{\bar n,\omega_2})\,(Y^\dagger_{\bar n} Y_n)\gamma^\perp_\mu(\chi^f_{n,\omega_1})\big]_\mu, \tag{6.104}$$

where $C_f^{(0)}$ is the Wilson coefficient encoding virtual hard interactions at any order in α_s, and renormalization is carried out in the $\overline{\mathrm{MS}}$ scheme, inducing dependence on the renormalization scale μ. In Eq. (6.104) the usoft Wilson lines $Y^\dagger_{\bar n} Y_n$ appear from the BPS field redefinition in Eq. (6.98). Also, the leptonic vector current is $J^{\mu}_{\bar\ell\ell} = (-1)\bar\ell\gamma^\mu\ell$, and we sum over quark flavors f. At any order in α_s the Wilson coefficient $C_f^{(0)}(\omega_1\omega_2)$ encodes virtual corrections from the hard scale $\omega_1\omega_2 \sim Q^2$. For hard Lagrangians with only a single field in a given collinear direction, the large collinear momentum factors ω_i are fixed by the overall kinematics of the hard process, and thus remain unchanged by perturbative corrections. For example, $\omega_1 = \omega_2 = Q$ for $e^+e^- \to$ 2-jets. At tree level $C_f^{(0)} = Q_f + \mathcal{O}(\alpha_s)$, where the quarks have charge $Q_f|e|$. To calculate $C_f^{(0)}$ at higher orders we carry out loop level matching calculations, comparing hard scattering Feynman diagrams separately computed and renormalized in full QCD and in $\mathrm{SCET_I}$, while using the same states and infrared (IR) regulators. Since SCET captures all the IR physics, the difference between these calculations determines $C_f^{(0)}$ order by order, and implies it encodes hard effects. For the particular example in Eq. (6.104), $C_f^{(0)}$ is related to the IR finite part of the $\overline{\mathrm{MS}}$ massless quark form factor with $Q^2 \gg \Lambda^2$. (In general when carrying out loop calculations in SCET with both (u)soft and collinear loops, one must include 0-bin subtractions which ensure there is not double counting of IR regions [1890]. For some choices of IR regulators these subtractions are scaleless in dimensional regularization, and hence can be dropped, up to interpreting the divergence structure.)

For $\mathrm{SCET_{II}}$ processes like the broadening event shape for $e^+e^- \to$ 2-jets, or transverse momentum dependent (TMD) distributions for Drell–Yan, SIDIS, or $e^+e^- \to h_1h_2X$, the leading hard scattering Lagrangian is

$$\mathcal{L}^{\mathrm{II}(0)}_{\mathrm{hard}}(0) = \frac{ie^2}{Q^2} J^{\mu}_{\bar\ell\ell} \int \mathrm{d}\omega_1 \mathrm{d}\omega_2\, C_f^{(0)}(\omega_1\omega_2,\mu) \times \big[(\bar\chi^f_{\bar n,\omega_2})\,(S^\dagger_{\bar n} S_n)\gamma^\perp_\mu(\chi^f_{n,\omega_1})\big]_\mu, \tag{6.105}$$

with the same Wilson coefficient $C_f^{(0)}$ as Eq. (6.104). The only difference is the appearance of soft Wilson lines S instead of usoft Y. This operator can be obtained immediately from Eq. (6.104) by matching $\mathrm{SCET_I} \to \mathrm{SCET_{II}}$.

As a final example we consider $\bar B^0 \to D^+\pi^-$ mediated by the weak W-boson flavor changing transition $b \to c\bar u d$. Here the matching is from the electroweak Hamiltonian $\mathcal{H}_W = 2\sqrt{2}G_F V^*_{ud}V_{cb}\sum_{i=0,8} C_i^{\mathrm{F}} O_i$, with 4-quark operators $O_0 = [\bar c\gamma^\mu P_L b][\bar d\gamma_\mu P_L u]$ and $O_8 = [\bar c\gamma^\mu T^A P_L b][\bar d\gamma_\mu P_L T^A u]$, onto coefficients and operators in SCET. The heavy quark fields are matched onto HQET fields $h_v^{(Q)}$ for $Q = b, c$, while the light quarks become collinear. The leading power hard scattering Lagrangian in SCET is [1834]

$$\mathcal{L}^{\mathrm{II}(0)}_{\mathrm{hard}} = \int \mathrm{d}\omega_1 \mathrm{d}\omega_2\, C^{j(0)}_{BD\pi}(\omega_1,\omega_2,m_b,m_c,\mu) \times \big\{\big[\bar h^{(c)}_{v'}\Gamma^j_h h^{(b)}_v\big]\big[(\bar\chi^{\bar d}_{n,\omega_2})\,\Gamma_\xi(\chi^u_{n,\omega_1})\big]\big\}_\mu, \tag{6.106}$$

where we sum over $j = 1, 5$ with Dirac structures $\Gamma^{1,5}_h = \not{n}\{1,\gamma_5\}/2$ and $\Gamma_\xi = \not{\bar n}(1-\gamma_5)/4$. Here the hard coefficients $C^{j(0)}_{BD\pi}$ depend on multiple hard scales as in Eq. (6.106). There are no soft Wilson lines because the n-collinear quark pair is a color singlet and $S^\dagger_n S_n = 1$. An analogous SCET operator with color structure $T^A \otimes T^A$ exists and does involve soft Wilson lines. Since it can be factorized into a product of soft and collinear octet operators, it does not contribute to the physical process: a factorized octet collinear bilinear operator can not produce a color singlet pion.

Let us return to the leading power Glauber Lagrangian. It involves interactions between soft and collinear modes in the

form of potentials, and has the form [1850]

$$\mathcal{L}_G^{(0)} = \sum_{n,\bar{n}} \mathcal{O}_n^{iB} \frac{1}{\mathcal{P}_\perp^2} \mathcal{O}_s^{BC} \frac{1}{\mathcal{P}_\perp^2} \mathcal{O}_{\bar{n}}^{jC} + \sum_n \mathcal{O}_n^{iB} \frac{1}{\mathcal{P}_\perp^2} \mathcal{O}_s^{j_n B}, \tag{6.107}$$

in both SCET_I and SCET_II. Further details and the definitions for the operators $\mathcal{O}$ can be found in Ref. [1850]. Many of the steps involved in deriving factorization at leading power are manifest in the construction of SCET; in particular we arrive at hard scattering Lagrangians $\mathcal{L}_{\text{hard}}^{(0)}$ that can be written as products of gauge invariant collinear and soft operators, and we have a direct sum of independent Lagrangians for soft and collinear fields in $\mathcal{L}_{\text{dyn}}^{(0)}$. With just these terms the SCET Hilbert space of states factorizes as direct products, and matrix elements of collinear and soft operators with their Wilson coefficients define independent collinear, soft and hard functions (examples given below). Since $\mathcal{L}_G^{(0)}$ can be inserted any number of times without power suppression, and couples different sectors, it breaks factorization.

Thus proving factorization reduces to demonstrating that contributions from $\mathcal{L}_G^{(0)}$ either cancel out, or can be absorbed into other interactions. Both of these occur. For example, in $e^+e^- \to$ 2-jets the non-trivial interactions from $\mathcal{L}_G^{(0)}$ can be absorbed into the *direction* of the (u)soft and collinear Wilson lines, which in that case then run from $[0, \infty)$ rather than $(-\infty, 0]$, see Ref. [1850]. The same absorption is true for the exclusive $B \to D\pi$ process, with the common feature being that these processes involve only active partons and do not involve forward scattering configurations (see also Refs. [1891,1892]). In a process like Drell–Yan the cancellation of $\mathcal{L}_G^{(0)}$ is much more complicated due to interactions involving spectator partons in the initial protons, but these still cancel out. Low order demonstrations can be found in Refs. [1850,1893,1894], while the all order statement was made in the classic CSS proof of Glauber region cancellations in Ref. [1806]. For cases where factorization is known to be violated [1892,1895–1899], it is not possible to absorb or cancel the effects of $\mathcal{L}_G^{(0)}$ in this manner. The factorization of Glauber effects in SCET can also be used to sum so-called superleading logarithms [1900].

It is worth noting that in SCET the proof of factorization for cross sections and decay rates at subleading power follows the same steps as at leading power. Higher power $\mathcal{L}_{\text{hard}}^{(i\geq 1)}$ simply involve more complicated products of factorized soft and collinear operators. While terms in $\mathcal{L}_{\text{dyn}}^{(i\geq 1)}$ also involve products of soft and collinear fields, they are always inserted only a finite number of times at any given order in the power counting, and hence still lead to factorized matrix elements, albeit with time ordered products of operators. For a gauge invariant description of power suppressed SCET_I operators see Refs. [1901,1902]. Many of these observations go back to the beginnings of SCET, since the processes that people focused on at the time involved exclusive B decays that only start at subleading power [1831–1833,1835–1837,1903–1907], primarily because the soft spectator quark in the B had to be converted into a collinear quark, a subleading power process.

Finally, we remark that $\mathcal{L}_G^{(0)}$ is interesting in its own right, because for processes involving forward scattering rather than hard scattering, it does not cancel but instead provides the dominant contributions, yielding Reggeization, BFKL evolution, and the shockwave picture. For more work in this direction see Refs. [1850–1853]. It is also worth noting that this implies that SCET can potentially provide a framework to parameterize and describe spectator factorization violating contributions to certain hard scattering processes from first principles, though so far very little work has been done in this direction.

6.4.4 *Examples of factorization*

To connect theory and experiment, consider a few examples of factorization formulae that have been derived or studied with SCET. A key attribute of these formulas is that they are determined using only the SCET power expansion, and do not rely on any α_s expansion. First consider $e^+e^- \to$ 2-jets, with a measurement of $\tau = 1 - T$ where T is thrust, working in the dijet limit $\tau \ll 1$. We can relate τ to the sum of the two hemisphere jet masses, $\tau = (m_{Ja}^2 + m_{Jb}^2)/Q^2$, where m_{Ja}^2 and m_{Jb}^2 are each determined by the particles on one side of the plane perpendicular to the thrust axis. Thus $\tau \ll 1$ restricts the invariant mass of the radiation in both hemispheres and forces us into a dijet configuration. Squaring the SCET_I leading power amplitudes obtained from $\mathcal{L}_{\text{hard}}^{(0)}$ in Eq. (6.104), Fierzing the fields of distinct types into independent matrix elements, integrating over phase space with the measurement function, and renormalizing the resulting factorized functions, gives [302,1809–1812]

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\tau} = \sigma_0 H(Q,\mu) Q \int \mathrm{d}\ell \mathrm{d}\ell' J_T(Q^2\tau - Q\ell, \mu) \times S_T(\ell - \ell', \mu) F(\ell', \Lambda). \tag{6.108}$$

Here $H(Q,\mu) = |C^{(0)}(Q,\mu)|^2$ is a hard function encoding virtual corrections (magenta line in Fig. 141), the thrust jet function $J_T = J \otimes J$ combines two jet functions J obtained from the n-collinear or $\bar{n}$-collinear matrix elements (dots on the blue line in Fig. 141), and the full soft function is defined from a vacuum matrix element of usoft Wilson lines. This soft function can be further factorized into two parts, $S_T \otimes F$, where S_T is perturbative (green line in Fig. 141) and F is nonperturbative (brown line in Fig. 141) [1840,1908]. Renormalization group evolution of H_T, J_T, and S_T enables a summation of large Sudakov double logarithms, $\alpha_s \ln^2 \tau$. The state-of-the-art for this resummation is next-to-next-to-next-to-leading logarithmic order (N^3LL), and was first achieved

with SCET [1812]. For $\Lambda \ll Q\tau \ll Q$ the nonperturbative effects from F are power corrections, so the spectrum is dominated by perturbation theory, and is used to obtain high precision fits for $\alpha_s(M_Z)$ [302,303,1909].

DIS, $e^- p \to e^- X$, provides another useful SCET factorization example [1807]. It is simplest to consider in the Breit frame where the virtual photon has $q^\mu = (0,0,0,Q)$. Here the factorization theorem is between hard and collinear modes with $\lambda = \Lambda/Q$, and soft contributions cancel out. A feature of this process is that the hard contributions come from both virtual effects as well as real radiation in X. Therefore matching onto Wilson coefficients $C^{i(0)}_{j\mathrm{DIS}}(\omega_i, Q, \mu)$ takes place at the level of the amplitude squared, and so does the construction of the appropriate SCET operators. These operators involve collinear quarks in $\bar\chi^q_{n,\omega_1}\bar{n}\!\!\!/\chi^q_{n,\omega_2}$ with flavor q, or collinear gluons in $\mathcal{B}^\nu_{n\perp,\omega_1}\mathcal{B}_{n\perp,\omega_2\nu}$. The proton matrix elements $\langle p|\cdots|p\rangle$ of these operators define the well known quark parton distribution functions (PDFs) $f_{q/p}(\xi,\mu)$ and gluon PDFs $f_{g/p}(\xi,\mu)$, respectively. Carrying out the same steps listed above to arrive at Eq. (6.108) now gives factorization theorems for DIS structure functions. For example

$$W_1(x,Q^2) = \frac{-1}{x}\int_x^1 \mathrm{d}\xi\; H_1^{(i)}(\xi/x, Q, \mu)\, f_{i/p}(\xi,\mu^2), \tag{6.109}$$

where we sum over parton types i, and the hard function $H_1^{(i)} = (1/\pi)\,\mathrm{Im}\, C^{i(0)}_{1\mathrm{DIS}}$. There is a similar formula for $W_2(x,Q^2)$. Equation (6.109) factorizes perturbative short distance contributions in $H_1^{(i)}$ at the scale Q from the nonperturbative PDFs $f_{i/p}$ at the scale Λ. Here the renormalization group evolution (RGE) sums up single logs $\alpha_s \ln(Q/\mu_0)$, for a hadronic scale $\mu_0 \simeq 1\,\mathrm{GeV} > \Lambda$. Thus SCET reproduces classic DIS results in a very simple manner. For example, the fact that $\bar\chi^q_{n,\omega_1}\bar{n}\!\!\!/\chi^q_{n,\omega_2} \sim \lambda^2$ is related to the PDFs being built from twist-2 operators. The operator with Wilson lines in SCET captures the full tower of twist-2 operators simultaneously.

To provide a SCET$_{\mathrm{II}}$ example, we consider the Higgs transverse momentum q_T in $pp \to H + X$ in the region where $Q = m_H \gg q_T \gg \Lambda$. Due to the measurement of $q_T \sim Q\lambda$ there is a restriction on the final state X. It can involve collinear and soft particles which individually have $p_T \sim Q\lambda$, but can no longer involve hard particles. Due to this restriction, the hard matching takes place at the amplitude level in this case, giving $\mathcal{L}^{(0)}_{\mathrm{hard}} \propto C^{(0)}_H(\omega_i,\mu)\mathrm{tr}[\mathcal{B}^\nu_{n\perp,\omega_1}\mathcal{S}_n^T\mathcal{S}_{\bar n}\mathcal{B}_{\bar n\perp,\omega_2\nu}]_\mu$, where $\mathcal{S}$ are soft Wilson lines in the adjoint representation. Since this only involves one field of each collinear type, the ω_i momenta are fixed by Q and the Higgs rapidity Y to be $\omega_1 = Qe^Y$ and $\omega_2 = Qe^{-Y}$. Here the factorization is simplest in Fourier space

$$\begin{aligned}
\frac{\mathrm{d}\sigma}{\mathrm{d}Q\mathrm{d}Y\mathrm{d}^2\vec p_T^{\,H}} &= 2H_{ggH}(Q,\mu)\int \mathrm{d}^2\vec b_T\, e^{i\vec b_T\cdot\vec p_T^{\,H}}\,\mathcal{S}_H(b_T,\mu,\nu)\\
&\quad\times B^{\alpha\beta}_{g/p}(x_a,\vec b_T,\mu,\zeta_a/\nu^2)B_{g/p\alpha\beta}(x_b,\vec b_T,\mu,\zeta_b/\nu^2)\\
&= H_{ggH}(Q,\mu)\int \mathrm{d}^2\vec b_T\, e^{i\vec b_T\cdot\vec p_T^{\,H}}\\
&\quad\times\big[f_{1g/p}(x_a,b_T,\mu,\zeta_a)f_{1g/p}(x_b,b_T,\mu,\zeta_b)\\
&\quad + h^\perp_{1g/p}(x_a,b_T,\mu,\zeta_a)h^\perp_{1g/p}(x_b,b_T,\mu,\zeta_b)\big],
\end{aligned} \tag{6.110}$$

where $x_a = Qe^Y/\sqrt{s}$, $x_b = Qe^{-Y}/\sqrt{s}$, s is the invariant mass of the colliding protons, and $\zeta_{a,b}$ are Collins–Soper parameters satisfying $\zeta_a\zeta_b = Q^4$. Here the hard function is $H_{ggH} \propto |C^{(0)}_H|^2$ (leaving out simple kinematic prefactors), the squared $\langle p|\cdots|p\rangle$ matrix element of n-collinear fields yields the beam function $B^{\alpha\beta}_{g/p}$ (and likewise for $\bar n$), and the squared vacuum matrix element of soft Wilson lines yields the soft function $\mathcal{S}_H$. In the final line of Eq. (6.110) we did two things in one step: (i) grouped a $\sqrt{\mathcal{S}_H}$ together with each beam function to absorb the soft function symmetrically, and (ii) decomposed the Lorentz indices $\alpha\beta$ into two possible structures, $g_T^{\alpha\beta} f_{1g/p}$ and $(b_T^\alpha b_T^\beta + \vec b_T^{\,2} g_T^{\alpha\beta}/2)h^\perp_{1g/p}$. This yields definitions for the TMD PDFs $f_{1g/p}$ (unpolarized gluon TMD PDF) and $h^\perp_{1g/p}$ (linearly polarized gluon TMD PDF).

A novel feature of this factorization theorem is the appearance of the rapidity scale ν in the collinear and soft functions, which is associated to the need to regulate rapidity divergences in many SCET$_{\mathrm{II}}$ processes [1890,1910–1913], and the presence of the associated rapidity renormalization group equations [1912,1914]. The result in the first line of Eq. (6.110) is presented with the rapidity regulator defined in [1912] and may look somewhat different with other choices of the rapidity regulator, such as in the original Collins construction [1267]. However the result in the final line will be the same. Evolution in both μ and ν is needed to sum the large logs, $\alpha_s \ln^2(Q/q_T)$, in this process, and the state of the art is resummation at N^{3}LL. This resummation may also be done at the level of the TMD PDFs, where the rapidity RGE is replaced by the Collins–Soper evolution in $\zeta_{a,b}$ [1347].

As our final example, we consider the measurement of jet mass in inclusive jet production, $pp \to$jet$+X$, where the jet has radius R and is defined with the anti-k_T algorithm. To make this example more interesting (and more phenomenologically relevant) we also carry out jet grooming to remove soft contaminating radiation in the jet, using the soft drop algorithm [1915,1916]. Examples of contaminating radiation in the jet include initial state radiation from the protons, underlying event effects due to radiation from spectator partons, and pileup effects due to radiation from the interaction among other protons in the colliding beams. The soft-drop

grooming is defined by iteratively applying a test on transverse momentum p_T and angular separations ΔR of branches i and j in an angular ordered tree formed from particles in the jet: $\min(p_{Ti}, p_{Tj})/(p_{Ti}+p_{Tj}) > z_{\rm cut}(\Delta R_{ij}/R_0)^\beta$ where $z_{\rm cut}, \beta$, and R_0 are soft drop parameters. Branches that fail this test are removed from the tree, thus grooming soft radiation. This causes the soft function for this process to split itself into two parts [1826]: a global soft function sensitive to the scale $Qz_{\rm cut}$ associated with the groomed soft radiation, and a collinear-soft function, S_c^κ, that describes soft radiation that is collimated enough with the jet axis to have been retained by the grooming. The groomed jet mass cross section can be factorized as [1826,1917,1918]

$$\frac{\mathrm{d}\sigma}{\mathrm{d}m_J^2 \mathrm{d}\Phi_J} = N_\kappa(\Phi_J, R, z_{\rm cut}, \beta, \mu) Q_{\rm cut}^{\frac{1}{1+\beta}} \int \mathrm{d}\ell \times J_\kappa(m_J^2 - Q\ell, \mu) S_c^\kappa\left[\ell Q_{\rm cut}^{\frac{1}{1+\beta}}, \beta, \mu\right], \quad (6.111)$$

with a sum on $\kappa = q, g$ for quark and gluon jets and $Q_{\rm cut} = p_T R z_{\rm cut}(R/R_0)^\beta$. Here J_κ is the usual jet function since the collinear radiation is not affected by the grooming. The normalization factor N_κ is a short hand for a combination of terms that include PDFs, a hard-collinear function describing the production of the parton κ, and the global soft function. This is an example of a SCET$_+$ factorization formula due to the presence of soft-collinear modes that make up the Wilson lines that appear in S_c^κ. Groomed observables have become widely used in predictions at hadron colliders due to the fact that they are much more robust to contamination, and have reduced hadronization corrections. Other examples of soft drop groomed calculations with SCET are found in Refs. [1828,1830,1918–1933].

6.4.5 State-of-the-art and attractive directions

The nature of a short review is that key ideas can be highlighted, but it is hard to do credit to the depth of work in the field. Let me close by giving a brief overview of some of the interesting centers of activity currently going on with SCET, with an eye to the future.

SCET continues to have a significant impact on the field of high precision calculations for collider cross sections, in particular for the resummation of large logarithms. This activity is motivated by the clear universality of anomalous dimensions and factorized functions in SCET, giving results of broad utility. Below I summarize the highest order results achieved to date for various processes which exploit these perturbative achievements, referring to references therein for further background and details. This list includes: e^+e^- thrust to N^{3}LL$'$ [302,1812] and massive thrust to N^{3}LL [1934], e^+e^- heavy jet mass to N^{3}LL [1935], e^+e^- C-parameter to N^{3}LL$'$ [1936], e^+e^- Energy-Energy-Correlator (EEC) to N^{3}LL$'$ [1937], e^+e^- oriented event shapes to N^{3}LL [1938], e^+e^- groomed jet mass to N^{3}LL [1939], $e^+e^- \to t\bar{t}$ thrust to N^{3}LL [1940], e^-p DIS thrust to N^{3}LL [1941–1944], the Drell–Yan $p_T^{\ell\ell}$ spectrum to N^{3}LL$'$ [1945,1946], the pp Higgs p_T^H spectrum [1947–1949] and rapidity spectrum [1950] to N^{3}LL$'$, and LHC processes with a jet-veto [1813,1951–1956]. Recently the first N^{4}LL resummed calculation has been carried out for the EEC [1957] (with an approximation for the 5-loop cusp anomalous dimension). Key ingredients are the four-loop hard (collinear) anomalous dimensions [1958,1959], the four-loop rapidity anomalous dimension for TMDs [1957,1960], the four-loop cusp anomalous dimension [1961] and five-loop approximation [1962], and calculations of three loop boundary conditions [1963–1966]. Many more processes have been resummed to NNLL or NNLL$'$ order with SCET; for example in Refs. [1514,1814,1815,1818,1918,1931,1963,1967–1999]. Factorized functions remain important targets for future perturbative calculations, with the anticipated reward of simultaneously impacting multiple processes.

Power corrections are another lively topic in SCET, from the continued activity around B-decays, to recent significant results for collider physics. A key strength of SCET is its systematic nature, ensuring one can target the desired terms without missing contributions. Recent collider physics literature on subleading power results in SCET includes: formalism such as enumerating operator bases [1886–1888,2000,2001], hard renormalization and evolution [2000,2002–2004], collinear and soft renormalization and evolution [2005–2008], subleading power factorization [1808,2009–2013], and resummation for collider observables, including for event shapes [2005,2014,2015], for threshold resummation [2016–2018], and for the EEC [2019]. These results provide bright prospects for the future, with the ultimate goal of building a complete story for the structure of gauge theories like QCD beyond leading power, and thus generalizing the leading power picture of collinear splittings and soft eikonal radiation.

One popular method for carrying out fixed order calculations at higher orders, is that of slicing, whereby a resolution variable is used to act as a physical regulator for infrared divergences, enabling analytic and numerical calculations to be combined in a systematic way. SCET has contributed to this program with the invention of N-jettiness subtractions [2020,2021] based on the N-jettiness event shape variable [1816]. It has also been used to calculate power suppressed large logarithms, enabling order-of-magnitude improvements to slicing techniques [1889,2006,2022–2029]. Further improvements to such techniques will be important as theorists continue to move towards calculating experimentally accessible fiducial cross sections.

Other interesting applications of SCET include: the generalization of threshold factorization formulae to include collinear limits [2030,2031], the computation of non-global logarithms and associated effects [1824,2032–2045], the parametrization of hadronization corrections with field theory matrix elements [2046–2048], studying fragmentation inside a jet [1921,1925,1997,2049–2061], and to studying double parton distributions and fragmentation [2062–2065]. A particularly interesting direction with many connections to other fields is the study of EECs. Results from SCET include deriving factorization for the back-to-back limit [2066], and collinear limit [2067,2068], jet analyses with charged tracks [2069], generalizing factorization to the back-to-back limit at hadron colliders [2070], and deriving factorization formula for jet substructure applications of the EEC [2071]. The prospects for new applications of SCET technology remain bright.

A final hallmark of SCET is the use of the physical picture it provides to construct novel observables. Past examples of this type include: beam thrust and functions [1844, 2072,2073], N-jettiness [1816], N-subjettiness [2074], jet substructure for disentangling color and spin in J/Ψ production [2054,2075], D_2 and related jet-substructure observables [1821,2076–2078], the winner-takes-all-axis for jets [1929,2079–2081], track functions [2082–2085], the XCone jet algorithm [2086], collinear drop [1927,2087], an EEC probe of top mass [2088], and measuring initial state tomography with a Nuclear EEC [2089]. I look forward to many more examples of such new observables in the future.

6.5 Hard thermal loop effective theory

Michael Strickland

In this section we review progress in understanding QCD at finite temperature and density. Unlike QCD in vacuum new classes of physical infrared divergences appear which cause naive perturbation theory to break down. Luckily, at least at leading order in the coupling constant, it is possible to identify a class of diagrams that must be resummed in order to cure these divergences.

6.5.1 The breakdown of naive perturbation theory at finite temperature

There are two fundamental formalisms for computing the properties of QCD at high temperature: (1) the real-time formalism and (2) the imaginary-time formalism [2090–2092]. The former is necessary when considering systems that are out of equilibrium, while the second is more convenient for computing bulk thermodynamic quantities. Here we will focus on the imaginary-time formalism and progress that has been made in understanding how to reorganize the perturbative expansion of finite temperature QCD in order to deal with infrared singularities which emerge in this case using self-consistent inclusion of Debye screening and Landau damping. This is accomplished through an all-orders resummation of a class of diagrams referred to as the hard-thermal-loop (HTL) diagrams. For an introduction to the real-time formalism and applications to real-world calculations we refer the reader to Sect. 6.6 and Ref. [2091].

In thermal and chemical equilibrium with temperature T and quark chemical potentials μ_i with $\pi T \gg \mu_i$, one finds that the naive loop expansion of physical quantities is ill-defined and diverges beyond a given loop order, which depends on the quantity under consideration. In the calculation of QCD thermodynamics, this stems from uncanceled infrared (IR) divergences that enter the expansion of the partition function at three-loop order. These IR divergences are due to long-distance interactions mediated by static gluon fields and result in contributions that are non-analytic in $\alpha_s = g^2/4\pi$, e.g., $\alpha_s^{3/2}$ and $\log(\alpha_s)$, unlike vacuum perturbation expansions which involve only powers of α_s.

A simple way to understand at which perturbative orders terms non-analytic in α_s appear is to start from the contribution of non-interacting static gluons to a given quantity. For the pressure of a gas of gluons one has $P_{\rm gluons} \sim \int d^3p\, p\, n_B(E_p)$, where n_B denotes a Bose–Einstein distribution function and E_p is the energy of the in-medium gluons. The contributions from the momentum scales πT, gT and g^2T can be expressed as

$$P_{\rm gluons}^{p\sim\pi T} \sim T^4 n_B(\pi T) \sim T^4 + \mathcal{O}(g^2), \tag{6.112}$$

$$P_{\rm gluons}^{p\sim gT} \sim (gT)^4 n_B(gT) \sim g^3 T^4 + \mathcal{O}(g^4), \tag{6.113}$$

$$P_{\rm gluons}^{p\sim g^2T} \sim (g^2T)^4 n_B(g^2T) \sim g^6 T^4, \tag{6.114}$$

where we have using the fact that $n_B(E) \sim T/E$ if $E \ll T$. This fact is of fundamental importance since it implies that when the energy/momentum are *soft*, corresponding to electrostatic contributions, $p_{\rm soft} \sim gT$, one receives an enhancement of $1/g$ compared to contributions from *hard* momenta, $p_{\rm hard} \sim T$, due to the bosonic nature of the gluon. For *ultrasoft* (magnetostatic) momenta, $p_{\rm ultrasoft} \sim g^2T$, the contributions are enhanced by $1/g^2$ compared to the naive perturbative order. As the Eqs. (6.112)–(6.114) demonstrate, it is possible to generate contributions of the order $g^3 \sim \alpha_s^{3/2}$ from soft momenta and, in the case of the pressure, although perturbatively enhanced, ultrasoft momenta only start to play a role at order $g^6 \sim \alpha_s^3$.

Note that the expansion parameters in Eqs. (6.112)–(6.114) are of order $g^2 n_B(\pi T) \sim g^2$, $g^2 n_B(gT) \sim g$, and $g^2 n_B(g^2T) \sim 1$, implying in particular that the contribution of magnetostatic gluons to the pressure is fundamentally non-perturbative in nature at $\mathcal{O}(\alpha_s^3)$, which for the pressure corresponds to four-loop order. This complete breakdown of the loop expansion at the ultrasoft scale is called the *Linde*

problem [2093,2094]. The specific order at which the expansion breaks down depends on the quantity under consideration and is not universal. For example, in Ref. [2095], the authors demonstrated that a certain second-order transport coefficient, λ_1, receives a leading-order contribution from the ultrasoft scale. We also note that, in the case of the $\mathcal{O}(\alpha_s^3)$ contribution to the pressure it is possible to isolate the purely non-perturbative contribution and compute this numerically using a three-dimensional lattice calculation [2096]. Paradoxically, the difficult part then becomes computing the perturbative contributions at this order [2097]. Beyond four-loop order, all contributions are once again perturbatively computable.

As a result of the infrared enhancement of electrostatic contributions it was shown that a class of diagrams called hard-thermal-loop (HTL) graphs which have soft external and hard internal momenta need to be resummed to all orders in the strong coupling coupling [2098–2100]. In the high temperature limit, there exist several schemes for carrying out such resummations, see e.g. [2101–2119]. Here we will briefly review the method of dimensionally reduced effective theories (EFTs), which take advantage of the scale hierarchies and the manifestly gauge-invariant hard thermal loop perturbation theory (HTLpt) resummation. This makes use of the HTL effective action to reorganize the perturbative expansion of finite temperature and density QCD [2120].

6.5.2 Dimensional reduction and QCD EFT

The method of dimensional reduction is based on the fact that, at weak-coupling, there is a hierarchy of scales between the three energy scales (hard, soft, and ultrasoft or, equivalently, hard, electric, and magnetic) which contribute to bulk thermodynamic observables. Specifically, if $g \ll 1$, one has

$$m_{\text{magnetic}} \sim g^2 T \ll m_{\text{electric}} \sim gT \ll m_{\text{hard}} \sim \pi T. \tag{6.115}$$

Above we have denoted the magnetostatic and electrostatic screening scales by m_{magnetic} and m_{electric}, respectively, and the hard or thermal one, corresponding to the lowest non-zero Matsubara frequency, by m_{hard}. To leading order, the electrostatic screening mass can be computed from the IR limit of the A_0 one-loop self energy, however, the magnetic screening mass cannot be computed perturbatively [2104,2105]. In the high temperature limit with $\pi T \gg \mu_i, m_i, \Lambda_{\text{QCD}}$, the above three scales are the only ones appearing and the two non-trivial scales m_{magnetic} and m_{electric} are connected to the static sector corresponding to the zero Matsubara mode ($n = 0$). As a result, in the effective field theory language it is natural to integrate out the hard scale, yielding a three-dimensional effective field theory which is valid for long-distance static field modes. Another way to see this is recognize that, in four-dimensional Euclidean space, a system in thermal equilibrium has its time direction compactified to a circle of radius $1/T$ [2090]. In the high-temperature limit, the Euclidean time direction has zero extent and the parent field theory becomes effectively three dimensional. Since fermionic modes have odd Matsubara frequencies, they become super massive and decouple from the theory in this limit, as do all non-zero gluonic Matsubara modes.

The construction of dimensionally reduced effective theories for high-temperature field theory began with the work of Ginsparg [2121] and was quickly followed by Appelquist and Pisarski [2122]. In the mid-1990s, Kajantie et al. were the first to apply this formalism to the study of the electroweak phase transition [2123]. Around the same time Braaten and Nieto demonstrated how to apply these ideas to thermal QCD [2104,2105]. Recently, these methods have been extended to the computation of the thermodynamics of $\mathcal{N} = 4$ supersymmetric Yang–Mills theory to order λ^2, where $\lambda = g^2 N_c$ is the t'Hooft coupling [2124].

In the EFT technique, the Lagrangian densities of the three- and four-dimensional theories can be obtained by writing down the most general local Lagrangians respecting all necessary symmetries. One then orders all operators in terms of their dimensionality and truncates the Lagrangians at the desired order. For electrostatic QCD (EQCD), this procedure results in [2104,2105]

$$\begin{aligned}\mathcal{L}_{\text{E}} = &\frac{1}{2}\text{Tr}\mathbf{F}_{ij}^2 + \text{Tr}[\mathbf{D}_i, \mathbf{A}_0]^2 + m_{\text{E}}^2 \text{Tr}\mathbf{A}_0^2 \\ &+\lambda_{\text{E}}^{(1)}(\text{Tr}\mathbf{A}_0^2)^2 + \lambda_{\text{E}}^{(2)}\text{Tr}\mathbf{A}_0^4 \\ &+i\lambda_{\text{E}}^{(3)}\text{Tr}\mathbf{A}_0^3 + \cdots,\end{aligned} \tag{6.116}$$

where the adjoint fields $\mathbf{A}_i \equiv A_i^A T^A$, $\mathbf{A}_0 \equiv A_0^A T^A$ are three dimensional, $F_{ij}^A = \partial_i A_j^A - \partial_j A_i^A + g_{\text{E}} f^{ABC} A_i^B A_j^C$, $\mathbf{F}_{ij} \equiv F_{ij}^A T^A$, and $\mathbf{D}_i = \partial_i - i g_{\text{E}} \mathbf{A}_i$. Integrating out the temporal gauge field, one can obtain the magnetostatic effective theory (MQCD) with $\mathcal{L}_{\text{M}} = \frac{1}{2}\text{Tr}\mathbf{F}_{ij}^2 + \cdots$ and $F_{ij}^A = \partial_i A_j^A - \partial_j A_i^A + g_{\text{M}} f^{ABC} A_i^B A_j^C$ [2104,2105].

At leading order in g, the degrees of freedom in the above effective theories are the $n = 0$ Matsubara modes of the four-dimensional A_i and A_0 fields. The former transforms as a three-dimensional adjoint gauge field and the latter as a scalar in the adjoint representation of SU(N_c). By computing the contributions from the hard scale in the four-dimensional theory (non-resummed), the massless two-loop self-energy in the four-dimensional theory, and the massive three-loop vacuum graphs and matching the two theories, one obtains the following result for the QCD free energy through $\mathcal{O}(\alpha_s^{5/2})$

$$\frac{\mathcal{F}_{\text{QCD}}}{\mathcal{F}_{\text{ideal}}} = 1 - \frac{15}{4}\frac{\alpha_s}{\pi} + 30\left(\frac{\alpha_s}{\pi}\right)^{3/2}$$

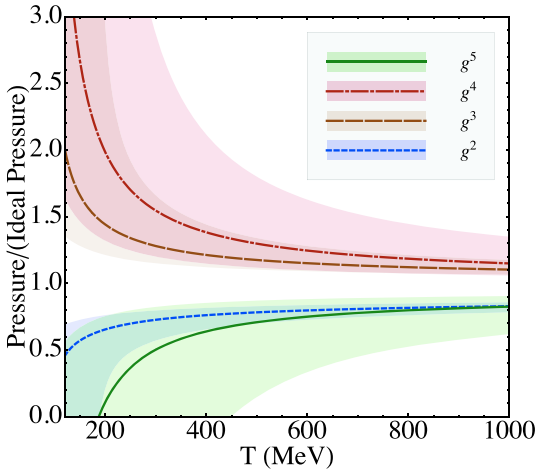


Fig. 143 Naive weak-coupling expansion of the scaled QCD pressure for $N_f = 3$. Shaded bands show the result of varying the renormalization scale Λ by a factor of 2 around the central renormalization scale $\Lambda = 2\pi T$

$$+\frac{135}{2}\left(\log\frac{\alpha_s}{\pi} - \frac{11}{36}\log\frac{\Lambda}{2\pi T} + 3.51\right)\left(\frac{\alpha_s}{\pi}\right)^2 + \frac{495}{2}\left(\log\frac{\Lambda}{2\pi T} - 3.23\right)\left(\frac{\alpha_s}{\pi}\right)^{5/2} + \mathcal{O}(\alpha_s^3\log\alpha_s), \tag{6.117}$$

where $\mathcal{F}_{\text{ideal}} = -(8\pi^2/45)T^4$ is the free energy of an ideal gas of massless gluons and $\alpha_s = \alpha_s(\Lambda)$ is the running coupling constant in the $\overline{\text{MS}}$ scheme. Note, importantly, the appearance of non-analytic terms in α_s. Logarithms of α_s appear as ratios of the electric screening scale over the temperature. In order to avoid notational overlap with the chemical potential μ, here Λ is used to indicate the renormalization scale. There is also a residual dependence on the renormalization scale Λ at orders α_s^2 and $\alpha_s^{5/2}$. The result obtained when this expression is truncated at various orders in the coupling constant is shown in Fig. 143. As can be seen from this figure, at phenomenologically relevant temperatures the resulting weak coupling expansion shows poor convergence and an increasing sensitivity to the renormalization scale as the perturbative truncation order is increased. The reason for this poor convergence is that one is expanding around the $T = 0$ QCD vacuum, which does not include the effects of Debye screening and Landau damping. In order to improve the convergence of this series, HTLpt was introduced to reorganize the calculation instead around the $T \to \infty$ limit. We will discuss this reorganization in the next subsection.

6.5.3 Hard-thermal loop perturbation theory

Hard-thermal-loop perturbation theory is a reorganization of perturbative QCD. The HTLpt Lagrangian density is written as [2114,2115]

$$\mathcal{L} = (\mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{HTL}})\big|_{g\to\sqrt{\delta}g} + \Delta\mathcal{L}_{\text{HTL}}, \tag{6.118}$$

where $\Delta\mathcal{L}_{\text{HTL}}$ collects all necessary renormalization counterterms and δ is a formal expansion parameter, which will be taken to be unity in the end of the calculation. The HTL improvement term appearing above is

$$\mathcal{L}_{\text{HTL}} = (1-\delta)im_q^2\bar{\psi}\gamma^\mu\left\langle\frac{y_\mu}{y\cdot\mathbf{D}}\right\rangle_{\hat{\mathbf{y}}}\psi - \frac{1}{2}(1-\delta)m_D^2\text{Tr}\left[\mathbf{F}_{\mu\alpha}\left\langle\frac{y^\alpha y_\beta}{(y\cdot\mathbf{D})^2}\right\rangle_{\hat{\mathbf{y}}}\mathbf{F}^{\mu\beta}\right]. \tag{6.119}$$

Above $y^\mu = (1, \hat{\mathbf{y}})$ is a light-like four-vector with $\hat{\mathbf{y}}$ being a three-dimensional unit vector and the angular bracket indicates an average over the direction of $\hat{\mathbf{y}}$. The parameters m_D and m_q can be identified with the gluonic screening mass and the thermal quark mass. In HTLpt one treats δ as a formal expansion parameter. By including the HTL improvement term (6.119) HTLpt shifts the perturbative expansion from being around an ideal gas of massless particles to being around a gas of massive quasiparticles. This shift dramatically improves the convergence of the successive loop approximations to QCD thermodynamics [2107–2109,2114–2119].

The HTLpt Lagrangian (6.118) reduces to the QCD Lagrangian when $\delta = 1$. Physical observables are calculated in HTLpt by expanding in powers of δ, truncating at some specified order in δ, and then setting $\delta = 1$. This defines a reorganization of the perturbative series in which the effects of m_D^2 and m_q^2 terms in (6.119) are included to leading order but then systematically subtracted out at higher orders in perturbation theory by the δm_D^2 and δm_q^2 terms in (6.119). To obtain leading order (LO), next-to-leading order (NLO), and next-to-next-leading order (NNLO) results for the QCD pressure, one expands to orders δ^0, δ^1, δ^2, respectively. Note, importantly, that HTLpt is gauge invariant order-by-order in the δ expansion.

In order to obtain analytically tractable sum-integrals, in addition to the δ expansion, one must also make a Taylor expansion in the mass parameters scaled by the temperature, m_D/T and m_q/T. The final result obtained at NNLO is completely analytic, however, it is too lengthy to list here, instead we refer the reader to the most recent works using HTLpt, which apply this technique at finite temperature and quark chemical potentials [2118,2119,2131]. In Fig. 144 we compare the NNLO EQCD and HTLpt results for the scaled pressure (negative of the free energy). As can be seen from this figure, for the central choice of the renormalization scale, namely $\Lambda_g = 2\pi T$ and $\Lambda_q = \pi T$, there is excellent agreement between HTLpt and the lattice data. The same is true to a lesser extent for EQCD. Both, however, have a large uncertainty related to the variation with respect to the renor-

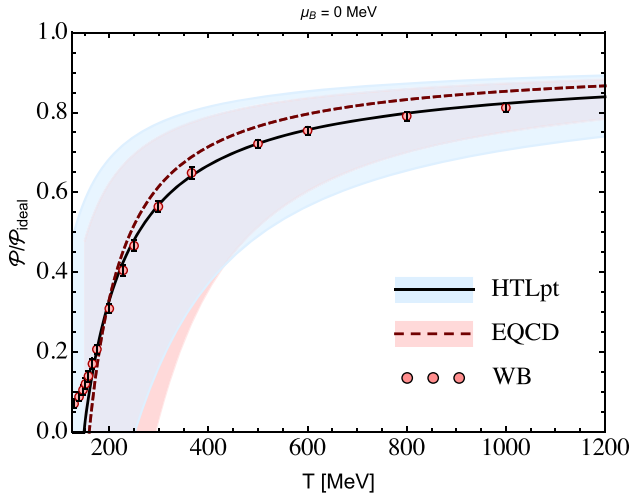


Fig. 144 The resummed QCD pressure for $\mu_B = 0$. We compare the three-loop EQCD and HTLpt results with lattice data from the Wuppertal–Budapest (WB) collaboration [2125]

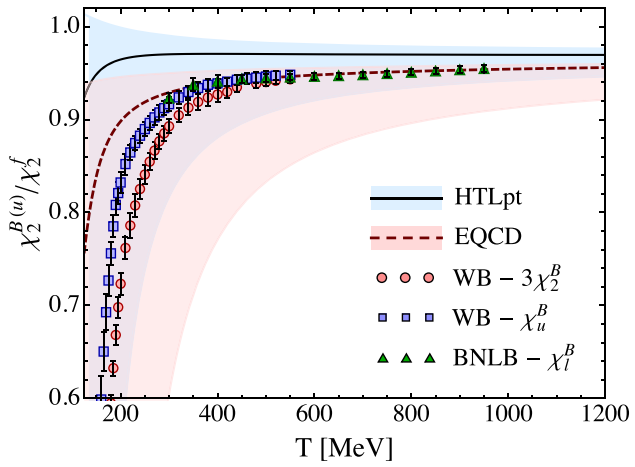


Fig. 145 The second-order light quark (and baryon) number susceptibilities. Lattice data are from the Wuppertal–Budapest (WB) [2126,2127] and BNLB collaborations [2128]

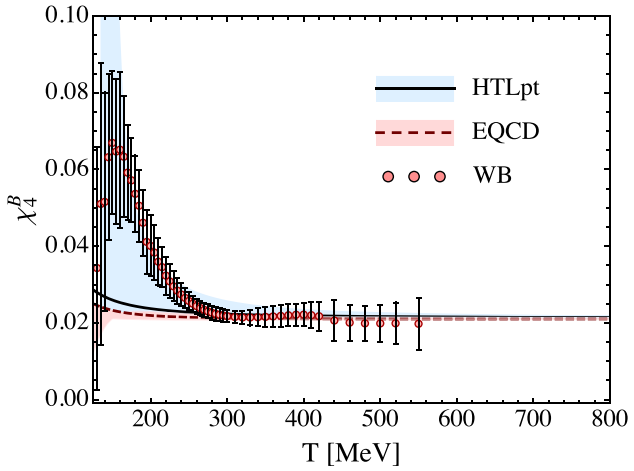


Fig. 146 The 4th baryon number susceptibility. Lattice data sources are the same as Fig. 145

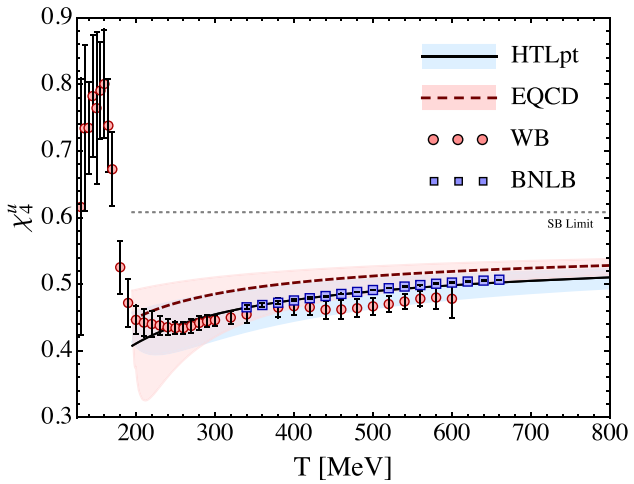


Fig. 147 The 4th light quark number susceptibility. Lattice data sources are the same as Fig. 145

malization scale. This sensitivity is particularly large for the free energy; however other quantities show much less renormalization scale dependence. From the NNLO results, one can obtain predictions for various quark and baryon number susceptibilities.

In Figs. 145, 146, 147 we present the NNLO resummed perturbative predictions for the second-order baryon number susceptibility, the fourth-order baryon number susceptibility, and the fourth-order light quark susceptibility, respectively. As these figures demonstrate HTLpt and EQCD to a only slightly lesser extent, have reasonable agreement with lattice extractions of these susceptibilities down to temperatures on the order of $T \sim 250$ MeV which is only slightly higher than the QCD phase transition temperature of $T_c \sim 155$ MeV. The lone exception is $\chi_2^{B(u)}$ where EQCD seems to perform better than HTLpt, although the results are consistent within the scale uncertainties. Finally, in Figs. 148 and 149 we present results recently presented in Ref. [2131] for the second- and fourth-order curvatures of the QCD phase transition line obtained from the analytical NNLO HTLpt result and the world's compiled lattice QCD data. We display three different physical cases which correspond to (1) equal quark chemical potentials, (2) zero strange quark chemical potential, and (3) the case $\langle S \rangle = 0$ and $Q/B = 0.4$, which corresponds to the case appropriate to heavy-ion collisions. As can be seen from these figures, NNLO HTLpt agrees quite well with the existing lattice data in each case. The horizontal error bars (which are sometimes not even visible) indicate the renormalization scale dependence of these curvatures.

To close this section, we have demonstrated that although naive perturbative expansions applied to QCD thermodynamics fail dramatically, it is possible reorganize the cal-

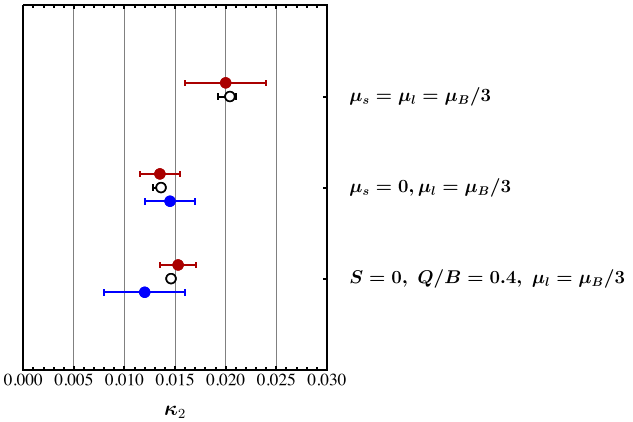


Fig. 148 Filled circles are lattice calculations of κ_2 [485–487,2129, 2130], from top to bottom, respectively. Red filled circles are results obtained using the imaginary chemical potential method and blue filled circles are results obtained using Taylor expansions around $\mu_B = 0$. Black open circles are the NNLO HTLpt predictions. The error bars associated with the HTLpt predictions result from variation of the assumed renormalization scale

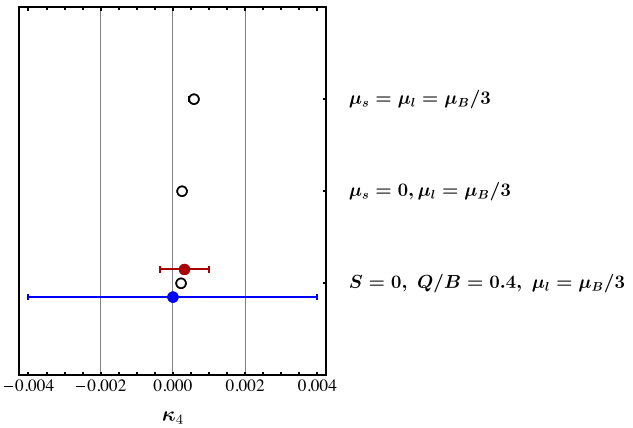


Fig. 149 Filled circles are lattice calculations of κ_4 from Refs. [487, 2130], from top to bottom, respectively. The color coding etc. for the symbols is the same as in Fig. 148

culation of the QCD free energy in such a way as to achieve improved convergence at phenomenologically relevant temperatures. Interestingly, we find excellent agreement between the resummed approaches and lattice data down to rather low temperatures and are even able to predict the curvature of the QCD phase transition line using perturbation theory.

6.6 EFT methods for nonequilibrium systems

Miguel Escobedo

6.6.1 Introduction

There are many situations in which we are interested in describing non-equilibrium phenomena that involve the strong interaction. An example is the study of the medium

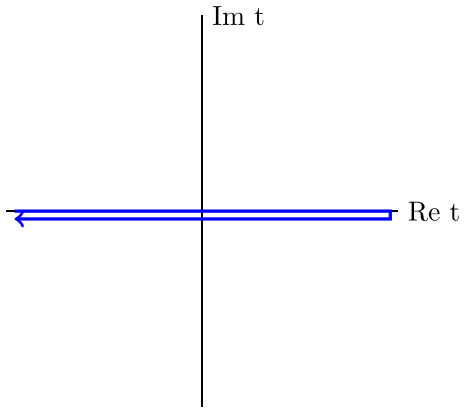


Fig. 150 Schwinger–Keldysh contour of the real time formalism

created when colliding heavy ions at ultrarelativistic speeds. This kind of experiment is nowadays performed at facilities like the Large Hadron Collider (LHC) in Geneva and the Relativistic Heavy Ion Collider (RHIC) in Brookhaven. The motivation is to study a new state of matter that appears at high temperatures and densities, the Quark Gluon Plasma (QGP). More details are given in Sect. 4.4. The medium created in heavy ion collisions can be regarded as an out-of-equilibrium system. Soft particles in the medium are able to approximately thermalize [2132,2133]; however, this thermalization is only local. Looking at length and time scales much larger than the inverse of the temperature the bulk properties of the medium are well described by relativistic hydrodynamics [2134–2136].

One important way to obtain information about the QGP created in heavy ion collisions is by studying its effects on hard probes, for example, heavy quarkonium suppression and jet quenching [1601,2137]. We can regard these particles as out-of-equilibrium probes interacting with a thermal equilibrium environment of soft particles. Precisely because they do not have time to thermalize inside of the medium, they allow us to characterize the QGP in a way that would not be possible otherwise. For example, they are sensitive to transport properties of the medium such as the heavy quark diffusion coefficient [2138] and the jet broadening parameter $\hat{q}$ [2137]. Note also that the problem of a hard probe interacting with a soft medium is one in which a hierarchy of well-separated energy scales appear. This is precisely the situation in which EFTs are useful. In summary, the study of hard probes in heavy ion collisions provides a clear motivation to study EFTs far from equilibrium.

The theoretical description of a QFT out of equilibrium requires the use of the real-time formalism [2139]. When dealing with $T = 0$ scattering process, we are used to assuming that the system is in the ground state both at the remote past and in the distant future. This is what is done to obtain the LSZ reduction formula [2140]. The consequence of this

is that, when computing amplitudes, all field insertions are chronologically ordered. The situation is completely different when the initial state of the system is described by a given density matrix. In this case, the state of the system in the distant future is unknown, so we have to average over all the possible outcomes imposing that the system is described by the initial density matrix in the remote past. For this reason, the real-time formalism is sometimes called an *in–in* formalism while the formalism leading to the LSZ reduction formula is called an *in–out* formalism. As a consequence, the path integral needs to go from the remote past to the distant future and back again around a path called the Schwinger–Keldysh contour (see Fig. 150) [2139]. Fields in the upper (lower) branch of the contour are chronologically (anti-chronologically) ordered and are customary labelled fields of type 1 (2).[73]

The doubling of degrees of freedom discussed in the previous paragraph can affect the construction and use of an EFT in two different ways depending on whether or not medium degrees of freedom are integrated out when going from the full theory to the EFT. If the matching is not affected by the medium, then we can apply the real time formalism in exactly the same way as it is done for a normal QFT. However, if the matching is affected by the medium, we can not assume that the EFT does not contain terms mixing the two branches of the Schwinger–Keldysh contour. Recently, this issue has been discussed in detail in the context of the construction of an EFT for hydrodynamics [2141–2144]. However, regarding the study of hard probes, the complications arising from the doubling of degrees of freedom are substantially diminished when we take into account that only few of them are created in each heavy ion collision. We will discuss in detail how the dilute nature of heavy quarks and high-energy partons simplify their study in the real time formalism.

Lastly, there is another aspect of the application of EFTs to the study of hard probes out of equilibrium that we would like to highlight. This is the connection that naturally appears with the formalism of open quantum systems (OQSs) [2145]. The OQS formalism studies the evolution of quantum systems interacting with an environment, that at the same time is also a quantum system. The central object of study is the reduced density matrix, obtained from the density matrix of the combination of the system plus the environment after performing a trace over the degrees of freedom of the environment. The evolution of the reduced density matrix is not necessarily of the quantum Liouville type as there might appear terms that increase its von Neumann entropy. It happens that, when studying hard probes interacting with a medium using EFTs, one typically finds equations that are well known in the context of OQSs. This is not surprising since when we compute how thermal propagators influence the evolution of a hard probe we are actually making a trace over environment degrees of freedom.

In summary, in this section we are going to discuss the application of EFTs to study nonequilibrium phenomena. In particular, we will focus on interesting problems that appear in the study of heavy ion collisions in which a large separation of energy scales appear. First, we will review the open quantum system formalism. This will allow us to discuss the Lindblad equation, which will play a key role in the later discussion. Then, we will discuss the application of EFTs to the study of quarkonium suppression. More specifically, we will study the evolution of the reduced density matrix of heavy quarks using pNRQCD. In another subsection, we will review the description of jet broadening based on the study of the reduced density matrix using SCET. Finally, we will review applications of the EFT to study hydrodynamics and the interesting structure regarding the doubling of degrees of freedom that have been discovered in this context.

6.6.2 Open quantum systems

Let us consider a *universe* formed by a system plus an environment. Let us assume that at some initial time the density matrix of the *universe* $\rho_U(t_0)$ fulfills

$$\rho_U(t_0) = \rho_S(t_0) \otimes \rho_E(t_0), \tag{6.120}$$

where S corresponds to the system and E to the environment. The motivation for this assumption is twofold. On one hand, when studying a dilute hard probe interacting with a medium, it is natural to assume that the medium acts as a thermal reservoir that is not affected by the probe. On the other hand, even if the assumption is not true, any density matrix for the *universe* can be decomposed as a sum of density matrices that do fulfil this structure. The reduced density matrix at time t_0 is also $\rho_S(t_0)$.

$$\rho_S(t_0) = \mathrm{Tr}_E(\rho_U(t_0)). \tag{6.121}$$

Let us now look at what happens at $t > t_0$. If $U(t, t_0)$ is the time evolution operator of the *universe*, then

$$\rho_U(t) = U(t, t_0)\rho_U(t_0)U^\dagger(t, t_0), \tag{6.122}$$

and it follows that

$$\rho_S(t) = \mathrm{Tr}_E(U(t, t_0)\rho_U(t_0)U^\dagger(t, t_0)). \tag{6.123}$$

However, in general it is not true that

$$\rho_S(t) \neq U_S(t, t_0)\rho_S(t_0)U_S^\dagger(t, t_0), \tag{6.124}$$

so in this sense we can say that the evolution is non-unitary.

The equation that describes the time evolution of ρ_S is called a master equation. In general, it is not trivial to determine the form of this equation. However, an important result

[73] From a complementary point of view, fields of type 1 (2) act on the left (right) of the density matrix.

of the OQS formalism is that the master equation of a Markovian evolution that preserves the fundamental properties of a density matrix (Hermitian positive-definite operator with trace equal to 1) takes the form of a GKSL or Lindblad equation [2146,2147]

$$\frac{d\rho_S}{dt} = -i[H, \rho_S] + \sum_i \left(C_i \rho_S C_i^\dagger - \frac{1}{2}\{C_i^\dagger C_i, \rho_S\} \right), \tag{6.125}$$

where C_i are the collapse or Lindblad operators. They are operators that encode the dissipative part of the Lindblad equation and will depend on the problem we are studying. Let us note that it is very computationally expensive to solve the GKSL equation, as it is generally the case for any master equation. The reason is that the cost scales with N^2, where N is the dimension of the Hilbert space. This means that, if we discretize the QCD system in a lattice, doubling the lattice size multiplies the numerical cost by four. This problem can be solved by using techniques called unravelling of the master equation. Examples of unravellings used to study quarkonium suppression are the Quantum State Diffusion [2148] and the Quantum Trajectories method [1602,2149–2151].

6.6.3 EFTs for quarkonium suppression

Quarkonium suppression was proposed as a probe of the formation of a QGP in the pioneering work of Matsui and Satz [2152]. The original proposal was based on the phenomenon of color screening. Chromoelectric fields are screened at large distances in the presence of a QGP. This modifies the heavy quarkonium potential and, if the screening length is smaller than the size of the bound state, inhibits bound state formation. Later on, it was realized that the potential develops an imaginary part in the presence of a QGP [2153]. This is related to the appearance of a thermal induced decay width which can dissociate quarkonium in many cases more efficiently than screening. However, before asking which phenomenon more substantially modifies the heavy quarkonium potential, we should understand whether quarkonium's evolution follows a Schrödinger equation at all in the presence of a medium and what is the definition of the potential. In Sect. 6.1, we have seen that similar issues can be addressed using non-relativistic EFTs such as NRQCD and pNRQCD at $T = 0$. Therefore, it is reasonable to expect that the finite temperature versions of these EFTs will allow us to answer the previous questions.

In order to construct an EFT, we should first discuss the energy scales and the symmetries of the problem. In addition to the hard, soft and ultrasoft scales that already appear when studying quarkonium at $T = 0$, we should also consider the energy scales induced by the presence of the medium. One of the energy scales that obviously appears is the temperature itself. However, in a weakly-coupled plasma ($g \ll 1$), other dynamically generated energy scales appear. For example, the Debye mass (of order gT) and the non-perturbative magnetic mass (of order g^2T). More details about these scales can be found in Sect. 6.5. Depending on the relation between the medium induce energy scales and those that already appear at $T = 0$, we will find different physical situations. For example, if the Debye mass is much larger than the inverse of the Bohr radius, there would be no bound state formation due to screening. On the other hand, if the temperature is smaller than the inverse of the Bohr radius, thermal effects are a perturbation compared with the binding energy because the medium sees quarkonium as a small color dipole.

Regarding the symmetries of the problem, we will focus on the scenario in which quarkonium is co-moving with the medium. Note, however, that there are EFT studies considering the finite velocity case [2154,2155]. In the co-moving case, the medium only breaks Lorentz symmetry. Note that in $T = 0$ NRQCD and pNRQCD, Lorentz symmetry is not explicit. It manifests through relations between the Wilson coefficients of different operators [1437,1438]. These relations are broken in the presence of a medium [2156].

Now, let us discuss how the doubling of degrees of freedom influences the use of non-relativistic EFTs. First, consider the thermal equilibrium case. Since the mass of the heavy quark M is much larger than the temperature T, it follows that the thermal modifications of the heavy quark propagator in NRQCD or the singlet propagator in pNRQCD are suppressed by the Boltzmann factor $e^{-M/T}$. This reflects the fact that physically heavy particles are dilute in a thermal equilibrium medium that has a temperature much lower than M. We are interested in the more general case in which the heavy particles are not in thermal equilibrium. However, we will still consider that heavy particles are dilute. This is clearly the case for bottom quarks at LHC since only a few of them are produced in each heavy ion collision.[74] A direct consequence of the dilute limit is that the 12 propagator of a heavy particle is suppressed. This corresponds to a propagator involving a field of type 1 (upper branch of the Schwinger–Keldysh contour) and a field of type 2 (lower branch). Therefore, if we are interested in Green's functions involving only heavy quark fields of type 1, we can ignore the doubling of degrees of freedom and proceed in the same way as we would do at $T = 0$ (the doubling of degrees of freedom still affects the propagators of light particles). The reason is that, in any Green's function in which they appear at the same time heavy fields of type 1 and 2, there will appear at least one 12 propagator. In conclusion, if we are interested in spectroscopy at finite temperature, we can ignore the doubling of degrees of freedom. This is what we will do for the

[74] The situation could be different for charm quarks. For pNRQCD studies of the non-dilute limit for charmonium see [2157,2158].

moment. Later on, we will discuss the evolution of reduced density matrix of quarkonium that involves discussing the 12 propagator.

The first applications of NRQCD and pNRQCD at finite temperature can be found in Refs. [1592,1593]. Reference [1592] considers the infinite mass limit while Ref. [1593] discusses the Abelian analogue of quarkonium, the hydrogen atom. In both works the issue of the doubling of degrees of freedom is discussed in detail. Later on, the results were generalized to the case of real quarkonium [2159]. Let us summarize the main results found by studying quarkonium spectroscopy at finite temperature using EFTs

- The leading thermal effect can only be encoded as a modification of the potential when the Debye mass m_D is much larger than E. In the EFT framework, we only talk about a potential when we are dealing with an interaction that is non-local in space but local in time. When the condition $m_D \gg E$ is not fulfilled, thermal corrections are sensitive to E in a non-polynomial way and this signals that the interaction is non-local in time. In summary, potential models are suitable when $m_D \gg E$.
- We can consider thermal effects a perturbation if $1/r \gg T$ (where r is the radius). In this case, the medium does not modify the matchings from QCD to NRQCD and from NRQCD to pNRQCD. The medium sees quarkonium as a small color dipole. This manifests in the pNRQCD Lagrangian in the following way. The coupling between the singlet fields and the ultrasoft gluons of the medium is proportional to r. This implies that thermal effects are always multiplied by a factor of rT.
- In a qualitative way, we can say that quarkonium dissociates at the temperature at which thermal effects are of the same order of magnitude as the binding energy. The logic behind this statement is the following. If thermal effects are smaller than the binding energy, then they are a perturbation. If thermal effects are much larger than the binding energy it is impossible for a bound state to exist. Therefore, the transition between these two regimes must be found when the thermal effects and E are of the same size. In the weakly-coupled scenario, the imaginary part of the potential is larger than the screening corrections to the real part. Therefore, dissociation occurs when $T \sim Mg^{4/3}$. At this temperature, screening is a perturbation as it only becomes important when $T \sim mg$. This is at odds with the original proposal of Matsui and Satz [2152] in which the mechanism responsible for quarkonium suppression was believed to be color screening.
- There are two processes that contribute to the thermal decay width of quarkonium: gluo-dissociation and inelastic scattering with medium partons. Gluo-dissociation is the process in which a singlet state absorbs a medium gluon and becomes an octet state. It was first computed in Ref. [2160] using the Operator Product Expansions and the large-N_c limit. Within pNRQCD, this process was studied in detail in Ref. [1595], where the expression of Ref. [2160] was generalized to a finite number of colors. Inelastic scattering with medium partons is a process in which a singlet scatters with a medium quark or gluon through the exchange of an off-shell gluon [1596]. Gluo-dissociation is a leading-order process in the coupling constant expansion but it has a smaller phase space since the gluon is required to be on-shell. The pNRQCD power counting correctly predicts that gluo-dissociation is the dominant process if $E \gg m_D$. On the contrary, if $m_D \gg E$, it is inelastic parton scattering that dominates.

6.6.4 *The master equation in pNRQCD*

Previously, we have discussed the information that can be obtained from the time-ordered propagator of quarkonium. This includes the values of the binding energies and decay widths. However, since we were using the dilute limit, we did not obtain any information about how the density of heavy quarkonium evolves inside of a medium. This is needed in order to compute the probability that a bound state is detected in a heavy-ion collision.

The information about the density of heavy quarkonium is contained in the 12 singlet propagator of pNRQCD. This is zero at leading order in the dilute limit; therefore, we need to go to next-to-leading order in this expansion; i.e. we need to consider all diagrams in which the 12 propagator appears only once.

Until now, all of the studies concerning the evolution of the density of heavy quarkonium inside a medium using non-relativistic EFTs have focused on the $1/r \gg T$ regime. In this case, we can use the $T = 0$ pNRQCD Lagrangian as a starting point. It has been demonstrated that computing the evolution of the 12 singlet and octet propagator gives a system of coupled equations that resembles very closely the master equations that appear in the OQS framework [1599,1600]. This is not surprising, because we can regard $\langle S_1(t, \mathbf{r}_1) S_2^\dagger(t, \mathbf{r}_2)\rangle$ as the reduced density matrix of heavy quarks projected into the sub-space in which there is a singlet state. In general, the master equation is a complex non-Markovian equation. However, there are two limits in which simpler Markovian equations can be obtained. These limits are the ones that have been studied up to now in phenomenological applications.

In the limit $1/r \gg T, m_D \gg E$, we obtain a Lindblad equation in which all of the information about the medium is encoded in two non-perturbative parameters, κ and γ. This equation has been used to predict the nuclear modification factor in heavy ion collisions using as additional input the initial distribution of heavy quarkonium previous to the formation of the QGP and how the temperature evolves with time. However, early studies were limited due to the high

computational cost. This problem was solved by the application of the Monte Carlo wave function method [2149]. Thanks to this, it was possible to combine the solution of the master equation with state-of-the-art modelling of the time evolution of the medium to obtain results compatible with the observations at LHC [1602,2149].

Another interesting limit is the one in which thermal effects are much smaller than the binding energy. In this case, we can use the rotating wave approximation,which assumes only the diagonal elements of the density matrix in the basis that diagonalizes the leading order Hamiltonian need be considered. Using this, the master equation simplifies into a Boltzmann equation [2157,2158]. Moreover, using the molecular chaos assumption, it is possible to use the derived formulas outside of the dilute limit. Thanks to this, the authors of Refs. [2157,2158] were able to successfully reproduce experimental data for charmonium suppression at LHC.

The application of pNRQCD to the computation of the nuclear modification factor has been a very active and successful approach in recent years. However, at the moment, all of the studies have focused on the case $1/r \gg T$ for the reasons discussed in the introduction. This limits the applicability of the approach to excited states that are expected to be of larger size. In Sect. 6.6.6 we are going to discuss some recent developments that might be used to improve the situation.

6.6.5 EFT description of jet broadening

A jet is a collimated ensemble of particles with a large momentum and a small opening angle. They are useful in the context of QCD because the definition of a jet is constructed in such a way that the sensitivity to non-perturbative low-energy physics is minimized. More details can be found in Sects. 6.4 and 11. The interest in jets in heavy-ion collisions is due to a phenomenon called jet quenching [2137]. Jets lose energy when traversing a QGP. Therefore, by observing how opaque the medium is to high-energy particles allows us to infer some of its properties.

Jets might lose energy due to two different mechanisms: collisional and radiative energy loss. In the first case, the jet loses energy because it collides with the particles forming the medium. In the case of radiative energy loss, the collisions in the medium provide the high energy parton with additional transverse momentum and virtuality (a process called jet broadening). Due to this increase, the high energy parton is more likely to radiate energy outside of the jet cone. The amount of virtuality that a parton gains while traversing a given length in the medium is controlled by the transport coefficient $\hat{q}$. At the moment, it is generally believed that radiative energy loss is the dominant mechanism at high momentum while at low momentum both processes have to be taken into account.

The problem of a high-energy parton traversing a medium is one in which widely separated energy scales appear. First, we have the energy Q of the high energy particle. This is the highest energy scale that appears in the problem. Additionally, we have the transverse momentum of the particle $p_\perp$. If we use light-cone coordinates, with $p_\pm = (p^0 \pm p^3)/\sqrt{2}$, and we choose the 3 direction such that $p^+ \sim Q$, then an on-shell particle must have $p_- \sim p_\perp^2/Q \ll p_\perp$. On top of this, we have to consider the energy scales induced by the presence of the medium, which by construction are always much smaller than Q. The EFT that is suitable to study this problem is SCET (see Sect. 6.4). Note that Glauber gluons (those with momentum $p = (p_+, p_-, p_\perp)$ of order $(T, T^2/Q, T)$) play a prominent role in the physics of a jet traversing a medium. Inclusion of Glauber gluons in the SCET formalism was discussed in Refs. [1855,2161]. A more recent and extended discussion can be found in Ref. [1850].

There have been many studies of jet quenching using SCET [2162] and jet broadening [1854,1857]. In contrast to the case of quarkonium suppression, at the moment all applications use SCET as a starting point, without constructing an EFT in which medium degrees of freedom have been integrated out. This may be due to the fact that there is no information relevant to jet quenching in the time ordered propagator of a high energy particle. Instead, we need to focus on the distribution of high-energy particles that requires an approach similar to the study of the 12 propagator of heavy quarkonium. Some of the results that have been obtained from the application of SCET to the study of jet quenching are the following:

- The non-perturbative expression of $\hat{q}$ in terms of an expectation value of gauge fields was re-obtained in Ref. [1854] for the case of a Feynman or Coulomb gauge and generalized to a gauge invariant expectation value in Ref. [1857]. This result is important because it allows one to compute $\hat{q}$ using non-perturbative approaches such as lattice QCD.
- The use of SCET including Glauber gluons made it possible to derive a medium-modified parton shower in a model in which the medium is approximated as an ensemble of static scattering centers [1855,2161].

In recent years, SCET has been combined with the OQS approach to study jet quenching [1858,2163] similarly to how pNRQCD was combined with OQS to study quarkonium suppression. In this case, one considers a high-energy particle (system) that is interacting with the soft particles that comprise the medium (environment). The interaction between the two is mediated by the Glauber part of the SCET Hamiltonian [1850]. The evolution of the reduced density matrix of the system (high-energy particle) has been studied first ignoring all radiation (only considering jet broadening)

[1858] and, later on, incorporating the leading-order radiative corrections [2163]. In both cases, a master equation of the Lindblad type is found. The advantages of proceeding in this way is that the information about the medium is encoded in expectation values of gauge invariant operators of soft fields. This allows separating the physics of jet-medium interaction from the way in which the medium is modeled. In addition, it opens the way for future determinations of the influence of the medium using lattice QCD.

6.6.6 *EFTs for hydrodynamics*

We have previously discussed the difficulties encountered when constructing an EFT in which medium degrees of freedom are integrated out. In a few words, terms that mix the two branches of the Schwinger–Keldysh contour appear and this changes the properties of the EFT in a profound way in comparison with the EFT at $T = 0$. Let us summarize how this challenge has been avoided until now in the study of hard probes of the QGP:

- In the case of quarkonium we could use the dilute limit and focus on the time-ordered propagator. In this case, we know that the terms that mix the two branches of the SK contour give a small contribution and proceed as it is done at $T = 0$. The problem with this is that there is valuable information that can not be obtained from the time-ordered propagator in the dilute limit, as for example, the nuclear modification factor.
- We can choose to integrate out only the energy scales higher than the temperature. This is what has been done to study quarkonium suppression in the limit $1/r \gg T$ and jet quenching using SCET. However, this limits the applicability of the approach. Moreover, many of the simplifications introduced by the EFT framework come from being able to threat each energy scale separately from the others. This can not always be done if we are unable to integrate out medium-induced energy scales.

Recently, this issue has been addressed in the context of the construction of an EFT for hydrodynamics [2141–2144]. Going from a $T = 0$ EFT to an EFT living in the SK contour implies a doubling of degrees of freedom, but this is compensated by the fact that additional symmetries must be fulfilled. There are two symmetries that have been largely discussed:

- The SK symmetry. This symmetry must be fulfilled by any system, in or far from thermal equilibrium. It implies that the *largest time equation* [2164] must be fulfilled. This means that the difference of two Green's functions that only differ in the SK sub-index of the field evaluated at the latest time must be zero. It is obvious that this must be the case because the trace of a commutator is zero. For example, in the case of a two-point Green's function

$$\langle \phi_1(t)\phi_1(0)\rangle - \langle \phi_2(t)\phi_1(0)\rangle = \mathrm{Tr}([\phi(t), \phi(0)\rho]) = 0. \tag{6.126}$$

  One consequence of this symmetry is that in the limit of exactly classical fields ($\phi_1 = \phi_2$) the action of the EFT must be zero [2141].
- The KMS symmetry. This is a symmetry that must be fulfilled by system in thermal equilibrium. A well-known consequence of this symmetry is the fluctuation-dissipation theorem. It is akin to an *earliest time equation* in which, if t is the earliest time, a Green's function in which the operator that appears just at the right of the density matrix is evaluated at time $t - i/T$ is equal to another Green's function that is equal except that the operator appears now just at the right of the density matrix and evaluated at time t. For the case of a two-point Green's function

$$\begin{aligned}&\langle \phi_1(t_2)\phi_1(t)\rangle - \langle \phi_1(t_2)\phi_2(t - i/T)\rangle \\ &\quad = \mathrm{Tr}\left(\phi(t_2)\left(\phi(t)\rho - \rho\phi(t - i/T)\right)\right) = 0.\end{aligned} \tag{6.127}$$

  Note that the previous equation is only valid if $\rho = \mathrm{e}^{-H/T}$.

At tree level it is relatively easy to write an EFT that fulfills these conditions. However, it is more difficult to ensure them when higher-order quantum loops are involved. A solution to this is to expand the theory by introducing ghost fields and using the BRST formalism.

We note that, apart from the theoretical importance as an example of an EFT in which medium degrees of freedom are integrated out, hydrodynamics is also very important in the field of heavy ion collisions. Among other important predictions, it describes the evolution of the soft medium in which the hard probes discussed in this subsection evolve [2134–2136].

7 QCD under extreme conditions

Conveners:
Johanna Stachel and Eberhard Klempt

In nucleus–nucleus collisions at ultra-relativistic energies a new kind of matter is created, the Quark–Gluon Plasma. Peter Braun-Munzinger, Anar Rustamov and Johanna Stachel report on the phase diagram of hadronic matter at high temperature and low net baryon density. A connection is made between the experimentally determined chemical freeze-out points and the pseudo-critical temperature for the chiral cross over transition computed in lattice QCD. The role of fluctu-

ations giving experimental access to the nature of the chiral phase transition will be summarized. Azimuthal anisotropies of hadron distributions show that the Quark–Gluon Plasma formed in high energy collisions is strongly coupled, allowing to deduce bulk and shear viscosities. In the hot and dense plasma partons lose a large fraction of their energy and this observation leads to the determination of another medium parameter, a jet transport coefficient. Quarkonia and their role as a probe of deconfinement form the final topic of their contribution.

The phase structure of strongly interacting matter a low temperature and high density is discussed by Kenji Fukushima. In this region of the phase diagram that is probed e.g. in neutron stars, different phases and phase transitions are expected on theoretical grounds. Astrophysical observations and the observation of gravitational waves lead to important constraints for calculations modeling the transitions into a quarkyonic regime, into quark matter or color-superconducting states. The theoretical challenges to locate a conjectured critical end point in the QCD phase diagram are discussed.

7.1 QGP

Peter Braun-Munzinger, Anar Rustamov and Johanna Stachel

7.1.1 Introduction

The infrared slavery and asymptotic freedom properties of QCD, discussed in previous sections, form the theoretical basis that strongly interacting matter at finite temperature and/or density exists in different thermodynamic phases. This was realized [463,464] already shortly after these properties of QCD were introduced. The term quark–gluon plasma (QGP) was coined soon after by Shuryak [1392] for the high temperature/density phase where confinement is lifted and a global symmetry of QCD, the chiral symmetry, is restored. The first lattice QCD (lQCD) calculations of the equation of state were performed soon thereafter [470]. Already in early lQCD calculations a close link between deconfinement and restoration of chiral symmetry was found [462].

7.1.2 Lattice QCD

For deconfinement there is an order parameter for the phase transition, the so-called Polyakov loop, in the limit without dynamical quarks. For chiral symmetry restoration the chiral condensate $\langle \bar{\psi}\psi \rangle$ forms an order parameter for vanishing but also for finite quark masses. Indeed, recent numerical lQCD calculations [478] provide, in the limit of massless u and d quarks, strong indications for a genuine second-order chiral transition between a hadron gas and a QGP at a critical

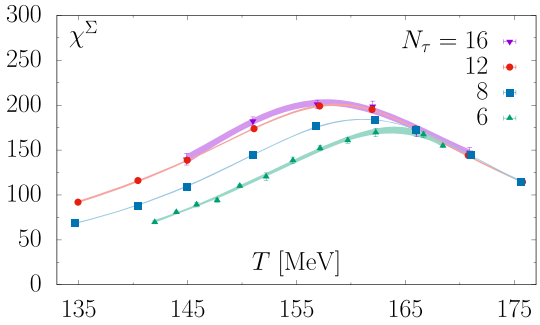


Fig. 151 Susceptibility of the chiral u, d- and s-quark condensate as a function of temperature computed in 2+1 flavor lQCD (Fig. from [484])

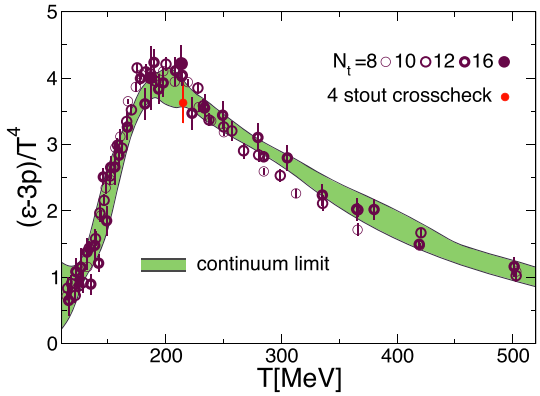


Fig. 152 The interaction measure or trace anomaly normalized to the fourth power of the temperature as a function of temperature, computed in 2+1 flavor lQCD (Fig. from [502])

temperature of $T_c \approx 132^{+3}_{-6}$ MeV. For realistic u,d,s-quark masses, chiral symmetry is restored in a crossover transition at vanishing net-baryon density and a precisely determined pseudo-critical temperature of $T_{pc} = 156.5 \pm 1.5$ MeV [484]. Consistent with this result, a transition temperature of 158.0 $\pm$ 0.6 MeV was recently reported in [487]. This pseudo-critical temperature is found as a maximum in the susceptibility (derivative with respect to mass) of the chiral condensate as displayed in Fig 151. Contrary to early ideas, the system remains strongly coupled over a rather large temperature range above T_{pc}. This is reflected in the interaction measure computed in lQCD as the difference between the energy density and three times the pressure, $I = \epsilon - 3P$, which by definition vanishes for an ideal gas of massless quarks and gluons. Figure 152 shows that this interaction measure, normalized to the fourth power of the temperature, peaks at about 20% above T_{pc} and falls off only slowly towards higher temperature values.

The lQCD calculations have been extended into the region of finite net baryon density quantified by a baryon chemical potential μ_B [484,487]. Current lQCD expansion techniques are valid in the regime of $\mu_B/T \leq 3$. The so obtained line of pseudo-critical temperatures is shown in the QCD phase dia-

gram displayed in Fig. 159 below. Because of the sign problem, the lQCD technique cannot be applied for still larger values of μ_B, see e.g. [2165], and one has to resort to models of QCD for theoretical guidance in the high net baryon density region.

7.1.3 Hadron production, nuclear stopping and global observables

Experimentally, this regime of the QCD phase transitions is accessible by investigating collisions of heavy nuclei at high energy. It was conjectured already in [467] that, in such hadronic collisions, after some time local thermal equilibrium is established and all properties of the system (fireball) are determined by a single parameter, the temperature T, depending on time and spatial coordinates. This is exactly the regime probed by collisions of nuclei at the Large Hadron Collider (LHC), as will be outlined in the following. The region of finite to large μ_B is accessed by nuclear collisions at lower energies.

In the following, we describe the experimental efforts, principally at the LHC and at RHIC (Relativistic Heavy Ion Collider), to provide from analysis of relativistic nuclear collision data quantitative information on the QCD phase diagram by studying hadron production as a function of the nucleon–nucleon center of mass energy $\sqrt{s_{\rm NN}}$. We can only touch a small fraction of the physics of the quark–gluon plasma (QGP) in this brief review. Excellent summaries of the many other interesting topics can be found in recent review articles [2166–2169].

In the early phase of the collision, the incoming nuclei lose a large fraction of their energy leading to the creation of a hot fireball characterized by an energy density ϵ and a temperature T. This stopping is characterized by the average rapidity shift of the incident nucleons, with $\Delta y = -\ln(E/E_0)$. Quantitative information is contained in the experimentally measured net-proton rapidity distributions (i.e. the difference between proton and anti-proton rapidity distributions). These distributions are presented for different collision energies from the SPS to RHIC energy range in [2170]. There it can be seen that the rapidity shift saturates at approximately two units from $\sqrt{s_{NN}} \approx 17.3$ GeV upwards, implying a fractional energy loss of $1 - \exp(-\Delta y) \approx 86\%$. In fact, the same rapidity shift was already determined for p-nucleus collisions at Fermilab for 200 GeV/c proton momentum [2171]. With increasing collision energy, the target and projectile rapidity ranges are well separated, leaving at central rapidity a net-baryon depleted or even free high energy density region. Figure 153 shows the distribution of slowed down beam nucleons, after subtracting the tail of the target distribution and plotted against rapidity minus beam rapidity. It is apparent that up to $\sqrt{s_{NN}} = 62.4$ GeV the concept of limiting fragmentation [2172] is well realized. At higher energies, this rapidity region is very hard to reach experimentally for identified particles.

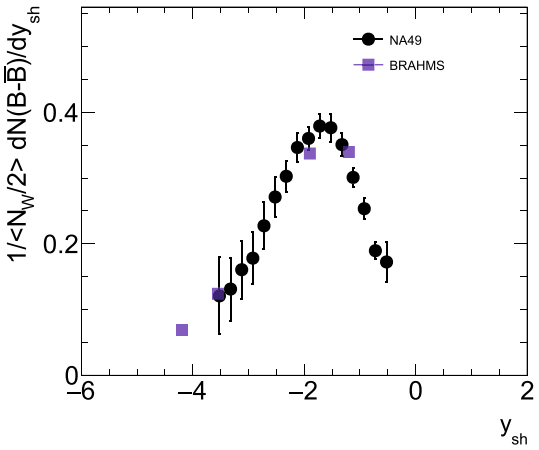


Fig. 153 Normalized net-baryon rapidity densities for $\sqrt{s_{NN}} = 17.3$ GeV [2173] and 62.4 GeV [2174] after subtracting the corresponding target contributions using the limiting fragmentation concept. Here $y_{sh} = y - y_b$ with y_b the beam rapidity

The rapidity shift of the incident nucleons leads to high energy densities at central rapidity, i.e., in the center of the fireball. These initial energy densities can be estimated, after fixing the kinetic equilibration time scale τ_0, using the Bjorken model [2175]:

$$\epsilon_{BJ} = \frac{1}{A\tau_0}\frac{d\eta}{dy}\frac{dE_T}{d\eta}, \tag{7.1}$$

where $A = \pi r^2$ is the overlap area of two nuclei. Eq. (7.1) is evaluated at a time $\tau_0 = 1$ fm and the resulting energy densities are displayed in Table 8 for central Au–Au and Pb–Pb collisions. For central Pb–Pb collisions ($A = 150\,\text{fm}^2$) at $\sqrt{s_{NN}} = 2.76$ TeV this yields an energy density of about 14 GeV/fm^3 [2176], more than a factor of 30 above the critical energy density for the chiral phase transition as determined in lQCD calculations. In fact, for all collision energies shown the initial energy density significally exceeds the energy density computed in lQCD at the pseudo-critical temperature, indicating that the matter in the fireball is to be described with quark and gluon degrees of freedom rather than as hadronic matter. The corresponding initial temperatures can be computed using the energy density of a gas of quarks and gluons with two quark flavors, $\epsilon = 37\frac{\pi^2}{30}T^4$, yielding $T \approx 307$ MeV. Temperature values for lower collision energies are also quoted in the Table.[75] It can be seen that already at AGS

[75] The values reported in the table are all for vanishing chemical potentials. We have evaluated the differences if one assumes values for chemical potentials as determined at chemical freeze-out, see below. The resulting temperature values differ by less than 5% from those reported in Table 8. Owing to the proportionality of energy density to the fourth power of temperature, inclusion of a bag pressure only mildly changes the calculated temperature values.

Table 8 Collision energy, measured transverse energy pseudo-rapidity density at mid-rapidity [2176–2179], energy density, and initial temperature estimated as described in the text for central Pb–Pb and Au–Au collisions at different accelerators

	$\sqrt{s_{NN}}$ [GeV]	$dE_t/d\eta$ [GeV]	ϵ_{BJ} [GeV/fm^3]	T [GeV]
AGS	4.8	200	1.9	0.180
SPS	17.2	400	3.5	0.212
RHIC	200	600	5.5	0.239
LHC	2760	2000	14.5	0.307

energies the estimated values of ϵ and T are significantly above the values for the chiral cross over transition.

7.1.4 Hydrodynamic flow and QGP transport parameters

Depending on energy, collisions of heavy ions populate different regimes falling into two categories: (i) the stopping or high baryon density region reached at $\sqrt{s_{NN}} \approx$ 3–20 GeV and (ii) the transparency or baryon-free region reached at $\sqrt{s_{NN}} >$ 100 GeV. The net-baryon-free QGP presumably existed in the early Universe after the electro-weak phase transition and up to a few microseconds after the Big Bang.[76] On the other hand, a baryon-rich QGP could be populated in neutron star mergers or could exist, at very low temperatures, in the center of neutron stars[2180,2181].

For the system considered to come into local thermal equilibrium and, more importantly, for the development of a phase transition, the presence of interactions is necessary. In fact, close to the phase transition, the system has to be strongly coupled. As mentioned above, quarks and gluons under the extreme conditions reached in nuclear collisions are indeed strongly coupled. The large values of the interaction measure from lQCD calculations $(\epsilon - 3P)/T^4$, introduced above in Fig. 152, lend support to the strong coupling scenario. Further, the energy and entropy densities ϵ/T^4 and s/T^3, as calculated in lQCD, fall significantly short (by about 20 %) of the Stefan-Boltzmann limit for an ideal gas of quarks and gluons up to a few times the pseudo-critical temperature. The conclusion about a strongly coupled QGP close to T_{pc} also follows from experimental results at the colliders, and even at the SPS, on the coefficients of azimuthal anisotropies of hadron distributions in combination with a viscous hydrodynamic description.

For non-central nuclear collisions the distributions in transverse momentum p_T of hadrons exhibit modulations with respect to the azimuthal angle ϕ in the reaction plane. These anisotropies can be characterized by p_T dependent Fourier coefficients. The dominant term is the 2nd order Fourier coefficient v_2, also called the elliptic flow coefficient. This modulation has been predicted to arise from the anisotropy of the gradient of the pressure P in the early phase of the collision due to the geometry of the nuclear overlap region, leading to correspondingly larger expansion velocities in the reaction plane and hence large v_2 coefficients.

The strength of the coupling can be quantified by introducing transport parameters for the QGP such as the shear viscosity η, which is related to the mean free path of quarks and gluons inside the QGP, and the bulk viscosity ζ, with its connection to QGP expansion dynamics and speed of sound. The smaller the transport coefficients the stronger the coupling. Larger values of the shear viscosity, e.g., suppress the magnitude of the elliptic flow.

For a strongly coupled system with small enough values of mean free path (comparable to or lower than the corresponding de Broglie wavelength of particles), treatment as a fluid is more appropriate. One then describes its properties by solving hydrodynamic equations. The shear viscosity enters the hydrodynamic equations as $\eta/(\epsilon + P) = \eta/(Ts)$, hence the quantity characterizing the medium is η/s. By comparing flow observables measured in experiments at RHIC [2184,2185] and LHC [2186] to the corresponding calculations in viscous hydrodynamics, accompanied with converting the fluid into thermal distributions of hadrons at the freeze-out hyper-surface, remarkably low values for η/s are obtained. Figure 154 shows as an example the elliptic flow coefficients v_2 for different identified hadrons at the LHC. A mass ordering characteristic for a hydrodynamically expanding medium is observed very clearly. And indeed, the mass ordering and its p_T dependence are described quantitatively by a relativistic viscous hydrodynamic calculation [2182] as indicated by the lines in Fig. 154 employing a small ratio of η/s.

In fact, a lower bound of $\eta/s = 1/(4\pi)$ (in units of $\hbar = k_B = 1$) can be obtained for a large class of strongly coupled field theories from the quantum mechanical uncertainty principle [2187] and using the AdS/CFT correspondence [2095,2188,2189]. Recently, the values and the temperature dependence of the shear and bulk viscosities employed in hydrodynamic codes were extracted by fitting spectra and azimuthal anisotropies of hadrons measured at the LHC and RHIC using Bayesian estimation methods [2183,2190]. An example is shown in Fig. 155. Inspection of this figure indicates that, at T_{pc}, the estimated value of η/s is close to the lower bound of $1/(4\pi)$, indicating that the observed matter is a nearly perfect fluid. Above the transition temperature, the extracted band for η/s is rising, reflecting a weakening of the coupling, although even at twice T_{pc} the medium is still strongly coupled. On the other hand, as presented in Fig. 152, near the phase transition the lQCD results exhibit a maximum in the interaction measure, which is an indication for interactions in the system. In the hydrodynamic calcu-

[76] In the QGP of the early universe, particles interacting via the strong and electro-weak force are part of the system, while an accelerator-made QGP only contains strongly interacting particles.

lations the breaking of scale invariance is accounted for by introducing a bulk viscosity ζ along with the shear viscosity. While increasing sheer viscosity reduces the momentum anisotropy, hence lowering the elliptic flow coefficients, the bulk viscosity reduces the overall rate of the radial expansion. The right panel of Fig. 155 shows the temperature dependence of ζ/s, which exhibits a peak just above the transition temperature [2183]. This location of the maximum is consistent with the temperature dependence of the interaction measure from lQCD.

We also note here that signatures for collective behavior such as anisotropic azimuthal distributions have rather surprisingly also been found in small collision systems such as pp or pPb [2191,2192]. A detailed understanding of this rather unexpected result is currently debated in the literature, in particular since another signature of a dense, strongly coupled and colored medium, significant parton energy loss, is not observed for such small systems, see, e.g., [2193].

7.1.5 Jet quenching and parton energy loss

Important information on the structure of the QGP is also obtained by studying the interaction of high-momentum partons with the thermalized quarks and gluons in the QGP. A strongly coupled QGP is opaque to high momentum partons, leading to the phenomenon of 'jet quenching' [2169]. In fact, the theoretical foundation for strong jet quenching by QCD bremsstrahlung was laid by [2194]. There it was shown that, for sufficiently energetic quarks and gluons, such that the radiation does not decohere, the radiative energy loss scales quadratically with the length traversed, leading to very large values. An important experimental observable linked to jet quenching is the observed suppression ('quenching') of high-momentum hadrons in central nuclear collisions at high collision energy. This suppression is quantified by the p_T dependence of the ratio R_{AA} of inclusive hadron production in collisions of nuclei with mass number A to that in proton–proton collisions, taking into account the collision geometry by scaling to the number of binary collisions [2169].

In Fig. 156 we present the evolution with cm energy of the transverse momentum dependence of R_{AA} for leading particles as obtained from measurements at the SPS, RHIC, and LHC accelerators. Note that, by construction, $R_{AA} = 1$ for hard binary collisions in the absence of nuclear effect such as jet quenching. At very low p_T one observes R_{AA} values less than unity and increasing with p_T since soft particle production scales with the number of participating nucleons and not the number of binary collisions. For RHIC and LHC energies the jet quenching is born out by a decreasing trend observed for $p_T > 2.5$ GeV/c reaching a broad minimum near $p_T = 7$ GeV/c of $R_{AA} = 0.1-0.2$: high momentum hadrons are quenched by about a factor of 5 or more. At LHC energies R_{AA} increases again for higher p_T values until a plateau is

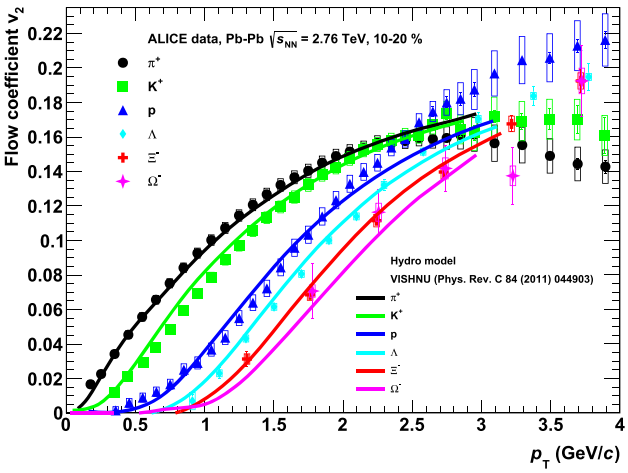


Fig. 154 Elliptic flow coefficient v_2 for identified hadrons as a function of transverse momentum measured by ALICE and compared to results from viscous hydrodynamics calculations [2182]

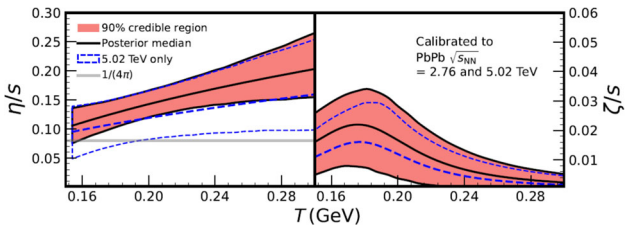


Fig. 155 Temperature dependence of the shear (left panel) and bulk (right panel) viscosity to entropy density ratios. Figure taken from [2183]

reached above $p_T \approx 100$ GeV/c. Measurements for fully reconstructed jets have been performed by the ATLAS collaboration. The results demonstrate [2196] that the quenching by about a factor of 2 persists to the highest available jet p_T values of 1 TeV/c. Recently, also detailed measurements became available from measurements of heavy flavor jets, where for the first time in-medium transport coefficients for charm quarks were determined [2197].

The data on jet quenching have been modeled in terms of elastic and inelastic collisions of partons in the dense QGP, taking into account important coherence effects [2198,2199]. For a recent summary see [2200] and Ref. cited there.

To model experimental data with QCD-based jet quenching calculations one has to take into account that the jet is created as a product of an initial hard parton–parton collision with large momentum transfer Q. That implies that the parton initiating the jet is highly virtual. The magnitude of its 4-momentum Q as reflected in the total jet energy E can be hundreds of GeV (or even a few TeV at the LHC) while, for a real parton, $Q^2 \approx 0$. The highly virtual parton will evolve into a parton shower which eventually hadronizes to form a collimated jet of hadrons. During the entire evolution the highly virtual initial parton and the parton shower compo-

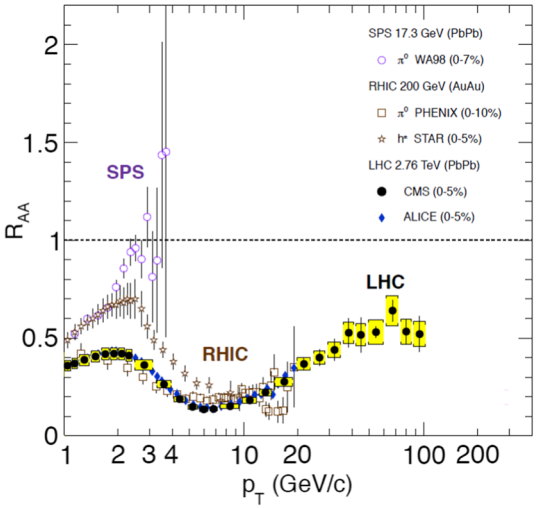


Fig. 156 Evolution of the transverse momentum dependence of R_{AA} for leading particles for central nuclear collisions with collision energy [2195]

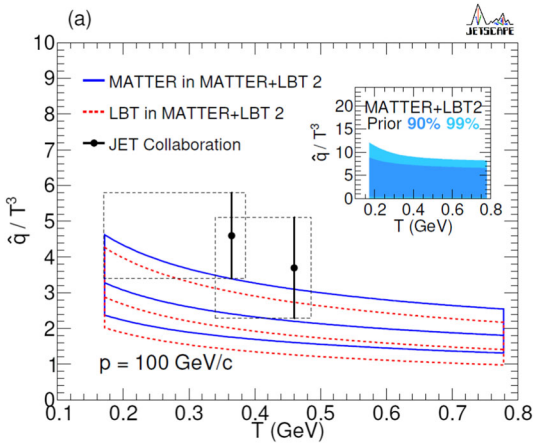


Fig. 157 QGP jet transport parameter $\hat{q}/T^3$ obtained by the JETSCAPE collaboration [2200]

nents lose energy by interactions with the QGP constituents, resulting in a medium-modification of the entire parton fragmentation pattern, i.e. the jet [2198]. The most modern jet quenching analyses take into account the different regimes of parton virtuality as described in [2200]. The calculations have as leading input parameter a jet transport coefficient $\hat{q}$ that is determined by the differential mean squared momentum transfer $\langle k_t^2 \rangle$ between jet parton and the QGP constituents with respect to the length traversed, i.e. $\hat{q} = d\langle k_t^2 \rangle / dL$.

The recent analysis by the JETSCAPE collaboration [2200] uses data on inclusive hadron suppression from central Au–Au collisions at RHIC and Pb–Pb collisions at LHC, applying a Bayesian parameter estimation to determine the temperature dependence of the dimensionless, renormalized jet transport parameter $\hat{q}/T^3$. The calculations are based on two different models for parton energy loss, called MATTER and LBT, to effectively cover the large range of parton virtualities. A switch-over between the virtuality-ordered splitting dominated regime and the time-ordered transport dominated regime happens at low virtualities of $Q_0 = 2{-}2.7$ GeV. The results are shown in Fig. 157. Note that the plot is for a parton momentum of 100 GeV/c, but as demonstrated in [2200] the momentum dependence is rather mild. To put the results into context, a value of $\hat{q}/T^3 = 4$ implies that, at temperature T = 0.4 GeV, $\hat{q} \approx 1.3\,\mathrm{GeV}^2/\mathrm{fm}$. This value should be compared to what was determined for parton energy loss in cold nuclear matter. Analysis of data for deep inelastic scattering off large nuclei [2201] yielded a value of $\hat{q} = 0.024 \pm 0.008\,\mathrm{GeV}^2/fm$. A global analysis of the jet transport coefficient for cold nuclear matter was performed recently in [2202]. These authors obtain values of $\hat{q} < 0.03\,\mathrm{GeV}^2/fm$ over a wide range of (x_B, Q^2) values (here, x_B is the Bjorken x parameter). We conclude that, for high energy partons, the stopping power of a QGP formed at RHIC or LHC energy is increased by more than a factor of 40 compared to that for cold nuclear matter. The dramatic jet quenching observed experimentally as displayed in Fig. 156 finds its natural explanation in the large values of the transport coefficient $\hat{q}$ of the QGP.

7.1.6 The statistical hadronization model and the QGP phase diagram

Direct experimental access to the QCD phase diagram is obtained from the measurement of the yields of hadrons produced in (central) high energy nuclear collisions. Analysis of these data in terms of the Statistical Hadronization Model (SHM), see [2203] and Refs. given there, established that, at hadronization, the fireball formed in the collision is very close to a state in full (hadro-)chemical equilibrium.

The essential idea in the SHM is to approximate the partition function of the system by that of an ideal gas composed of all stable hadrons and resonances, hence also referred to as the Hadron Resonance Gas (HRG) model, see [2203]. From this partition function one can calculate the first moments (mean values) of densities of hadrons as a function of a pair of thermodynamic parameters, the temperature T_{chem} and the baryon chemical potential μ_B at chemical freeze-out. To go beyond the ideal gas approximation, attractive and repulsive interactions between hadrons can be taken into account in the S-matrix formulation of statistical mechanics [2204] by including the first term in the virial expansion. Ideally, the relevant coefficients are obtained from measured phase shifts. For the pion–nucleon interaction this was implemented in [2205] and the proton yield for LHC energy was corrected accordingly [2206]. The predictions of the SHM for hadron yields are compared to experimental data at LHC energy for $T_{chem} = 156.5$ MeV in Fig. 158. The agreement is excellent for the yields of all measured hadrons, nuclei

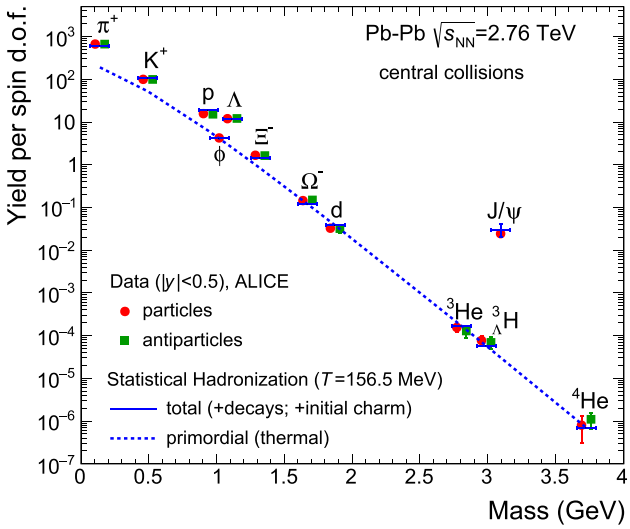


Fig. 158 Primordial and total (anti-)particle yields, normalized to the spin degeneracy, as calculated within the SHMc [2203]

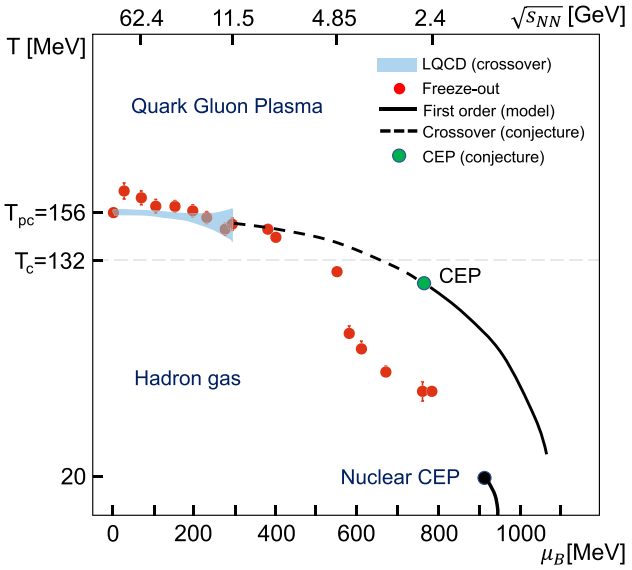


Fig. 159 Phase diagram of strongly interacting matter. The red symbols correspond to chemical-freezeout parameters, temperature T_{chem} and baryon chemical potential μ_B determined from experimental hadron yields [2203,2207,2208]. The blue band represents the results of lQCD computations of the chiral phase boundary [484,487]. Also shown are a conjectured line of first order phase transition with a critical end point as well as the nuclear liquid-gas phase boundary

and hyper-nuclei and their anti-particles, with yields varying over 9 orders of magnitude. Remarkably, the description works equally well for loosely bound states. This has led to the conjecture of hadronization into compact multi-quark bags with the right quantum numbers evolving into the final nuclear wave functions in accordance with quantum mechanics [2203].

The values of the hadro-chemical freeze-out parameters at lower collisions energies are similarly obtained by fitting the SHM results to the measured hadron yields. The extracted freeze-out parameters T_{chem} and μ_B [2203,2207] are presented as red symbols in the QCD phase diagram shown in Fig. 159. Also included is a freeze-out point from the HADES collaboration in Au–Au collisions at $\sqrt{s_{\rm NN}} \approx 2.4$ GeV [2208]. They can be compared to the crossover chiral phase transition line as computed in lQCD (blue band). From LHC energies down to about $\sqrt{s_{\rm NN}} = 12$ GeV, i.e., over the entire range covered by lQCD, there is a remarkable agreement between T_{chem} and the pseudo-critical temperature for the chiral cross over transition T_{pc}. We note that, along this phase boundary, the energy density computed (for 2 quark flavors) from the values of T_{chem} and μ_B exhibits a nearly constant value of $\epsilon_{crit} \approx 0.46$ GeV/fm^3.

The finding that the hadro-chemical freeze-out temperature is very close to T_{pc} has a fundamental consequence: because of the very rapid temperature and density change across the phase transition and the resulting low hadron densities in the fireball combined with its size, the produced hadrons cease to interact inelastically within a narrow temperature interval [2209] after hadron formation.

This is very different from particle freeze-out in the early universe where for temperatures $T > 10$ MeV even the mean free path for neutrinos is much smaller than its size, see section 22.3 of [513].

For large values of baryon chemical potential, experimental data for hadron-chemical freeze-out exist but the phase structure of strongly interacting matter remains uncertain; various model calculations suggest the appearance of a line of first order phase transition, which in combination with the crossover transition at smaller values of μ_B, would imply the existence of a critical end point (CEP) in the QCD phase diagram as indicated in Fig. 159. The experimental discovery of the CEP would mark a major break-through in our understanding of the QCD phase structure. The location of the CEP is most likely in the region $\mu_B > 470$ MeV, based mostly on results from lQCD. Searching for the CEP is the subject of a very active research program, at RHIC and the future FAIR facility at GSI. The importance of this research is underlined by the realization that we have currently no experimental evidence for the order of the chiral phase transition at any value of baryon chemical potential.

7.1.7 Fluctuations and the search for critical behavior

Important further information on the phase structure of QCD matter is expected by measuring, in addition to the first moments of hadron production data, also higher moments as such data can be directly connected to the QCD partition function via conserved charge number susceptibilities in the Grand Canonical Ensemble (GCE) [2210,2211]. For a ther-

mal system of volume V and temperature T the susceptibilities in the GCE are defined as the coefficients in the Maclaurin series of the reduced pressure $\hat{P} = P(T, V, \vec{\mu})/T^4$

$$\chi_n^q \equiv \frac{\partial^n \hat{P}}{\partial \hat{\mu}_q^n} = \frac{1}{VT^3} \frac{\partial^n ln Z(V, T, \vec{\mu})}{\partial \hat{\mu}_q^n} = \frac{\kappa_n(N_q)}{VT^3}, \qquad (7.2)$$

where $\vec{\mu} = \{\mu_B, \mu_Q, \mu_S\}$ is the chemical potential vector that is introduced to conserve, on average, baryon number, electric charge and strangeness. Here, $\hat{\mu}_q = \mu_q/T$ is the reduced chemical potential for the conserved charges $q \in \{B, Q, S\}$. The partition function $Z(V, T, \vec{\mu})$ encodes the Equation of State (EoS) of the system under consideration. Equation (7.2) establishes a direct link between susceptibilities and fluctuations of conserved charge numbers. By measuring the cumulants $\kappa_n(N_q)$ of net-charge number (N_q) distributions one can, using Eq. (7.2), further probe and quantify the nature of the QCD phase transition.

Important at this point is to define a non-critical baseline, which is done by using the ideal gas EoS, extended such as to account for event-by-event charge conservation and correlations in rapidity space [2170,2212,2213], see also [2214]. In addition, non-critical contributions arising, e.g., from fluctuations of wounded nucleons [2215,2216] need to be corrected for. Deviations from this non-critical baseline, for example leading to negative values of κ_6 for net-baryons would arise due to the closeness of the cross over transition to the O(4) 2nd order critical phase transition for vanishing light quark masses [2217].

In Fig. 160 the ALICE results on the normalized second order cumulants of net-proton distributions are presented as function of the experimental acceptance. The acceptance is quantified via the pseudo-rapidity coverage around mid-rapidity $\Delta\eta$ [2218–2220]. The measured cumulant values approach unity at small values of $\Delta\eta$, essentially driven by small number Poisson statistics. With increasing acceptance, the data progressively decrease from unity. For small but finite acceptance the decrease can be fully accounted for by overall baryon number conservation in full phase space. Hence, after correcting for baryon number conservation, the experimental data would be consistent with unity over the range of the experimental acceptance.

This observation has three important consequences. (i) It shows that, up to second order, cumulants of the baryon number distribution functions follow a poissonian distribution, a posteriori justifying the assumptions underlying the construction of the partition function used in the SHM. (ii) This is the first experimental verification of lQCD results which also predict unity for the second order scaled cumulants of baryon distributions. (iii) Compared to the different calculations, the data imply long range correlations in rapidity space, calling into question the baryon production mechanism implemented in string fragmentation models. Indeed,

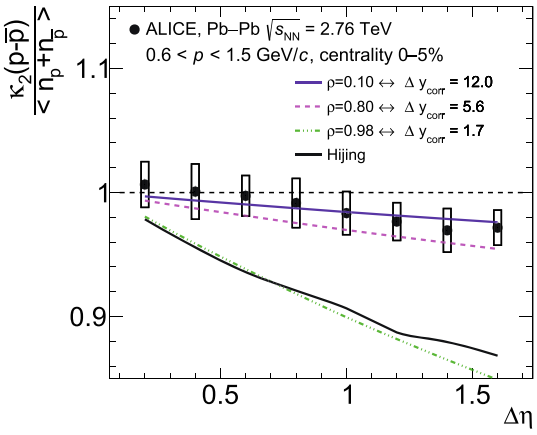


Fig. 160 Scaled second order cumulants of the net-proton distribution as a function of the pseudo-rapidity acceptance measured by the ALICE experiment (black symbols) [2219]. The colored lines correspond to calculations accounting for baryon number conservation with different correlation length in rapidity space [2213]. The results of the HIJING event generator are presented with the black solid line

the results from the HIJING event generator based on the Lund String Fragmentation model shown in Fig. 160, due to the typical correlation over about one unit of rapidity, grossly overpredict the suppression due to baryon number conservation [2221].

Contrary to the detailed predictions for signals in the cross-over region of the transition covered by lQCD, no quantitative signals are available for the existence of a possible critical end point in the phase diagram. All predicted signals are of generic nature and mostly based on searching for non-monotonic behavior in the excitation function of fourth order cumulants of, e.g., net-protons [2222]. A compilation of the respective measurements [2223,2224] is presented in Fig. 161. The search for non-monotonic behavior needs a starting point. In Fig. 161 two possibilities are presented, one corresponding to calculations in HRG within GCE (dashed line at unity) and the other the non-critical baseline introduced above where baryon number conservation is explicitly accounted for (red solid line or blue symbols). With respect to unity the data indeed exhibit an indication for non-monotonic behavior with a significance corresponding to 3.1 standard deviations [2224]. However, a significant part of this deviation from unity is induced by non-critical effects, such as baryon number conservation. Therefore, one must search for non-monotonic behavior with respect to the red solid line. Analysis of the data shows that there are no significant deviations from a statistical ensemble with event-by-event baryon number conservation, i.e, within the current precision of the data there is not yet evidence for the presence of a critical end point [2170,2212]. The analysis of fourth order cumulants from a much higher statistics data set has just started and will be essential for a possible discovery of the critical point.

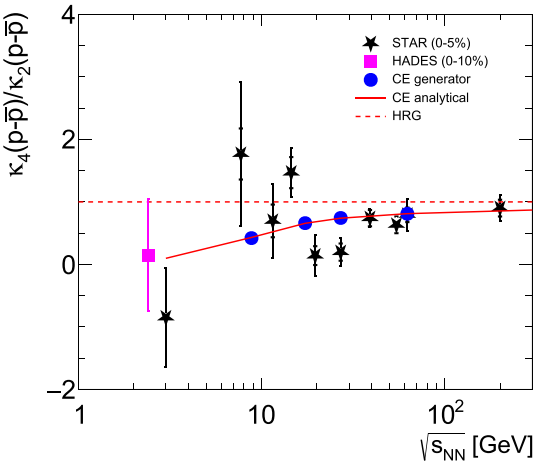


Fig. 161 Collision energy dependence of the fourth to second order cumulants of net-proton distributions as measured by experiments. The STAR data are for $|y| < 0.5$ and $p_t = 0.4$–2 GeV/c, the HADES data for $|y| < 0.4$ and $p_t = 0.4$–1.6 GeV/c. The non-critical baseline induced by global baryon number conservation is indicated by the blue circles and the red line

The current status on experimental verification of the nature of the chiral cross-over transition at vanishing or moderate μ_B is still rather open. Within QCD inspired model calculations [2225,2226], based on O(4) scaling functions the predicted sixth order cumulants for net-baryon distributions exhibit negative values at T_{pc} due to a singular term in the pressure. Similarly, the sixth order susceptibilities of baryon number resulting from lQCD calculations are also negative [2217,2227] and this sign change (relative to the HRG baseline in GCE) has been linked to the critical component in the pressure present as a residue of the 2nd order chiral phase transition for vanishing (u,d) quark masses, due to the smallness of the physical masses. First experimental results on sixth order net-proton cumulants were reported by the STAR collaboration [2228] for Au–Au collisions, albeit with sizeable statistical uncertainties since the data analysis to determine high order cumulants is extremely statistics hungry. Qualitatively, the STAR results at $\sqrt{s_{NN}} = 200$ GeV are indeed consistent with the expectations for the crossover transition. At the same time, the experimentally measured energy dependence of κ_6 [2228] is at odds with both model and lQCD calculations. For a quantitative conclusion, in any case, the effects of baryon number conservation [2170] and transformation from net-protons (experiment) to net-baryons (theory) [2229] are still to be performed. So far, experimental insight into the nature of the chiral cross-over transition and the development towards low net-baryon densities remains inconclusive. It can be expected that ongoing and future high statistics measurement campaigns by the STAR and ALICE collaborations will elucidate the situation.

7.1.8 Hadrons with heavy quarks

There is now significant experimental information, from relativistic nuclear collisions, not only on the production of hadrons composed of light (u,d,s) quarks, but also of open and hidden charm and beauty hadrons. In particular, there is good evidence, mainly from results obtained at the CERN Large Hadron Collider (LHC) [2230–2232], that charm quarks reach a large degree of thermal equilibrium, although charm quarks in the system are chemically far out of equilibrium. This is supported by heavy quark diffusion coefficients from lQCD [2233]. A strong indication for equilibration is the fact that J/ψ mesons participate in the collective, anisotropic hydrodynamic expansion [2234,2235].

To microscopically understand the production mechanism of charmed hadrons for systems ranging from pp to Pb–Pb, various forms of quark coalescence models have been developed [2236–2240]. This provides a natural way to study the dependence of production yields on hadron size and, hence, may help to settle the still open question whether the many exotic hadrons that have been observed recently are compact multi-quark states or hadronic molecules (see [2241,2242] and Refs. cited there). Conceptual difficulties with this approach are that energy is not conserved in the coalescence process and that color neutralization at hadronization requires additional assumptions about quark correlations in the QGP [2243].

Another approach, named SHMc, has been made possible by the extension of the SHM to also incorporate charm quarks. This was first proposed in [2244] and developed further in [2203,2231,2245–2248] to include all hadrons with hidden and open charm. The key idea is based on the recognition that, contrary to what happens in the (u,d,s) sector, the heavy (mass ~ 1.2 GeV) charm quarks are not thermally produced. Rather, production takes place in initial hard collisions. The produced charm quarks then thermalize in the hot fireball, but the total number of charm quarks is conserved during the evolution of the fireball [2248] since charm quark annihilation is very small. In essence, this implies that charm quarks can be treated like impurities. Their thermal description then requires the introduction of a charm fugacity g_c [2231,2244]. The value of g_c is not a free parameter but experimentally determined by measurement of the total charm cross section. For central Pb–Pb collisions at LHC energy, $g_c \approx 30$ [2231]. The charmed hadrons are, in the SHMc, all formed at the phase boundary, i.e. at hadronization, in the same way as all (u,d,s) hadrons.

In Fig. 158 it can be seen that, with that choice, the measured yield for J/ψ mesons is very well reproduced, the uncertainty in the prediction is mainly caused by the uncertainty in the total charm cross section in Pb–Pb collisions. We note here that, because of the formation from deconfined charm quarks at the phase boundary, charmonia are unbound inside

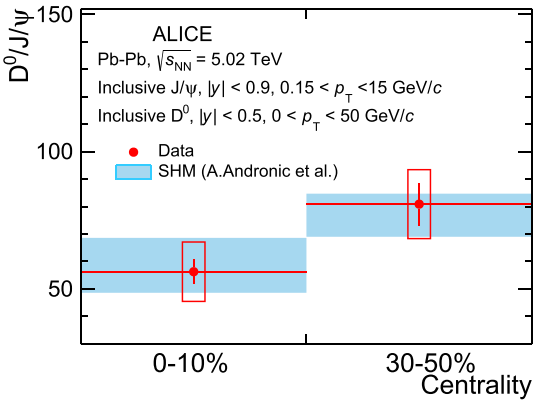


Fig. 162 D^0 to J/ψ yield ratio measured in Pb–Pb collisions at the LHC and predicted by the Statistical Hadronization Model with charm SHMc. Figure from [2249]

the QGP but their final yield exhibits enhancement compared to expectations using collision scaling from pp collisions, contrary to the original predictions based on [2152]. For a detailed discussion see [2203].

For the description of yields of charmonia, feeding from excited charmonia is very small because of their strong Boltzmann suppression. For open charm mesons and baryons, this is not the case and feeding from excited D^* and Λ_c^* is an essential ingredient for the description of open charm hadrons [2231]. Even though the experimental delineation of the mass spectrum of excited open charm mesons and baryons is currently far from complete, the prediction of yields for D-mesons and Λ_c baryons compares very well with the measurements,[77] both concerning transverse momentum and centrality dependence.

A particularly transparent way to look at the data for Pb–Pb collisions is obtained by analyzing the centrality dependence of the yield ratio $D^0/(J/\psi)$ and comparing the results to the predictions of the SHMc. Recently, both the D^0 and J/ψ production cross sections have been well measured down to $p_t = 0$. The yield ratio $D^0/(J/\psi)$ is reproduced with very good precision for both measured centralities, as demonstrated in Fig. 162. This result lends strong support to the assumption that open and hidden charm states are both produced by statistical hadronization at the phase boundary. A more extensive comparison between SHMc and data for open charm hadrons is shown in [2231].

Impressive experimental results were also obtained mainly by the CMS collaboration ([2250] and Refs. there) on the suppression of excited states of Υ mesons in Pb–Pb collisions at LHC energy. These data have led to intense theoretical work and, indeed, to a break-through of our understanding of the complex heavy quark potential in a hot medium [2149,2251].

From the successful comparison of measured yields for the production of (u,d,s) as well as open and hidden charm hadrons obtained from the SHM or SHMc with essentially only the temperature as a free parameter at LHC energies, one may draw a number of important conclusions.

– First, we note that hadron production in relativistic nuclear collisions is described quantitatively by the chemical freeze-out parameters (T_{chem}, μ_B). Note that the fireball volume appearing in the partition function is determined by normalization to the measured number of charged particles. At least for energies $\sqrt{s_{NN}} \geq 10$ GeV these freeze-out parameters agree with good precision with the results from lQCD for the location of the chiral cross over transition. Under these conditions, hadronization is independent of particle species and only dependent on the values of T and μ_B at the phase boundary. At LHC energy, the chemical potential vanishes, and only $T = T_{pc}$ is needed to describe hadronization.
– The mechanism implemented in the SHMc for the production of charmed hadrons implies that these particles are produced from uncorrelated, thermalized charm quarks as is expected for a strongly coupled, deconfined QGP (see also the discussion in [2231]). At LHC energy, where chemical freeze-out takes place for central Pb–Pb collisions in a volume per unit rapidity of $V \approx 4000$ fm^3, this implies that charm quarks can travel over linear distances of order 10 fm (see [2203,2231] for more detail).

One may ask whether there is a possible contribution to the production of charmed hadrons (in particular of J/ψ) from the hadronic phase. At the phase boundary, assembly of J/ψ from deconfined charm quarks or from all possible charmed hadrons is indistinguishable, as discussed in detail in [2203]. In fact, in [2209] it was demonstrated that multi-hadron collisions lead to very rapid thermal population, while within very few MeV below the phase boundary, the system falls out of equilibrium. Both is driven by the rapid drop of entropy and thereby particle density in the vicinity of T_{pc}. In the confined hadronic phase, i.e. for temperatures lower than T_{pc}, the hadron gas is off-equilibrium, and any calculation via reactions of the type $D\bar{D}^* \leftrightarrow n\pi$J/$\psi$ has to implement the back-reaction [2252]. Since predictions with the SHMc agree very well with the data for J/ψ production at an accuracy of about 10%, and since any possible hadronic contribution has to be added to the SHMc value, we estimate any contribution to J/ψ production from the confined phase to be less than 10%.

Future measurement campaigns at the LHC will yield detailed information on the production cross sections of

[77] For Λ_c baryons on has to augment the currently measured charm baryon spectrum with additional states to achieve complete agreement with experimental data [2231].

hadrons with multiple charm quarks as well as excited charmonia. The predictions from the SHMc for the relevant cross sections exhibit a rather dramatic hierarchy of enhancements [2231] for such processes. Experimental tests of these predictions would lead to a fundamental understanding of confinement/deconfinement and hadronization.

7.2 QCD at high density

Kenji Fukushima

7.2.1 QCD phase diagram

The QCD vacuum has rich contents, very different from an empty "vacuum" but rather close to a medium. The relevant physical degrees of freedom can change according to the probe resolution to the medium. As long as the typical momentum scale in physical processes is large compared to the QCD scale, i.e., $\Lambda_{\rm QCD} \sim 200\,{\rm MeV}$, observed particles – all hadrons including mesons and baryons – are only color-singlet composites. The typical scale of hadronic masses and radii is characterized by $\Lambda_{\rm QCD}$ or $\Lambda_{\rm QCD}^{-1} \sim 1\,{\rm fm}$. Therefore, if hadronic matter is compressed so that the interparticle distance becomes comparable to $\Lambda_{\rm QCD}^{-1}$, wavefunctions of hadrons overlap each other. Then, hadrons are no longer isolated and more elementary particles should take over the physical degrees of freedom.

High compression of QCD matter is achieved by increasing the particle number density. Actually, if matter is heated up, the density of massless thermal excitations increases as $\sim T^3$ which corresponds to the scaling of interparticle distance $\propto T^{-1}$. If the baryon density, $n_{\rm B}$, is increased in the same way, the average distance between baryons should scale as $\propto n_{\rm B}^{-1/3}$. It is hence natural to expect a phase boundary in the plane of T and $n_{\rm B}$ from hadronic matter to a new state of matter composed of quarks and gluons, which portrays the QCD phase diagram.

The idea of the QCD phase diagram was first cast into a concrete shape by Cabibbo and Parisi [464] based on the conjecture of Hagedorn's limiting temperature. Let us briefly look over the theory foundations according to explanations in Ref. [2253]. The thermal partition function at finite T but zero density reads:

$$\mathcal{Z}_{\rm M} = \int dm\, \rho_{\rm M}(m)\, e^{-m/T}, \tag{7.3}$$

where $dm\rho_{\rm M}(m)$ represents the number of mesonic states within the mass window, $m \sim m + dm$. The last exponential factor appears from the thermal Boltzmann weight. The density of states associated with degeneracy is increasing for larger eigen-energies, and so $\rho_{\rm M}(m)$ is an increasing function of m. It is empirically known that $\rho_{\rm M}(m) \sim e^{m/T_{\rm H}}$ with a phenomenological parameter $T_{\rm H}$ called the Hagedorn temperature. Because the logarithm of the combinatorial factor for a given energy is nothing but the entropy, this exponentially increasing $\rho_{\rm M}(m)$ means that the entropy grows linearly with m. As seen from Eq. (7.3), the m integration in $\mathcal{Z}_{\rm M}$ blows up for $T > T_{\rm H}$ for which the entropy enhancement overwhelms the energy suppression and the free energy is bottomlessly pushed down with increasing m. Hagedorn proposed that $T_{\rm H}$ is interpreted as the upper bound of the physically possible temperature. Later on, a physically sensible interpretation was clarified that the singularity in $\mathcal{Z}_{\rm M}$ should be overridden by a phase transition, possibly the one to a state with more fundamental degrees of freedom. The critical temperature from mesonic matter to deconfined matter with quarks and gluons is thus $T_{\rm c}^{\rm (M)} = T_{\rm H}$.

The above mentioned argument can be generalized to the case at finite baryon density. Then, the partition function is

$$\mathcal{Z}_{\rm B} = \int dm\, \rho_{\rm B}(m)\, e^{-(m-\mu_{\rm B})/T}, \tag{7.4}$$

where the Boltzmann factor depends on the baryon chemical potential $\mu_{\rm B}$. The experimental data imply that the baryonic spectrum exhibits $\rho_{\rm B}(m) \propto e^{m/T'_{\rm H}}$ with the baryonic Hagedorn temperature, $T'_{\rm H}$, that is slightly different from $T_{\rm H}$. The critical temperature for baryons is deduced from the singularity as $T_{\rm c}^{\rm (B)} = T'_{\rm H} - (T'_{\rm H}/m_0)\mu_{\rm B}$, which is derived from an approximation that the Boltzmann factor is replaced by $e^{-m(1-\mu_{\rm B}/m_0)/T}$ with a phenomenological parameter, m_0 (see Ref. [2253] for detailed discussions).

Now, supposing that $T'_{\rm H} > T_{\rm H}$, the critical temperature for the deconfinement transition is dominantly characterized by mesonic $T_{\rm c}^{\rm (M)}$ in the low density region at $\mu_{\rm B} \ll T$. With increasing $\mu_{\rm B}$, the two lines of constant $T_{\rm c}^{\rm (M)}$ and decreasing $T_{\rm c}^{\rm (B)}$ cross eventually. This consideration leads us to a picture of the phase diagram on the plane of the baryon density (along the horizontal axis) and the temperature (along the vertical axis) as illustrated in Fig. 163. This QCD phase diagram handwritten by Gordon Baym (see Ref. [2254] for more historical backgrounds) has played a role of prototype of the contemporary QCD phase diagram.

So far, we addressed only the deconfinement phase transition associated with the liberation of quarks and gluons in hot and dense media. The theoretical description of deconfinement in the presence of dynamical quarks is subtle, however. One may think that each phase separated by phase boundaries should be distinctly defined by a different realization of some global symmetry but for the deconfinement phenomenon, the symmetry corresponding to quark confinement/deconfinement (known as *center symmetry* [475] that is a 1-form symmetry in finite-T quantum field theory [1263]) is not exact but approximate. Still, as long as the approximate symmetry is barely broken, an approximate value of critical temperature called the *pseudo-critical temperature*

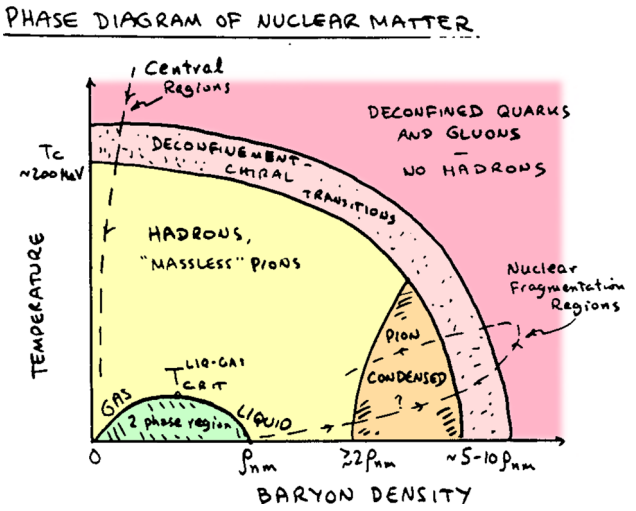


Fig. 163 A prototype of the QCD phase diagram. The handwritten phase diagram recaptured in Ref. [2254] was adapted and colorized here

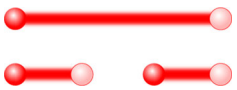


Fig. 164 The QCD string extends between a pair of test quark and antiquark (upper figure). The string breaking occurs once the fluxtube energy exceeds the meson mass (lower figure)

can be prescribed. Therefore, the temperature at which the deconfinement takes place is not uniquely defined but the pseudo-critical temperature is inevitably scheme dependent. This is why some theoretical QCD phase diagrams show the phase boundary with an uncertainty band associated with non-unique pseudo-critical temperatures.

The deconfined phase of gluons corresponds to the vacuum with spontaneous breaking of center symmetry, while quarks explicitly break this symmetry. Here, we shall avoid cumbersome mathematics and limit ourselves to pedagogical explanations about center symmetry. Let us consider a free energy gain, $f_{\rm q}(\boldsymbol{x})$, in response to a test quark placed at $\boldsymbol{x}$, and then construct a quantity called the Polyakov loop:

$$L(\boldsymbol{x}) = e^{-f_{\rm q}(\boldsymbol{x})/T}. \tag{7.5}$$

If the gluonic medium confines quarks, on the one hand, $f_{\rm q} \to \infty$ leads to $L \to 0$. Here we note that in a homogeneous system the $\boldsymbol{x}$ dependence of the expectation value can be safely dropped. On the other hand, L remains nonvanishing in the deconfined phase with $f_{\rm q} < \infty$. Thus, L serves as an order parameter for quark confinement, and in quantum field theory L is expressed by an expectation value of a 1-form line operator (i.e., Wilson line), see Ref. [1263] for generalized higher-form symmetries.

In reality with dynamical quarks, $f_{\rm q}$ would never diverge. The reason is easy to understand. A single test quark is a source to which a color fluxtube is attached. The color field energy is proportional to the squeezed fluxtube length. Thus, in a purely gluonic medium, a test quark cannot be screened and the fluxtube goes to spatial infinity, yielding $f_{\rm q} \to \infty$ and thus $L \to 0$ if confined. Fluctuations of dynamical quarks allow for the creation of a quark and an antiquark pair once the fluxtube energy exceeds the mesonic mass threshold as illustrated in Fig. 164. Then, the color field energy stored between a test quark and a test antiquark cannot become greater than twice the mesonic mass M_V. That is, the clustering decomposition property indicates $L(0)L^\dagger(\boldsymbol{x}) \sim e^{-2M_V/T}$ for sufficiently large $|\boldsymbol{x}|$, and so $L \sim e^{-M_V/T} > 0$ even in the confined hadronic phase.

This argument implies that the QCD string is further breached by fluctuations of surrounding quarks and holes at finite density. In other words the explicit breaking of center symmetry is enlarged in high-density matter and the mathematical concept of quark confinement would be obscure. Not that we do not yet find a better order parameter. The absence of the deconfinement order parameter could be attributed to the profound nature of dense QCD matter; namely, duality from hadronic to quark matter.

Now, let us shift gear to another aspect of the QCD vacuum and the QCD phase transition. The QCD Lagrangian contains quark mass parameters $m_{\rm q}$. The bare values of up and down (i.e., u and d) quark masses are only a few MeV, accounting for an only small fraction of the nucleon mass composed of u and d quarks. This huge discrepancy in the masses of quarks and baryons is explained by spontaneous breaking of *chiral symmetry*. Its order parameter is the chiral condensate $\langle\bar{q}q\rangle$ that gives rise to the dynamical mass, $M_{\rm q} \sim \Lambda_{\rm QCD} \gg m_{\rm q}$.

Almost all textbooks on quantum field theory affirm that the divergent zero-point oscillation energy is harmlessly discarded, but this common assertion is not valid for QCD because the mass is not a physical constant but is dynamically rooted in the QCD interactions. That is, the zero-point oscillation of quarks and antiquarks with $N_{\rm c}$ colors and $N_{\rm f}$ flavors gives [59]

$$\begin{aligned} E_{\rm zero} &= -2N_{\rm f}N_{\rm c}\int^{\Lambda}\frac{d^3p}{(2\pi)^3}\sqrt{p^2+M_{\rm q}^2} \\ &\simeq -N_{\rm f}N_{\rm c}\frac{\Lambda^4}{8\pi^2}\big[2+\xi^2+\mathcal{O}(\xi^4)\big], \end{aligned} \tag{7.6}$$

where Λ is a ultraviolet (UV) cutoff and the dimensionless parameter, $\xi = M_{\rm q}/\Lambda$, is assumed to be small. We see that the UV divergent term $\propto \Lambda^4$ is irrelevant to the dynamics, but we cannot drop another UV divergent term $\sim M_{\rm q}^2\Lambda^2 \sim \Lambda^4\xi^2$. Because $M_{\rm q}$ is related to the chiral condensate in the QCD vacuum, $\langle\bar{q}q\rangle$, the value of $M_{\rm q}$ is dynamically determined to minimize the vacuum energy. The coefficient of the quadratic term, ξ^2, is negative in Eq. (7.6), so that $E_{\rm zero}$ energetically favors larger $M_{\rm q}$. It is the condensation energy, $E_{\rm cond}$, that

competes the zero-point oscillation energy. Let us postulate that gluon mediation induces a four-fermionic interaction term $\sim \lambda\bar{q}q\bar{q}q$ in the low-energy Lagrangian where the mass dimension of the coupling constant, λ, is -2. Thus, a dimensionless coupling, $\hat{\lambda} = \Lambda^2\lambda$, is useful, and the dimensional analysis hints at a relation $M_\mathrm{q} = -2\lambda\langle\bar{q}q\rangle$. (In QCD $\langle\bar{q}q\rangle$ is known to take a negative value.) Then, the condensation energy from the interaction term is parametrically written as

$$E_\mathrm{cond} = N_\mathrm{f}N_\mathrm{c}\lambda\langle\bar{q}q\rangle^2 = N_\mathrm{f}N_\mathrm{c}\frac{M_\mathrm{q}^2}{4\lambda} = N_\mathrm{f}N_\mathrm{c}\frac{\Lambda^4}{4\hat{\lambda}}\xi^2. \qquad (7.7)$$

Now, the balance between two energies gives a condition for the spontaneous generation of $M_\mathrm{q} \neq 0$; that is, $\hat{\lambda} > 2\pi^2$, as first derived by Nambu and Jona-Lasinio [59,2255]. For the four-fermionic interaction stronger than this threshold, the QCD vacuum accommodates a nonvanishing chiral condensate.

From the Dirac mass term $m_\mathrm{q}\bar{q}q$ in the Lagrangian we see that the mass and the chiral condensate are conjugate to each other. It is thus evident that a nonzero $\langle\bar{q}q\rangle$ is a source to generate M_q even from a massless theory with $m_\mathrm{q} = 0$. The massless Dirac fermions are split into the right-handed and the left-handed components and they do not communicate. Therefore, for the theory with N_f flavors of massless quarks, a unitary rotation in flavor space is a symmetry in each of the right-handed and the left-handed sectors, i.e., the system enjoys the symmetry of $U_\mathrm{R}(N_\mathrm{f}) \times U_\mathrm{L}(N_\mathrm{f})$. Actually, the chiral condensate is decomposed as $\langle\bar{q}q\rangle = \langle q_\mathrm{R}^\dagger q_\mathrm{L} + q_\mathrm{L}^\dagger q_\mathrm{R}\rangle$ and it breaks the symmetry down to the vectorial one only, $U_\mathrm{V}(N_\mathrm{f})$. Among these symmetries, conventionally, $SU_\mathrm{R}(N_\mathrm{f}) \times SU_\mathrm{L}(N_\mathrm{f})$ is called *chiral symmetry* that is spontaneously broken so as to generate the dynamical mass, $M_\mathrm{q} \sim \Lambda_\mathrm{QCD}$, out from the bare mass, $m_\mathrm{q} \ll \Lambda_\mathrm{QCD}$.

We can expect, as elaborated below, that $\langle\bar{q}q\rangle$ should melt at high density and chiral symmetry should be restored then, which is commonly referred to as the *chiral phase transition*. It is the zero-point oscillation energy (7.6) that favors the symmetry breaking, and its expression involves the phase-space integration. At finite quark chemical potential μ_q which takes a larger value with increasing quark number density, the Fermi sphere is excluded from the phase-space integration due to the Pauli exclusion principle. Accordingly the symmetry breaking effect is diminished at finite μ_q. Therefore, it is a reasonable educated guess that the chiral phase transition makes a boundary curve on the density-temperature plane just like the deconfinement phase transition, as already depicted in Fig. 163.

The exact relation between the deconfinement phase transition with an approximate order parameter L and the chiral phase transition with another approximate order parameter $\langle\bar{q}q\rangle$ is a longstanding problem in QCD, and the satisfactory answer has not been found especially at finite density. As a function of m_q, actually, the deconfinement phase transition is exact only in the limit of $m_\mathrm{q} \to \infty$, while the chiral phase transition is exact only in the opposite limit of $m_\mathrm{q} \to 0$. The lattice-QCD data at finite T suggest that these two conceptually distinct phase transitions at opposite limits be interpolated by a single line for arbitrary m_q [2256].

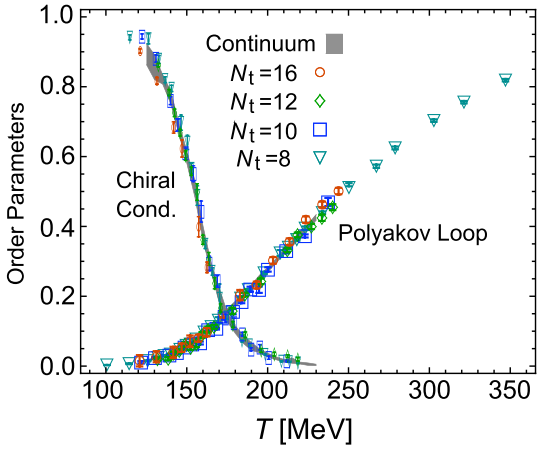


Fig. 165 Two order parameters as functions of the temperature at zero density as measured in the lattice-QCD simulation. N_t represents the site number along the temporal direction and the extrapolation to $N_t \to \infty$ defines the continuum limit. The figure and the lattice data are adapted from Ref. [2257]

Figure 165 shows the Polyakov loop and the chiral condensate as functions of T, normalized by the $T = 0$ values. We clearly notice that chiral symmetry is restored around $T_\mathrm{c} \sim 150\,\mathrm{MeV}$, and at the same time the Polyakov loop starts increasing from nearly zero, indicating a simultaneous deconfinement crossover. Thus, the lattice-QCD simulation at finite T has led us to a conclusion that two phase transitions of chiral restoration and deconfinement are somehow locked together. Actually, the prototype phase diagram in Fig. 163 assumes such tight locking of two transitions on the entire plane. However, as mentioned earlier, the barrier for the QCD string breaking would be eased by the density effect and the deconfinement would be more and more blurred at higher density, which implies a modernized version of the phase diagram as shown in Fig. 166. Here, as compared to the prototype in Fig. 163, there are three new ingredients added to Fig. 166; namely, the color superconductivity, the quarkyonic regime, and the QCD Critical Point (QCP). Moreover, Fig. 166 shows a new label, "sQGP" at high T and zero density, that refers to strongly correlated quark–gluon plasma. We will address only high-density aspects of QCD in this section, and for the physical interpretation of sQGP and the experimental characterization, see the previous section.

7.2.2 *Quark matter*

There is no clear definition that distinguishes nuclear and quark matter. In one working definition, quark matter is a

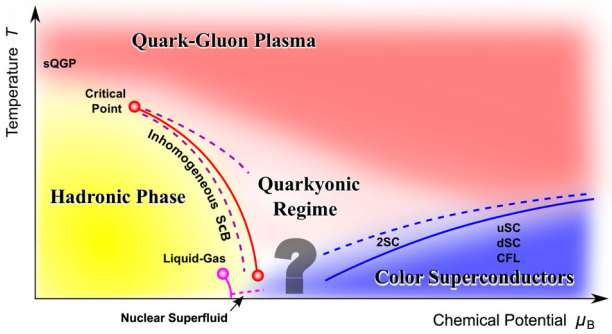


Fig. 166 A modern phase diagram of QCD with blurred deconfinement at higher density represented by the color gradation. Near the QCD Critical Point inhomogeneous spontaneous chiral-symmetry breaking (SCB) is predicted in the mean-field calculation, while the quasi-long-range order should take it over with fluctuation effects. The possible family of color-superconducting states includes not only the CFL and the 2SC phases (see the text) but also the uSC (where $\Delta_2 \neq 0$ and $\Delta_3 \neq 0$) and the dSC (where $\Delta_3 \neq 0$ and $\Delta_1 \neq 0$); see Refs. [2258,2259] for details. Unfortunately we are still unable to remove a big question mark

state of matter whose properties are reasonably approximated by perturbative QCD (pQCD) calculations. The presence of quark matter in the neutron star (NS) has been proposed by Collins and Perry [463] based on the asymptotic freedom at high baryon density (see also Ref. [2260] for a preceding hypothesis on quark matter). If the momentum scale associated with the running strong coupling constant, α_s, is characterized by the baryon chemical potential μ_B or the quark chemical potential $\mu_q = \mu_B/3$, asymptotically free quarks should be liberated from hadrons as the density goes above a certain threshold. In Ref. [463] the leading order (LO) contributions, i.e., thermodynamic quantities of free massless quarks, were considered:

$$P_{(0)} = \frac{N_c}{12\pi^2}\sum_{i=1}^{N_f}\mu_i^4. \tag{7.8}$$

The next-to-leading order (NLO) diagrams add corrections of $\mathcal{O}(\alpha_s)$, i.e.,

$$P_{(2)} = -\left(\frac{\alpha_s}{\pi}\right)\frac{N_g}{16\pi^2}\sum_{i=1}^{N_f}\mu_i^4, \tag{7.9}$$

where the adjoint color factor, $N_g = N_c^2 - 1$, was introduced. The next-to-next leading order (N^2LO) calculations produce a logarithmic term with μ_q dependence in the argument. All the terms are not listed up here (see Ref. [2261] for the formulation and Ref. [2262] for the QCD application); the logarithmic term looks like

$$P_{(4)} = +\left(\frac{\alpha_s}{\pi}\right)^2\frac{N_g\beta_0}{64\pi^2}\sum_{i=1}^{N_f}\mu_i^4\ln\frac{\mu_i^2}{\mu_0^2} + (\text{non-log terms}), \tag{7.10}$$

where $\beta_0 = (11N_c - 2N_f)/3$. Non-logarithmic terms are omitted. Even if α_s is sufficiently small, $\alpha_s \ln(\mu_i^2/\mu_0^2)$ may become large, and then the perturbative expansion breaks down. A remedy for this problem of the singular logarithm is the resummation over the leading-log terms. For simplicity let us assume that all the quark chemical potentials are identical. (More generally one can introduce a flavor averaged value of the chemical potential.) Actually, it is easy to confirm that, if α_s is upgraded to the running one, i.e.,

$$\frac{\alpha_s(\mu_q)}{\pi} = \frac{\alpha_s}{\pi}\left[1 + \left(\frac{\alpha_s}{\pi}\right)\frac{\beta_0}{4}\ln(\mu_q^2/\mu_0^2)\right]^{-1}, \tag{7.11}$$

an expansion of Eq. (7.9) can reproduce Eq. (7.10). In other words, such dangerous logarithmic terms are absorbed into the density-dependent running coupling, $\alpha_s(\mu_q)$ (see Ref. [2262] for more details). In this way the perturbative calculation is justified at high enough density.

From this construction of the running coupling constant, one can easily imagine that the resummation is not free from an arbitrary choice of irrelevant constants. Instead of $\ln(\mu_q^2/\mu_0^2)$, one could try to make a resummation of $\ln(\mu_q^2/\mu_0^2) + C = \ln(\mu_q^2/\mu_1^2)$ with $\mu_1^2 = \mu_0^2 e^{-C}$. In principle, an optimal choice of C could exist to reduce higher-order corrections. If C is close to the optimal point, the results are expected to be flat against changes of C, and it is customary to check the stability of the results by changing X of $\alpha_s(X\mu_q)$. Here, the logarithmic term in $\alpha_s(X\mu_q)$ takes the form of $\ln(X^2\mu_q^2/\Lambda^2_{\overline{\rm MS}})$ in the $\overline{\rm MS}$ scheme [2263]. It is then found that such variation of $X = 1 \sim 4$ leads to huge uncertainty unless μ_q becomes unphysically large. This is sometimes referred to as the slow-convergence problem. The next correction, i.e., the N^3LO contribution is expected to stabilize the results better, and indeed the soft N^3LO part has been shown to cure the slow-convergence problem partially [2264,2265].

7.2.3 Color-superconducting phases

The pQCD calculation is not capable of describing dynamical generation of $\langle\bar{q}q\rangle$, which is apparently consistent with melting chiral condensate at high density. However, even at high density, high enough to validate pQCD, the chiral condensate is not simply gone.

Quarks carry a fundamental charge in color SU(3), and so two charges of a pair of quarks (i.e., a diquark) connected by one-gluon exchange are coupled via

$$(t^a)_{ij}(t^a)_{kl} = -\frac{N_c+1}{4N_c}(\delta_{ij}\delta_{kl} - \delta_{il}\delta_{kj}) + \frac{N_c-1}{4N_c}(\delta_{ij}\delta kl + \delta_{il}\delta_{kj}) \tag{7.12}$$

corresponding to $\mathbf{3} \otimes \mathbf{3} = \bar{\mathbf{3}} \oplus \mathbf{6}$ in the group theoretical language. Interestingly, as implied from the sign of each term in the above decomposition, the inter-quark interaction in the $\bar{\mathbf{3}}$ channel is attractive, while the $\mathbf{6}$ channel interaction is repulsive. This attractive nature is intuitively understood as follows: Suppose that two quarks are infinitely separate (in the deconfined phase), then the total field energy is just twice of the field energy associated with a single quark. If two quarks approach and make a composite of $\bar{\mathbf{3}}$, the total field energy is the same as that of a single quark, that is, a half of the original total energy. So, the energy decreases as two quarks touch. Consequently two quarks in the $\bar{\mathbf{3}}$ channel should feel an attractive force to minimize the total energy.

The most favored diquark channel is color anti-triplet (anti-symmetric) and spin singlet (anti-symmetric) and thus the flavor must be anti-symmetric. The diquarks generally carry two color indices and two flavor indices, but the diquark matrix in the most favored color-flavor channel simplifies to

$$\Phi_{i\alpha} = \epsilon_{ijk}\epsilon_{\alpha\beta\gamma}\, q^T_{j\beta} C\gamma_5 q_{k\gamma}. \tag{7.13}$$

Here, $C = i\gamma^0\gamma^2$ is the charge conjugation matrix necessary to make the diquark a Lorentz scalar. The Latin and the Greek letters represent the indices in flavor and color space, respectively.

In the three-flavor symmetric limit with $m_u = m_d = m_s$, the flavor rotation as well as the color rotation is a symmetry of the system. Then, it is possible to choose the flavor and the color bases to diagonalize $\Phi_{i\alpha}$. Without loss of generality we can parametrize the diquark condensate as

$$\langle \Phi_{i\alpha} \rangle = \delta_{i\alpha}\, \Delta_i. \tag{7.14}$$

Under the identification of $i = 1, \alpha = 1$ for up (u) and red (r), $i = 2, \alpha = 2$ for down (d) and green (g), and $i = 3, \alpha = 3$ for strange (s) and blue (b), for example, Δ_1 involves pairings of gd-bs and gs-bd quarks. A state of quark matter with $\Delta_1 \neq 0$, $\Delta_2 \neq 0$, and $\Delta_3 \neq 0$ is known as the color-flavor locking (CFL) phase. The CFL phase is considered to be the ground state as long as the strange quark mass is ignored. In the opposite limit of infinitely heavy strange quark mass, we can regard quark matter as composed from only light flavors. In this case only Δ_3 (involving ru-gd and rd-gu quark pairings) can take a nonzero value, while $\Delta_1 = \Delta_2 = 0$ due to suppression of strange quarks. Such a state of $\Delta_1 = \Delta_2 = 0$ and $\Delta_3 \neq 0$ is called the two-flavor color-superconducting (2SC) phase.

Which symmetry should spontaneously be broken by the diquark condensate is a nontrivial question. Let us first consider the 2SC phase. We note that the local gauge symmetry is never broken. Then, the baryon $U_{\rm V}(1)$ symmetry is not broken in the 2SC phase since its rotation on Δ_3 can be always canceled by unbroken electromagnetic transformation. The same argument concludes that flavor (chiral) symmetry is not broken, either. Therefore, in the 2SC phase, all global symmetries are unbroken, only modified with a mixture of local symmetry. One might think that color-superconducting phases assume deconfined quark matter, but as shown in Ref. [2266], the low-energy physics in the 2SC phase is governed by ungapped gluons in the unbroken $SU(2)$ sector and color confinement persists. Theoretically speaking, there is no gauge-invariant order parameter to define the 2SC phase.

In reality, however, the 2SC phase is anyway taken over by the CFL phase at high density where the strange quark mass is negligible. The $N_{\rm f} = 3$ world is drastically different from the 2SC phase. The $U_{\rm V}(1)$ symmetry can no longer be restored by the electromagnetic symmetry because Δ_1 and $\Delta_{2,3}$ are differently charged. Thus, the CFL phase has a superfluid, and a vortex configuration is topologically stabilized. Also, chiral symmetry is spontaneously broken. We note that the diquark condensate has both the left-handed and the right-handed components; that is, $\langle qq \rangle = \langle q_{\rm R} q_{\rm R} \rangle + \langle q_{\rm L} q_{\rm L} \rangle \neq 0$, and $\langle q_{\rm R,L} q_{\rm R,L} \rangle$ breaks $SU_{\rm R,L}(3)$. The vectorial rotation in flavor space can still be canceled by unbroken color rotation, so the symmetry breaking pattern in the CFL phase turns out to be: $SU_{\rm R}(3) \times SU_{\rm L}(3) \to SU_{\rm V}(3)$. Interestingly, this is identical to the symmetry breaking in the hadronic phase. Actually the gauge-invariant order parameter of the CFL phase is, $\langle (\bar{q}\bar{q})(qq) \rangle \sim \langle (\bar{q}q)^2 \rangle$ that induces $\langle \bar{q}q \rangle \neq 0$ unless the anomalous $U_{\rm A}(1)$ is restored. The observation of exactly the same symmetry properties has led to a conjecture of continuity between the hadronic phase (i.e., the confined phase) with superfluidity and the CFL phase (i.e., the Higgs phase) [2267].

We can develop an intuitive understanding of the continuity. In the case of electron superconductivity, there is no gauge-invariant order parameter, and one might think that the theoretical characterization is as problematic. In this case, however, the solution has already been known. Because the Cooper pairs have twice the elementary charge, they cannot completely screen a single elementary charge. This would lead to an emergent Z_2 symmetry and the superconducting state is unambiguously defined by the symmetry.

This argument makes it clear why the CFL phase is so special. As mentioned earlier, the most favored diquark is found in the color triplet (and the anti-triplet) channel made from $\mathbf{3} \otimes \mathbf{3} \to \bar{\mathbf{3}}$. So, the Cooper pairs (i.e, the diquarks) are charged just like the fundamental (anti-)charge. Thus, a fundamental charge can be screened by Cooper pairs and the definition of the CFL phase is obscured, which underlies the continuity scenario between hypernuclear matter and CFL quark matter.

The continuity scenario cannot be applied to the 2SC phase as it is, but it was pointed out in Ref. [2268] that the NS environment can realize continuity within the two-flavor sector only. The idea is that the electric neutrality requires twice more d-quarks than u-quarks, and free d-quarks (not paired

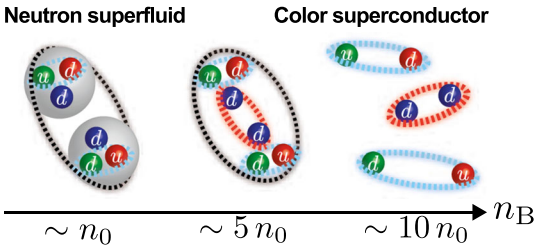


Fig. 167 An illustration of the two-flavor continuity scenario between nuclear matter and 2SC+d quark matter in the NS environment in β equilibrium. Figure taken from Ref. [2268]

with u-quarks) may form a condensate of $\langle dd \rangle$. In this exotic phase that may be called the 2SC+d phase, the electromagnetic rotation cannot cancel rotations in Δ_3 and $\langle dd \rangle$ simultaneously, and so it is a superfluid with $U_{\rm V}(1)$ breaking, and also, it spontaneously breaks chiral symmetry. In this way, as illustrated in Fig. 167, the continuity can be formulated.

Recently, the quark–hadron continuity scenario is encountering a fatal challenge. As mentioned, the emergent Z_2 symmetry characterizes ordinary electron superconductivity, and to see it mathematically, a Wilson loop as a symmetry generator is acted on a magnetic vortex operator. The magnetic flux in a superconducting cylinder is quantized in units not of $2\pi\hbar c/e$ but of $\pi\hbar c/e$ due to doubly charged Cooper pairs. The symmetry operation with the Wilson loop, hence, results in a Z_2 phase. The same exercise in the CFL phase replaces the magnetic vortex with the non-Abelian CFL vortex [2269] that carries both the global phase as well as the chromo-magnetic flux. From the explicit expression of the non-Abelian CFL vortex, it has been shown in Ref. [2270] that a Z_3 symmetry emerges (see also Ref. [2271] for more mathematical discussions). The hadronic phase presumably confines any color degrees of freedom, and it is a natural anticipation (but not proven yet) that this Z_3 symmetry operation is merely trivial in the confined phase. If so, the spontaneous breaking of emergent Z_3 symmetry should result in a phase transition from nuclear to quark matter. It is not yet settled theoretically whether a phase transition really separates nuclear and quark matter. The symmetry arguments are convincing, but the calculations are feasible only at high enough density, not at intermediate density where a transitional change may occur. As we will argue later, astrophysical observations constrain the strength of first-order phase transition for the neutron-rich NS matter, and for the moment it disfavors the first-order phase transition.

7.2.4 Quarkyonic regime

In the large-$N_{\rm c}$ limit the duality between nuclear and quark matter has been recognized by McLerran and Pisarski [2272] and they named the dual regime of matter Quarkyonic Matter after a combination of "quark" and "baryonic". It should be noted that Quarkyonic Matter is not a novel phase of matter but it refers to a regime in which the duality is manifested.

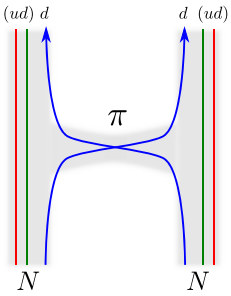


Fig. 168 A quark description of two-body NN interaction

The conjectured duality is based on the large-$N_{\rm c}$ counting of the pressure. Along the temperature axis at zero baryon density, the pressure jumps from $\mathcal{O}(1)$ in the confined phase to $\mathcal{O}(N_{\rm c}^2)$ in the deconfined gluonic phase, which defines a first-order phase transition even with dynamical quarks. Then, along the axis of the baryon/quark chemical potential at zero temperature, one might also think of a phase transition from $\mathcal{O}(1)$ in confined nuclear matter to $\mathcal{O}(N_{\rm c})$ in deconfined quark matter. This naïve order counting implicitly neglects the contribution from interactions that could be dropped in the dilute/dense limits, but not in the intermediate density region. Actually, in the large-$N_{\rm c}$ limit, the amplitude of meson scattering is suppressed so that mesons can be regarded as non-interacting particles, while baryons interact strongly. It is immediately understood why the strength of baryon interaction scales as $\mathcal{O}(N_{\rm c})$. The one pion exchange process for the two-body nucleon–nucleon (NN) interaction can be viewed microscopically as a quark hopping from one to the other baryon as shown in a schematic picture in Fig. 168. There are $N_{\rm c}^2$ combinations of quark exchanges, among which color singlets are of $\mathcal{O}(N_{\rm c})$. In contrast, the n-point interaction vertices of mesons scale as $\mathcal{O}(N_{\rm c}^{1-n/2})$ that goes to zero as $N_{\rm c} \to \infty$ for $n \geq 3$. All the multi-body interactions of nucleons turn out to scale as $\mathcal{O}(N_{\rm c})$ which coincides with the scaling property peculiar to quark matter. In this way, in Quarkyonic Regime, the system is still in the confined phase and the relevant degrees of freedom are baryons, but the pressure is sensitive to quark degrees of freedom through inter-baryonic interactions.

Now, we see that the deconfinement phenomenon induced by baryons at high density is far more nontrivial than the high temperature situation dominated by mesons. For weakly interacting mesons the onset of deconfinement can be approximated as an overlap of wavefunctions, that agrees with a picture of site percolation. For baryons, however, the onset of deconfinement would be rather located at the density where the NN, NNN, and arbitrary multi-body interactions become comparably strong, building a connected network of interacting bonds. In the language of percolation, hence, it would not be the site percolation but the bond percolation that is appropriate for high-density deconfinement. It

has also been speculated that the deconfinement onset could be delayed toward higher density by quantum fluctuations as suggested in a quantum percolation picture [2273].

In Quarkyonic Regime the state of matter is not simply quark-like nor baryon-like, but something that shares both features. It is unlikely that there is any sharp deconfinement boundary in the phase diagram as drawn in the prototype in Fig. 163. This is why deconfinement is represented by smooth gradation over Quarkyonic Regime in Fig. 166. It is quite suggestive that both the CFL phase and Quarkyonic Regime favor smooth continuity between nuclear and quark matter in parallel, even though the diquark condensate is suppressed in the large-N_c limit and color-superconducting matter and Quarkyonic Matter seem not to coexist.

7.2.5 Critical point vs. inhomogeneous states

So far, we have focused on deconfinement, and we shall now turn to the chiral phase transition at finite density. It has been established that the chiral phase transition at physical quark masses is a smooth crossover if the chiral restoration is induced by the temperature effect [2274]. Most chiral models predict that, as the baryon density increases, the behavior of the chiral condensate as a function of increasing T becomes steeper. Eventually, in some chiral models, the chiral restoration occurs with a discontinuous jump in the chiral condensate, and the separation point between the smooth crossover and the first-order phase transition corresponds to the exact second-order critical point, which is commonly called the QCD Critical Point (QCP). It is sometimes referred to as the critical end point (CEP) of QCD matter as well. The presence of the QCP was first recognized in the Nambu–Jona–Lasinio model by Asakawa, Yazaki [2275], and in a QCD-like model by Barducci, Casalbuoni, De Crutis, Gatto, Pettini [2276], independently. For a comprehensive review on the order of chiral restoration at the early stage, see Ref. [2277].

In the language of the Ginzburg–Landau theory, the grand potential has an expansion,

$$\Omega = \frac{\alpha_2}{2}M^2 + \frac{\alpha_4}{4}M^4 + \frac{\alpha_6}{6}M^6 + \mathcal{O}(M^8), \tag{7.15}$$

with respect to an order parameter $M \sim \langle \bar{q}q \rangle$ (proportional to the constituent quark mass). For simplicity the bare quark mass effect that induces a symmetry-breaking term $\propto M$ is dropped. The coefficients, α_i, are functions of T and μ_B. If α_2 changes its sign while $\alpha_4 > 0$ is kept, a second-order phase transition is derived. If $\alpha_2 = 0$ and α_4 changes its sign for $\alpha_6 > 0$, a tricritical point appears.

Interestingly, the QCP has nothing to do with the original chiral symmetry of QCD, and the universality class belongs to the same as the three-dimensional Ising model. Only when the bare quark mass is vanishing, as mentioned above, the QCP is located on the chiral phase transition, which exhibits tricriticality. At finite bare quark mass that explicitly breaks chiral symmetry, the QCP is identified as the Z_2 liquid-gas transition whose order parameter is the density, i.e., a conserved quantity coupled with the energy–momentum tensor, resulting in the dynamical universality class of the model H [2278].

The QCP can be an unambiguous landmark, if experimentally confirmed, on the QCD phase diagram. It is, however, quite nontrivial what plays a role of a signature. The most well-investigated quantity in the search for the QCP is the fluctuation observable. Because the correlation functions scale with the correlation length ξ, that ideally diverges near the critical point but does not in reality due to the critical slowing down, one can make a robust prediction for the critical behavior characterized by ξ to the power of critical exponent. Although the time evolution away from the QCP may wash out the critical signature, the fluctuation of the conserved quantities such as the baryon number, the electric charge, and the strangeness (within the time scale of the strong interaction) could retain a trace of criticality if its value is frozen inside the critical region. This means that, to probe the QCP in a heavy-ion collision experiment, the created hot and dense matter must cool down along the trajectory hitting the critical region of the QCP, and the chemical freezeout (that fixes the ratio of the particle species) must be located sufficiently near the QCP. Such requirements may hold or may not.

There is no reliable QCD-based prediction for the location of the QCP due to the sign problem, but the virtue of the QCP search is that the critical theory provides us with unique theoretical prediction once its location on the phase diagram is experimentally constrained. We have already learnt a lot about fluctuations from nearly zero baryon density (at high collision energy $\sqrt{s_{NN}} \gtrsim 100\,\mathrm{GeV}$) to high density (at $\sqrt{s_{NN}} \sim 3\,\mathrm{GeV}$) from the heavy-ion collision experiments. See discussions in the previous section and the figure to show the data of κ_4/κ_2. For a review including related topics, especially the kurtosis (fourth order fluctuation of the proton number) and the skewness (third order fluctuation), see Ref. [2279] and references therein.

It is also mentioned that constructing an effective description of low-energy dynamics near the critical point is an intriguing theoretical challenge. Typically the time evolution of locally equilibrated matter is governed by undamped zero modes associated with conservation laws, which constitutes the hydrodynamic description based on the derivative expansion. In the vicinity of the critical point, the critical slowing down breaks the clear scale separation. Then, the correlation of the diffusive mode, that is the slowest one, should be coupled in the hydrodynamic equations, and such a generalized framework – called the "Hydro+" – has been proposed [2280].

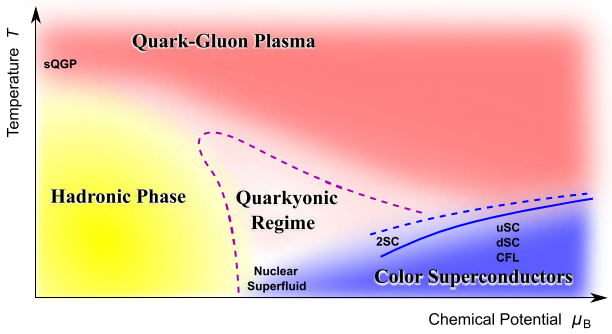


Fig. 169 Another phase diagram of QCD matter without the QCD Critical Point, which may be the case in the neutron star matter in β equilibrium for which the nuclear liquid-gas critical point is known to disappear

One subtlety remains. The grand potential, Ω, can be generally expressed as a functional of the order parameter, which is denoted by $M(\boldsymbol{x})$ here. If the spatial variation is smaller than other scales, the derivative expansion makes sense and the spatial profile of $M(\boldsymbol{x})$ is optimized to minimize the energy locally. It was Nickel [2281] who first recognized that in a typical chiral model in the mean-field approximation the local energy takes the following structured form:

$$\begin{aligned}\Omega &= \tfrac{\alpha_2}{2} M(\boldsymbol{x})^2 + \tfrac{\alpha_4}{4}\left[M(\boldsymbol{x})^4 + (\nabla M(\boldsymbol{x}))^2\right] \\ &+ \tfrac{\alpha_6}{6}\left[M(\boldsymbol{x})^6 + 5(\nabla M(\boldsymbol{x}))^2 M(\boldsymbol{x})^2 + \tfrac{1}{2}(\Delta M(\boldsymbol{x}))^2\right].\end{aligned} \tag{7.16}$$

This is a striking result. At the tricritical point (and near the QCP also) α_4 should change its sign. For $\alpha_4 < 0$, the coefficient of the first derivative correction turns out to be negative too, which means that $\nabla M(\boldsymbol{x}) \neq 0$ would lower the local energy. Therefore, the above form of the expanded energy indicates that the ground state should be spatially inhomogeneous.

The onset of inhomogeneity is called the Lifshitz point and Nickel's calculation was the first clarification for an explicit relation between the QCP and the Lifshitz point, though there were preceding works to hint at the possibility of inhomogeneous ground states [2282]. Whether the QCP and the Lifshitz point exactly coincide or not depends on the model choice and the approximation, and in more realistic model studies the QCP is overridden by the inhomogeneous states (see Ref. [2283] for a comprehensive review). Interestingly, such an inhomogeneous state is favored also in Quarkyonic Regime; the large-N_c limit justifies an approximation of nuclear matter by a Skyrme crystal that inevitably gives rise to inhomogeneous chiral condensate.

Therefore, another view of the QCD phase diagram may look like Fig. 169 on which the QCP is taken over by an approximate triple point where the hadronic phase, the quark–gluon plasma, and Quarkyonic Regime (or the inhomogeneous state) meet [2253]. Once the large-N_c approximation is relaxed, however, the thermal fluctuations of phonons and pions should be taken into account. It is known by now that inhomogeneous condensates are unstable and the quasi-long-range order (i.e., not exponential but algebraic decay of the order parameter correlation) could survive there [2284,2285]. In contrast to the QCP on Fig. 166, it is a demanding question what can be an experimental signature to detect Quarkyonic Regime (or the quasi-long-range order) if the genuine phase diagram is like Fig. 169. Even without inhomogeneous condensates, for example for the theory proposal, the order parameter modes could be modified nontrivially to have a damped dispersion relation similar to the roton, which was discussed as a candidate for the observable signature [2286].

7.2.6 Astrophysical constraints

Figure 169 looks like one variant of conjectured phase diagrams, but a special realization of dense matter in accord to Fig. 169 is known. That is, the state of dense matter in deep cores of a neutron star (NS) satisfies the β equilibrium condition and contains more neutrons than protons due to the Coulomb interaction. This makes a sharp contrast to the heavy-ion collision whose time scale is shorter than the weak interaction, and flavor changing processes are negligible. It is important to note that the isospin contents would significantly affect the phase structure of QCD matter. A well-known example is that the first-order liquid-gas phase transition of symmetric nuclear matter in Fig. 166 does not exist any more in the NS matter; that is, pure neutron matter is not a self-bound fermionic system unlike symmetric nuclear matter. Then, it would be conceivable that the β equilibrium condition simplifies the phase diagram from the conventional one as in Fig. 166 into a smoother shape without any solid phase boundary as in Fig. 169.

In fact, as we saw already before, the quark–hadron continuity scenario of the color-superconducting phase and the large-N_c Quarkyonic Regime supports a picture of smooth crossover from nuclear to quark matter. Here, we discuss astrophysical constraints about the phase transition of QCD matter. The internal structure of the NS follows from the hydrostatic condition (called the Tolman–Oppenheimer–Volkoff equation) between the inward gravitational force and the outward pressure gradient. To this end, the calculation of the pressure gradient needs the relation of the pressure as a function of the baryon density, i.e., $p = p(\rho)$, or as a function of the energy density, $p = p(\varepsilon)$, which is referred to as the equation of state (EOS).

There is no first-principles derivation of the EOS except for the zero-density and the high-density limits and the EOS is the most crucial source of uncertainty in NS phenomenology. For a given ε, the EOS with larger p (and smaller p) is

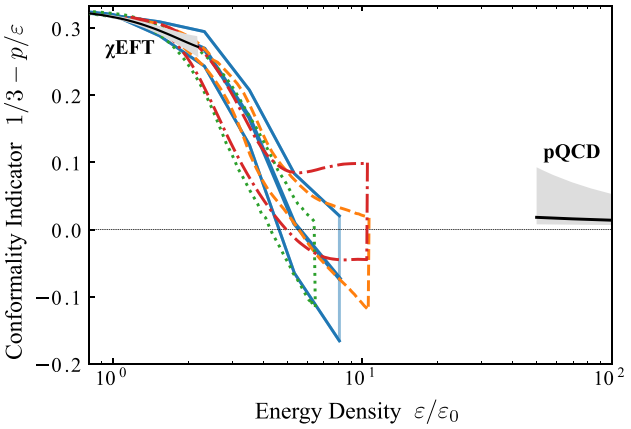


Fig. 170 Conformality indicator deduced from the neutron star data as a function of the energy density normalized by $\varepsilon_0 = 150\,\mathrm{MeV/fm^3}$. Bands with different colors refer to the results from Refs. [2268,2292–2294]. Figure adapted from Ref. [2295]

called "stiff" (and "soft", respectively). Generally speaking, stiffer EOSs can support heavier NSs, and so the heaviest NS can give us the information about the EOS stiffness. If an assumed model cannot predict a stiff EOS enough to explain the experimentally confirmed largest NS mass, this model is falsified. In the presence of the first-order phase transition, $p = p(\varepsilon)$ should have a plateau, i.e., a window of ε with a constant p, in the mixed-phase region, which generally makes the EOS softer.

In 2010 the mass measurement in a binary system (an NS and a white dwarf) by means of the Shapiro time delay established the existence of a two-solar-mass NS (PSR J1614-2230 [2287]). Later, similar massive NSs (PSR J0348+0432 [2288] and PSR J0740+6620 [2289]) have been discovered. These observations are extremely useful to make strict constraints and to exclude some of soft EOSs. In particular, the first-order phase transition is disfavored; it should be weak if the first-order phase transition takes place at moderate density reachable in the NS environment, or the first-order phase transition can occur only at large density beyond the NS region [2290]. In principle, a very rapid stiffening before/after the first-order phase transition could also yield an EOS that supports the massive NSs, but justification of the underlying mechanism needs further investigations. Actually, the ab initio estimates based on the chiral effective theory (χEFT) and the pQCD suggest that the nuclear EOS near the saturation density n_0 and quark EOS for high density $\gtrsim 5n_0$ are both softer than empirically adopted EOSs, and the stiffening should occur around 1.5–1.8 times n_0 [2291].

It is quite suggestive that such behavior of rapid stiffening from a low-density soft EOS is inferred from the experimental data, irrespective of any theoretical conjecture. The distribution of masses and radii of the observed NSs can be analyzed by probabilistic methods and the preferred EOS can be constructed from the observational data only. Figure 170 shows a specific combination of the EOS, i.e., $1/3 - p/\varepsilon$, as a function of dimensionless energy density $\varepsilon/\varepsilon_0$ with $\varepsilon_0 = 150\,\mathrm{MeV/fm^3}$, that approaches zero in the conformal limit at high density. In Fig. 170 the orange, the green, and the red lines represent the results from the Bayesian analyses of the observational data in Refs. [2292], [2293], and [2294], respectively. The blue line represents the results from the neural network analysis in Ref. [2268]. An intriguing finding is that the system seems to restore the conformality quite rapidly as first quantified in Ref. [2295]. The pQCD results at asymptotically high density as indicated in Fig. 170 are nearly conformal because the density scale is sufficiently larger than Λ_{QCD} and the system is weakly interacting. Thus, the NS experimental data imply the realization of strongly correlated conformal matter far earlier at not asymptotically high but just intermediate density. The microscopic origin of early conformality is to be identified by future studies.

Finally, let us briefly mention the impact of the gravitational wave signal from the NS merger. So far, the LIGO-Virgo collaboration reported two events of GW170817 and GW190425 as candidates of the NS–NS merger. In particular the former happened at a short distance of only 40 Mpc, and the electromagnetic counterpart (called the "kilonova") was also detected. For the moment only the signal from the inspiral stage before the merger has led to an EOS constraint in terms of the tidal deformability coefficient [2296,2297], which turned out to be consistent with preceding constraints from the symmetry energy measurement [2298] as well as the NS mass-radius distributions. In the future the post-merger stage after the merger might be detected, and if so, an extremely dense state of matter, even denser than the largest density in the central core of the NS, could be probed, which will eventually clarify the nature of Quarkyonic Regime, quark matter, and hopefully color-superconducting states.

8 Mesons

Conveners:
Eberhard Klempt and Curtis Meyer

The Particle Data Group lists 78 light mesons with u and d quarks, 50 of them are "established", with 3* or 4* ratings. 25 mesons carry strangeness, 16 of them are established. Most mesons show a regular pattern, their masses are mostly compatible with a Regge behavior in L and N. Curtis Meyer introduces the meson quantum numbers and their regularities. The scalar mesons of lowest mass have resisted for a long time an undisputed acceptance with proper poles in the complex energy plane. José Pelaez shows how unitarity, analyticity and dispersion relations are exploited to determine the scalar partial wave and to extract the poles with high precision. A driving force in meson spectroscopy is since long the quest

for hybrids, in particular those with exotic quantum numbers, and for glueballs. Boris Grube and Eberhard Klempt present old and recent evidence for these states. 12 (7) established (candidate) charmed mesons are known at present, 7 (5) mesons with a bottom and a light quark, 6 (5) with a strange and 2 with a charm quark. Charmonium (and bottomonium) played a crucial role for the general acceptance of the quark model. Nowadays, 39 $c\bar{c}$ states are known, 25 of them established. The so-called XYZ states, unexpected states or states with unexpected properties, play an important role to understand the richness of QCD. Marco Pappagallo reports on the crime story of $X(3872)$ with its dual nature, on the unexpected $Y(4260)$ and the discovery of $Z_c^+(4430)$ and $T_{cc}(3875)$, both with minimal four-quark content. Nora Brambilla outlines the different approaches to identify the degrees of freedom driving the exotic states.

8.1 The meson mass spectrum, a survey

Curtis Meyer

8.1.1 Introduction

In the quark model, mesons are states containing quarks, antiquarks and gluons such that the net baryon number of the state is zero. Conventional mesons are described as bound states of a quarks and an antiquark ($q\bar{q}$) and can be viewed as similar to positronium (e^+e^-). Exotic mesons can include hybrids, which are $q\bar{q}g$ states with valence glue, four-quark states containing two quarks and two antiquarks, and glueballs containing only glue. These more exotic forms will be discussed in later sections, this section will deal with the ordinary mesons, referred to here as simply mesons. Mesons containing only u, d and s quarks are known as *light-quark mesons*. Given three quarks and three antiquarks, nine possible $q\bar{q}$ combinations can be made. These nine mesons form a so-called nonet where the members have the same well-defined quantum numbers: total spin J, parity P, and C-parity C, represented as J^{PC}.

8.1.2 Meson quantum numbers

The J^{PC} quantum numbers of quark–antiquark systems are functions of the total spin, S, and the relative orbital angular momentum, L. The spin S and angular momentum L combine to yield the total spin J,

$$\vec{J} = \vec{L} \oplus \vec{S}, \tag{8.1}$$

where L and S add as two angular momenta.

Parity is the result of a mirror reflection of the wave function, taking $\vec{r}$ into $-\vec{r}$. It can be written as

$$P\left[\psi(\vec{r})\right] = \psi(-\vec{r}) = \eta_P \psi(\vec{r}), \tag{8.2}$$

Table 9 The allowed J^{PC} quantum numbers for light-quark mesons with L up to 4

L	S	J^{PC}	L	S	J^{PC}
0	0	0^{-+}	3	0	3^{+-}
0	1	1^{--}	3	1	2^{++}
1	0	1^{+-}	3	1	3^{++}
1	1	0^{++}	3	1	4^{++}
1	1	1^{++}	4	0	4^{+-}
1	1	2^{++}	4	1	3^{--}
0	2	2^{-+}	4	1	4^{--}
1	2	1^{--}	4	1	5^{--}
1	2	2^{--}			
1	2	3^{--}			

where η_P is the eigenvalue of parity. An application of parity twice must return the original state, $\eta_P = \pm 1$. In spherical coordinates, the parity operation reduces to the reflection of a Y_{LM} function,

$$Y_{LM}(\pi - \theta, \pi + \phi) = (-1)^L Y_{LM}(\theta, \phi). \tag{8.3}$$

From this, we conclude that $\eta_P = (-1)^L$. For a $q\bar{q}$ system, the intrinsic parity of the antiquark is opposite to that of the quark, which yields the total parity of a $q\bar{q}$ system as

$$P(q\bar{q}) = -(-1)^L. \tag{8.4}$$

Charge conjugation, C, is the result of a transformation that takes a particle into its antiparticle. For a $q\bar{q}$ system, only electrically-neutral states can be eigenstates of C. In order to determine the eigenvalues of C (η_C), we need to consider a wave function that includes both spatial and spin information

$$\Psi(\vec{r}, \vec{s}) = R(r) Y_{lm}(\theta, \phi) \chi(\vec{s}). \tag{8.5}$$

If we consider a $u\bar{u}$ system, the C operator reverses the meaning of u and $\bar{u}$ which has the effect of mapping the vector $\vec{r}$ to the u quark into $-\vec{r}$. Thus, following the arguments for parity, the spatial part of C yields a factor of $(-1)^L$. The C operator also reverses the two individual spins. For a symmetric χ, we get a factor of 1, while for an antisymmetric χ, we get a factor of -1. For two spin-$\frac{1}{2}$ particles, the $S = 0$ singlet is antisymmetric, while the $S = 1$ triplet is symmetric. Finally, there is an additional factor of -1 when we interchange two fermions. Combining all of this, we find that the C-parity of (a neutral) $q\bar{q}$ system is

$$C(q\bar{q}) = (-1)^{L+S}. \tag{8.6}$$

In Table 9 are shown the J^{PC}s and the possible values of L and S up to L of 3. From the list, the J^{PC} values of 0^{--}, 0^{+-}, 1^{-+}, 2^{+-} and 3^{+-} are missing. These missing J^{PC} are referred to as *exotic*.

Because C-parity is only defined for neutral meson, we define G-parity to extend this to all non-strange $q\bar{q}$ states,

independent of charge. For isovector states ($I = 1$), C would transform a charged member into the oppositely charged state (e.g. $\pi^+ \to \pi^-$). In order to transform this back to the original charge, we would need to perform a rotation in isospin ($\pi^- \to \pi^+$). For a state of whose neutral partner has C-parity C, and whose total isospin is I, the G-parity is defined to be

$$G = C \cdot (-1)^I, \tag{8.7}$$

which can be generalized to

$$G(q\bar{q}) = (-1)^{L+S+I}. \tag{8.8}$$

The latter is valid for all of the $I = 0$ and $I = 1$ non-strange members of a nonet. In the limit of exact SU(3) symmetry, G is conserved. Mesons with $G = +1$ decay into an even number of pions while mesons with $G = -1$ decay into an odd number of pions. From this, mesons have the following well defined quantum numbers: total angular momentum J, isospin I, parity P, C-parity C, and G-parity G. These are represented as $(I^G)J^{PC}$, or simply J^{PC} for short. For the case of $L = 0$ and $S = 0$, we have $J^{PC} = 0^{-+}$, while for $L = 0$ and $S = 1$, $J^{PC} = 1^{--}$.

8.1.3 Light-quark meson names

Prior to 1986, there was no systematic naming scheme for mesons. Those who discovered new states often proposed what those states would be called. In 1986, the Particle Data Group [2299] proposed a naming scheme for mesons that is still in use today. This scheme is based on the total spin J, parity P and charge conjugation C of the nonet, and then the isospin of the nonet members. The base name is the same for all mesons of a given I and PC, where there is a subscript denoting the total spin J. For the kaons, ($I = \frac{1}{2}$), those with $J^P = 0^-, 1^+, 2^-, \ldots$ are named K_J, while those with $J^P = 0^+, 1^-, 2^+, \ldots$ are named K_J^*. Table 10 lists the names of the light-quark mesons up to $L = 3$.

8.1.4 SU(3) flavor and light-quark mesons

Given three flavors of light quarks, there are nine possible $q\bar{q}$ combinations. SU(3) flavor groups these mesons into eight members of an SU(3) octet and one SU(3) singlet. Figure 171 shows these $q\bar{q}$ combinations plotted on a graph where the strangeness S is plotted against the third component of isospin, I_3. There are four mesons with $S = 0$, three with isospin 1 and one with isospin 0. The SU(3) singlet state also has $I = 0$. The SU(3) singlet state with $I = 0$ is

$$|\,1\rangle = \frac{1}{\sqrt{3}}\left(u\bar{u} + d\bar{d} + s\bar{s}\right), \tag{8.9}$$

Table 10 The naming scheme for light-quark mesons [616]

L	S	J^{PC}	$I=1$	G	$I=0$		G	K
0	0	0^{-+}	π	-1	η	η'	$+1$	K
0	1	1^{--}	ρ	$+1$	ω	ϕ	-1	K^*
1	0	1^{+-}	b_1	$+1$	h_1	h_1'	-1	K_1
1	1	0^{++}	a_0	-1	f_0	f_0'	$+1$	K_0^*
1	1	1^{++}	a_1	-1	f_1	f_1'	$+1$	K_1
1	1	2^{++}	a_2	-1	f_2	f_2'	$+1$	K_2^*
2	0	2^{-+}	π_2	-1	η_2	η_2'	$+1$	K_2
2	1	1^{--}	ρ_1	$+1$	ω_1	ϕ_1	-1	K_1^*
2	1	2^{--}	ρ_2	$+1$	ω_2	ϕ_2	-1	K_2
2	1	3^{--}	ρ_3	$+1$	ω_3	ϕ_3	-1	K_3^*
3	0	3^{+-}	b_3	$+1$	h_3	h_3'	-1	K_3
3	1	2^{++}	a_2	-1	f_2	f_2'	$+1$	K_2^*
3	1	3^{++}	a_3	-1	f_3	f_3'	$+1$	K_3
3	1	4^{++}	a_4	-1	f_4	f_4'	$+1$	K_4^*
4	0	4^{-+}	π_4	-1	η_4	η_4'	$+1$	K_4
4	1	3^{--}	ρ_3	$+1$	ω_3	ϕ_3	-1	K_3^*
4	1	4^{--}	ρ_4	$+1$	ω_4	ϕ_4	-1	K_4
4	1	5^{--}	ρ_5	$+1$	ω_5	ϕ_5	-1	K_5^*

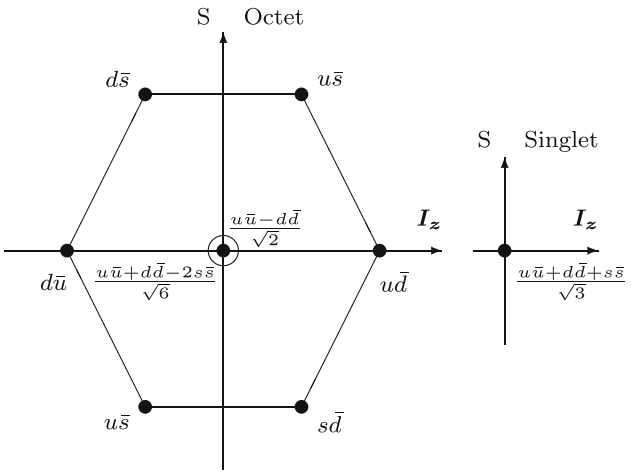


Fig. 171 The SU(3) quark structure of the light-quark mesons. The mesons are plotted against strangeness S on the vertical axis and the z component of isospin, I_z on the horizontal axis The left-hand plot shows the octet mesons, while the right-hand plot shows the singlet meson

while the SU(3) octet state with $I = 0$ is

$$|\,8\rangle = \frac{1}{\sqrt{6}}\left(u\bar{u} + d\bar{d} - 2\,s\bar{s}\right). \tag{8.10}$$

The two physical $I = 0$ states are mixtures of the two SU(3) states. Following the Particle Data Group [616] notation with nonet mixing angle θ_n, the physical isospin-zero states are

$$\begin{pmatrix} f \\ f' \end{pmatrix} = \begin{pmatrix} \cos\theta_n & \sin\theta_n \\ -\sin\theta_n & \cos\theta_n \end{pmatrix} \begin{pmatrix} |\,1\rangle \\ |\,8\rangle \end{pmatrix}. \tag{8.11}$$

Table 11 The nonet mixing angles as reported in reference [616]. The linear formula is given by Eq. (8.12) while the quadratic angle is given by Eq. (8.13)

J^{PC}	θ_{lin}	θ_{quad}
0^{-+}	$-24.5°$	$-11.3°$
1^{--}	$36.5°$	$39.2°$
2^{++}	$28.0°$	$29.6°$
3^{--}	$30.8°$	$31.8°$

Many of the known nonets have physical states that separate the light-quark states, $u\bar{u} + d\bar{d}$, and the states with hidden strangeness, $s\bar{s}$. This is known as ideal mixing and corresponds to $\tan\theta_n = \frac{1}{\sqrt{2}}$, or $\theta_n \approx 35.26°$. Contrary to this, the ground state mesons are almost pure SU(3) states. The η' is nearly pure singlet and the η is the octet state. The nonet mixing angles can be determined from masses of the member states. In the following, m_a is the mass of the isospin 1 state, m_K is the mass of the isospin-$\frac{1}{2}$, and m_f and $m_{f'}$ are the masses of the two isospin-0 states. θ in Eq. (8.12) is known as the linear mixing angle,

$$\tan\theta = \frac{4m_K - m_a - 3m_{f'}}{2\sqrt{2}\,(m_a - m_K)} \tag{8.12}$$

while Eq. (8.13) is known to define the quadratic mixing angle.

$$\tan^2\theta = \frac{4m_K - m_a - 3m_{f'}}{-4m_K + m_a + 3m_f} \tag{8.13}$$

The Particle Data Group quotes mixing angles for four nonets which are listed in Table 11. With exception of the pseudoscalar $J^{PC} = 0^{-+}$ nonet, the others nonets are all fairly close to being ideally mixed.

The most comprehensive predictions for nonet mixing angles comes from lattice QCD [521]. Those predictions are in good agreement with the known values. The mixing angles can also be determined using relative decay rates for the physical isospin 0 states to pairs of mesons in the same nonets, to two pseudoscalar mesons, or to a pseudoscalar and a vector $J^{PC} = 1^{--}$ meson. Determinations exploiting decay rates exist for several nonets [2300]. In Table 12 are listed the lattice QCD predictions as well as several determinations of mixing angles from decay measurements. The one discrepancy between lattice and decay rate predictions are in the 2^{-+} nonet; this may be due to incorrect assignments and is discussed later.

8.1.5 Light-quark mesons

The pseudoscalar mesons

The $J^{PC} = 0^{-+}$ mesons are spin singlets with 0 orbital angular momentum and are known as the pseudoscalar mesons. They are listed in Table 13. These are the lightest mesons, and with the exception of the η', all their decays are either weak or electromagnetic. In addition, the mixing of this nonet is quite different from other nonets in that the mixing angle is small, and the η and η' are very close to being SU(3) octet and singlet states respectively.

Table 12 The nonet mixing angles for mesons with orbital angular momentum less than 4. The lattice results are reported in reference [521], while the references for the decay rate determinations are given in the table

J^{PC}	θ_n lattice	θ_n decays
0^{-+}	$-11°$	$-9.3°$ [2301]
1^{--}	$33°$	
1^{+-}	$35°$	
1^{++}	$8°$	
2^{++}	$28°$	$32.1°$ [2302]
2^{-+}	$33°$	$-6.7°$ [2303]
1^{--}	$30°$	
2^{--}	$33°$	
3^{--}	$33°$	$31.8°$ [2304]
3^{-+}	$34°$	
2^{++}	$26°$	
3^{++}	$33°$	
4^{++}	$29°$	

Table 13 The pseudoscalar mesons

Isospin	State(s)	Mass [MeV]	Width or lifetime
1	π^0	134.9768	8.52×10^{-17}s
1	$\pi^{\pm}$	139.57039	2.6033×10^{-8}s
$\frac{1}{2}$	$K^{\pm}$	493.677	1.238×10^{-8}s
$\frac{1}{2}$	$K^0/\bar{K}^0$	497.611	
0	η	547.862	0.00131 MeV
0	η'	957.78	8.49 MeV

In addition to the ground state pseudoscalar mesons, there can also be radially excited states. Both excited πs, the $\pi(1300)$ and $\pi(1800)$, and ηs, the $\eta(1295)$, the $\eta(1405)$, the $\eta(1475)$, $\eta(1760)$ and the $\eta(2225)$ have been observed. The $K(1460)$ and $K(1830)$ are the observed $J^P = 0^-$ states. The lighter is consistent with the first radial excitation. The Particle Data Group [616] identifies the states listed in Table 14 as the nonet of radially excited pseudoscalar mesons. One could also associate the $\pi(1800)$, $\eta(1760)$ and $K(1830)$ together as a third nonet as listed in Table 15. However in addition to the second radial excitation, there is a predicted pseudoscalar glueball (see Sect. 8.4) as well as a nonet of $J^{PC} = 0^{-+}$ hybrid mesons (see Sect. 8.3).

With regard to the $\eta(1295)$ state, we believe that its status deserves some scrutiny and that the $\eta(1405)$ and $\eta(1475)$ should be the two $I = 0$ members of the radially excited pseudoscalar mesons. For the $\eta(1295)$, there is a single report in radiative J/ψ decays [2305,2306], but there is ambigu-

Table 14 The radially excited pseudoscalar mesons according to reference [616]

Isospin	State(s)	Mass [MeV]	Width [MeV]
1	$\pi(1300)$	1300	200 to 600
0	$\eta(1295)$	1294	55
0	$\eta(1475)$	1475	90
$\frac{1}{2}$	$K(1460)$	1482	335

Table 15 A possible third nonet of pseudoscalar mesons

Isospin	State(s)	Mass [MeV]	Width [MeV]
1	$\pi(1800)$	1810	215
0	$\eta(1760)$	1751	240
0			
$\frac{1}{2}$	$K(1830)$	1874	168

Table 16 The vector mesons

Isospin	State(s)	Mass [MeV]	Width [MeV]
1	$\rho(770)$	775.26	149.1
0	$\omega(782)$	782.65	8.49
0	$\phi(1020)$	1019.461	4.249
$\frac{1}{2}$	$K^{*\pm}(892)$	891.66	50.8
$\frac{1}{2}$	$K^{*0}(892)$	895.5	47.3

ity about whether the signal is $\eta(1295)$ or $f_1(1285)$. It has not been reported in other J/ψ measurements since then, while there has been extensive results of the η', $\eta(1405)$ and $\eta(1475)$. The majority of the observations have been in pion production [2307–2311] where there are generally contributions from both the $\eta(1295)$ and the $f_1(1285)$. In $p\bar{p}$ annihilation, both the $\eta(1405)$ [2312] and the $\eta(1475)$ [2313] have been observed, but no observation of the $\eta(1295)$ has been reported.

The vector mesons

A spin triplet system with $L = 0$ forms the $J^{PC} = 1^{--}$ nonet, its members are known as vector mesons. These mesons are shown in Table 16. The dominant decay modes of the vector mesons are through the strong interaction to two or three pseudoscalar mesons and the states are nearly ideally mixed with the ω nearly all $u\bar{u}$ and $d\bar{d}$, while the ϕ is nearly all $s\bar{s}$.

In addition to the expected radial excitations of the vector mesons, the $L = 2$, $S = 1$ $q\bar{q}$ system can also have $J^{PC} = 1^{--}$. Finally, there is a nonet of hybrid mesons expected with the same J^{PC}. Thus, we expect disentangling of the excited vector meson spectrum to be tricky. The reported states in $I = 1$ are $\rho(1450)$, $\rho(1570)$, $\rho(1700)$, $\rho(1900)$ and $\rho(2150)$. In the $I = 0$ system, $\omega(1420)$, $\omega(1680)$, $\phi(1680)$ and $\phi(2170)$ have been reported. Finally, for $I = \frac{1}{2}$, the $K^*(1410)$ and $K^*(1680)$ are known. The Particle Data Group identifies the radially excited states as in Table 17. The states identified with the D-wave nonet are listed in Table 24 and discussed later. Finally, the $\rho(1900)$ and $\phi(2170)$ could be part of another nonet; either the hybrid nonet or the second radial excitation of the ground-state vector mesons (Table 18).

Table 17 The radially excited vector mesons

Isospin	State(s)	Mass [MeV]	Width [MeV]
1	$\rho(1450)$	1465	400
0	$\omega(1420)$	1410	290
0	$\phi(1680)$	1680	150
$\frac{1}{2}$	$K^*(1410)$	1414	232

Table 18 A possible fourth nonet of vector mesons

Isospin	State(s)	Mass [MeV]	Width [MeV]
1	$\rho(1900)$		
0			
0	$\phi(2170)$	2162	100
$\frac{1}{2}$			

Table 19 The pseudo vector mesons

Isospin	State(s)	Mass [MeV]	Width [MeV]
1	$b_1(1235)$	1229.5	142
0	$h_1(1170)$	1166	375
0	$h_1(1415)$	1416	90
$\frac{1}{2}$	K_{1A}		

The pseudo vector mesons

Spin singlet states with $L = 1$ form the $J^{PC} = 1^{+-}$ nonet, and are known as the pseudo vector mesons. These mesons are listed in Table 19. There is one known state beyond those listed in the table, the $h_1(1595)$ which has been reported in pion production [2314]. There is also an interesting complication with the kaonic states where C-parity is not defined. The states with open strangeness have $J^P = 1^+$ which is the same as those in the $J^{PC} = 1^{++}$ axial vector mesons. Because of this, the two states can mix, and it is believed that the physical states, $K_1(1270)$ and $K_1(1400)$, are mixtures of the SU(3) states, K_{1A} and K_{1B}, with a mixing angle $\theta_{K_1} = -(33.6 \pm 4.3)^\circ$ [2315] conventionally defined by

$$\begin{pmatrix} |\, K_1(1270)\rangle \\ |\, K_1(1400)\rangle \end{pmatrix} = \begin{pmatrix} \sin\theta_{K_1} & \cos\theta_{K_1} \\ \cos\theta_{K_1} & -\sin\theta_{K_1} \end{pmatrix} \begin{pmatrix} |\, K_{1A}\rangle \\ |\, K_{1B}\rangle \end{pmatrix}.$$

The scalar mesons

Table 20 The axial vector mesons

Isospin	State(s)	Mass [MeV]	Width [MeV]
1	$a_1(1260)$	1230	250 to 600
0	$f_1(1285)$	1281.9	22.7
0	$f_1(1420)$	1426.3	54.5
$\frac{1}{2}$	K_{1B}		

Table 21 The tensor mesons

Isospin	State(s)	Mass [MeV]	Width [MeV]
1	$a_2(1320)$	1316.9	107 to 600
0	$f_2(1270)$	1275.5	186.7
0	$f_2'(1525)$	1517.4	86
$\frac{1}{2}$	$K_2^*(1430)$	1427	100

Table 22 The radial excitations of the tensor mesons

Isospin	State(s)	Mass [MeV]	Width [MeV]
1	$a_2(1700)$	1698	265 to 600
0	$f_2(1640)$	1639	99
0	$f_2(1950)$	1936	464
$\frac{1}{2}$	$K_2^*(1980)$	1995	349

A spin triplet with $L = 1$ can form three possible J^{PC}s: 0^{++}, 1^{++} and 2^{++}. The 0^{++} states are known as scalar mesons and are discussed in Sect. 8.2 because there are added complications which make it difficult to discuss them with the other mesons. There is also significant discussion of the scalar states and their relation to the scalar glueball, see Sect. 8.4).

The axial vector mesons

The $L = 1$ $J^{PC} = 1^{++}$ mesons are known as the axial vector mesons and are listed in Table 20. As noted earlier in the discussion of the pseudo vector mesons, the SU(3) K_{1B} state is a mixture of the physical $K_1(1270)$ and $K_1(1400)$ states. In addition to the states listed, two additional states have been reported. The $f_1(1510)$ has been seen in kaon production [2316,2317] as well as pion production [2318] decaying to K^*K. These productions and decay would favor an $s\bar{s}$ interpretation of the $f_1(1510)$, but it is probably too light to be the radial excitation. A second state, the $a_1(1640)$ is identified as the radial excitation of the $a_1(1270)$. This has been observed in pion production with the most significant observation in reference [2319]. It has also been reported in D decays [2320].

The tensor mesons

The last $L = 1$ nonet contains the $J^{PC} = 2^{++}$ tensor mesons, where the states are listed in Table 21. This well-established nonet is close to ideally mixed as noted in Table 11. As with the vector mesons, there is a second L, S combination that can exist for $J^{PC} = 2^{++}$, $L = 3$ and $S = 1$. In addition, one of the lightest glueballs is also expected to have these quantum numbers, and of course radial excitations should be present.

With regard to excited states, there is a second a_2 state, the $a_2(1700)$, which the Particle Data Group associates with the radial excitation of the tensor mesons. This assignment is based on mass, where we would expect the radial a_2 state to be close in mass to the $a_1(1640)$, the radial excitation of the a_1. The $L = 3$ state is expected to have similar mass to other $L = 3$ states, where here the $a_4(1970)$ anchors these nonet around 2 GeV. The $a_2(1700)$ has been observed in many production mechanisms including pion production [2319,2321–2324], $p\bar{p}$ annihilation [2325–2328], two-photon production [2329, 2330] and ψ' radiative decays [2331].

For the isospin 0 states, there is an overpopulation of f_2 states, with 10 additional states beyond the two ground state tensors reported. These include the $f_2(1430)$, $f_2(1565)$, $f_2(1640)$, $f_2(1810)$, $f_2(1910)$, $f_2(1950)$, $f_2(2010)$, $f_2(2150)$, $f_2(2300)$ and $f_2(2340)$. For $I = \frac{1}{2}$, there is a single state, the $K_2^*(1980)$. The Particle Data Group identifies the radially excited states as listed in Table 22. With the radial states accounting for 2 of the 10 extra states, a second pair in the $L = 4$ mesons, probably above 2 GeV in mass, and a glueball state, there are still 5 states. Presumably several of the reported states are all the same state, with low statistics and differences in production mechanisms accounting for the differences. Three of the isoscalar tensor states were observed in the OZI rule suppressed reaction $\pi^- p \to \phi\phi n$ [2332] and were discussed as one or three glueballs. This interpretation is supported by a recent analysis of BESIII data on radiative J/ψ decays (see Sect. 8.4). In any case, a careful examination of the $I = 0$ $J^{PC} = 2^{++}$ data with high statistics experiments is merited.

The pseudo tensor mesons

Mesons formed with $S = 0$ and $L = 2$ have $J^{PC} = 2^{-+}$ and are known as the pseudo tensor mesons. The known states are listed in Table 23. In addition to the radial excitations of these states, there is also a nonet of hybrid mesons expected. The latter are likely slightly heavier than the mesons in Table 23. There are three known states beyond those in the table, the $\pi_2(1880)$, $\pi_2(2005)$ and the $\pi_2(2100)$. It is also interesting that the decay patterns of the $\eta_2(1645)$ and the $\eta_2(1870)$ both look like those for a $u\bar{u}/d\bar{d}$ state and not an $s\bar{s}$ state [2333]. This suggests that the $\eta_2(1870)$ might be paired with the $\pi_2(1880)$ in a third nonet. However, studies of the axial anomaly [2300] favor the assignment in Table 23, but with an unusual mixing angle that is inconsistent with lattice, as shown in Table 12.

The D-state vector mesons

Table 23 The pseudo tensor mesons

Isospin	State(s)	Mass [MeV]	Width [MeV]
1	$\pi_2(1670)$	1670.6	258
0	$\eta_2(1645)$	1617	181
0	$\eta_2(1870)$	1842	225
$\frac{1}{2}$	$K_2(1770)$	1773	186

Table 24 The $L = 2$ 1^{--} vector mesons

Isospin	State(s)	Mass [MeV]	Width [MeV]
1	$\rho(1700)$	1720	250
0	$\omega(1650)$	1670	315
0			
$\frac{1}{2}$	$K^*(1680)$	1718	322

Table 25 The $L = 2$ 3^{--} vector mesons

Isospin	State(s)	Mass [MeV]	Width [MeV]
1	$\rho_3(1690)$	1688.8	161
0	$\omega_3(1670)$	1677	168
0	$\phi_3(1850)$	1854	87
$\frac{1}{2}$	$K_3^*(1780)$	1776	159

The mesons formed from an $S = 1$, $L = 2$ $q\bar{q}$ system can have $J^{PC} = 1^{--}, 2^{--}$ and 3^{--} and are referred to as vector mesons. The Particle Data Group identifies the states listed in Table 24 with the 1^{--} states, where there is no candidate for the ϕ state which is probably expected with a mass in the 1.8 to 1.9 GeV mass region. For the $J^{PC} = 2^{--}$ states, very little is known with the only assignment made by the Particle Data Group being the $K_2(1820)$. However, similar to the K_{1A} and K_{1B} of the 1^{+-} and 1^{++} nonets, the kaonic states from the 2^{-+} and 2^{--} nonets can also mix.

The $J^{PC} = 3^{--}$ nonet is one of the well established nonets where a mixing angle is also reported. These states are listed in Table 25. In addition to the listed states, there are two additional ρ_3 states reported in literature. The $\rho_3(1990)$ is reported in $p\bar{p}$ annihilation [2334,2335], and the $\rho_3(2250)$ reported in both $\bar{p}p$ annihilation and in ψ' decays [2334, 2336]. The lighter state could be a radial excitation of the $L = 2$ $\rho_3(1690)$. The higher mass state is of similar mass to the $\rho_5(2350)$ and could be an $L = 4$ meson.

Higher excitations

Going beyond the $L = 2$ mesons, less is known, with the most information tending to be on the nonets with the largest J. For the case of $L = 3$, there are candidates for the $J^{PC} = 4^{++}$ mesons as shown in Table 26. There should also be a 2^{++} and 3^{++} nonet as well as a $J^{PC} = 3^{+-}$ nonet. While as noted in the tensor meson section, there are a large number of reported f_2 states, in particular the $f_2(2010)$, $f_2(2150)$, $f_2(2300)$ and $f_2(2340)$, assigning any of these to an $L = 3$ nonet is not clear. There is also a $J^P = 3^+$ kaonic state, the $K_3(2320)$ which could be a member of either of the spin 3 nonets.

Table 26 The $L = 3$ 4^{++} mesons

Isospin	State(s)	Mass [MeV]	Width [MeV]
1	$a_4(1970)$	1967	324
0	$f_4(2050)$	2018	237
0	$f_4(2300)$	2320	260
$\frac{1}{2}$	$K_4^*(2045)$	2048	199

Table 27 The $L = 4$ 5^{--} vector mesons

Isospin	State(s)	Mass [MeV]	Width [MeV]
1	$\rho_5(2350)$	2350	400
0			
0			
$\frac{1}{2}$	$K_5^*(2380)$	2382	178

For the $L = 4$ mesons, the highest spin is $J^{PC} = 5^{--}$, and a few states with these quantum numbers are known, as listed in Table 27. There should also be a 3^{--}, 4^{--} and 4^{-+} nonet for which a few states are reported. For $I = \frac{1}{2}$ the $K_4(2500)$ which could be a member of either of the $J = 4$ nonets. There are also two ρ_3 states reported, the $\rho_3(1990)$ and the $\rho_3(2250)$. The latter state is of similar mass to the $\rho_5(2350)$ and could be an $L = 4$ meson. The lighter state could be a radial excitation of the $L = 2$ $\rho_3(1690)$.

8.1.6 The leading Regge trajectories

The original meson Regge trajectories[78] described a linear relation between the mass squared and the orbital angular momentum of mesons [1068,2338], where the trajectories include $J^{PC} = 0^{-+}, 1^{+-}, 2^{-+}, 3^{+-}, \ldots$ and $1^{--}, 2^{++}, 3^{--}, 4^{++}, 5^{--}, \ldots$. In reality, the trajectories are often more complicated than the simple linear form. In a simplified picture when the quarks can be regarded as ultrarelativistic, a linear confining potential leads to linear Regge trajectories, while in the nonrelativistic regime, the trajectories would be nonlinear, and the intermediate regime would lead to a transition in the slope of the Regge trajectories. In the ultrarelativistic regime, the Regge slope depends on the string tension, while more generally it depends on both the quark masses and the tension. See reference [2337] for a more detailed discussion on this. In addition to the trajectories in orbital angular momentum l, there are also trajectories in the

[78] See paragraph *The Regge approach and QCD* in Sect. 12.6 for an introduction to Regge phenomenology

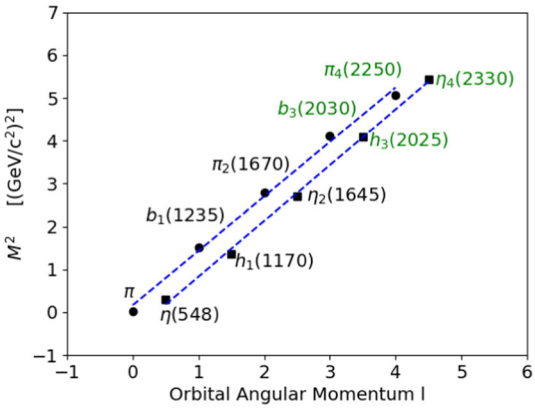


Fig. 172 The pseudoscalar meson Regge trajectories as a function of the orbital angular momentum l. The isospin 0 trajectory has been shifted by $\frac{1}{2}$ unit in l to the right. The established states, π, $b_1(1235)$, $\pi_2(1670)$, η, $h_1(1170)$ and $\eta_2(1645)$ are shown in blue. The states shown in green, $b_3(2030)$, $\pi_4(2250)$, $h_3(2025)$ and $\eta_4(2330)$, need confirmation. The fitted slopes are consistent with $1.20\,\text{GeV}^2$ for $I = 1$ and $1.37\,\text{GeV}^2$ for $I = 0$ as in reference [2337]

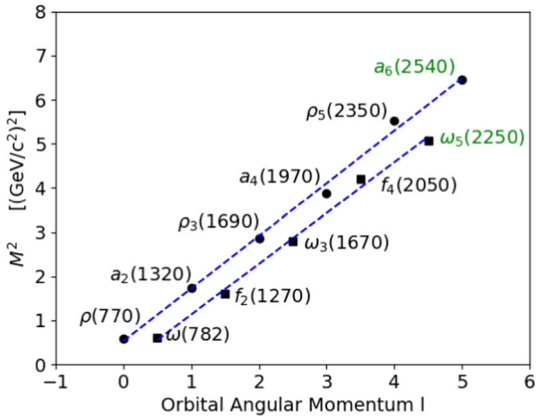


Fig. 173 The vector meson Regge trajectories as a function of the orbital angular momentum l. The isospin 0 trajectory has been shifted by $\frac{1}{2}$ unit in l to the right. The established states, $\rho(770)$, $a_2(1320)$, $\rho_3(1690)$, $a_4(1970)$, $\rho_5(2350)$, $\omega(781)$, $f_2(1270)$, $\omega_3(1670)$ and $f_4(2050)$ are shown in blue. The states shown in green, $a_6(2540)$ and $\omega_5(2250)$ need confirmation. The fitted slopes are consistent with $1.10\,\text{GeV}^2$ for $I = 1$ and $1.09\,\text{GeV}^2$ for $I = 0$ as in reference [2337]

radial excitation quantum number, n. From a simple linear confining potential with string tension σ, the orbital trajectory is given in Eq. (8.14) and the radial as in Eq. (8.15), where one would expect universal slopes in both cases, with the slopes related by a factor of $\frac{\pi}{2}$.

$$M^2 = 8\sigma l + c_1, \tag{8.14}$$

$$M^2 = 4\pi\sigma n + c_2. \tag{8.15}$$

For light-quark mesons, the slopes are similar, but not universal. The orbital trajectories starting with the pseudoscalar mesons are shown in Fig. 172. The slopes for the two are found to be $1.20\,\text{GeV}^2$ and $1.37\,\text{GeV}^2$ respectively. The orbital trajectories starting with the vector mesons are shown in Fig. 173. In these cases, the slopes are found to be $1.10\,\text{GeV}^2$ and $1.09\,\text{GeV}^2$ respectively.

8.2 The light scalars

José R. Peláez

8.2.1 Introduction

Light scalar mesons are treated in a separate subsection because, on the one hand, both their existence and nature have been the subject of a six-decade-long debate that predates QCD. On the other hand, they are particularly interesting because they play a very relevant role in several aspects, gathered below in seven items for concreteness, some of them already present before QCD, some others after.

First of all, in 1955, well before QCD was formulated, Johnson and Teller [2339] proposed the existence of a light scalar-isoscalar field to explain the attractive part of the nucleon–nucleon interaction. Two years later, Schwinger suggested that such a field, which he named σ, could be an isospin singlet, difficult to observe due to its huge width caused by its most likely very strong coupling to two pions.

Second, in the early sixties, Gell-Mann and Levy [2340] considered this field as the fourth member of a multiplet together with the three pions to build the popular "Linear sigma model" (LσM). Such a state could also be generated dynamically in the Nambu and Jona-Lasinio (NJL) models [59,2255]. These relatively simple models were able to explain the light masses of the pions, kaons and eta, and their mass gap with respect to the other hadrons, since they are the Nambu–Goldstone Bosons (NGB) of a spontaneous chiral symmetry breaking observed in the spectrum. Actually, they are pseudo-NGB, because they are not strictly massless. The masses of light-scalar mesons are closely related to the size of the non-zero vacuum expectation value, particularly those that share its same quantum numbers. Details of their interactions are also related to the specifics of the spontaneous breaking mechanism. The consequences of chiral symmetry were initially worked out with current–algebra methods as described in Sect. 1. Of particular interest for us will be the derivative interactions of NGB among themselves and the requirement of an Adler-zero below threshold in the NGB scattering amplitudes [20]. The leading order at low energies of those amplitudes was obtained by Weinberg in [22].

Third, since light scalars are the lightest states in the QCD spectrum that are not pseudo-NGB, we may expect them to fit as ordinary quark–antiquark mesons within the Quark Model that was proposed in the mid 1960s [17,18]. However, they do not, as we will see repeatedly below. Moreover, within the Quark Model, another scalar strange state, relatively sim-

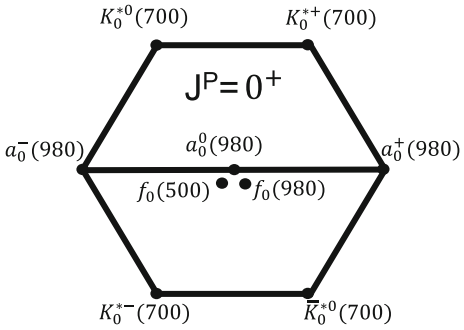


Fig. 174 Light scalar nonet. Note the inverted hierarchy with respect to the naive $q\bar{q}$ assignment in Fig. 171, according to which the $a_0(980)$ should be ~ 200 MeV lighter than the $K_0^*(700)$

ilar to the σ and called κ, was proposed by Dalitz in 1965 [2341], with a quark–antiquark assignment in a simple potential model, or more generally "simply on the basis of SU(3) symmetry". The existence of these two states, the σ and κ, nowadays known as $f_0(500)$ and $K_0^*(700)$, has been very controversial until very recently because they are extremely wide and difficult to observe. Actually, since they were first proposed, there were many experimental and phenomenological claims of such states, sometimes narrow, sometimes wide, sometimes lighter than 1 GeV, sometimes heavier, and sometimes absent. The list of references is huge and we refer the reader to the Review of Particle Properties (RPP) [616] and the evolution of its "Note on Light Scalars" over the years, as well as the historical accounts in relatively recent reviews [2342,2343]. An additional pair of scalar mesons, sitting very close to the $K\bar{K}$ threshold at 980 MeV, were soon identified, presently known as $f_0(980)$ and $a_0(980)$. These are narrower and their existence has not been controversial, although their mass and width values have changed slightly over the years. All in all they form the lightest scalar SU(3) nonet in Fig. 174. Note the largely broken flavor symmetry since the difference in the nominal masses is as large as 480 MeV. In addition, the mass hierarchy is inverted with respect to the naive expectations for an ordinary nonet of quark–antiquark states as in Fig. 171. For example, since in such a scheme the $a_0(980)$ would contain no strange valence quarks or antiquarks it should be about 200 MeV lighter than the $K_0^*(700)$, with one valence strange quark or antiquark. But this is precisely the opposite of what is found for the lightest scalars.

Fourth, light scalars, and particularly the σ and κ, are difficult to include in the linear Regge trajectories that the other ordinary mesons follow [2344–2346] . These linear trajectories are related to the confinement mechanism. This difficulty became clear only around the time when the existence of the lightest scalars was being settled, and although QCD had already been formulated, it played no direct role in this discussion.

With the advent of QCD new interesting perspectives arose. In particular:

Fifth, one of the most attractive possibilities of a non-abelian gauge theory like QCD is the existence of glueballs, discussed in Sect. 8.4. The lightest one is expected to have scalar-isoscalar quantum numbers and to appear as an "extra state" beyond the quark SU(3) multiplets. It is therefore important to identify all states within some light-scalar meson SU(3) nonets. For this, strange states are important, since they do not mix with glueballs and count how many quark nonets exist.

Sixth, given the quark constituent masses, tetraquarks would be naively expected to appear naturally around 1.4 GeV, if they appear at all. However, based on the dominance of the magnetic contribution of gluon interactions, Jaffe [2347] was able to build, within the "bag" model, tetraquarks well below 1 GeV. This suggests the existence of two 0^+ nonets, one made of such tetraquarks, below 1 GeV, that on a first approximation could be identified with the nonet in Fig. 174 and another one made of ordinary $q\bar{q}$ above 1 GeV. This is how light scalars became the first non-ordinary-meson candidates, in the form of tetraquarks, or meson molecules. Still, they are not usually considered "exotics", but "cryptoexotics", since their quantum numbers can also be built with ordinary quark–antiquark configurations, with which they will necessarily mix, thus complicating this simple picture.

Seventh and final, despite QCD being non-perturbative at low energies, its symmetries, and particularly the spontaneous symmetry breaking of chiral symmetry leading to a mass gap between NGB and other hadrons, allow for a systematic low-energy (and low-mass) expansion of amplitudes involving pions, kaons, and the eta. The mathematical formulation in terms of an Effective Theory [1426] in the meson sector, has been presented in Sect. 6.2 and is called Chiral Perturbation Theory (ChPT) [69,1610]. Being the next less massive states after the NGB, one would naively expect the lightest scalars to saturate the ChPT parameters at NLO discussed in Sect. 6.2. Once again, they do not, but the vector mesons do instead. This suggests once more that the dynamics that govern the formation of light scalars might be different from that of ordinary mesons like vectors.

With all those pieces of motivation in mind, the rest of the section is divided into two parts. First, we will describe the light scalars present status, paying attention to the dispersive and analytic methods used to settle the controversy about their existence, and other dispersive applications that are of relevance for the next part. Since the purpose of this work is to celebrate the 50th anniversary of QCD, we apologize for discussing in the second part only the most direct connections with it. Namely, their description in terms of (unitarized) ChPT and their dependence on the number of colors and quark masses. We will then discuss what can we conclude from these results.

Table 28 Scalar light mesons below 1.9 GeV as listed in the RPP[300]. Note that for the first nonet we have taken the "T-matrix pole" parameters, not available for the rest. Also, there seems to be one f_0 state too many to form a second nonet

Isospin	State(s)	Mass [MeV]	Width [MeV]
0	$f_0(500)$	400–550	400–700
1/2	$K_0^*(700)$	630–730	520–680
0	$f_0(980)$	980–1010	40–70
1	$a_0(980)$	960–1030	40–140
0	$f_0(1370)$	1200–1500	200–500
1/2	$K_0^*(1430)$	1425 ± 50	270 ± 80
1	$a_0(1450)$	1474 ± 19	265 ± 13
0	$f_0(1500)$	1506 ± 6	112 ± 9
0	$f_0(1710)$	1704 ± 12	123 ± 18

Present status

As it is customary, and given the present precision, we consider the isospin limit. At present, 19 well-established scalar mesons are identified in the RPP [300] below 1.9 GeV, which we list in Table 28 with their present names. We have already classified nine of them in the lightest scalar nonet in Fig. 174. The other ten are three isoscalars, $f_0(1370)$, $f_0(1500)$, $f_0(1710)$, the a_0 (1450) isovectors with their three different charges, and the $K_0^*(1430)$ in four different combinations of strangeness and charge. There are more scalar mesons, but they all lie nominally at or above 1.95 GeV. Hence, given the number of strange scalars and isovectors, there must be two nonets below 1.75 GeV. Looking at their masses, one lies below 1 GeV and the other one around 1.4 GeV. Note, however, that above 1 GeV there seems to be one scalar state too many. This agrees with the expectation for the lightest QCD glueball. On the other hand, the newly proposed $a_0(1700)$ [2348–2351] points to an interpretation of the $f_0(1710)$ as isoscalar partner of the $a_0(1700)$.

Compared with other mesons made of light quarks discussed in Sect. 8.1, we see that, for similar masses, they tend to have larger widths. The exceptions are the $f_0(980)$ and $a_0(980)$, which are narrower than the rest because their decay into $K\bar{K}$ is suppressed due to their proximity to the $K\bar{K}$ threshold. Given the $O(100)$ MeV width of most of these resonances, there must be some mixing between states with the same quantum numbers in different nonets. This mixing most likely distorts the mass hierarchies expected if they were narrow. Many mixing schemes have been proposed, but they only make sense for the flavor part of the wave function. We will see below one such treatment. Unfortunately, they are often used for the spatial or momentum part, which would only make sense in the narrow width approximation for almost stable mesons, which is not the case of any pair of light scalars with the same quantum numbers, and should be avoided (see Section 4.6.2 in [2342] and references therein).

In general, light-scalar-meson parameters have much bigger uncertainties than those of other mesons. This is because their large widths make them often overlap with one another as well as with other analytic features like thresholds. As a consequence, in many analyses they do not show up as clean resonance peaks and their observed shapes can vary strongly, depending on specific features of their production, becoming dips or being even completely masked. It is therefore essential to determine light-scalar-meson parameters from process-independent quantities. In particular, resonances are rigorously defined through their associated poles in the complex plane, that we briefly describe next.

8.2.2 Resonance poles and dispersive determinations

Resonance poles

These are poles appearing in the complex s-plane of any T-matrix element describing a process where a resonance R is produced as an intermediate state. As a technical remark, these poles appear in conjugated pairs in the Riemann sheet that is reached when crossing continuously from above the square-root cuts associated with the center-of-mass (CM) momenta of the particles in the physically available intermediate states. This sheet is sometimes called "adjacent", "proximal" and in the elastic case "second" sheet. Out of the conjugate pair, it is the pole in the lower-half plane that most influences the behavior of the amplitude on the real axis. Then, its position s_R is related to the resonance mass and width as $\sqrt{s_R} \equiv M - i\Gamma/2$. The familiar peak shape in the modulus squared of the amplitude is clearly observed for real-physical values of s only when the resonance is narrow and well isolated from other singularities. Only in such cases the simple Breit–Wigner (BW) approximation, or models like K-matrices or isobar sums, etc. may be justified. However, this is not the case for most scalars and definitely not for the $f_0(500)$ and $K_0^*(700)$, which have been the most controversial and latest states to be accepted as well-established in the RPP. This is the reason why in Table 28 we provide the "T-matrix pole" mass and widths, avoiding "Breit–Wigner" parameters.

Let us then briefly comment on how the poles of those states are determined by means of model-independent dispersive and analytic techniques, although we first need to define partial waves.

Partial waves

Resonances and their quantum numbers are most easily identified using partial waves of definite isospin and angular momentum ℓ. For rigorous determinations of the lightest scalar mesons, the most relevant process is meson–meson scattering, whose partial waves are defined as follows:

$$f_\ell^I(s) = \frac{1}{32\pi K}\int_{-1}^{1} dz_s\, P_\ell(z_s) F^I(s, t(z_s)), \tag{8.16}$$

where $F^I(s,t)$ are the amplitudes, or elements of the T-matrix, of definite isospin I; s,t are the usual Mandelstamm variables, P_ℓ the Legendre polynomials and z_s the scattering angle in the s channel. Note that $K = 1,2$ for $K\pi$ and $\pi\pi$, respectively, because, for hadron interactions, pions are identical particles in the isospin limit that we will use.

It is convenient to recast partial waves in terms of the phase shift δ_ℓ^I and elasticity η_ℓ^I as follows:

$$f_\ell^I(s) = \frac{\eta_\ell^I(s)e^{i2\delta_\ell^I(s)} - 1}{2i\sigma(s)}, \quad \sigma(s) = \frac{2q(s)}{\sqrt{s}}, \tag{8.17}$$

where q is the CM momentum of the scattering particles. In the elastic regime $\eta_\ell^I = 1$ and we can write:

$$f_\ell^I(s) = \frac{e^{i\delta_\ell^I(s)}\sin\delta_\ell^I(s)}{\sigma(s)}. \tag{8.18}$$

For later purposes, it is important to recall that we are interested in poles in the second Riemann sheet. Let us illustrate the elastic case, where the analytic continuation to the second sheet through the physical cut is very simple. Moreover, it is the most relevant for the $\sigma/f_0(500)$ and $\kappa/K_0^*(700)$, since they appear in elastic $\pi\pi$ and πK scattering, respectively, well below the next open threshold. For elastic partial waves, the following relation holds $S_\ell = 1 + 2i\sigma(s)f_\ell$. Note that, in the partial-wave context, the T-matrix is actually called f. In addition, above threshold, the unitarity of the S-matrix implies

$$\text{Im}\, f_\ell^I(s) = \sigma(s)|f_\ell^I(s)|^2, \quad \text{Im}\, f_\ell^I(s)^{-1} = -\sigma(s), \tag{8.19}$$

which in turn imposes the following unitarity bounds:

$$|f_\ell^I(s)| \leq 1/\sigma(s). \tag{8.20}$$

Knowing the imaginary part of f_ℓ^I on the cut allows us to write a very simple relation between the S-matrix in the first (I) and second (II) sheet:

$$S_\ell^{(II)} = \frac{1}{S_\ell^{(I)}}, \quad f^{(II)}(s) = \frac{f^{(I)}}{1 + 2i\sigma(s)f^{(I)}}, \tag{8.21}$$

where the isospin and angular-momentum indices have been momentarily suppressed for convenience. Note that in the second sheet $\sigma(s^*) = -\sigma(s)^*$.

Nevertheless, in the σ and κ case, we still need to know the value of f_ℓ^I in the first Riemann sheet, but very deep in the complex plane. Unfortunately, the continuation to the complex plane is a hard and unstable mathematical problem. Different parameterizations or models, seemingly equivalent when describing data in a given region, may lead to different analytic continuations and different poles. The rigorous way of extending the amplitudes to the complex plane is through dispersion relations, if available, and analytic continuation techniques.

Analyticity and dispersion relations

Relativistic causality implies that the amplitude $F(s,t)$, for fixed t, must be analytic in the first Riemann sheet of the complex s-plane except for the real axis. In the absence of bound states in meson–meson scattering, only singularity cuts are present on the real axis. First of all, a right-hand-cut (RHC) appears from threshold to $+\infty$. Crossing this RHC continuously leads to the adjacent Riemann sheet, where resonance poles may exist. In turn, crossing symmetry implies that there is a left-hand-cut (LHC) from $-\infty$ to $s = -t$ due to cuts in the u channels. In particular, the LHC extends up to $s = 0$ for forward scattering ($t = 0$) and for partial-wave amplitudes. Finally, for scattering of two particles with different masses, the $P_\ell(\cos\theta)$ integration in the partial wave definition yields a circular cut of radius $|m_1^2 - m_2^2|$ centered at $s = 0$. Then, Cauchy's integral formula relates the amplitude at any s in the first Riemann sheet with integrals over the amplitude imaginary part along the cuts. These are called dispersion relations.

Since Cauchy's Integral formula applies to functions that depend on one variable, say s, the other variables have to be fixed or integrated over. Of particular interest are forward dispersion relations (FDRs), which correspond to the fixed-t case with $t = 0$. Also of interest for our discussion below are hyperbolic dispersion relations, obtained when s,t,u are fixed to lie on an hyperbola $(s-a)(u-a) = b$. Any of these relations can also be integrated in t as in Eq. (8.16) to obtain a partial-wave dispersion relation. In principle, forward dispersion relations are applicable at any s, but for different fixed-t and hyperbolic cases the applicability is reduced. These applicability domains affect those of the partial waves, depending on how they have been obtained (see the appendix in [2343] for details).

Generically, the most complicated parts of the calculation are the left and circular cuts. Within the context of light scalars, partial-wave dispersion relations are the most relevant and we can crudely group their most frequent uses into two categories: precision dispersive approaches and unitarization techniques.

Before discussing these two uses in detail, let us just mention that dispersive approaches also constrain Regge trajectories and they hence can be used to calculate, not fit, the Regge parameters of resonances using their poles as input. While the resulting trajectories for ordinary mesons like the $\rho(770)$, $K^*(892)$, $f_2'(1525)$, $f_2'(1525)$ come out [2345,2352] with a rather small imaginary part and a dominant real part, whose s dependence is almost a straight line, as expected, those for the $f_0(500)$ and κ come at odds with the ordinary behavior [2345,2346]. This explains why those two resonances do not fit well in the usual phenomenological Regge plots.

Precision dispersive approaches:

We aim at mathematical rigor to establish the existence of the σ and κ poles and at precision to determine their param-

eters. Note that these are the poles closest to the left and circular cuts. Therefore, those cuts must be accurately evaluated using the partial-wave expansion of the crossed channels. This complicates the integrands, and the new relations then couple different partial waves and channels. These relations are generically called Roy-like equations [2353]. There are variations like Roy–Steiner ([2354,2355] for different masses and hyperbolic relations), GKPY ([2356] with minimal subtractions), etc. Their applicability is reduced in practice to energies around 1.1 GeV for $\pi\pi$ [2357,2358] and πK scattering [2359]. The inelastic, higher-energy, and higher-wave contributions are calculated from phenomenological fits. They have been used with two approaches:

- *Solving the equations* for the lowest partial waves $\ell = 0, 1$, in the region of interest, without using any data in that region. All other contributions come from phenomenological fits. Sometimes these are supplemented with ChPT constraints, which reduce considerably the uncertainties. Thus, poles and results in the resonance region could be considered as predictions from the equations and the other terms (and ChPT if used). The proof of the applicability of this approach to determine the existence of the $\sigma/f_0(500)$ and κ/K_0^* and their resulting parameters were provided in [2360] and [2361], respectively.
- *Data driven approach*. Here Roy-like equations are used as constraints on fits to the S and P partial-wave data [2362]. Data sets that are largely inconsistent with these constraints are discarded. Additional contributions from higher energies and partial waves are constrained with forward dispersion relations and sum rules. Simple parameterizations are then fitted to the remaining data, but constrained to satisfy Roy-like equations in different versions and/or number of subtractions, as well as forward dispersion relations up to 1.42 GeV for $\pi\pi$ [2356] and up to 1.6 GeV for πK [2363]. The latter was later coupled to $\pi\pi \to K\bar{K}$ and studied in [2343] with Roy–Steiner equations. With this approach the $\sigma/f_0(500)$ and $\kappa/K_0^*(700)$ poles were obtained in [2362] and [2343,2364], respectively.

Recall that dispersion relations are written in the first Riemann sheet. However, in both approaches above, poles can be determined within a fully dispersive approach, because the second sheet can be easily obtained using Eq. (8.21). In contrast, accessing the "contiguous" sheet in the inelastic regime requires additional analytic continuation methods. Detailed reports on the dispersive determinations of the $\sigma/f_0(500)$ and κ/K_0^* (700) poles can be found in [2342] and [2343], respectively. For convenience we have gathered their resulting poles in Tables 29 and 30. We also provide the modulus of the coupling to the dominant decay channel. Note

Table 29 $\sigma/f_0(500)$ pole determinations using Roy–Steiner equations and the conservative dispersive estimate [2342] which covers them. For the latter, we have corrected a typo in the error of Im$\sqrt{s_{pole}}$ which read ± 12 MeV instead of ± 15 MeV

$\sigma/f_0(500)$	$\sqrt{s_{pole}}$ (MeV)	$\lvert g\rvert$ (GeV)
Refs. [2360,2365]	$(441^{+16}_{-8}) - i(272^{+9}_{-12.5})$	$3.31^{+0.35}_{-0.15}$
Ref. [2366]	$(442^{+5}_{-8}) - i(274^{+6}_{-5})$	-
Ref. [2362]	$(457^{+14}_{-13}) - i(279^{+11}_{-7})$	$3.59^{+0.11}_{-0.13}$
Conservative Dispersive Estimate		
Ref. [2342]	$(449^{+22}_{-16}) - i(275 \pm 15)$	$3.45^{+0.25}_{-0.29}$

Table 30 $\kappa/K_0^*(700)$ dispersive pole determinations using Roy–Steiner equations

$\kappa/K_0^*(700)$	$\sqrt{s_{pole}}$ (MeV)	$\lvert g\rvert$ (GeV)
Ref. [2361]	$(658 \pm 13) - i(279 \pm 12)$	
Ref. [2343]	$(648 \pm 7) - i(280 \pm 16)$	3.81 ± 0.09

that the uncertainty and spread of the dispersive results are much smaller than the RPP estimates in Table 28. This is because other non-dispersive and model-dependent determinations are included in the RPP estimate. However, the existence of two independent dispersive approaches was decisive to consider both resonances as well established in the RPP 2012 and 2020 editions, respectively, changing their nominal masses in their names to be closer to their pole values.

Note that the $f_0(980)$ pole was obtained simultaneously within the same framework [2356,2360]. However, being a narrow resonance and further away from left cuts, its pole is more similar to those obtained with other methods. Finally, some of these analytic continuation methods – using dispersively constrained input – have been applied to determine the poles of further mesons in the inelastic regime, including the scalars $K_0(1430)$ [2367], $f_0(1370)$ and $f_0(1500)$ [2368]. In such cases Eqs. (8.18), (8.19) and (8.21) do not hold and the use of analytic continuation methods is unavoidable to suppress any model dependence.

8.2.3 Light scalars and QCD

In the previous section, we have discussed how the rigorous dispersive approach was instrumental in settling the controversy about the existence and parameters of the σ and κ. Once this is settled, we now concentrate on light scalars within the context of QCD, which is the subject of this volume.

Unitarized chiral perturbation theory (UChPT)
Being so light, these resonances lie in the non-perturbative region of QCD, and thus an effective treatment with ChPT seems appropriate. However, the ChPT series by itself cannot generate poles and also violates unitarity as the energy reaches the resonance region. The most successful approach

is thus a combination of Chiral Perturbation Theory (ChPT) with dispersion relations. This is generically known as unitarized ChPT.

ChPT, which is the low-energy theory of QCD, and is formulated as an expansion in momenta or masses of the NGB, has been introduced for the meson sector in Sect. 6.2.2. Meson–meson scattering partial waves are then expanded as $f(s) = f_2(s) + f_4(s) + \cdots$, where $f_{2n}(s) = O(p^2/F_\pi^2) \times O(p/\Lambda_\chi)^{2n-2}$, p are the meson CM momenta or masses, $\Lambda_\chi = 4\pi F_0$ and F_0 is the NGB decay constant at LO, common to all mesons at that order. Up to higher orders, F_0 can be approximated by $F_\pi, F_K \ldots$ Note that we have suppressed momentarily the isospin and angular momentum indices I, ℓ. As an example, the $O(p^2)$ or LO $\pi\pi$ and πK elastic partial waves in the scalar channel with lowest isospin are:

$$f_0^0(s) = \frac{2s - M_\pi^2}{32\pi F_\pi^2}, \tag{8.22}$$
$$f_0^{1/2}(s) = \frac{5s^2 - 2(M_K^2 + M_\pi^2)s - 3(M_K^2 - M_\pi^2)^2}{128\pi F_\pi^2 s}.$$

ChPT amplitudes are an expansion in powers of p and cannot satisfy the unitarity condition in Eq. (8.19) exactly, but just perturbatively:

$$\mathrm{Im}\, f_2(s) = 0, \ \mathrm{Im}\, f_4(s) = \sigma(s) f_2(s)^2, \ldots \tag{8.23}$$

When p/Λ_χ is very small, this is not a problem, but the violation of unitarity grows with momenta or energy. This violation then becomes a severe caveat to describe resonances, since, in typical cases, resonant effects saturate the unitarity bound in Eq. (8.20). Even worse, the ChPT series cannot generate poles in s and thus, in principle, cannot generate resonances.

Therefore, if we want to describe resonances, we need to implement unitarity, but also analyticity if we want to study their associated poles. Let us now provide a simple, but formal, derivation of ChPT unitarization methods. The elastic unitarity condition in Eq. (8.19) fixes the imaginary part of the inverse partial wave. Hence, naively, we just have to use ChPT to calculate the real part of the inverse amplitude, and write: $\mathrm{Re}(1/f) = \mathrm{Re}\, 1/(f_2 + f_4 + \cdots) \simeq (1/f_2)(1 - \mathrm{Re}\, f_4/f_2 + \cdots)$, since f_2 is real from Eq. (8.22). Then we write a unitarized elastic partial wave at different orders as:

$$f_{LO}^U(s) = \frac{1}{1/f_2(s) - i\sigma(s)}, \tag{8.24}$$
$$f_{NLO}^U(s) = \frac{1}{1/f_2(s) - \mathrm{Re}\, f_4(s)/f_2(s)^2 - i\sigma(s)}, \ldots \tag{8.25}$$

and similar expressions for NNLO, etc. Note that the ChPT series is recovered if re-expanding again. These expressions are unitary and can be recast in explicitly analytic forms. For instance, using Eq. (8.23), the second one is $f_{NLO}^U = f_2^2/(f_2 - f_4)$, which is known as the NLO Inverse Amplitude Method (IAM). Similar analytic formulas for higher orders exist [2369–2371]. Thus these methods can be analytically continued to the complex plane and the second sheet using Eq. (8.21). This derivation is formal because, strictly speaking, we could still not use the expansion of the real part beyond the applicability realm of ChPT into the resonance region. However, there are derivations [2372–2374] from partial-wave dispersion relations for the inverse partial wave and ChPT is only used in the subtraction constants at $s = 0$ or in the left and circular cuts. The use of several subtractions makes those cuts to be dominated by the low energies, where ChPT is applicable, thus justifying the use of the Inverse Amplitude Method.

Interestingly, with the simplest possible calculation, i.e. using just the LO in Eqs. (8.24) and (8.22) in the chiral limit $M_\pi, M_K \to 0$, we find the following poles in the second Riemann sheets of the partial waves where the $\sigma/f_0(500)$ and $\kappa/K_0^*(700)$ are seen:

$$f_0^0 : \sqrt{s_\sigma} = (1 - i)\sqrt{8\pi} F_0 \simeq (463 - i463)\,\mathrm{MeV}, \tag{8.26}$$
$$f_0^{1/2} : \sqrt{s_\kappa} = (1 - i)8\sqrt{\pi/5} F_0 \simeq (638 - i638)\,\mathrm{MeV},$$

where for the numerical values of F_0^2 we have taken $F_\pi^2 \simeq 92.3\,\mathrm{MeV}$ for $\pi\pi$ and $F_\pi F_K$ for πK scattering, with $F_K = 1.19 F_\pi$ [300]. Taking into account that this is the most naive LO unitarized calculation, with no free parameters, the lightest scalar masses come remarkably close to their actual values, while their widths are about a factor of 2 too wide. Note that the only dynamical information is the scale of the spontaneous chiral symmetry breaking, given by F_0. In contrast, if the same procedure is followed with the vector $\ell = 1$ channels, the resulting poles for the $\rho(770)$ and $K^*(892)$ come almost twice too heavy, and their widths more than 16 times too wide. This is already an indication that the LO low-energy chiral dynamics plays a predominant role in the formation of light scalar resonances, and very little for other ordinary mesons.

The description of meson–meson scattering at NLO in UChPT is very successful for both scalar and vector partial waves in all isospin combinations (tensor waves start at NNLO). In particular, now not only the pole width of the scalars comes right, but also the vector meson poles and their parameters. Recall that, as explained in Sect. 6.2, the NLO ChPT calculations contain several Low Energy Constants $L_i(\mu)$, which multiply the terms in the NLO Lagrangian allowed by symmetry. They are scale-dependent because they absorb, through renormalization, the loop divergences at previous orders. In addition, they contain information about the underlying quark and gluon dynamics, namely, QCD. Only when these L_i are taken into account it is possible to describe the "ordinary" quark–antiquark vector mesons with UChPT. However, the L_i combinations that appear in the scalar channels are much less relevant numerically and that is why scalar

poles come out fairly decent with just the LO UChPT and just information on the chiral breaking scale.

So far we have only discussed elastic unitarization. But exactly the same naive derivation can be followed in matrix form to obtain a coupled channel T-matrix formalism [2375–2377], only slightly more complicated. When this is done, besides the $f_0(500)$ and $K_0^*(700)$ poles, those associated with the $a_0(980)$ and $f_0(980)$ resonances also appear in the Inverse Amplitude Method [2377], completing the lightest scalar nonet, as well as those of the $\rho(770)$ and $K^*(892)$ vectors.

Many variations of ChPT unitarization techniques exist in the literature of which, together with the IAM, the simplest and most popular is the Chiral Unitary Approach [2378, 2379] (for other variations, see the reports [2342,2380–2382]), which usually raises the caveat about some arbitrariness. However, all unitarization methods just correspond to finer or more crude approximations to $\mathrm{Re}(1/f)$ and its ChPT series or to different treatments of the left cut, or even including some additional heavier states. But as long as they contain the basic information about the chiral scale, or are equivalent to the ChPT LO, they all obtain a similar description of light scalars, whereas vectors or other resonances can be accommodated only when including enough NLO information.

Of course, since unitarization methods involve some truncation of ChPT and approximations, they are not competitive in precision and rigor with the precise dispersive approaches discussed before. They have, however, another advantage, which is that we can study the dependence of the resonances on QCD parameters, which we will describe next.

Leading QCD $1/N_c$ behavior

At leading order in the $1/N_c$ expansion [1162,2383], ordinary $q\bar{q}$ mesons behave as $M \sim O(1)$ and $\Gamma \sim O(1/N_c)$. Genuine tetraquark states [2384,2385] have at least that same N_c behavior, which is even more suppressed for glueballs.

First of all, using meson–meson scattering and the light-resonance pole parameters it is possible to build observables whose sub-dominant N_c corrections are highly suppressed [2386]. When evaluated for the $f_0(500)$ and $K_0^*(700)$ the resulting values are at odds with the ordinary meson or glueball behavior by several orders of magnitude.

Next, using the effective theory approach, the $1/N_c$ leading order of the NLO ChPT parameters is known from a model-independent analysis: $M_\pi, M_K \sim O(1)$, $F_0 \sim \sqrt{N_c}$ and the L_i behavior is either $O(1)$ or $O(N_c)$ [69,2387]. If, in the UChPT amplitudes, we then call p a parameter whose behavior is $O(N_c^k)$ and change its value to $p \to p(N_c/3)^k$, we will obtain the leading $1/N_c$ behavior of resonances following their associated poles as N_c is increased. Thus, already with the substitution $F_0 \to F_0\sqrt{N_c/3}$ in the LO UChPT results in Eqs. (8.26), we obtain a non-ordinary behavior for both the σ and κ. Namely, their $M, \Gamma \sim O(\sqrt{N_c})$.

That was just the naive LO estimate, but the leading $1/N_c$ dependence within UChPT has been studied to NLO in [2377,2388] and NNLO in [2389]. It is then possible to study the light vectors as well, and they come remarkably compatible with the expected ordinary behavior. This is shown for the $\rho(770)$ in the top panel of Fig. 175. However, the σ and κ poles, shown in the center and bottom panels of Fig. 175, respectively, display again a non-ordinary behavior, at least near the physical value of $N_c = 3$. This is a robust result also found in other approaches. Of course, if N_c is made very large, the dominance of meson loops governed just by F_0, which are suppressed by $1/N_c$, fades away. Then, even the tiniest mixture with an ordinary meson could dominate at sufficiently large N_c. We should remark that there is some uncertainty, that grows with N_c due to the scale dependence of the L_i, illustrated for the σ in Fig. 175. Indeed, Fig. 175 shows that the sigma pole could turn back [2390] to the real axis, well above 1 GeV. This could be a small mixture with an "ordinary" state around or above 1 GeV. This is also found in NNLO UChPT [2389]. Similarly, in other phenomenological approaches the σ and κ only appear when the unitarized meson–meson interaction is included, showing up as an additional pole due to unitarization, in addition to ordinary states above 1 GeV that are present even if meson–meson interactions are turned off (this was first proposed in [2391], for additional references see [2342]). Back to UChPT, the ordinary subdominant component restores the semi-local duality sum-rules [2390] that would be violated if the light scalars just disappeared from the spectrum by becoming too massive and wide. However, other analyses [2392,2393], challenged in [2394], yield a σ behavior closer to the one of the opposite side of the scale uncertainty in Fig. 175, reaching the third quadrant at very large N_c, which lacks a clear interpretation. One should nevertheless recall that the large-N_c regime, although of mathematical interest, is not the one of relevance for the observed meson, but the leading $1/N_c$ behavior near $N_c = 3$.

Quark-mass dependence and light-scalar multiplets

The study of quark-mass dependence is of interest to understand the dynamics of their formation, to provide a guideline for lattice studies, and to check that the light scalar states that we have grouped in an octet are degenerate when the strange and non-strange quark masses are equal.

We have seen in Sect. 6.2 the relation between quark and meson masses. This allows us to study the quark mass dependence of the σ at NLO [2395] and NNLO [2396] and κ at NLO [2397]. A slight IAM modification is used to deal with subthreshold Adler zeros [2398]. Figure 176 thus shows the resulting σ and κ pion mass dependence. Note that beyond 300–350 MeV the results are at most qualita-

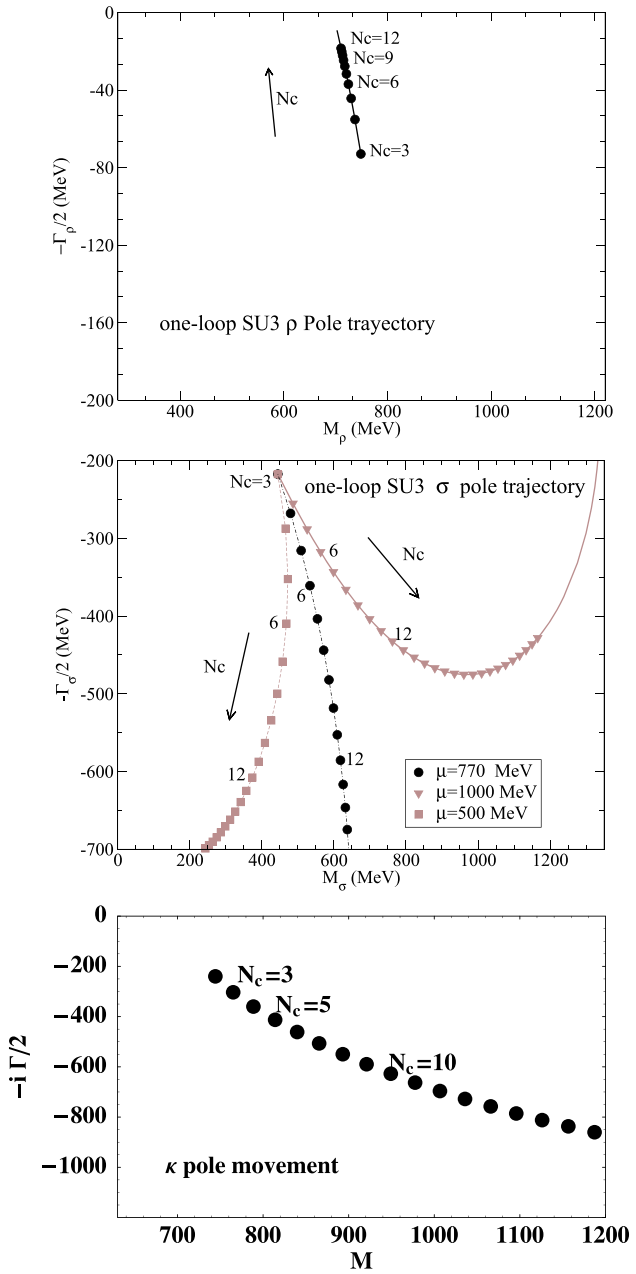


Fig. 175 Trajectories of the $\rho(770)$ (top), $\sigma/f_0(500)$ (center) and $\kappa/K_0^*(700)$ (bottom) poles in the complex plane as N_c is varied away from 3 within NLO ChPT unitarized with the IAM. The lighter curves in the center plot indicate the uncertainties when varying the regularization scale μ in the usual range, as recalculated in [2390]. In the case of the $\rho(770)$ the three lines almost overlap and are not plotted. Top and center figures taken from [2390] and bottom figure from [2377]

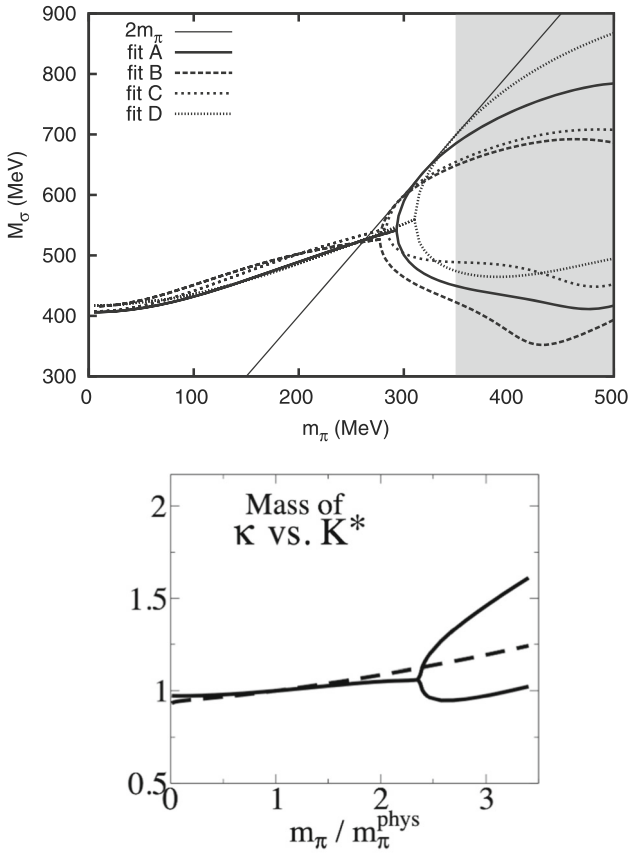


Fig. 176 Top: Dependence of the sigma mass M_σ on the pion mass, from the NNLO (two-loops) IAM [2389]. Different curves represent different fits on [2389]. The thin continuous line shows the $2m_\pi$ threshold. Bottom: m_π dependence of the κ (solid line) and $K^*(892)$ (dashed line) masses [2397]. All masses and widths are defined from the pole positions as obtained from NLO IAM fits. Figures taken from [2399] (top) and [2400] (bottom)

tive. With increasing pion mass, the meson masses grow, although slower than the two-pion threshold, and their pole widths decrease. When the pion mass is 2–3 times its physical value, the 2π threshold is above the pole mass of these resonances. Then, their behavior differs dramatically from that of the $\rho(770)$ and $K^*(892)$ (the latter shown in Fig. 176).

The width of these non-scalar mesons would tend to zero, and their conjugated pair of poles would meet at threshold [2395,2401]. Right after that, one of their poles would jump to the first sheet, whereas the other would remain at a symmetric position in the second sheet, both below threshold. This is a bound state. In contrast, the σ and κ conjugated poles meet in the second sheet below threshold. The two branches observed in Fig. 176 correspond to these two poles in the second sheet, where at first one moves towards threshold and the other away from it. The closest one to threshold, influencing the most the physical region, is known as a "virtual" or quasi-bound state. Eventually, it reaches threshold and jumps to the first sheet, becoming a bound state. However, its second-sheet counterpart lies in a rather different position. The more asymmetric their positions, the more predominant their "molecular" or "meson cloud" nature is. Hence, UChPT suggests that, at high pion masses, both the $f_0(500)$ and $K_0^*(700)$ are closer to two-meson states than to ordinary mesons.

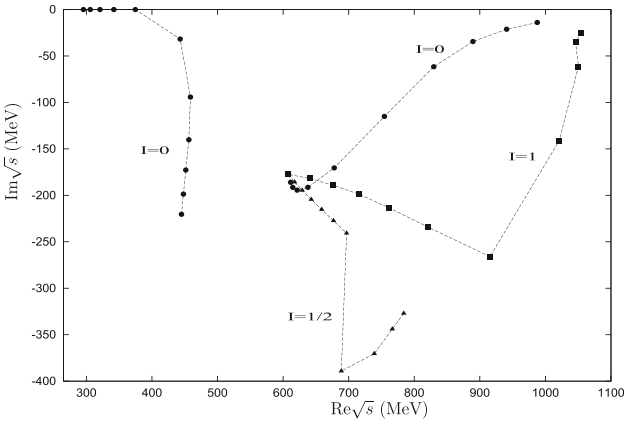


Fig. 177 Trajectories of the poles that appear in coupled-channel unitarized amplitudes of different isospin as the pion, kaon, and eta masses are varied from their physical values to a common value of 350 MeV [2406]. This shows that the lightest scalars actually belong to a nonet in the SU(3) limit. The two trajectories with $I = 0$ correspond to the singlet and octet states, not directly to the poles of the σ or $f_0(980)$ resonances, which are a mixture of these two. Figure taken from [2406]

Quark masses can be changed on the lattice. Actually, calculations are not often done at physical masses, which are expensive numerically. Note also that analytic continuation to reach poles would be required, although models are often used to reach poles. There are lattice calculations for the σ [2402,2403], supporting its molecular picture at very large pion masses, where it is a bound state. The σ is also found at moderately large pion masses [558,2404] qualitatively consistent with UChPT. For $m_\pi = 236\,\text{MeV}$ [558,559] lattice results are consistent with a pole now in the second Riemann sheet, also consistent with UChPT. A virtual state was found for κ in πK scattering on the lattice [551,577], again in qualitative agreement with UChPT. However, as the pion mass becomes lighter, the σ and κ poles are plagued again with instabilities [552,559,560]. In [2405], Adler zeros, i.e. chiral symmetry, were found to be very relevant in the κ determination. A dispersive "data-driven" approach of the kind explained above may be relevant for a robust extraction of light scalar poles from lattice-QCD. We refer to Sect. 4 for further details.

The strange-quark mass can also be varied [2397], but not much since it is already quite high, and thus the observed changes on scalars are very smooth. However, when changing both quark masses one can reach the degenerate pion–kaon mass limit. Figure 177 shows that the trajectories of the κ pole, and a combination of the σ and $f_0(980)$ become degenerate with the $a_0(980)$ pole in that limit. This result has been obtained [2406] within the Unitary Chiral Approach, where the left cut is neglected and the effect of the L_i is mimicked by a mass-independent cutoff. Still, this provides strong support for the assignment of these states to the same lightest scalar octet.

8.2.4 Summary

Despite their relevant role in numerous aspects of hadron physics and QCD, the controversy about the existence and the parameters of the lightest scalar nonet, particularly for the $\sigma/f_0(500)$ and $\kappa/K_0^*(700)$, predated the establishment of QCD. The settling of this controversy was hindered by the conflicting available data sets and by the use of models. We have provided here a brief account of how it has been settled recently by using rigorous dispersive techniques to constrain data analyses and to determine the poles associated with the light-scalar resonances. Many phenomenological approaches were able to describe to different degrees of accuracy these states. Here, we have focused on those most directly linked to QCD through the unitarization of Chiral Perturbation Theory, the $1/N_c$ behavior, and the dependence on the quark masses. The general picture that arises is that there is one light scalar nonet below 1 GeV. Their non-ordinary N_c behavior, quark mass dependence, Regge trajectories, and the fact that they do not saturate the ChPT constants strongly support that these mesons are not of the ordinary quark–antiquark type. Rather their predominant component would be of the meson–meson type (molecule, meson cloud, etc). Still, they are most likely mixed with some companion bare or preexistent quark–antiquark state above 1 GeV. Indeed, a second scalar multiplet can be identified between 1.2 and 1.8 GeV. There is still ample room for refining this picture and a high expectation of further experiments and developments from lattice-QCD.

8.3 Exotic mesons

Boris Grube

8.3.1 Introduction

Already when Gell-Mann [17] and Zweig [18] formulated the constituent quark model they presumed that additional states beyond the baryonic qqq and the mesonic $q\bar{q}$ combinations exist.[79] For a long time, the search for such states was unsuccessful and hence all hadronic states going beyond the constituent quark model were labelled *exotic*. However, rather recently experiments have found compelling evidence that exotic states indeed exist. Here, we will focus on exotic mesons, which can be divided into three categories:

[79] In Ref. [17], Gell-Mann writes: "Baryons can now be constructed from quarks by using the combinations (qqq), $(qqqq\bar{q})$, etc., while mesons are made out of $(q\bar{q})$, $(qq\bar{q}\bar{q})$, etc." Similarly, Zweig writes in a footnote in Ref. [18]: "In general, we would expect that baryons are built not only from the product of three aces, AAA, but also from $\bar{A}AAAA$, $\bar{A}\bar{A}AAAAA$, etc., where $\bar{A}$ denotes an anti-ace. Similarly, mesons could be formed from $\bar{A}A$, $\bar{A}\bar{A}AA$ etc."

(i) *spin-exotic states*, which have J^{PC} quantum number combinations that are not possible for ordinary $q\bar{q}$ states (cf. Table 9),[80] (ii) *flavor-exotic states*, which have flavor quantum numbers, such as isospin and/or strangeness, that are not possible for $q\bar{q}$ states, and (iii) *crypto-exotic states*, which have quantum numbers of ordinary $q\bar{q}$ states and are therefore able to mix with them.

Possible exotic mesonic configurations beyond $q\bar{q}$ are four-quark combinations such as tightly bound $qq\bar{q}\bar{q}$ *tetraquark states*, where the constituents are bound directly by the strong force, or more loosely bound $(q\bar{q})(q\bar{q})$ *molecular states*, which consist of a pair of mesons bound by nuclear forces. Also the gluon fields are expected to manifest themselves in the meson spectrum either in the form of *hybrid states*, where, in addition to a $q\bar{q}$ pair, excited gluonic field configurations contribute to the quantum numbers of the meson, or in the form of *glueballs*, which are color-singlet bound states of gluons (see Sect. 8.4). However, in general, physical mesons are not pure realizations of single configurations but are instead mixtures of all possible configurations that are allowed for the given quantum numbers. Disentangling these different contributions is a highly difficult experimental and theoretical problem.

Crypto-exotic states will manifest themselves as supernumerary states compared to the spectrum expected from the quark model. This makes them rather difficult to establish. And even if experimental data unambiguously show an overpopulation of states in a certain mass range, the determination of the internal configuration of these states is an even harder problem. The prime example for such a situation is the sector of isoscalar scalar mesons discussed in Sects. 8.2 and 8.4. Therefore, the cleanest way to unambiguously establish the existence of exotic mesons is to search for spin- and/or flavor-exotic states. Presently, the clearest evidence for the existence of such states comes from the heavy-quark sector (see Sects. 8.5 and 8.6), where experiments have found several flavor-exotic states with a minimum quark content of four, for example, the charged charmonium and bottomonium states, $Z_c^{\pm}$ and $Z_b^{\pm}$ [1427,2407], or the doubly-charmed state, T_{cc}^{+} [1067].

Although mesons from the light-quark sector, i.e. mesons composed of up, down, or strange quarks, are usually easier to produce in experiments, the picture is less clear in this sector. This is mainly because light mesons have relatively large decay widths compared to their masses. As a consequence, these mesons usually do not appear as isolated and narrow peaks in the invariant mass spectra of their decay products. Instead, they often overlap and interfere with neighboring states, which makes their extraction from experimental data challenging. In addition, in most analyses models are required in order to extract resonances from the data and the results therefore depend on the employed model assumptions and approximations. In the following, we will confine the discussion to spin-exotic light mesons. More details on exotic light mesons can be found in the reviews in Refs. [420,2408–2414].

8.3.2 Predictions

Model predictions

Various models have been employed to study the light-meson spectrum. Some of these model approaches are discussed in more detail in Sect. 5. Further discussions can be found, e.g., in Refs. [420,2414]. Most of the models that include exotic mesons predict the lightest spin-exotic state to be a hybrid meson with $J^{PC} = 1^{-+}$ quantum numbers.

The first detailed studies of hybrid light mesons were based on the bag model [785–787,2415,2416]. In this model, quarks and gluons are described by cavity modes in a confining vacuum bubble (see Sect. 5.1.3). Detailed predictions for the decays of hybrid light mesons were obtained using, for example, the fluxtube model [2417–2423]. This model extends the conventional quark model by explicitly modeling the gluonic fields in form of an oscillating flux tube described by single-phonon excitations. Decays of hybrid mesons were also studied in constituent-glue models [2424–2427], where one assumes that a massless gluon with $J^{P} = 1^{-}$ interacts with quarks via potentials that depend linearly on the distance of the constituents. Recently, also the Dyson–Schwinger/Bethe–Salpeter approach (see, e.g., Refs. [896,2428,2429] and also Sect. 5.2), basis light-front quantization (see, e.g., Ref. [950] and also Sect. 5.3), as well as the AdS/QCD correspondence (see, e.g., Ref. [1003] and also Sect. 5.4) were applied to study hybrid light mesons.

The models predict the mass of the lightest 1^{-+} state to be in the range from about 1.3 to 2.2 GeV and most model calculations find that $f_1(1285)\pi$ and $b_1(1235)\pi$ are the dominant decay modes for the lightest isovector 1^{-+} state. However, for the $\eta\pi$, $\eta'\pi$, and $\rho(770)\pi$ decay modes, discussed in Sect. 8.3.4 below, the model predictions diverge.

Lattice QCD calculations

In recent years, lattice QCD calculations of the hadron excitation spectrum have made tremendous progress (see Sect. 4, in particular Sect. 4.5). Currently, calculations that study the excitation spectrum of light mesons still have to be performed in an unphysical world, where the up and down quarks are much heavier in the simulation than in nature.[81] The main reason for this is that decays into multi-body hadronic final states, which for most excited states are the dominant decay

[80] More correctly, these states have forbidden J^{PG} quantum numbers. However, here we use the common convention that the C-parity of a charged meson in an isospin triplet is given by the C-parity of its neutral partner state.

[81] This is often expressed in terms of an unphysically large pion mass.

modes, cannot yet be calculated on the lattice. By setting the quark masses to sufficiently high values and neglecting multi-hadron operators, the excited states become quasi stable and can be extracted from the simulation. Consequently, such calculations cannot predict widths and decay modes and also cannot take into account coupled-channel effects. Despite these limitations, lattice calculations have already provided important insights by making predictions for light-meson spectra and for two-body scattering processes [514,532].

For example, the seminal calculation performed by the Hadron Spectrum collaboration [521] showed for the first time a nearly complete spectrum of isoscalar and isovector mesons covering a wide range of J^{PC} quantum numbers up to $J = 4$ (see Fig. 46). The lattice spectrum is qualitatively similar to the one obtained from quark-model calculations. However, the lattice calculation in addition revealed a whole supermultiplet of extra states [527] that lie about 1.3 GeV above the lightest $J^{PC} = 1^{--}$ state and that have quantum numbers of 0^{-+}, 1^{--}, 2^{-+}, and 1^{-+}, where the latter one is spin-exotic. Studying the overlap of these states with various operators used in the calculation allowed to probe their internal structure. All states in the supermultiplet have large overlaps with operators that correspond to a chromomagnetic gluonic excitation coupled to a color-octet $q\bar{q}$ pair in an S-wave and were therefore identified as hybrid states. Intriguingly, the spin-exotic 1^{-+} state was predicted to be the lightest hybrid state confirming many model calculations (see Sect. 8.3.2).

Recently, the Hadron Spectrum collaboration published results of the first lattice QCD calculation of the hadronic decays of the lightest 1^{-+} resonance using a two-body approximation for the decay [582]. They performed this calculation at the $SU(3)_{\text{flavor}}$ symmetric point, where up, down, and strange-quark masses are chosen to approximately match the physical strange-quark mass, corresponding to a large unphysical pion mass of about 700 MeV. Using a coupled-channel approach, the Hadron Spectrum collaboration studied the scattering amplitudes of eight meson–meson systems and extrapolated the extracted 1^{-+} resonance pole and its couplings to the physical light-quark masses. Doing so and assuming a 1^{-+} resonance mass of 1564 MeV (value taken from Ref. [2324]), they found a broad π_1 resonance with a total width ranging between 139 and 590 MeV. The dominant decay mode of this resonance is $b_1(1235)\pi$ (partial width ranging from 139 to 529 MeV), in qualitative agreement with most model calculations (see Sect. 8.3.2). Compared to the $b_1(1235)\pi$ channel, the partial widths for the decays into $f_1(1285)\pi$, $\rho(770)\pi$, $\eta'\pi$, and $\eta\pi$ are much smaller. Although these results still have large uncertainties, they provide important guidance for experiments.

The next great leap for lattice QCD is the calculation of three-body systems, which is already looming on the horizon (see Sect. 4.5.8 and Ref. [593]). First proof-of-principle calculations of three-body systems that do not contain any resonances (see, e.g., Fig. 53) demonstrate the feasibility of the approach and are paving the way towards calculations of more interesting systems that contain two- and/or three-body resonances.

8.3.3 Experimental methods

Excited light mesons can be studied in many reactions. They are copiously produced in high-energy scattering reactions of meson beams on nucleon or nuclear targets, such as diffractive dissociation or charge exchange, as well as in central-production reactions in hadron–hadron scattering. Also, antiproton–nucleon annihilations are a source of light mesons. Complementary to these purely strong-interaction processes are photoproduction reactions, which are induced by photon or lepton beams, and e^+e^- scattering reactions such as annihilation, initial-state radiation, or two-photon fusion. Finally, also multi-body decays of heavy particles, such as τ, J/ψ, or D, are good laboratories to study light mesons. Conservation laws, couplings, and the available energy impose constraints that determine which excited states are allowed to be produced from the various initial states in these reactions. The study of the light-meson spectrum is a world-wide effort with experiments performed at all major particle-accelerator labs covering all the above reactions.

Excited light mesons decay via the strong interaction and are hence extremely short-lived. This is why these states are usually referred to as *resonances*, which are characterized by their nominal mass m_0, their total width Γ_0, and their quantum numbers. In the simplest case of an isolated resonance, its experimental signature is a *peak* at m_0 in the distribution of the invariant mass m of the system of daughter particles that the resonance decays into. This peak is accompanied by a *phase motion*, i.e. an increase of the phase of the quantum mechanical amplitude of the studied process by 180° with increasing m, reaching 90° at m_0 (see Fig. 178). If the resonance is in addition narrow and m_0 is far away from kinematical thresholds, the resonance amplitude is well approximated by a Breit–Wigner amplitude. However, in general resonances are described by amplitudes that are analytical functions of m^2 and the resonance parameters are defined by the position of pole singularities of this amplitude in the complex m^2 plane (see, e.g., Ref. [2430] for more details).

Depending on its mass and quantum numbers, a resonance may have several decay modes, which for highly excited states often lead to multi-body hadronic final states consisting mostly of π, K, η, and/or η'. Due to their short-lived nature, any information about resonances has to be inferred from the kinematic distribution of their decay products. To this end, *partial-wave analysis* (PWA) techniques are often employed, which take into account possible interferences of

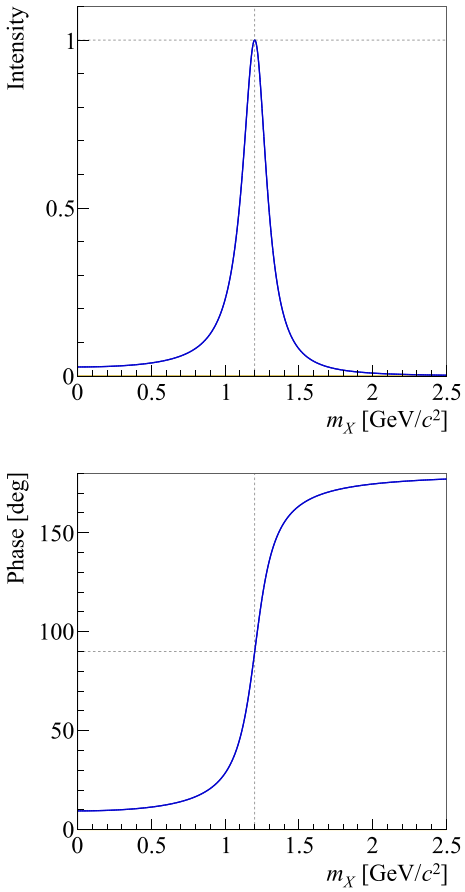


Fig. 178 Example of a relativistic Breit–Wigner amplitude with constant width for a fictitious resonance with a mass of $m_0 = 1200\,\text{MeV}$ and a total width of $\Gamma_0 = 200\,\text{MeV}$. (Top) intensity, i.e., absolute value squared of the amplitude, (bottom) phase of the amplitude

all the intermediate resonances produced in the reaction and exploit the full kinematic information contained in the data. For an n-body final state with given mass m, a set τ of $3n-4$ kinematic variables is needed to completely define the decay kinematics. In a simplified picture, a PWA model describes the measured *intensity distribution* $\mathcal{I}(m,\tau)$, i.e. the density distribution of the events in the $(3n-4)$-dimensional phase space of the final-state particles, in terms of *partial-wave amplitudes* $\mathcal{T}_i(m)$, which describe the strength and phase with which an intermediate state with given quantum numbers $i = \{J^{PC}\, M\}$ and mass m is produced, and *decay amplitudes* $\Psi_i(m,\tau)$, which describe the decay of this intermediate state into the observed final state. Here, M is the projection of the spin J along the chosen quantization axis. High-energy scattering reactions, for which examples will be discussed below, are known to be dominated by natural-parity exchange.[82]

[82] The *naturality* is defined as $\epsilon = P\,(-1)^J$, i.e. $\epsilon = +1$ corresponds to the natural-parity series with $J^P = 0^+, 1^-, 2^+, \ldots$ and, correspondingly, $\epsilon = -1$ corresponds to the unnatural-parity series with $J^P = 0^-, 1^+, 2^-, \ldots$.

When analyzing data from these reactions, it is hence advantageous to perform the PWA in the *reflectivity* basis [2431], where the spin state of a resonance is characterized by M^ϵ with $M \geq 0$ and $\epsilon = \pm 1$ such that the multiplicity of $2J+1$ of the spin state remains unchanged. Here, ϵ corresponds to the naturality of the exchange particle in the scattering reaction. By performing the PWA in this basis, it is therefore possible to separate the contributions from natural- and unnatural-parity exchange to the scattering reaction.

Since production and decay of a resonance are independent of each other, the total amplitude for an intermediate state i is given by $\mathcal{T}_i(m)\,\Psi_i(m,\tau)$. In the simplest case, the amplitudes of the various allowed intermediate states i are assumed to be fully coherent so that

$$\mathcal{I}(m,\tau) = \left| \sum_i \mathcal{T}_i(m)\,\Psi_i(m,\tau) \right|^2, \tag{8.27}$$

where the sum runs over all allowed states. It is important to note that in the above equation, the intensity is given by the sum of the contributing amplitudes, i.e. all intermediate states may interfere with each other. The decay amplitudes can be calculated using first principles and models. The analyses that will be discussed in Sect. 8.3.4 below use a two-stage procedure, where in the first stage the known decay amplitudes Ψ_i are used to determine the partial-wave amplitudes $\mathcal{T}_i$ in narrow m bins by fitting the PWA model in Eq. (8.27) to the measured τ distributions. At this stage, no assumptions are made about the resonance content in the studied n-body system. In a second stage, a resonance model is fit to the m dependence of selected partial-wave amplitudes in order to extract the resonances and their parameters. For high-energy scattering data, the resonance model also has to take into account contributions from *non-resonant processes*, i.e. processes where the measured n-body final state is produced without going through an intermediate n-body resonance. Unfortunately, in most cases no detailed theoretical models exist for these non-resonant contributions and one has to revert to empirical models. More details on the PWA procedure and the involved model assumptions can be found, e.g., in Ref. [2432].

8.3.4 Experimental evidence

More than three decades ago the GAMS experiment claimed the first observation of a spin-exotic resonance with $J^{PC} = 1^{-+}$ [2433]. Since then, many other experiments reported such signals. Currently, the Particle Data Group (PDG) lists three spin-exotic light-meson states: the $\pi_1(1400)$, the $\pi_1(1600)$, and the $\pi_1(2015)$ [513]. However, despite the seemingly large body of evidence, which includes data from pion diffraction, antiproton–nucleon annihilation, photoproduction, and charmonium decays covering several decay channels, the experimental situation is still puzzling and the

interpretation of many of the observed signals is controversial.

The $\pi_1(1400)$ was observed nearly exclusively in the $\eta\pi$ decay channel produced in pion diffraction and antiproton–nucleon annihilation [2328,2433–2439]. Only the OBELIX and Crystal Barrel experiments claimed to see the $\pi_1(1400)$ also in the $\rho(770)\pi$ decay channel in their antiproton–nucleon annihilation data [2440,2441]. Surprisingly, the signal in the $\rho(770)\pi$ channel arises from antiproton–nucleon initial states with different quantum numbers than the signal in the $\eta\pi$ channel.[83] Since production and decay of a resonance are independent, the $\rho(770)\pi$ resonance claimed by OBELIX and Crystal Barrel cannot be the same $\pi_1(1400)$ state that is observed in $\eta\pi$ – a puzzling result. The $\pi_1(1400)$ masses quoted by the various experiments are in fair agreement; the width values, however, scatter over a larger range. The PDG estimates for the $\pi_1(1400)$ mass and width are $m_0 = (1354 \pm 25)$ MeV and $\Gamma_0 = (330 \pm 35)$ MeV [513].

Compared to the $\pi_1(1400)$, the $\pi_1(1600)$ was seen in a much wider range of decay channels produced in pion diffraction, antiproton–nucleon annihilation, and χ_{c1} decays. Signals were reported in the $\rho(770)\pi$ [2319,2442–2445], $\eta'\pi$ [2446–2450], $f_1(1285)\pi$ [2449,2451], and $b_1(1235)\pi$ [2322,2438,2447–2449,2452] decay channels. As for the $\pi_1(1400)$, the measured $\pi_1(1600)$ mass values are in better agreement with each other than the measured width values. The PDG estimates for the $\pi_1(1600)$ mass and width are $m_0 = (1661^{+15}_{-11})$ MeV and $\Gamma_0 = (240 \pm 50)$ MeV [513].

The $\pi_1(2015)$ was so far only observed by the BNL E852 experiment in the decay modes $f_1(1285)\pi$ [2451] and $b_1(1235)\,\pi$ [2322]. It hence still needs to be confirmed by other experiments and is listed as a "further state" by the PDG.

Although on first sight there seems to be strong experimental evidence for the $\pi_1(1400)$ and the $\pi_1(1600)$, some analyses have issues and some experimental results are disputed. From a phenomenological standpoint, the properties of the $\pi_1(1400)$ are problematic. Compared to most of the predictions (see Sect. 8.3.2), it is too light. Also, the $\pi_1(1600)$ is too close in mass to the $\pi_1(1400)$ in order to be an excitation of the latter. Additionally, the fact that the $\pi_1(1400)$ seems to decay only to $\eta\pi$ is hard to explain.[84]

The analyses of some channels also face technical issues. For example, in order to extract the $\pi_1(1400)$ in the $\eta\pi$ channel and the $\pi_1(1600)$ in the $\eta'\pi$ channel, the phase motions of the P-wave amplitudes need to be measured. Often, this can be done only relative to the D-wave amplitudes. However, in the mass region of interest the D-waves contain contributions from the $a_2(1700)$, which is the first radial excitation of the $a_2(1320)$ ground state. Unfortunately, the $a_2(1700)$ is a rather broad state and its resonance parameters are not well known. For the widely used simple Breit–Wigner based resonance models, this may lead to systematic uncertainties that are hard to control.

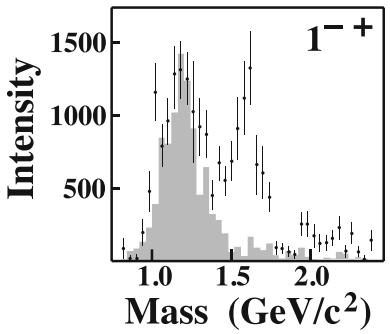


Fig. 179 Intensity of the $\rho(770)\pi$ P-wave with spin-exotic $J^{PC} = 1^{-+}$ quantum numbers produced in natural-parity exchange (points with statistical uncertainties) as a function of the $\pi^-\pi^-\pi^+$ mass obtained by the BNL E852 collaboration. (Adapted from Fig. 3(b) in Ref. [2442])

The analysis of the data of the BNL E852 experiment yielded inconsistent results on the production properties of the $\pi_1(1600)$. Whereas in the $\eta'\pi$ [2446] and $f_1(1285)\pi$ [2451] channels the $\pi_1(1600)$ is observed to be produced only via natural-parity exchange, i.e. with $M^\epsilon = 1^+$, it appeared in the $\rho(770)\pi$ [2442,2443] and $b_1(1235)\pi$ [2322] channels also in unnatural-parity exchange, i.e. in waves with $M^\epsilon = 0^-$ and 1^-, with similar strength as in the $M^\epsilon = 1^+$ wave. This is hard to explain as production and decay of a resonance are independent processes.

One of the deepest puzzles, however, concerns the seemingly contradictory conclusions on the existence of the $\pi_1(1600)$ in the $\rho(770)\pi$ decay channel that were drawn from similar analyses. The BNL E852 experiment was the first to claim the observation of $\pi_1(1600) \to \rho(770)\pi$ based on a sample of about 250 000 $\pi^- p \to \pi^-\pi^-\pi^+ p$ events and using a PWA model with 21 waves [2442,2443]. The measured intensity distribution of the spin-exotic wave with $J^{PC} = 1^{-+}$ quantum numbers is shown in Fig. 179. It exhibits a pronounced peak at about 1.6 GeV that is accompanied by significant phase motion with respect to other partial waves (see Fig. 19 in Ref. [2443]).[85] Based on a simultaneous resonance-model fit of the intensities of the 1^{-+} wave

[83] In the $\rho(770)\pi$ channel, the $\pi_1(1400)$ is seen predominantly in P-wave antiproton–nucleon initial states, whereas in the $\eta\pi$ channel it is seen mainly in the $^3S_1$ initial state.

[84] If one would take the $\pi_1(1400) \to \rho(770)\pi$ claims of OBELIX and Crystal Barrel [2440,2441] at face value, then even two mass-degenerate $\pi_1(1400)$ states would exist, one decaying to $\eta\pi$ the other to $\rho(770)\pi$ – an even more puzzling scenario.

[85] The second peak at about 1.2 GeV was explained as an analysis artifact caused by intensity leaking from the dominant 1^{++} wave into the spin-exotic wave because of a non-uniform detector acceptance in combination with the finite experimental resolution. The gray-shaded histogram in Fig. 179 represents an estimate of this effect from Monte Carlo simulations.

and of the $f_2(1270)\pi$ S-wave with $J^{PC} = 2^{-+}$ and their relative phase, the authors of Refs. [2442,2443] claimed the observation of the $\pi_1(1600)$. However, they also observed a strong dependence of the shape and strength of the $\pi_1(1600)$ signal on the PWA model.

Surprisingly, an analysis of a more than 20 times larger data sample (2.6×10^6 $\pi^- p \to \pi^-\pi^-\pi^+ p$ events plus 3.0×10^6 $\pi^- p \to \pi^-\pi^0\pi^0 p$ events) from the same experiment performed by Dzierba et al. came to a completely different conclusion [2453]. They performed the partial-wave analysis independently in 12 bins of the reduced four-momentum squared t' that is transferred from the beam to the target recoil particle[86] in the range from 0.08 to 0.53 GeV^2 using a larger PWA model of 36 waves. The observed intensity distribution of the 1^{-+} wave exhibits a broad and structureless enhancement (see black points in Fig. 180; cf. Fig. 179). The shape of this enhancement was found to change strongly with t' with intensity moving from the 1.2 GeV region towards higher masses with increasing t'. However, the peak at 1.6 GeV, which in Refs. [2442,2443] was attributed to the $\pi_1(1600)$, had disappeared. By applying the 21-wave PWA model from Refs. [2442,2443], Dzierba et al. were able to reproduce the results from Refs. [2442,2443] (see gray points in Fig. 180; cf. Fig. 179). They also showed that the omission of important 2^{-+} waves in the 21-wave PWA model causes leakage from the $\pi_2(1670)$ producing an artificial peak at 1.6 GeV in the 1^{-+} wave. Based on these findings, Dzierba et al. concluded that the BNL E852 data provide no evidence for the existence of the $\pi_1(1600)$ in the $\rho(770)\pi$ decay channel and that the signal reported in Refs. [2442,2443] was an artifact of a too restricted PWA model. However, this conclusion was not based on a resonance-model fit and did not take into account the phase motions of the 1^{-+} wave that were still present in the analysis of Dzierba et al.. In addition, Dzierba et al. only considered the kinematic region $t' < 0.53$ GeV^2, which will become important in the discussion below.

The first results from the COMPASS experiment only added to the confusion. The authors of Ref. [2444] performed a partial-wave analysis of 420 000 events for the reaction $\pi^-\text{Pb} \to \pi^-\pi^-\pi^+\text{Pb}$ in the kinematic range $0.1 < t' < 1.0$ GeV^2 using an even larger PWA model than Dzierba et al. consisting of 42 waves. This model is similar to the 36-wave PWA model used in Ref. [2453] and includes in particular the 2^{-+} waves that were found to cause leakage from the $\pi_2(1670)$ into the 1^{-+} wave. However, in contrast to Dzierba et al., COMPASS observed an enhancement at 1.6 GeV in the intensity distribution of the 1^{-+} wave (see data points in Fig. 181; cf. black data points in Fig. 180). In the performed resonance-model fit, which describes the intensities and mutual interference terms of six waves simultaneously, the 1^{-+} amplitude is well described by a coherent sum of a non-resonant and a Breit–Wigner amplitude for the $\pi_1(1600)$ (see curves in Fig. 181) and the resulting resonance parameters are compatible with the previous measurements of the $\pi_1(1600)$. Hence, COMPASS claimed the observation $\pi_1(1600) \to \rho(770)\pi$.

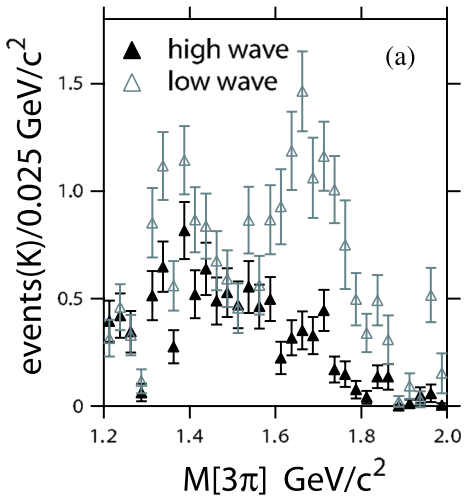


Fig. 180 Intensity distribution of the $\rho(770)\pi$ P-wave with spin-exotic $J^{PC} = 1^{-+}$ quantum numbers produced in natural-parity exchange as obtained by Dzierba et al. using BNL E852 data on $\pi^- p \to \pi^-\pi^-\pi^+ p$ in the kinematic range $0.18 < t' < 0.23$ GeV^2. The open gray points ("low wave") correspond to the 21-wave PWA model from Refs. [2442,2443] (cf. Fig. 179), the solid black points ("high wave") correspond to the 36-wave PWA model from Ref. [2453]. (Taken from Fig. 25(a) in Ref. [2453])

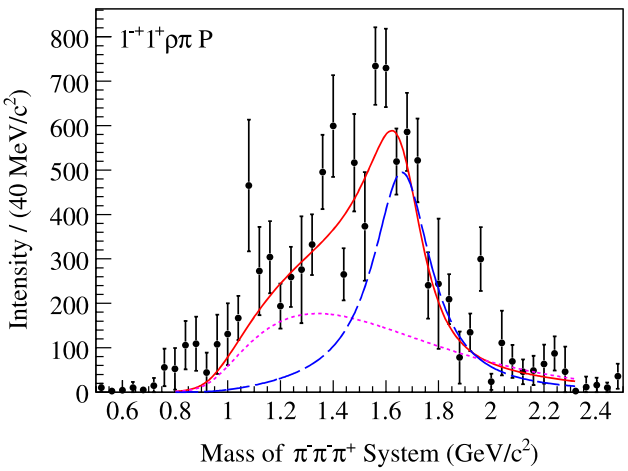


Fig. 181 Intensity distribution of the $\rho(770)\pi$ P-wave with spin-exotic $J^{PC} = 1^{-+}$ quantum numbers produced in natural-parity exchange as obtained by the COMPASS experiment using data on $\pi^-\text{Pb} \to \pi^-\pi^-\pi^+\text{Pb}$ (points with statistical uncertainties). The red curve represents the result of a fit with a resonance model, which is the coherent sum of a Breit–Wigner amplitude for the $\pi_1(1600)$ (blue) and a non-resonant amplitude (magenta). (Taken from Fig. 2(d) in Ref. [2444])

[86] Here, $t' \equiv |t| - |t|_{\text{min}}$ with $t = (p_{\text{beam}} - p_X)^2$ being the Mandelstam variable, p_{beam} the four-momentum of the beam pion, and p_X the total four-momentum of the produced 3π system.

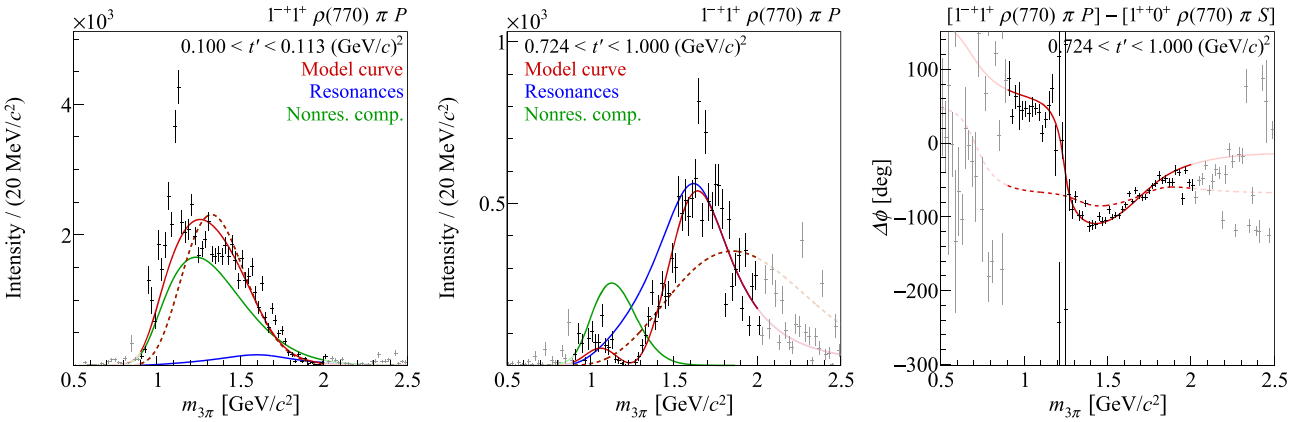


Fig. 182 (Left) and (center) Intensity distribution of the $\rho(770)\pi$ P-wave with spin-exotic $J^{PC} = 1^{-+}$ quantum numbers produced in natural-parity exchange as obtained by the COMPASS experiment using data on $\pi^- p \to \pi^-\pi^-\pi^+ p$ at low and high t'. (Right) Phase of the 1^{-+} wave relative to the $\rho(770)\pi$ S-wave with $J^{PC} = 1^{++}$ at high t'. In the three diagrams, the points with statistical uncertainties represent the measured values. The red curves represent the results of fits with two resonance models. The continuous red curve corresponds to the coherent sum of a Breit–Wigner amplitude for the $\pi_1(1600)$ (blue) and a non-resonant amplitude (green). The dashed red curve corresponds to a model that contains only the non-resonant amplitude. (Taken from Figs. 48(b), (c), and (d) in Ref. [2319])

These puzzling experimental findings were reconciled only recently by the results of a comprehensive partial-wave analysis performed on a highly precise sample of 46×10^6 $\pi^- p \to \pi^-\pi^-\pi^+ p$ events obtained by the COMPASS experiment [2319,2445,2454]. The PWA was performed independently in 11 t' bins in the range $0.1 < t' < 1.0\,\mathrm{GeV}^2$ using the so far largest PWA model with 88 waves. The intensity distribution of the 1^{-+} wave summed over the 11 t' bins exhibits a broad enhancement from about 1.0 to 1.8 GeV but no peak at 1.6 GeV. This is consistent with the distribution observed by the VES experiment in a similar t' range [2455]. The shape of the intensity distribution changes strongly with t' confirming a similar observation made by Dzierba et al. in the BNL E852 data [2453]. At low t', COMPASS observes a broad structure in the mass range from about 1.0 to 1.7 GeV (see Fig. 182(left)).[87] As t' increases, this structure becomes narrower and its maximum moves to about 1.6 GeV so that it becomes similar to the distribution observed in the first COMPASS data on the Pb target (see Fig. 182(center); cf. Fig. 181).

Since resonance parameters are independent of t', the observed strong modulation of the intensity distribution with t' hints at large contributions from non-resonant processes. This was confirmed by the resonance-model fit, which simultaneously describes the amplitudes of 14 selected partial waves. The large wave set provides tight constraints for the 1^{-+} amplitude via the mutual interference terms between the amplitudes. In addition, for the first time all 11 t' bins were fit simultaneously, forcing the resonance parameters to be the same across the t' bins. This t'-resolved approach leads to a much better disentanglement of the resonant and the non-resonant contributions, which have in general different dependences on t'. For $t' \lesssim 0.5\,\mathrm{GeV}^2$, the fit finds that the 1^{-+} intensity is almost saturated by the non-resonant component (green curve in Fig. 182(left)) with only a small $\pi_1(1600)$ contribution (blue curve). With increasing t' the strength of the non-resonant component decreases relative to that of the $\pi_1(1600)$, so that for $t' \gtrsim 0.5\,\mathrm{GeV}^2$ the $\pi_1(1600)$ becomes the dominant component (see Fig. 182(center)).

Applying the 21- and 36-wave PWA models from the two analyses of BNL E852 data [2445] to the COMPASS data yields results consistent with those reported in Refs. [2442,2443,2453] confirming the observations by Dzierba et al. that the 21-wave model produces an artificial peak at 1.6 GeV in the 1^{-+} waves for natural as well as unnatural-parity exchange due to leakage from the $\pi_2(1670)$. This explains the puzzling observation of a $\pi_1(1600) \to \rho(770)\pi$ signal in unnatural-parity exchange by the BNL E852 experiment [2442,2443] as an artifact caused by leakage. In addition, the t'-resolved analysis of the COMPASS data shows that for $t' \lesssim 0.5\,\mathrm{GeV}^2$ the $\pi_1(1600)$ signal is masked by the dominant non-resonant contribution. This explains why Dzierba et al., who considered only the range $t' < 0.53\,\mathrm{GeV}^2$, reported a non-observation of the $\pi_1(1600)$. However, in the kinematic range $t' \gtrsim 0.5\,\mathrm{GeV}^2$ COMPASS observes a clear $\pi_1(1600) \to \rho(770)\pi$ signal and a $\pi_1(1600)$ resonance is indeed required to explain the COMPASS data. This is demonstrated by the dashed red curve in Fig. 182, which represents the result of a resonance-model fit, where the 1^{-+} amplitude was described using

[87] The distribution also exhibits a narrow peak at about 1.1 GeV, which, however, has no associated phase motion and depends on the PWA model. According to Refs. [2319,2445] this peak is likely an artifact induced by imperfections of the analysis method.

only the non-resonant component. At low t', this model is able to describe the data fairly well (see Fig. 182(left)), but clearly fails at high t' (see Fig. 182(center) and (right)). The t'-resolved COMPASS results in Refs. [2319,2444,2445] therefore establish unambiguously the $\rho(770)\pi$ decay mode of the $\pi_1(1600)$ and in addition resolve a long-standing controversy by showing that the data of previous experiments are indeed consistent and that the BNL E852 puzzle was caused by a too restricted PWA model on the one hand [2442,2443] and a too restricted t' range on the other hand [2453].

Another big step towards a better understanding of the π_1 states was the coupled-channel analysis of the $\eta\pi$ and $\eta'\pi$ P- and D-wave amplitudes measured by the COMPASS experiment [2457], which was performed by the JPAC collaboration [2324]. Using a unitary model based on S-matrix principles they find in the D-wave amplitudes two resonance poles, the $a_2(1320)$ and the $a_2(1700)$ and in the P-wave amplitudes a single resonance pole. The parameters of the P-wave resonance pole are $m_0 = (1564^{+24}_{-86})$ MeV and $\Gamma_0 = (492^{+\,54}_{-102})$ MeV, consistent with the $\pi_1(1600)$. Apart from determining the $\pi_1(1600)$ pole position for the first time using an analytic and unitary model, this result is in so far remarkable as only a single resonance pole is required to simultaneously describe the $\eta\pi$ and the $\eta'\pi$ P-wave amplitudes despite their rather different intensity distributions (see green and blue points and curves in Fig. 183). This is in contrast to most previous analyses, which considered two different resonance components in their models: a $\pi_1(1400)$ to describe the broad peak at 1.4 GeV in the $\eta\pi$ P-wave intensity and a $\pi_1(1600)$ to describe the narrower peak at 1.6 GeV in the $\eta'\pi$ P-wave intensity. It is interesting to note that in the COMPASS data the latter peak is nearly identical to the one observed in the 1^{-+} intensity in the high-t' region of the $\pi^-\pi^-\pi^+$ data (cf. blue and red points and curves in Fig. 183). Since the COMPASS partial-wave data are consistent with previous experiments, the JPAC analysis raises serious doubts about the existence of the $\pi_1(1400)$ as a separate resonance. Recently, the JPAC results were confirmed by Kopf et al., who performed a coupled-channel analysis that in addition to the COMPASS $\eta\pi$ and $\eta'\pi$ P- and D-wave amplitudes also includes Crystal Barrel data on $\bar{p}p \to \pi^0\pi^0\eta$, $\pi^0\eta\eta$, and $K^+K^-\pi^0$ as well as $\pi\pi$ scattering data [2458].

Both coupled-channel analyses favor a much simpler and more plausible picture with only one π_1 state below 2 GeV, the $\pi_1(1600)$, decaying into (at least) $\eta\pi$, $\eta'\pi$, $\rho(770)\pi$, $f_1(1285)\pi$, and $b_1(1235)\pi$. This scenario resolves the long-standing puzzle of two spin-exotic states having peculiar decay modes and lying unexpectedly close to each other. If interpreted in terms of hybrid states, this would also remove the discrepancy with lattice QCD and most model calculations, which predict the lightest hybrid state to have a mass substantially higher than that of the $\pi_1(1400)$ (see Sect. 8.3.2).

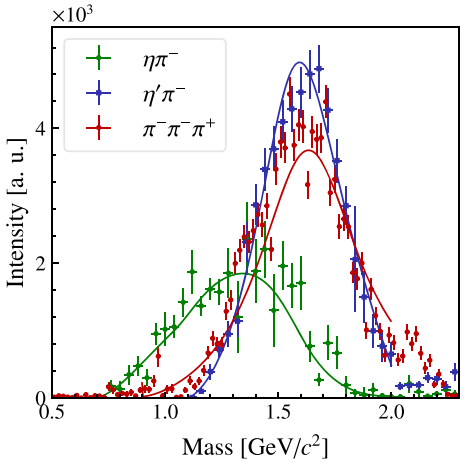


Fig. 183 Intensity distributions of spin-exotic waves with $J^{PC} = 1^{-+}$ from COMPASS data. (Green points) $\eta\pi$ P-wave, (blue points) $\eta'\pi$ P-wave, both for $0.1 < t' < 1.0\,\text{GeV}^2$. (Red points) $\rho(770)\pi$ P-wave for $0.449 < t' < 0.724\,\text{GeV}^2$. The curves represent the results of the resonance-model fits from Refs. [2319,2324]. (Taken from Fig. 2 in Ref. [2456])

Up to now only isovector spin-exotic states were observed in the light-meson sector. However, models and lattice QCD predict that $\text{SU(3)}_{\text{flavor}}$ partner states of the π_1, i.e. η_1 and η_1' as well as K^* states,[88] should exist. In order to establish exotic resonances it is therefore important to find these states. A first sign that they indeed exist is the very recent first observation of a spin-exotic isoscalar $\eta_1(1855)$ state in the $\eta\eta'$ decay channel produced in $J/\psi \to \gamma\eta\eta'$[89] reported by the BESIII experiment [2461,2462]. The challenge is now to confirm this state in other experiments.

8.3.5 *Summary and outlook*

The dust of more than three decades of research on spin-exotic light mesons is starting to settle. For a long time, the experimental data were confusing leading to contradictory conclusions on the existence and properties of π_1 mesons. Recently, high-precision data and more advanced theory models helped to resolve many of these puzzles and a more coherent picture seems to be emerging, where instead of two low-lying states, $\pi_1(1400)$ and $\pi_1(1600)$, with hard

[88] As kaons are neither eigenstates of C nor of G parity, there are no spin-exotic kaon states. Hence, the exotic K^* states can be identified only as supernumerary states and via their couplings.

[89] This is an example for a radiative J/ψ decay. Such decays are "gluon-rich" processes because in lowest order the $c\bar{c}$ pair in the J/ψ annihilates into the measured photon and a pair of gluons that hadronize into the measured final state, here the $\eta\eta'$ system. The production of mesons with explicit gluonic degrees of freedom, i.e. hybrids and glueballs, is expected to be enhanced in these decays.

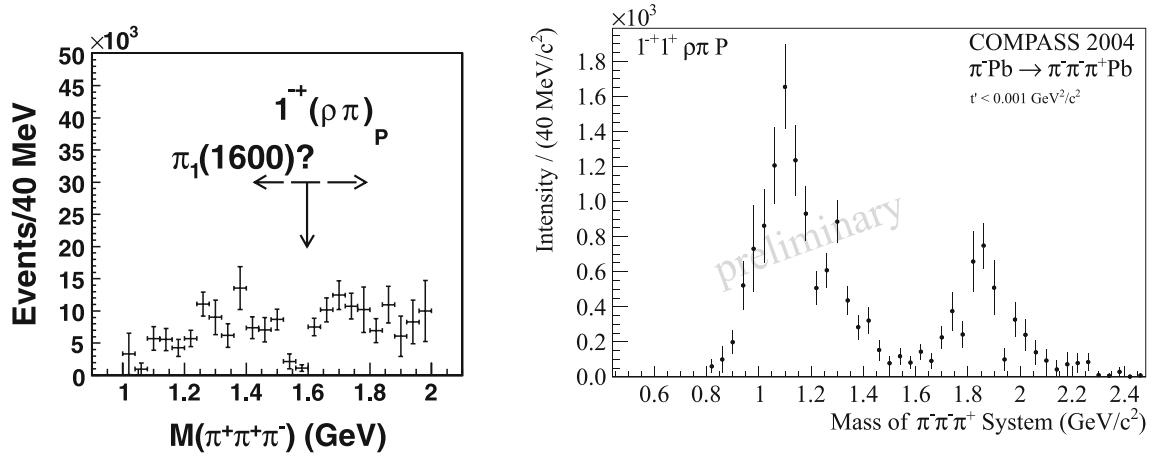


Fig. 184 Intensity distributions of the $\rho(770)\pi$ P-wave with spin-exotic $J^{PC} = 1^{-+}$ quantum numbers produced in $\gamma + \pi^{\pm} \to \pi^{\pm}\pi^{-}\pi^{+}$ reactions. (Left) Result from the CLAS experiment [2459], where the process is embedded into $\gamma + p \to \pi^{+}\pi^{-}\pi^{+} + (n)_{\text{miss}}$. (Right) Result from the COMPASS experiment [2413,2460], where the process is embedded into $\pi^{-} + \text{Pb} \to \pi^{-}\pi^{-}\pi^{+} + \text{Pb}$. (Taken from Fig. 5(d) of Ref. [2459] and Fig. 7(a) of Ref. [2413])

to explain properties only the $\pi_1(1600)$ remains. However, there are at least two puzzles to be solved. The first is the unexpected production of the $\pi_1(1600)$ in unnatural-parity exchange claimed by the BNL E852 experiment in the $b_1(1235)\pi$ channel [2322]. This can be clarified by the COMPASS experiment using data on the same reaction at higher energy. The second remaining puzzle is the seeming non-observation of the $\pi_1(1600)$ in photon-induced reactions. Since the $\pi_1(1600)$ is observed to decay into $\rho(770)\pi$, it should couple to $\gamma\pi$ via vector-meson dominance. However, in the $\gamma + \pi^{\pm} \to \pi^{\pm}\pi^{-}\pi^{+}$ reaction studied by the CLAS and the COMPASS experiments[90] nearly vanishing intensity was observed in the $J^{PC} = 1^{-+}$ wave in the mass range where a $\pi_1(1600)$ signal would be expected (see Fig. 184). The nearly vanishing intensity could be the result of a destructive interference of the $\pi_1(1600)$ amplitude with the one of non-resonant contributions. However, no resonance-model fits have been performed yet to test this hypothesis. In the future, much more precise photoproduction data from the GlueX experiment at JLab will help to clarify the situation.

Having established that spin-exotic 1^{-+} light-meson states do exist is, of course, only the starting point. The next goal is to study their properties in detail, in particular their couplings, by measuring them in various production and decay modes. Another goal is to find their excitations. A first step in this direction would be the confirmation of the $\pi_1(2015)$ signal in the $f_1(1285)\pi$ and $b_1(1235)\pi$ decay channels. In addition, it is important to search for the exotic $\text{SU}(3)_{\text{flavor}}$ partner states of the π_1. Here, the result by the BESIII experiment of a possible observation of an $\eta_1(1855)$ state could be a breakthrough. Last but not least, the search for states with other spin-exotic J^{PC} quantum numbers such as 0^{+-} and 2^{+-} continues. These searches will also yield a more complete picture of the spectrum of states with ordinary quantum numbers, which not only helps to identify supernumerary states, but is also an important input to theory in order to improve our understanding of the non-perturbative regime of QCD.

In turn, the analysis of the extremely high-precision data from running and upcoming experiments requires more advanced theoretical models and in particular a more accurate understanding of the dynamics of hadrons. Close collaboration of theorists and experimentalists will help us to formulate, test, and apply detailed models for production reactions and for the interactions of final-state hadrons in order to overcome limitations of the currently available analysis approaches. Together with refined statistical tools and novel approaches such as Machine Learning, this will enable us to leverage the full potential of the data.

8.4 Glueballs, a fulfilled promise of QCD?

Eberhard Klempt

8.4.1 Introduction

At the *Workshop on QCD: 20 Years Later* [78] held in 1992 in Aachen, Heusch [2463] reported on searches for glueballs, gluonium, or glue states as Fritzsch and Gell-Mann [35,55]

[90] CLAS measured the photoproduction reaction $\gamma + p \to \pi^{+}\pi^{-}\pi^{+} + (n)_{\text{miss}}$, where a pion is exchanged between the target and the beam photon producing the 3π final state. In COMPASS data, the $\gamma\pi \to 3\pi$ reaction is embedded into the reaction $\pi^{-} + \text{Pb} \to \pi^{-}\pi^{-}\pi^{+} + \text{Pb}$, which was measured at very low squared four-momentum transfer, where the beam pion predominantly scatters off quasi-real photons from the Coulomb field of the Pb target nucleus.

had called this new form of matter. Glueballs are colorless bound states of gluons and should exist when their newly proposed quark–gluon field theory yields a correct description of the strong interaction. The title of Heusch's talk *Gluonium: An unfulfilled promise of QCD?* expressed the disappointment of a glueball hunter: At that time there was some – rather weak – evidence for glueball candidates but there was no convincing case. In 1973, the e^+e^- storage ring SPEAR at the Stanford Linear Accelerator Center had come into operation and one year later, the J/ψ resonance was discovered [91] – this was the very first SPEAR publication on physics. The J/ψ resonance and its radiative decay became and still is the prime reaction for glueball searches.

One of the first glueball candidates was the $\iota(1440)$ [2464,2465]. The name ι stood for the "number one" of all glueballs to be discovered. It was observed as very strong signal with pseudoscalar quantum numbers in the reaction $J/\psi \to \gamma K\bar{K}\pi$. Its mass was not too far from the bag-model prediction (1290 MeV) [782]. Now the $\iota(1440)$ is supposed to be split into two states, $\eta(1405)$ and $\eta(1475)$, where the lower-mass meson is still discussed as glueball candidate even though its mass is incompatible with lattice gauge calculations. They find the mass of the pseudoscalar glueball above 2 GeV.

A second candidate was a resonance called $\Theta(1640)$ [2466,2467]. It was seen in the reaction $J/\psi \to \gamma\eta\eta$ and confirmed – as $G(1590)$ – by the GAMS collaboration in $\pi^- p \to \eta\eta n$ [2468]. Later, its quantum numbers shifted from $J^{PC} = 2^{++}$ to 0^{++}, and its mass changed to 1710 MeV. This resonance still plays an important role in the glueball discussion.

A third candidate, or better three candidates, were observed in the OZI rule violating process $\pi^- p \to \phi\phi n$ [2332, 2469]. Three $\phi\phi$ resonances at 2050, 2300 and 2350 MeV were reported. I remember Armenteros saying: *When you have found one glueball, you have made a discovery. When you find three, you have a problem.* Now I believe that this was a very early manifestation of the tensor glueball.

The situation was not that easy at that time as described here. Nearly for each observation, there were contradicting facts, and Heusch concluded his talk at the QCD workshop with the statement: *there is no smoking-gun candidate for gluonium* $\cdots$. At this workshop, I had the honor to present the results of the Crystal Barrel experiment at LEAR and to report the discovery of two new scalar mesons, $f_0(1370)$ and $f_0(1500)$, and I was convinced, Heusch was wrong: $f_0(1500)$ was the glueball! And I turned down my internal critical voice which told me that in my understanding of $\bar{p}N$ annihilation, this process is not particularly suited to produce glueballs [2470,2471]. Our glueball $f_0(1500)$ was not seen in radiative J/ψ decays where a glueball should stick out like a tower in the landscape. The $f_0(1500)$ as scalar glueball? That could not be the full truth!

8.4.2 QCD predictions

Glueball masses

First estimates of the masses of glueballs were based on bag models. The color-carrying gluon fields were required to vanish on the surface of the bag. Transverse electric and transverse magnetic gluons were introduced populating the bag. The lowest excitation modes were predicted to have quantum numbers $J^{PC} = 0^{++}$ and 2^{++} and to be degenerate in mass with $M = 960$ MeV [782,2472]. A very early review can be found in Ref. [2473].

The bag model is obsolete nowadays. Most reliable are presumably simulations of QCD an a lattice (see Sect. 4 and Ref. [2474] for an introduction). In lattice gauge theory, the spacetime is rotated into an Euclidean space by the transformation $t \to it$ and then discretized into a lattice with sites separated by a distance in space and time. The gauge fields are defined as links between neighboring lattice points, closed loops of the link variables (Wilson loops) allow for the calculation of the action density. Technically, gluons on a space-time lattice struggle against large vacuum fluctuations of the correlation functions of their operators, the signal-to-noise ratio falls extremely rapidly as the separation between the source and sink is increased. These difficulties can be overcome by anisotropic space-times with coarser space and narrow time intervals [2475,2476]. Fermion fields are defined at lattice sites. Different techniques have been developed to include fermions in lattice calculations [2477]. The effect of see quarks on glueball masses seems to be small [2478].

Recently, a number of different approaches were chosen to approximate QCD by a model that is solvable analytically. Szczepaniak and Swanson [2479] constructed a quasiparticle gluon basis for a QCD Hamiltonian in Coulomb gauge that was solved analytically. A full glueball spectrum was calculated with no free parameter. The authors of Ref. [2480] constructed relativistic two- and three-gluon glueball currents and applied them to perform QCD sum rule analyses of the glueball spectrum. The Gießen group calculated masses of ground and excited glueball states using a Yang–Mills theory and a functional approach based on a truncation of Dyson–Schwinger equations and a set of Bethe–Salpeter equations derived from a three-particle-irreducible effective action [2481,2482].

AdS/QCD relies on a correspondence between a five dimensional classical theory with an AdS metric and a supersymmetric conformal quantum field theory in four dimensions. In the bottom-up approach, models with appropriate operators are constructed in the classical AdS theory with the aim of resembling QCD as much as possible. Confinement is generated by a hard wall cutting off AdS space in the infrared region, or spacetime is capped off smoothly by a soft wall to break the conformal invariance. Rinaldi and Vento [1104]

Table 31 Masses of low-mass glueballs, in units of MeV. Lattice QCD results are taken from Refs. [2475,2477] (quenched) and Ref. [2478] (unquenched). Szczepaniak and Swanson [2479] construct of a quasi-particle gluon basis for a QCD Hamiltonian. Results from QCD sum rule results are given in Ref. [2480], from using Dyson–Schwinger equations in [2481,2482], and from a graviton-soft-wall model in Ref. [1104]

Glueball	Ref. [2475]	Ref. [2477]	Ref. [2478]	Ref. [2479]	Ref. [2480]	Ref. [2481]	Ref. [1104]
$\|0^{++}\rangle$	$1710 \pm 50 \pm 80$	1653 ± 26	1795 ± 60	1980	1780^{+140}_{-170}	1850 ± 130	1920
$\|2^{++}\rangle$	$2390 \pm 30 \pm 120$	2376 ± 32	2620 ± 50	2420	1860^{+140}_{-170}	2610 ± 180	2371
$\|0^{-+}\rangle$	$2560 \pm 40 \pm 120$	2561 ± 40	–	2220	2170 ± 110	2580 ± 180	

calculated the glueball mass spectrum within AdS/QCD. The results on glueball masses are summarized in Table 31.

The width of glueballs

Glueballs are often assumed to be narrow. ϕ decays into $\rho\pi$ are suppressed since the primary $s\bar{s}$ pair needs to annihilate and a new $q\bar{q}$ pair needs to be created. In glueball decays, there is no pair to be annihilated but a $q\bar{q}$ pair needs to be created. If the OZI rule suppresses the decay by a factor 10 to 100, we might expect the width of glueballs to be suppressed by a factor 3 to 10. Assuming a "normal" width of 150 MeV, a glueball at 1600 MeV could have a width of 15 to 50 MeV. This argument is supported by arguments based on the $1/N_c$ expansion (see, e.g., Ref. [2390]).

Narison applied QCD sum rules [2483]. Assuming a mass of 1600 MeV, he calculated the 4π width of the scalar glueball to 60 to 138 MeV while the partial decay width of the tensor glueball at 2 GeV to pseudoscalar mesons should be less than 155 MeV. Calculations on the lattice gave a partial decay width for decays into pseudoscalar mesons of 108 ± 29 MeV for a scalar glueball mass of 1700 MeV [2484]. In a semi-phenomenological model, Burakovsky and Page find that the width of the scalar glueball (at 1700 MeV) should exceed 250 to 390 MeV. A flux tube model predicted the mass of the glueball of lowest mass to 1680 MeV and its width to 300 MeV [2485]. In a field theoretical approach with an effective Coulomb gauge the glueball width was estimated to 100 MeV [2486].

Radiative yields

The study of radiative decays of the J/ψ meson is the prime path to search for glueballs with masses of less than ~2500 MeV.

Gui et al. [2487] calculated the yield of a scalar glueball having a mass of 1710 MeV on lattice and found

$$BR_{J/\psi\to\gamma G_{0^{++}}}(TH) = (3.8 \pm 0.9) \times 10^{-3}. \quad (8.28)$$

For higher glueball masses the yield increases.

Narison gave a mass dependent formula derived from sum rules. For a mass of 1865 MeV, a yield of about 10^{-3} is predicted [2483].

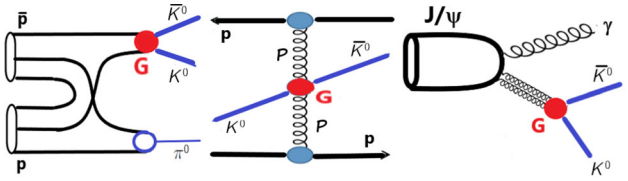


Fig. 185 Reactions most relevant for glueball searches. Left: $\bar{p}p$ annihilation; middle: Pomeron-Pomeron fusion; right: radiative J/ψ decays. The glueball is supposed to decay into $K^0\bar{K}^0$

The tensor glueball is expected [2488] to be observed with a branching ratio

$$BR_{J/\psi\to\gamma G_{2^{++}}}(TH) = (11 \pm 2) \times 10^{-3}. \quad (8.29)$$

Production of the pseudoscalar glueball seems to be considerably smaller. For a mass of 2395 (or 2560) MeV, the authors of Ref. [2489] find

$$BR_{J/\psi\to\gamma G_{0^{-+}}}(TH) = (0.231 \pm 0.080) \times 10^{-3} \text{ or } = (0.107 \pm 0.037) \times 10^{-3}. \quad (8.30)$$

These are very significant yields, and the glueballs must be found provided they can be identified convincingly as glueballs amidst their $q\bar{q}$ companions.

8.4.3 How to identify a glueball

Figure 185 shows the prime reactions in which glueballs have been searched for.

$N\bar{N}$ annihilation

A decisive step forward in the search for glueballs was the discovery of two scalar isoscalar states in $\bar{p}p$ annihilation at rest. With the large statistics available at the Low Energy Anitiproton Ring (LEAR) at CERN, $f_0(1370)$ and $f_0(1500)$ were identified in several final states. A large fraction of the data taken at LEAR is still used jointly with data on radiative J/ψ decays in a coupled-channel analysis. Glueballs decay via $q\bar{q}$ pair creation. Hence they can be produced via $q\bar{q}$ annihilation. Meson production in $\bar{p}p$ annihilation was studied by the ASTERIX, OBELIX and Crystal Barrel experiments at LEAR and is a major objective of the PANDA collaboration at the GSI.

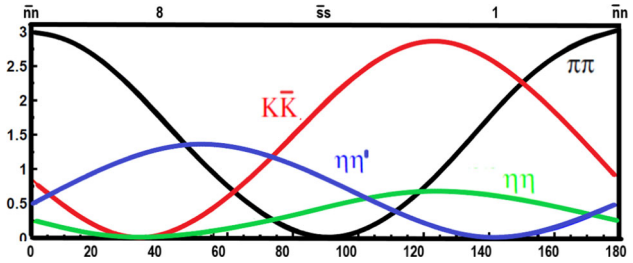


Fig. 186 Decay probabilities of mesons for decays into two pseudoscalar mesons as a function of the scalar mixing angle [2490]

Central production

In central production, two hadrons (e.g. two protons) scatter in forward direction via the exchange of Pomerons. Pomerons are supposed to be glue-rich, hence glueballs can be formed in Pomeron-Pomeron fusion. This process was studied extensively at CERN by the WA76 and WA102 collaborations and is now investigated with the STAR detector at BNL. In the WA102 experiment, $f_0(1370)$ and $f_0(1500)$ were confirmed and $f_0(1710)$ was added to the number of scalar resonances.

Radiative J/ψ decays

In radiative J/ψ decays, the primary $c\bar{c}$ pair converts into two gluons and a photon. The two gluons are mainly produced in S-wave, the two gluons can form scalar and tensor glueballs which should be produced abundantly. The large statistics available from BESIII at Beijing makes this reaction the most favorable one for glueball searches. Radiative decays of heavy mesons is the only process for which glueball yields have been calculated. The data will be discussed below in more detail.

Decay analysis

The decay of mesons into two pseudoscalar mesons is governed by $SU(3)_F$. In a meson nonet, there are two isoscalar mesons, one lower in mass the other one higher, which both contain a $n\bar{n} = (u\bar{u} + d\bar{d})/\sqrt{2}$ and a $s\bar{s}$ component and are mixed with the mixing angle φ. Figure 186 shows the $SU(3)_F$ squared matrix elements for meson decays into two pseudoscalar mesons as a function of the scalar mixing angle.

$$\begin{pmatrix} f^H \\ f^L \end{pmatrix} = \begin{pmatrix} \cos\varphi^s & -\sin\varphi^s \\ \sin\varphi^s & \cos\varphi^s \end{pmatrix} \begin{pmatrix} |n\bar{n}\rangle \\ |s\bar{s}\rangle \end{pmatrix} \tag{6}$$

Supernumery

The three scalar isoscalar mesons $f_0(1370)$, $f_0(1500)$ and $f_0(1710)$ played an important role in the glueball discussion. Amsler and Close [2491,2492] suggested to interpret these three states as the result of mixing of the two expected isoscalar states with the scalar glueball.

$$\begin{pmatrix} f_0(1370) \\ f_0(1500) \\ f_0(1710) \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix} \begin{pmatrix} |n\bar{n}\rangle \\ |s\bar{s}\rangle \\ |gg\rangle \end{pmatrix} \tag{8.31}$$

These papers led to a large number of follow-up papers, references can be found in Ref. [2490]. In all these papers, these three mesons contain the full glueball, $\sum_j x_{ij}^2 = 1$ is imposed. Note that the squared mass difference between $f_0(1370)$ and $f_0(1710)$ is slightly above 1 GeV^2, the $f_0(1710)$ could also be a radial excitation (and is interpreted as radial excitation below).

Conclusions

Identifying a glueball is a difficult task. The main argument in favor of a glueball interpretation is an anomalously large production rate in J/ψ decays. It turns out that scalar mesons are organized like pseudoscalar mesons, into mainly singlet and mainly octet mesons. A large production rate of a mainly-octet scalar isoscalar meson in radiative J/ψ decays directly points to a significant glueball content in its wave function. A second argument relies an meson decays into pseudoscalar mesons. In presence of a glueball, a pair of mesons assigned to the same multiplet should have a decay pattern that is incompatible with a $q\bar{q}$ interpretation for any mixing angle. Supernumery is a weak argument. It requires a full understanding of the regular excitation spectrum. Further studies are required to learn if central production is gluon-rich. The large production rates of $f_0(1500)$, $f_0(1710)$ and $f_0(2100)$ in $\bar{p}p$ annihilation at collision energies above 3 GeV encourages glueball searches at the FAIR facility with the PANDA detector (see Sect. 14.5).

8.4.4 *Evidence for glueballs from coupled-channel analysis*

We have performed a coupled-channel partial wave analysis of radiative J/ψ decays into $\pi^0\pi^0$, $K_s^0K_s^0$, $\eta\eta$, and $\omega\phi$, constrained by the CERN-Munich data on πN scattering, data from the GAMS collaboration at CERN, data from BNL on $\pi\pi \to K_s^0K_s^0$, and 15 Dalitz plots on $\bar{p}p$ annihilation at rest from LEAR. Data on K_{e4} decays constrain the low-energy region. Fitting details and references to the data can be found in Ref. [2493]. A similar analysis has been carried out by Rodas et al. [2494]. This will be discussed in Sect. 8.4.9.

Figure 187 shows the data on radiative J/ψ decays into $\pi^0\pi^0$, $K_s^0K_s^0$ and the fit. Ten scalar isoscalar resonances were included in the fit. Oller [2406] has shown that $f_0(500)$ is singlet-like, the $f_0(980)$ octet-like (see also [2495]). The $f_0(1500)$ is seen in Fig. 187 as a dip. This pattern was reproduced in Ref. [2493] assuming that $f_0(1370)$ is a singlet state and $f_0(1500)$ an octet state. Hence we assumed that the ten mesons can be divided into two series of states, mainly-singlet states with lower masses and mainly-octet states with higher masses.

In a (M^2, n) plot, the masses of singlet and octet states follow two linear trajectories (see Fig. 188). Remarkably, the slope (1.1 GeV^{-2}) is close to the slope of standard Regge trajectories. The separation between the two trajectories is

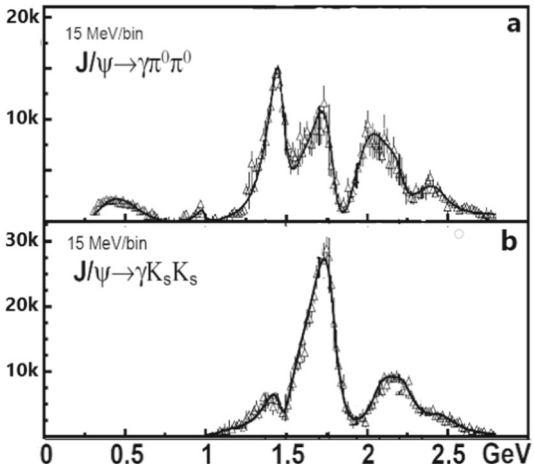


Fig. 187 The squared S-wave transition amplitudes for $J/\psi \to \pi^0\pi^0$ (**a**) and $J/\psi \to K_s^0 K_s^0$ (**b**). The data points are from an energy-independent partial-wave analysis [2496,2497], the curve represents our fit [2493]

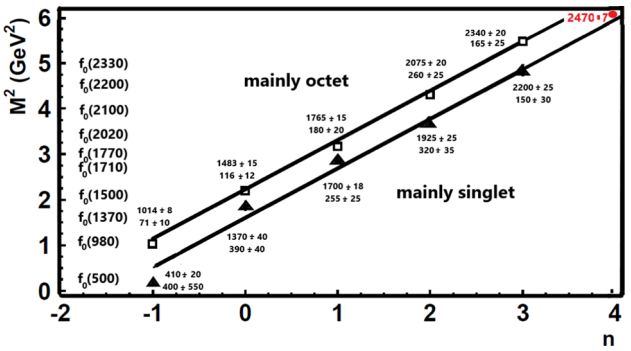


Fig. 188 M^2, n trajectories for mainly-singlet and mainly-octet scalar isoscalar resonances. The red dot at high masses represents a scalar state from $J/\psi \to \gamma\eta'\eta'$ [2498]. Adapted from Ref. [2493]

given by the mass square difference between η' and η-meson as suggested by instanton-induced interactions [2499]. The figure includes a meson reported by the BESIII collaboration studying $J/\psi \to \gamma\eta'\eta'$ [2498]. As $\eta'\eta'$ resonance, $f_0(2480)$ is very likely a SU(3) singlet state. Indeed, its mass is compatible with the "mainly-singlet" trajectory. The figure gives the pole positions of the eleven resonances as small inserts.

8.4.5 *Scalar multiplets*

The nonet of scalar mesons

$a_0(980)$, $K_0^*(700)$, $f_0(980)$, $f_0(500)$

of lowest mass were discussed by Pelaez (Sect. 8.2). In our interpretation, this nonet is followed by

$a_0(1450)$, $K_0^*(1430)$, $f_0(1500)$, $f_0(1370)$.

All known decay modes are compatible with SU(3) predictions except $K_0^*(1430) \to K\pi$: the SU(3) fit requires this

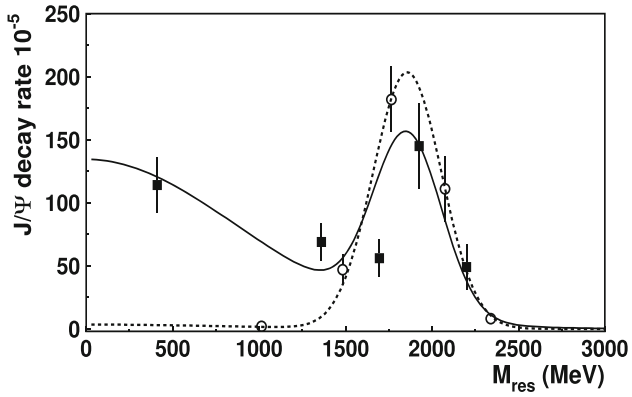


Fig. 189 Yield of scalar isoscalar mesons in radiative J/ψ decays into mainly-octet (open circles) and mainly-singlet mesons (full squares) as a function of their mass [2493]

resonance to have a large inelasticity. The large $K_0^*(1430) \to K\pi$ is incompatible with the known $a_0(1450)$ decay modes, unrelated of course with the assignment of the isoscalar mesons to this nonet.

The recently discovered $a_0(1700)$ [2348–2351] can be accommodated in a multiplet

$a_0(1700)$, $K_0^*(???)$, $f_0(1780)$, $f_0(1710)$,

where the $K_0^*(???)$, expected at about 1680 MeV, is missing.

The SU(3) fit to the next nonet

$a_0(1950)$, $K_0^*(1950)$, $f_0(2100)$, $f_0(2020)$

requires a larger $f_0(2100) \to \eta\eta$ coupling which is not observed. Too little is known about the $a_0(1950)$, $K_0^*(1950)$ decays modes.

It is obvious, that the assignments of the observed scalar mesons to multiplets is very tempting but tensions between an SU(3) analysis of the decay mode and experimental values remain.

8.4.6 *The yields of scalar mesons in radiative J/ψ decays*

The total yields of scalar mesons in radiative J/ψ decays – including decay modes not reported by the BESIII collaboration – was determined from the coupled-channel analysis [2493] that included also other data. The yield of mainly-octet and mainly-singlet mesons as a function of their mass is shown in Fig. 189. Mainly-octet mesons should not be produced (or at most weakly) in radiative J/ψ decays. However, they are produced abundantly, in a limited mass range centered at about 1865 MeV. Mainly-singlet mesons are produced over the full mass range but show a peak structure at the same mass. This enhancement must be due to the scalar glueball mixing into the wave functions of scalar mainly-octet and mainly-singlet mesons. A Breit–Wigner fit to these

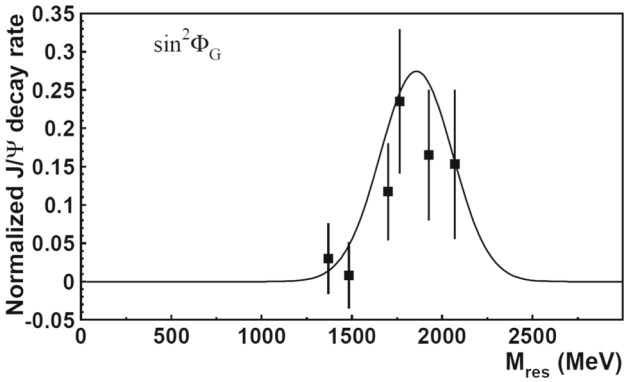


Fig. 190 The glueball content of scalar mesons. Black squares: $\sin^2\varphi_{\rm n}^{\rm s}$, solid curve: Breit–Wigner resonance with area 1 [2490]

distributions gives mass and width

$$M_G = (1865 \pm 25^{+10}_{-30})\,\text{MeV} \quad \Gamma_G = (370 \pm 50^{+30}_{-20})\,\text{MeV}, \tag{8.32}$$

and the (observed) yield is determined to

$$Y_{J/\psi\to\gamma G} = (5.8 \pm 1.0)\,10^{-3}. \tag{8.33}$$

8.4.7 Meson–glueball mixing

Earlier attempts to identify the glueball have in common that the full glueball is distributed among the three states $f_0(1370)$, $f_0(1500)$ and $f_0(1710)$. Inspecting Fig. 187, this seems not to be the case: Above 1 GeV, four peaks with three valleys are seen, and there is no reason why one particular region should be more gluish than the other ones. The yield of scalar mesons sees the glueball contribution distributed over several resonances.

We did not impose that the full glueball should be seen in these three states nor that we must see the full glueball at all. We fitted the decay modes of pairs of scalar mesons, one mainly-singlet one mainly-octet, and allowed for a glueball component [2490].

$$\begin{aligned} f_0^{\rm nH}(xxx) &= \left(n\bar{n}\cos\varphi_{\rm n}^{\rm s} - s\bar{s}\sin\varphi_{\rm n}^{\rm s}\right)\cos\phi_{\rm nH}^{G} + G\sin\phi_{\rm nH}^{G} \\ f_0^{\rm nL}(xxx) &= \left(n\bar{n}\sin\varphi_{\rm n}^{\rm s} + s\bar{s}\cos\varphi_{\rm n}^{\rm s}\right)\cos\phi_{\rm nL}^{G} + G\sin\phi_{\rm nL}^{G} \end{aligned} \tag{8.34}$$

$\varphi_{\rm n}^{\rm s}$ is the scalar mixing angle, $\phi_{\rm nH}^{G}$ and $\phi_{\rm nL}^{G}$ are the meson–glueball mixing angles of the high-mass state H and of the low-mass state L in the nth nonet. The fractional glueball content of a meson is given by $\sin^2\phi_{\rm nH}^{G}$ or $\sin^2\phi_{\rm nL}^{G}$.

With this mixing scheme and the SU(3) coupling constant (see Fig. 186), we have fitted the meson decay modes and have thus determined the glueball content of the eight high-mass scalar mesons. Figure 190 shows the glueball fraction in the scalar mesons.

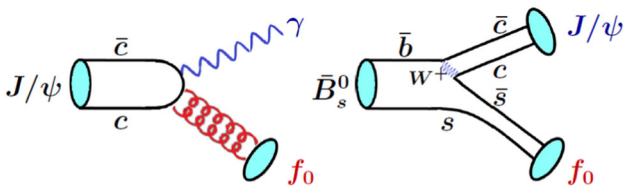


Fig. 191 In radiative J/ψ decays two gluons, in $\bar{B}_s^0 \to J/\psi + s\bar{s}$, a $s\bar{s}$ pair may convert into a scalar meson

The glueball fractions derived from the decay analysis of pairs of scalar mesons add up to a sum that is compatible with 1. The distribution of the glueball fraction in Fig. 190 is identical to the distribution of yields in Fig. 189. This is a remarkable confirmation that the scalar glueball of lowest mass has been identified and has mass and width as given in Eq. (8.32) and a yield as given in Eq. (8.33).

8.4.8 Comparison with LHCb data

Most striking is the mountain landscape above 1500 MeV in the data on radiative J/ψ decays. In these decays a $c\bar{c}$ pair converts into gluons which hadronize (see Fig. 191, left). The huge peak in the $K\bar{K}$ mass spectrum at 1750 MeV and the smaller one at 2100 MeV decay are produced with two gluons in the initial state. This is to be contrasted with data on B_s^0 and $\bar{B}_s^0$ decays into $J/\psi + \pi^+\pi^-$ [2500] and $K\bar{K}$ [2501]. In this reaction, a primary $s\bar{s}$ pair – recoiling against the J/ψ – converts into the final state mesons (see Fig. 191, right). We have included the spherical harmonic moments in the coupled channel analysis that describes the radiative J/ψ decays [2502]. High-mass scalar mesons are only weakly produced in B_s^0 decays with $s\bar{s}$ in the initial state. The strong peak in the $K\bar{K}$ invariant mass at 1750 MeV in Fig. 187 is nearly absent in $B_s^0 \to J/\psi\, K\bar{K}$!

Figure 192 shows the ratio of the decay frequencies of $J/\psi \to \gamma\, f_0$ and $B_s^0 \to J/\psi\, f_0$ with f_0 decaying into $\pi\pi$ or $K\bar{K}$. The $f_0(980)$ is likely a mainly-octet state, little produced in radiative J/ψ decays but strongly with $s\bar{s}$ in the initial state. On the contrary, $f_0(1770)$ is seen as strong peak in radiative J/ψ but very weakly only in B_s^0 decays. The uncertainties are large, but the ratio of the decay frequencies is fully compatible with the shape of the glueball derived above.

This is highly remarkable: the two gluons in the initial state must be responsible for the production of resonances that decay strongly into $K\bar{K}$ but are nearly absent when $s\bar{s}$ pairs are in the initial state. Also, there is a rich structure in the $\pi\pi$ mass spectrum produced in radiative J/ψ decays but little activity only when the initial state is an $s\bar{s}$ pair: The rich structure stems from gluon–gluon dynamics. Similar conclusions can be drawn [2495] from a comparison of the invariant mass distributions from radiative J/ψ decays with the pion

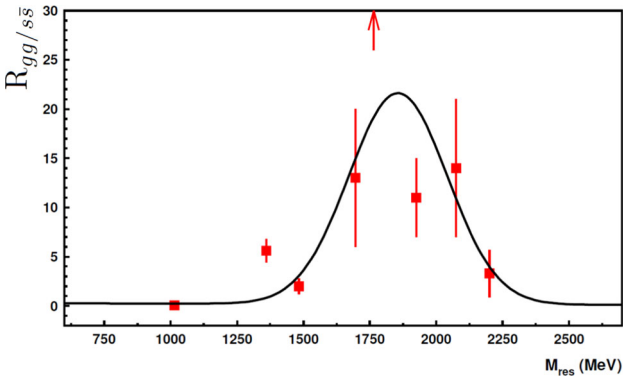


Fig. 192 The ratio $R_{gg/s\bar{s}}$ of the frequencies for $J/\psi \to \gamma f_0$ and $B_s^0 \to J/\psi f_0$ with f_0 decaying into $\pi\pi$ or $K\bar{K}$

Table 32 Comparison of the pole positions of the mainly-octet mesons from [2493] and the pole positions from [2494]. Masses are given in MeV

[2493]	[2494]
$(1483 \pm 15) - i(116 \pm 12)/2$	$(1450 \pm 10) - i(106 \pm 16)/2$
$(1765 \pm 15) - i(180 \pm 20)/2$	$(1769 \pm 8) - i(156 \pm 12)/2$
$(2075 \pm 20) - i(260 \pm 25)/2$	$(2038 \pm 48) - i(312 \pm 82)/2$
$(2340 \pm 20) - i(250 \pm 20)/2$	$(2419 \pm 64) - i(274 \pm 94)/2$

and kaon form factors [2503]. Their square is proportional to the cross sections. The $f_0(980)$ resonance dominates both formfactors but is nearly absent in radiative J/ψ decays: The $f_0(980)$ has large $n\bar{n}$ and $s\bar{s}$ components mixed to a dominant SU(3) octet component. The large intensity above 1500 MeV in radiative J/ψ decays is absent when not gluons but an $s\bar{s}$ pair are in the initial state: the mountainous structure in radiative J/ψ decays is produced by gluons and not by $q\bar{q}$ pairs: The structure is due to the scalar glueball.

8.4.9 Discussion

The BESIII data were also fitted by Rodas et al. [2494]. They describe the data above 1 GeV with four scalar and three tensor resonances only, and a "background" due to production of vector mesons decaying into $\pi^0\gamma$ and recoiling against a π^0, $J/\psi \to V\pi^0$, $V \to \pi^0\gamma$. The fit suggests vector meson masses outside of the available phase space leading to a smooth background. (For the $K_S K_S\gamma$ final state, the vector meson decays via $V \to K_S\gamma$.)

This fit is very interesting. It seems to identify the octet-like mesons and to dismiss the singlet-like mesons. In Table 32 we compare these masses. The similarity of the pattern is obvious!

The existence of the $f_0(1370)$ has been disputed since long, see, e.g. [420,2504,2505] and [2368,2506]. In the 1700–1800 MeV region, only one state is established, and the high-mass region is not constrained by the CERN-Munich data. Hence the existence of individual SU(3) singlet scalar resonances can be questioned. In [420] I had proposed that the mainly-singlet scalar mesons above the $f_0(500)$ merge into a continuous scalar "background". In the fits of Ref. [2493], the interference between a mainly-singlet and a mainly-octet amplitude is shown to be required to explain the dip in the scalar $K\bar{K}$ mass distribution and the rapid intensity change in the $\pi\pi$ mass distribution. I assume that the $J/\psi \to V\pi^0$ amplitude in Ref. [2494] simulates the SU(3)-singlet amplitude. The strong peaks in the $\pi\pi$ and $K\bar{K}$ mass distributions in Fig. 187 are still assigned to mainly-octet mesons that should not be produced in radiative J/ψ decays.

If these considerations are true, the structure in Fig. 187 still originates from an intermediate glueball mixed into the wave function of scalar mesons. Also the arguments presented in Sect. 8.4.7 remain valid: the "narrow" octet-like mesons and the singlet-like "background" contain identified fractions of the scalar glueball.

8.4.10 A trace of the tensor glueball

The tensor glueball is predicted with an even higher yield [2488]:

$$\Gamma_{J/\psi\to\gamma/G_{2^{++}}}/\Gamma_{\rm tot} = (11 \pm 2)10^{-3}. \quad (8.35)$$

The yield of $f_2(1270)$ in radiative J/ψ decays is $(1.64 \pm 0.12)10^{-3}$, about six times weaker than the predicted rate for the tensor glueball! Bose symmetry implies that the $\pi^0\pi^0$ or $K_S K_S$ pairs are limited to even angular momenta, practically, only S and D-waves contribute. The scalar intensity originates from the electric dipole transition $E0$. Three electromagnetic amplitudes $E1$, $M2$, and $E3$ excite tensor mesons. Figure 193 shows these three amplitudes and the relative phases.

Two fits were performed [2507]. One fit describes the mass distribution only. Apart from the well known $f_2(1270)$ and $f_2'(1525)$ the fit needs one high-mass resonance with

$$M = (2210 \pm 60)\,\text{MeV}; \quad \Gamma = (360 \pm 120)\,\text{MeV}, \quad (8.36)$$

where the error includes systematic studies with or without additional low-yield resonances. The enhancement was called $X_2(2210)$. In this fit, the phases are not well described. Figure 193 shows a fit in which the 2200 MeV region is described by three tensor resonances with masses and widths of about $(M, \Gamma) = (2010, 200)$, $(2300, 150)$, and $(2340, 320)$ MeV. These resonances had been observed by Etkin et al. [2332] in the reaction $\pi^- p \to \phi\phi n$. The unusual production characteristics were interpreted in Ref. [2332] as evidence that *these states are produced by* 1–3 *glueballs.*

The total observed yield of $X_2(2210)$ in $\pi\pi$ and $K\bar{K}$ is $(0.35 \pm 0.15)\ 10^{-3}$, far below the expected glueball yield. We assume the glueball is – like the scalar glueball – distributed

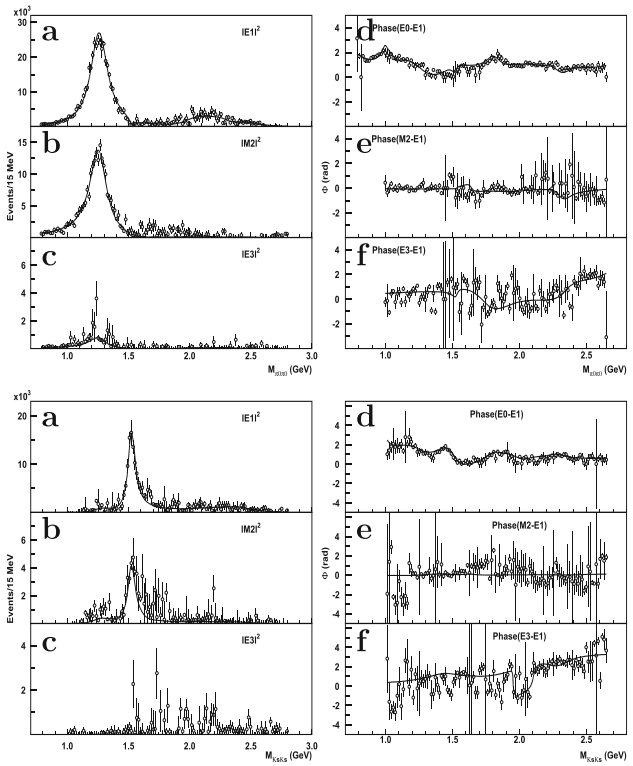


Fig. 193 D-wave intensities and phases for radiative J/ψ decays into $\pi^0\pi^0$ (top subfigures) and $K_s\,K_s$ (bottom subfigures) from Ref. [2496,2497]. The subfigures show the $E1$ (**a**), $M2$ (**b**) and $E3$ (**c**) squared amplitudes and the phase differences between the $E0$ and $E1$ (**d**) amplitudes, the $M2$ and $E1$ (**e**) amplitudes, and the $E3$ and $E1$ (**f**) amplitudes as functions of the meson–meson invariant mass. The phase of the $E0$ amplitude is set to zero. The curve represents our best fit

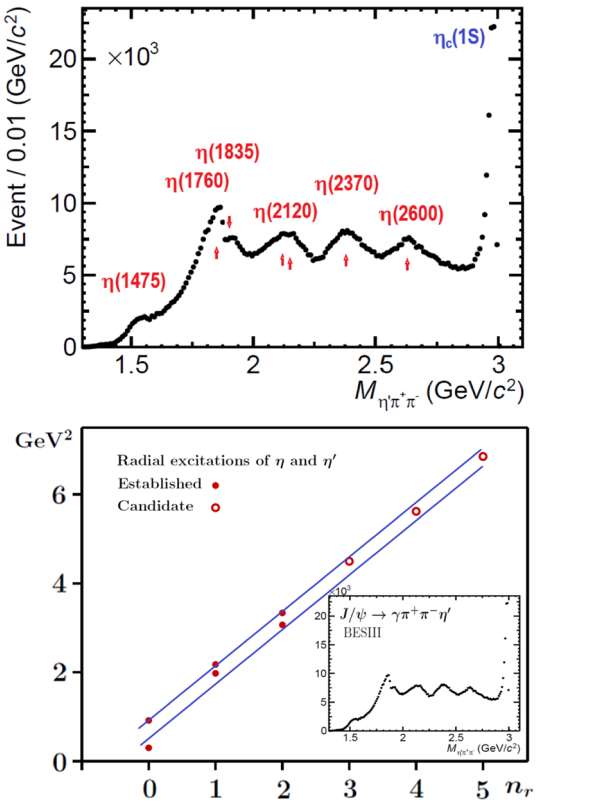


Fig. 194 Top: The $\pi^+\pi^-\eta'$ mass distribution from radiative J/ψ decays [2508]. The quantum numbers are not known. Bottom: M^2 versus n trajectories

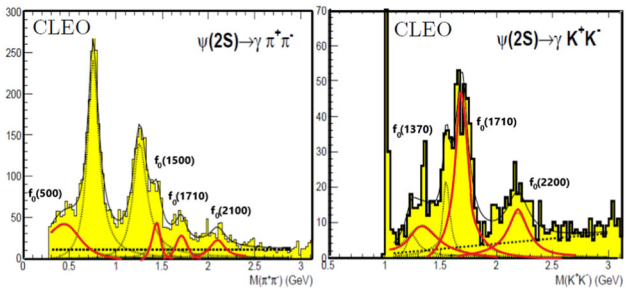


Fig. 195 $\pi^+\pi^-$ (left) and K^+K^- (right) invariant mass distributions from radiative decays of $\psi(2S)$. The red curves represent the S-wave contributions. Adapted from [2509]

over several tensor mesons. Adding up all contributions from tensor states above 1900 MeV seen in radiative J/ψ decays, one obtains

$$\sum_{M=1.9\,\mathrm{GeV}}^{M=2.5\,\mathrm{GeV}} Y_{J/\psi\to\gamma f_2} = (3.1\pm0.6)\,10^{-3}, \qquad (8.37)$$

which is a large yield even though still below the predicted yield.

8.4.11 How to find the pseudoscalar glueball

The BESIII collaboration has studied the reaction $J/\psi \to \pi^+\pi^-\eta'$ [2508]. The top panel of Fig. 194 shows the $\pi^+\pi^-\eta'$ invariant mass distributions with a series of peaks. Assuming that these are all pseudoscalar mesons, two trajectories can be drawn (see bottom panel of Fig. 194). The figure suggests that the higher-mass structures could house two mesons, possibly singlet and octet states in SU(3). If this is true, a cut in the $\pi^+\pi^-$ invariant mass at about 1480 MeV would partly separate the two isobars, $X(2600) \to f_0(1370)\eta'$ and $X(2600) \to f_0(1500)\eta'$. We may expect a slight mass shift in the two $\pi^+\pi^-\eta'$ invariant mass distributions. The two mesons $f_0(1370)$ and η' are both mainly singlet. The $f_0(1370)\eta'$ isobar as singlet meson in the $X(2600)$ complex should be slightly higher in mass than the $f_0(1500)\eta'$ mainly octet meson.

The total yields of the high-mass structures – including unseen decay modes – are not known. Nevertheless, their appearance above a comparatively low background is surprising. Personally, I suppose that the pseudoscalar glueball is rather wide, and that the structures are seen so clearly because of a small glueball content. More studies of theses

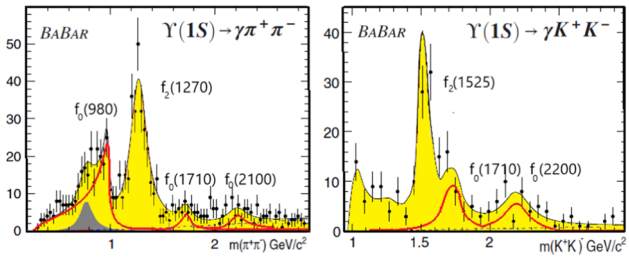


Fig. 196 $\pi^+\pi^-$ (left) and K^+K^- (right) invariant mass distributions from radiative deacys of $\Upsilon(1S)$. The $\Upsilon(1S)$ is observed in $\Upsilon(2S)/\Upsilon(3S)\to\pi^+\pi^-\Upsilon(1S)$. The red curves represent the S-wave contributions, the grey area the $\rho(770)$ contribution. Adapted from [2510]

data and of different channels are required to substantiate this conjecture.

8.4.12 Outlook

The data of the BESIII collaboration presented above are based on 1.3×10^9 events taken at the J/ψ. Presently available are 10^{10} events. Based on this large statistics, rare radiative decays like $J/\psi\to\gamma\eta\eta'$ [2461,2462] and $J/\psi\to\gamma\eta'\eta'$ [2498] have been analysed. Data on the different charge modes of $J/\psi\to\gamma4\pi$ would be extremely important. In an ideal world, these data would be publicly available after publication and would be included in different coupled-channel partial-wave analyses.

Radiative decays of $\psi(2S)$ and of $\Upsilon(1S)$ open a wider range in invariant mass. The authors of Ref. [2509] used the data of the CLEO collaboration on radiative $\psi(2S)$ decays into $\pi^+\pi^-$ and K^+K^-. The data are shown in Fig. 195. The data are fit with known resonances, no partial-wave analysis was performed. The BaBar collaboration studied radiative $\Upsilon(1S)$ decays into $\pi^+\pi^-$ and K^+K^- [2510]. Figure 196 shows the results. In all four distributions, there is not a single prominent peak in the S-wave contribution which would stick out as glueball candidate. The S-waves rather resembles the distributions observed in radiative J/ψ: three major enhancement in the 1500, 1750 and 2200 MeV region separated by dips. (With the larger statistics in J/ψ decays, a fourth enhancement is seen at about 2350 MeV.) In Fig. 195, a peak is found at 1447 MeV and assigned to $f_0(1500)$. At 1500 MeV, there is the dip. The wrong mass is due to the neglect of interference: The phase between $f_0(1500)$ and the "background" (due to the wider $f_0(1370)$) is 180° [2493]. This phase difference and the significant $f_0(1500)\to\eta\eta'$ branching ratio identify $f_0(1500)$ as mainly SU(3)$_F$ octet state. The different masses for the high-mass state in the $\pi^+\pi^-$ and K^+K^- invariant mass distributions point again to the neglect of interference between the prominent octet states and the singlet "background". Inspecting Figs. 195 and 196 shows: there is no striking isolated peak which could be interpreted as "the glueball". The glueball content must be distributed over a large number of states.

In $\psi(2S)$ radiative decays, the $f_0(1710)\to K\bar{K}$ is observed with a branching fraction of $(6.7\pm0.9)\times10^{-5}$, in $\Upsilon(1S)$ radiative decays, the $f_0(1710)\to K^+K^-$ is seen with a branching ratio of $(2.02\pm0.51\pm0.35)\times10^{-5}$. The comparison with the yield observed in Ref. [2493] allows us to calculated the branching ratio expected for $\psi(2S)$ and $\Upsilon(1S)$ decays when the full scalar glueball is covered, i.e. for $\Upsilon(1S)\to\gamma G_0(1865)$. The values are given in Table 33.

Clearly, a significant increase in statistics is required when these reactions should make an independent impact. The advantage of $\psi(2S)$ and $\Upsilon(1S)$ radiative decays is of course that phase space limitations play no role any more. This is particularly important for the search for the tensor and pseudoscalar glueball. The scalar glueball seems to be confirmed: there is not much intensity above 2500 MeV.

At the end I would like to give an answer to the question posed in the title: yes, I am convinced, the scalar glueball is identified, and the tensor glueball seems to have left first traces in the data.

8.5 Heavy quark–antiquark sector: experiment

Marco Pappagallo

8.5.1 Introduction

The term "quarkonium" is a collective name to denote heavy quark–antiquark bound states $Q\bar{Q}'$ (Q, $Q'=c,b$) where the masses of heavy (anti-)quarks are much larger than $\Lambda_{\rm QCD}$, the scale of non-perturbative physics. Therefore the velocities of the heavy (anti-)quark in quarkonium systems are small and a nonrelativistic potential between the heavy quark–antiquark can be employed to predict the properties of the quarkonium states. The spectra of the charmonium and bottomonium states, with quark content $c\bar{c}$ and $b\bar{b}$ respectively, have been extensively studied in the past years. All excited quarkonium states below the open-flavor $D\bar{D}^{(*)}$ or $B\bar{B}^{(*)}$ thresholds were predicted to be narrow. The observation of the J/ψ meson in 1974 and the success, to predict the electromagnetic and hadronic transitions among the narrow quarkonium states, established the potential models as a tool to unravel the complicated QCD dynamics.

Starting from 2003, new states with masses above the $D\bar{D}^{(*)}$ and $B\bar{B}^{(*)}$ thresholds were observed. A common feature is the presence of a heavy quark Q and anti-quark $\bar{Q}$ pair in the decay products. As a consequence, the constituent-quark content of the decaying meson has to include a heavy quark and a heavy anti-quark. However, the properties of many of these states did not match to those of any conventional quarkonium state. So, what are they?

Table 33 Radiative yields expected for $\psi(2S)$ and $\Upsilon(1S)$ radiative decays into the scalar glueball

	"Exp."	Theory	Refs.
$\psi(2S) \to \gamma G_0(1865)$	$\sim 5 \times 10^{-4}$	$(5.9^{+3.4}_{-1.4}) \times 10^{-4}$	[2511]
$\Upsilon(1S) \to \gamma G_0(1865)$	$\sim 3 \times 10^{-4}$	$(1.3^{+0.7}_{-0.3}) \times 10^{-4}$	[2511]
		$(1-2) \times 10^{-3}$	[2512]

In addition to the conventional $q\bar{q}$ mesons and qqq baryons, models based on QCD predict hadrons with different combinations of quarks q and gluons g, such as: pentaquarks ($q\bar{q}qqq$), tetraquarks ($q\bar{q}q\bar{q}$), six-quark H-dibaryons ($q\bar{q}q\bar{q}q\bar{q}$), hybrids ($q\bar{q}g$) and glueballs (ggg), see Sects. 8.3 and 8.4. The existence of such "exotic" hadrons has been debated for several years without reaching a general consensus. In the early 2000s new hadrons with unexpected features were observed, in particular the $D^*_{s0}(2317)^+$ [2513] and $\chi_{c1}(3872)$ [2514] mesons and the Θ^+ baryon [2515]. While the first two candidates are still consistent with being conventional $c\bar{s}$ and $c\bar{c}$ states, the latter one is manifestly exotic with a minimal quark content $uddu\bar{s}$ since it was observed in the nK^+ and pK^0_S final states. However, while the existence of the $D^*_{s0}(2317)^+$ and $\chi_{c1}(3872)$ mesons has been extensively confirmed by many experiments, the evidence of the Θ^+ baryon has faded away with time [2516]. The discovery of the $\chi_{c1}(3872)$ drew a lot of attention due to the narrowness of the signal and the proximity of the mass to the $m(D^0) + m(\bar{D}^{*0})$ threshold. Soon after many other charmonium-like and bottomonium-like states were observed. While it is still not possible to rule out firmly a conventional nature for the majority of them, the observation of the $Z_c(4430)^+$ meson, an electrically charged charmonium-like state, and of the T^+_{cc} state, a meson containing two charm quarks, established definitively the existence of QCD exotics. Many models have been proposed to explain the exotic nature of such a states: *hadronic molecules* [2517], whose constituents are color-singlet mesons bound by residual nuclear forces, *tetraquarks* [2518], bound states between a diquark and diantiquark, *hadro-quarkonium* [2519], a cloud of light quarks and gluons bound to a heavy $Q\bar{Q}$ core state via van-der-Waals forces, *threshold effects*, enhancements caused by threshold cusps [2520] or rescattering processes [2521].

The spectra of the conventional and exotic charmonium-like and bottomonium-like states are shown in Fig. 197. Many of them have been named X, Y and Z in the corresponding discovery papers without a consistent criterion as a consequence of their uncertain nature. With the number of X, Y, Z states growing, the need of an adequate naming scheme emerged. The current naming scheme in Particle Data Group extends the convention used for ordinary quarkonia by taking in account the isospin, spin and parity of the state [513]. The names are not related to the internal structure of the states given their nature is controversial. However, even the current scheme presents some limitation for the manifestly exotic states and a new scheme has been proposed recently [2522].

8.5.2 $\chi_{c1}(3872)$: *the renaissance of the exotic spectroscopy*

In 2003, the Belle collaboration, while studying the $B^+ \to J/\psi\pi^-\pi^+K^+$ decays, observed two peaking structures in the $J/\psi\pi^-\pi^+$ mass projection (Fig. 198): the well known $\psi(2S)$ meson and a new state, originally dubbed $X(3872)$ [2514]. The new meson has been confirmed by many experiments [2523–2529] and observed in prompt production in pp, $p\bar{p}$, pPb [2530] and PbPb [2531] collisions as well as in B and Λ^0_b hadron decays [2532,2533]. The invariant mass distribution of the dipion system is consistent with originating from $\rho(770)^0 \to \pi^+\pi^-$ decays [2527,2534]. Recently using a larger dataset the presence of a sizeable contribution of $\omega(782) \to \pi^+\pi^-$ decays has been established as well [2535]. For a pure charmonium state, the decays to $J/\psi\omega$ [2536,2537] and $J/\psi\rho^0$ are isospin conserving and violating, respectively. Therefore the latter should be strongly suppressed, in contrast to the measured branching ratios. Later, further decay modes have been reported: $D^0\bar{D}^0\pi^0$ [2538], $\bar{D}^0D^{*0}$ [2539], $\chi_{c1}\pi^0$ [2540], $J/\psi\gamma$ [2541] and $\psi(2S)\gamma$. The current branching-fraction measurements of the $\psi(2S)\gamma$ radiative decay [2541–2544] are, however, not fully consistent and further studies are needed to solve the emerging tension. Solving this puzzle will help to understand the nature of the $X(3872)$ meson, given that the predicted branching fractions span over a broad range of values depending if the $X(3872)$ state is a $D^{*0}\bar{D}^0$ molecule [2545,2546] or a pure charmonium state [2547,2548].

A study of the angular correlations among the final state particles from $X(3872) \to J/\psi\pi^+\pi^-$ decays constrained the possible J^{PC} assignments for the $X(3872)$ to $J^{PC} = 1^{++}$ and 2^{-+} [2549]. The latter, disfavoured by the observation of the radiative decays, was definitively ruled out by the LHCb experiment [2550,2551]. Once the quantum numbers $J^{PC} = 1^{++}$ have been firmly established, the name of $X(3872)$ turned into $\chi_{c1}(3872)$ according to the PDG naming scheme [513]. The identification of the $X(3872)$ with the the 2^3P_1 $c\bar{c}$ state is disfavoured by the large branching fraction of $X(3872) \to J/\psi\rho^0$ and the large mass splitting with respect to the 2^3P_2 state, identified with $\chi_{c2}(3930)$.

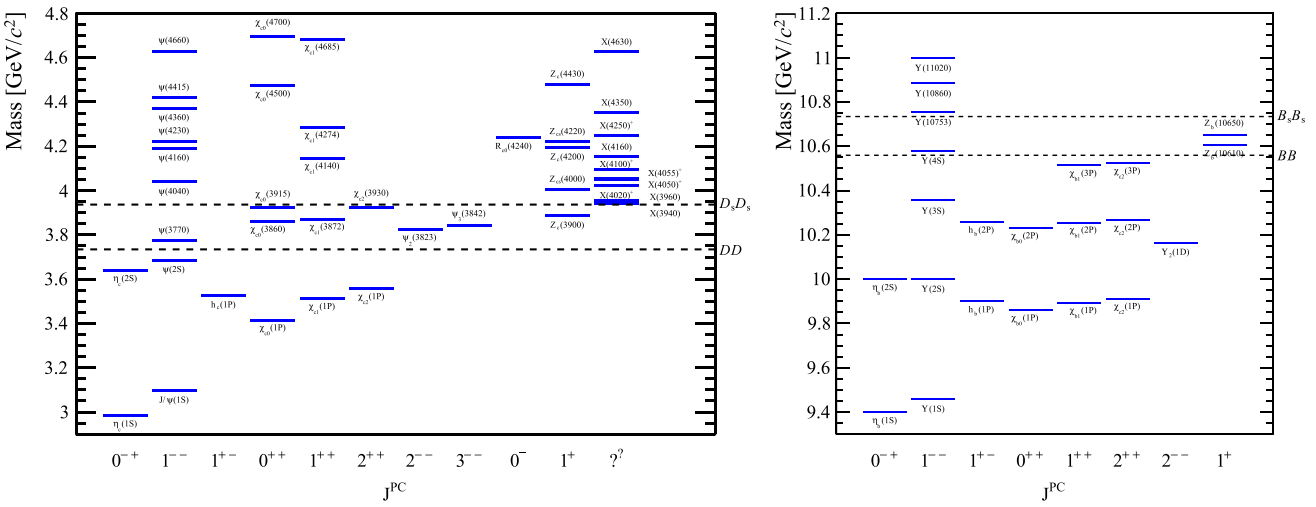


Fig. 197 The spectrum of charmonium(-like) (left) and bottomonium(-like) (right) states. States are labeled according to the PDG naming scheme. Dashed horizontal lines show some relevant open-charm or open-bottom thresholds. The states shown in the right columns are manifestly exotic, i.e. the quark content can not be only $c\bar{c}$ or $b\bar{b}$ given their non-zero electrical charge

An intriguing feature of the $\chi_{c1}(3872)$ meson is the proximity of its mass to the $m(D^{*0}) + m(D^0)$ threshold. This characteristic has led to speculate that the $\chi_{c1}(3872)$ is a molecular state [2552] where the D^{*0} and $\bar{D}^0$ mesons are bound by residual nuclear forces, similarly to a proton and a neutron in the nucleus of the deuterium. An important input for such an interpretation is the binding energy $E_b \equiv m_{D^0} + m_{D^{*0}} - m_{\chi_{c1}(3872)}$ which is still consistent with zero despite being measured with a precision of $\mathcal{O}(100)$ keV [2553,2554]. The analyses also reported a measurement of the natural width $\Gamma^{BW}_{\chi_{c1}(3872)} = (1.39 \pm 0.24 \pm 0.10)$ MeV by using a Breit–Wigner lineshape for the $\chi_{c1}(3872)$ signal. However, since the $|E_b| < \Gamma^{BW}_{\chi_{c1}(3872)}$, coupled channel effects might distort the lineshape. Indeed a Flatté-inspired model returned a significantly smaller full width at half-maximum FWHM $= (0.22 \pm^{0.07}_{0.06} \pm^{0.11}_{0.13})$ MeV, highlighting the relevance of a physically well-motivated lineshape parameterization (see Sect. 14.5).

The smallness of the binding energy $E_b = (0.07 \pm 0.12)$ MeV [2553,2554] implies a size of $\mathcal{O}(10)$ fm in a molecular scenario. The production of a large and weakly bound molecule is expected to be suppressed due to the interactions with comoving hadrons produced in the underlying event [2555]. The ratio of $\chi_{c1}(3872)$ to $\psi(2S)$ cross-sections for promptly produced particles has been measured at LHC [2556] and has been found to decrease with multiplicity. However the slope would seem not to agree with the expectations for a molecular state [2557]. In addition no enhancement of the $\chi_{c1}(3872)$ production has been observed in association to a pion [2558] as expected for a molecular state produced via the formation of a $D^*\bar{D}^*$ pair at short distance followed by the rescattering of the charmed mesons into $\chi_{c1}(3872)\pi$ [2559]. Finally, the relative production of $\chi_{c1}(3872)$ to $\psi(2S)$ mesons as a function of the transverse momentum and rapidity has shown a mild (or null) dependence for $\chi_{c1}(3872)$ (or $\psi(2S)$) mesons produced in prompt pp collisions and from b-hadron decays, respectively [2527,2560]. The CMS collaboration has measured a large production rate of the $\chi_{c1}(3872)$ mesons also at large transverse momenta while a suppression is expected for hadronic molecules [2561], as measured for the deuteron [2562].

In order to reconcile the molecular picture to the production measurements, it has been suggested that the physical $\chi_{c1}(3872)$ might be a quantum mechanical mixture of a $D^{*0}\bar{D}^0$ molecule and the 2^3P_1 $c\bar{c}$ charmonium state [2563], where the production is dominated by the charmonium component. Alternatively, an interpretation has been proposed where the $\chi_{c1}(3872)$ meson is a tightly bound diquark–diantiquark system [2518] with a size of a few fermis. In this scenario, isospin partner states are expected to exist. A search for charged X^- states has been carried out by studying the decays $B^0 \to X^-K^+$ and $B^- \to X^-K^0_S$, where $X^- \to J/\psi\pi^-\pi^0$ [2564,2565]. No charged X^- has been reported. Moreover no $X^- \to D^0D^{*-}$ signal has been observed in the $D^0\bar{D}^0\pi^-$ mass spectrum [2566]. Another firm prediction of the compact tetraquark models is that hidden charm states must form complete flavor-SU(3) multiplets with mass differences determined by the quark mass difference $m_s - m_u$ [2567]. The $\chi_{c1}(3872)$ meson could belong to the same flavor multiplet of the X(4140) given both states have $J^{PC} = 1^{++}$, where the two mesons are interpreted as $[\bar{c}\bar{q}][cq]$ ($q = u$ or d) and $[\bar{c}\bar{s}][cs]$ bound states, respectively. As a consequence, a $[cs][\bar{c}\bar{q}]$ state with mass $(m_{X(4140)} + m_{\chi_{c1}(3872)})/2 = 4009$ MeV should exist.

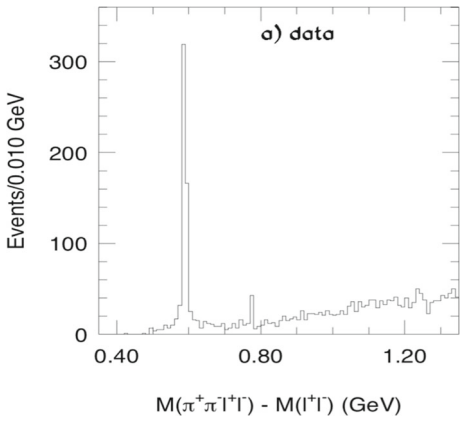


Fig. 198 Distribution of the mass difference $M(\pi^+\pi^-\ell^+\ell^-) - M(\ell^+\ell^-)$, where $\ell = e, \mu$ and the invariant mass of the dilepton system is within a range around the J/ψ mass. The two signals correspond to the $\psi(2S)$ and $\chi_{c1}(3872)$ mesons, respectively [2514]

Two exotic states, $Z_{cs}(3985)$ [2568] and $Z_{cs}(4000)$ [2569] have been observed in $D_s^+\bar{D}^{*0}$ and $J/\psi K^+$ mass spectra, respectively, with masses close to 4000 MeV, making them potential candidates to complete the $J^{PC} = 1^{++}$ tetraquark nonet, where $C = +1$ refers to the sign of charge conjugation of the neutral-non-strange members.

A C-odd partner of the $\chi_{c1}(3872)$ state, dubbed $\tilde{X}(3872)$, is expected as well [2518,2570]. Several experiments searched for a $\tilde{X}(3872)$ candidate in the $J/\psi\eta$ and $\chi_{c1}\gamma$ mass spectra in $B^+ \to J/\psi\eta/\chi_{c1}\gamma K^+$ decays but no signal was reported [2571–2574], even though many other charmonium states were observed. The COMPASS collaboration searched for muo-production of charmonia in the process $\mu^+ N \to \mu^+ X^0 \pi^\pm N'$ with $X^0 \to J/\psi\pi^+\pi^-$ where N denotes the target nucleon, N' the unobserved recoil system and X^0 an intermediate charmonium state [2575]. In addition to the observation of the $\psi(2S)$ meson, evidence of a narrow structure, peaking at about 3872 MeV in the $J/\psi\pi^+\pi^-$ spectrum, was reported. While the measured mass and width pointed to an interpretation of the signal as $\chi_{c1}(3872)$ meson, the $\pi^+\pi^-$ mass spectrum showed a rather flat distribution instead of the expected ρ^0-like shape thus disagreeing significantly with previous experimental results. This surprising result led the authors to speculate that the observed state might be the C-odd partner $\tilde{X}(3872)$ decaying to the $J/\psi f_0(500)$ final state.

Assuming heavy flavor symmetry, a bottomonium counterpart X_b of the $\chi_{c1}(3872)$ meson is expected. Searches for X_b, carried out by the CMS [2576] and ATLAS [2577] collaborations by studying the $\Upsilon\pi^+\pi^-$ final state, have not been successful. This result does not rule out the existence of an X_b state since, contrary to the $\chi_{c1}(3872)$ case, the $\Upsilon\pi^+\pi^-$ decay mode is expected to be suppressed due to the smaller isospin breaking effect: the mass difference between the neutral and charged B mesons is very small. Most likely, the X_b state would decay into the $\Upsilon\omega$ and $\chi_b\pi^+\pi^-$ final states. The former decay has been recently studied by the Belle-II collaboration. No X_b meson has been observed [2578].

8.5.3 $Z_c(4430)^+$ and the charmonium-like states

The observation of manifestly exotic candidates was a turning point in the discussion about the existence of non-conventional hadrons. Indeed, a peculiar characteristic of charmonium-like states is the possibility to observe states with non-zero electrical charge and quark content $c\bar{c}u\bar{d}$.

The first-ever candidate, the $Z_c(4430)^+$ meson, was observed by the Belle collaboration in the $\psi(2S)\pi^+$ projection of $\bar{B}^0 \to \psi(2S)K^-\pi^+$ and $B^+ \to \psi(2S)K_S^0\pi^+$ decays [2579]. The $m^2(K\pi^+)$ versus $m^2(\psi(2S)\pi^+)$ Dalitz-plot distributions show a continuous band (and a peak in the $m^2(\psi(2S)\pi^+)$ projection) together with two bands in the $m^2(K\pi^+)$ mass distributions corresponding to the $K^*(892)$ and $K_{0/2}^*(1430)$ resonances. After applying a veto on the K^* regions, a one-dimensional fit to the $\psi(2S)\pi^+$ projection returned the mass and width (Table 34) of a signal that was interpreted as the first charmonium-like state with non-zero electrical charge (Fig. 199). Given that the decay modes have four degrees of freedom, the claim of a new exotic state based on the study of a one-dimensional projection received some criticism. In addition, excluding regions in the $K\pi^+$ invariant mass does not imply that interference effects are removed which could lead to peaking structures in other projections.

A model-independent approach was pursued by the BaBar collaboration which investigated the extent to which the reflections of mass and angular distribution of structures in the $K\pi^+$ system might describe the associated $\psi(2S)\pi^+$ mass distributions in $\bar{B}^0 \to \psi(2S)K^-\pi^+$ and $B^+ \to \psi(2S)K_S^0\pi^+$ decays [2580]. For this purpose, the $K\pi^+$ angular distribution was represented, at a given $m(K\pi^+)$, in terms of a Legendre polynomial expansion. The combinations of the first $N = 2J_{max} + 1 = 7$ terms reproduces adequately the $\psi(2S)\pi^+$ mass distribution where $J_{max} = 3$ is the maximum spin of the excited K^* resonances expected in the $K\pi^+$ spectrum. This result provided a hint that an exotic contribution may not be needed, but it cannot rule out the presence of the $Z_c(4430)^+$ meson either. Later on, the Belle collaboration performed a fit to the $m^2(K\pi^+)$ versus $m^2(\psi(2S)\pi^+)$ Dalitz-plot [2581] and finally a complete four-dimensional amplitude analysis [2582], both confirming the observation of an exotic state. The latter analysis quotes a natural width for the $Z_c(4430)^+$ much larger than the one reported in the discovery paper (Table 34) which highlights the relevance of performing full amplitude analyses to measure the physical parameters.

The existence of the $Z_c(4430)^+$ exotic state was debated for many years until the LHCb collaboration also studied the

Table 34 Measurements of mass and natural width of the $Z_c(4430)^+$ meson. The last column reports the number of dimensions considered in the corresponding amplitude analysis

	$Z_c(4430)^+$		
	Mass [MeV/c^2]	Width [MeV]	
Belle [2579]	$4433 \pm 4 \pm 2$	45^{+18+30}_{-13-13}	1D
Belle [2581]	4443^{+15+19}_{-12-13}	107^{+86+74}_{-43-56}	2D
Belle [2582]	4485^{+22+28}_{-22-11}	200^{+41+26}_{-46-35}	4D
LHCb [2583]	$4475 \pm 7^{+15}_{-25}$	$172 \pm 13^{+37}_{-34}$	4D

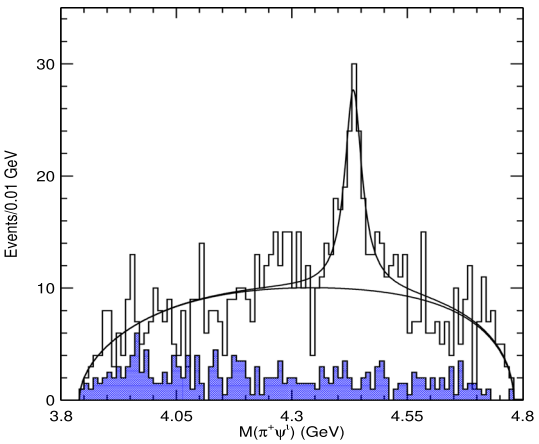


Fig. 199 Observation of the $Z_c(4430)^+$ meson in the $m(\psi(2S)\pi^+)$ distribution of the $\bar{B}^0 \to \psi(2S)K^-\pi^+$ and $B^+ \to \psi(2S)K^0_S\pi^+$ decays after applying a veto on the $K^*(892)$ and $K^*_{0/2}(1430)$ states in the $K\pi^+$ systems [2579]

$\bar{B}^0 \to \psi(2S)K^-\pi^+$ decays [2583]. The ten-fold increase in signal yield over the previous measurements allowed the collaboration to confirm the $Z_c(4430)^+$ state firmly with an improved measurement of mass and width (Table 34) and to establish its spin and parity to $J^P = 1^+$.

In addition, the resonant character of a charged four-quark state is demonstrated for the first time by representing the $Z_c(4430)^+$ amplitude as the combination of independent complex amplitudes at six equidistant points in the $m^2(\psi(2S)\pi^+)$ spectrum. The resulting Argand diagram, shown in Fig. 200, is consistent with a rapid change of the $Z_c(4430)^+$ phase when its magnitude reaches the maximum, a behavior characteristic of a resonance. Finally, an analysis of the data, using the model-independent approach developed by the BaBar collaboration, shows significant inconsistencies in the $Z_c(4430)^+$ region between the data and a model introducing K^* states with $J \leq 3$ [2584]. Evidence of the $Z_c(4430)^+ \to J/\psi\pi^+$ is also reported by an amplitude analysis of the $\bar{B}^0 \to J/\psi K^-\pi^+$ decays [2585].

After the discovery of the $Z_c(4430)^+$ meson, many further charged charmonium-like states have been reported

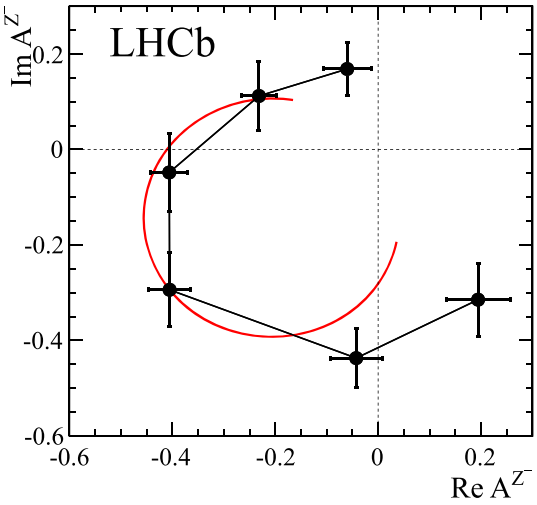


Fig. 200 Argand diagram of the $Z_c(4430)^+$ meson by a Dalitz analysis of the $\bar{B}^0 \to \psi(2S)K^-\pi^+$ decays where the $Z_c(4430)^+$ amplitude is fitted in six independent $m^2(\psi(2S)\pi^+)$ bins. The red curve is the expected shape according to a Breit–Wigner function with a resonance mass (width) of 4475 (172) MeV. Units are arbitrary [2583]

Table 35 Decay modes and quantum numbers of manifestly exotic charmonium-like states

State	Decay modes	$I^G(J^{PC})$
$Z_c(3900)^+$	$J/\psi\pi^+$ [2588–2590]	$1^+(1^{+-})$
	$\bar{D}^0D^{*+}$, $\bar{D}^{*0}D^+$ [2591,2592]	
$X(4020)^+$	$h_c\pi^+$ [2593], $D^{*+}\bar{D}^{*0}$ [2594]	$1^+(?^{?-})$
$X(4050)^+$	$\chi_{c1}(1P)\pi^+$ [2595]	$1^-(?^{?+})$
$X(4055)^+$	$\psi(2S)\pi^+$ [2596]	$1^+(?^{?-})$
$X(4100)^+$	$\eta_c(1S)\pi^+$ [2597]	$1^-(?^{??})$
$Z_c(4200)^+$	$J/\psi\pi^+$ [2585]	$1^+(1^{+-})$
$R_{c0}(4240)^+$	$\psi(2S)\pi^+$ [2583]	$1^+(0^{--})$
$X(4250)^+$	$\chi_{c1}(1P)\pi^+$ [2595]	$1^-(?^{?+})$
$X(3985)^+$	$D_s^+\bar{D}^{*0}$, $D_s^{*+}\bar{D}^0$ [2568]	$1/2(?^?)$
$Z_{cs}(4000)^+$	$J/\psi K^+$ [2569]	$1/2(1^+)$
$Z_{cs}(4220)^+$	$J/\psi K^+$ [2569]	$1/2(1^?)$

(Table 35), including candidates with strangeness and isospin partners [2586,2587].

8.5.4 *The bottomonium-like Z_b^+ states*

Few years after the discovery of the $Z_c(4430)^+$ meson, the Belle collaboration claimed the observation of two bottomonium-like states $Z_b(10610)^+$ and $Z_b(10650)^+$ in the $\Upsilon(nS)\pi^+$ (n = 1, 2, 3) and $h_b(mP)\pi^+$ (m = 1, 2) spectra by studying the exclusive processes $e^+e^- \to \Upsilon(nS)\pi^+\pi^-$ (n = 1, 2, 3) and $e^+e^- \to h_b(mP)\pi^+\pi^-$ (m = 1, 2) with data collected at the collision energy $\sqrt{s}$ = 10.865 GeV [2598], the $\Upsilon(5S)$ mass. Amplitude analyses of the three-body $\Upsilon(nS)\pi^+\pi^-$ decays were per-

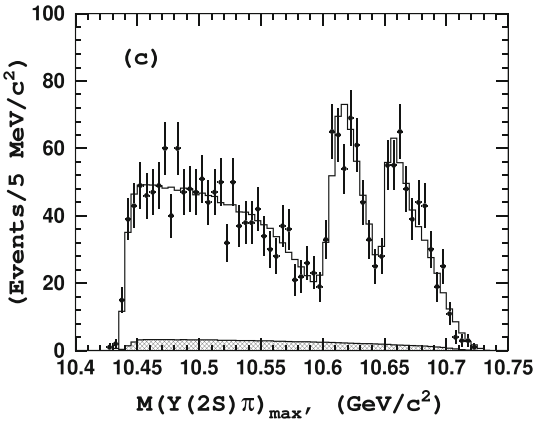


Fig. 201 The maximum invariant mass of the two $\Upsilon(2S)\pi$ combinations of the $e^+e^- \to \Upsilon(2S)\pi^+\pi^-$ process at $\sqrt{s} = 10.865$ GeV. The two peaking structures are interpreted as the $Z_b(10610)^+$ and $Z_b(10650)^+$ bottomonium-like states [2598]

formed by means of unbinned maximum likelihood fits to two-dimensional $m^2(\Upsilon(nS)\pi^+)$ versus $m^2(\Upsilon(nS)\pi^-)$ Dalitz distributions. Two narrow structures appear in the $m(\Upsilon(nS)\pi^\pm)$ spectrum (e.g. Fig. 201). The analyses of the $h_b(mP)\pi^\pm$ spectra returned compatible results. Weighted averages of mass and width measurements over all five channels yield for the $Z_b(10610)^+$

$$m = (10607.2 \pm 2.0)\,\text{MeV},\ \Gamma = (18.4 \pm 2.4)\,\text{MeV},$$

and for the $Z_b(10650)^+$

$$m = (10652.2 \pm 1.5)\,\text{MeV},\ \Gamma = (11.5 \pm 2.2)\,\text{MeV}.$$

Later on a six-dimensional amplitude analysis of the $\Upsilon(nS)\pi^+\pi^-$ (n = 1, 2, 3) three-body final states confirmed the existence of the two Z_b^+ states and strongly favored $I^G(J^P) = 1^+(1^+)$ quantum-number assignments for both of them [2599]. Finally, the two mesons $Z_b(10610)^+$ and $Z_b(10650)^+$ have been observed in the $B^+\bar{B}^{*0}$ and $B^{*+}\bar{B}^{*0}$ mass spectrum, respectively [2600]. Table 36 summarizes the branching fractions of $Z_b(10610)^+$ and $Z_b(10650)^+$ states by assuming that their sum is equal to one.

The large branching fractions of the $B^{(*)}\bar{B}^*$ decay modes and the measured quantum numbers are consistent with the interpretation of the two states as $B\bar{B}^*$ and $B^*\bar{B}^*$ loosely bound molecular hadrons. However the measured mass for the $Z_b(10610)^+$ and $Z_b(10650)^+$ are both above the nearby open-flavor thresholds. This might be the result of using Breit–Wigner functions to parameterize the amplitudes of very near-threshold states. Indeed, when amplitudes consistent with unitarity and analyticity are used instead, lower masses are measured, typically below the thresholds [2601].

Table 36 Branching fractions for the $Z_b(10610)^+$ and $Z_b(10650)^+$ decays. The first uncertainty is statistical while the second is systematic [2600]

Channel	Fraction (%)	
	$Z_b(10610)^+$	$Z_b(10650)^+$
$\Upsilon(1S)\pi^+$	$0.60 \pm 0.17 \pm 0.07$	$0.17 \pm 0.06 \pm 0.02$
$\Upsilon(2S)\pi^+$	$4.05 \pm 0.81 \pm 0.58$	$1.38 \pm 0.45 \pm 0.21$
$\Upsilon(3S)\pi^+$	$2.40 \pm 0.58 \pm 0.36$	$1.62 \pm 0.50 \pm 0.24$
$h_b(1P)\pi^+$	$4.26 \pm 1.28 \pm 1.10$	$9.23 \pm 2.88 \pm 2.28$
$h_b(2P)\pi^+$	$6.08 \pm 2.15 \pm 1.63$	$17.0 \pm 3.74 \pm 4.1$
$B^+\bar{B}^{*0} + \bar{B}^0 B^{*+}$	$82.6 \pm 2.9 \pm 2.3$	–
$B^{*+}\bar{B}^{*0}$	–	$70.6 \pm 4.9 \pm 4.4$

8.5.5 The B_c^+ mesons

Contrary to charmonium and bottomonium states, the B_c^+ mesons can not annihilate into gluons and thus these states are more stable. Indeed, apart from the ground state which decays weakly, all the excited states, with masses below the lowest strong decay $B^{(*)}D^{(*)}$ thresholds, are predicted to have narrow widths [2602,2603].

Before the start of LHC, only the ground B_c^+ state was observed [2604] via few decays modes: $B_c^+ \to J/\psi\pi^+$ and $B_c^+ \to J/\psi\ell^+\nu$. The LHCb and CMS experiments have observed 15 new decays modes and have largely improved the precision of the B_c^+ mass [2605] and lifetime [2606–2608]. The production of the B_c^+ meson has been observed in $p\bar{p}$, pp as well as in PbPb collisions [2609], where the measurement of the nuclear modification factor hints that effects of the hot and dense nuclear matter created in heavy ion collisions contribute to its production.

Despite the large number of expected excited states, only a few have been observed so far due to the small production cross sections of the B_c^+ mesons and the small branching ratios of the reconstructed decay chains. In 2014 the ATLAS collaboration reported the first observation of an excited B_c^+ state decaying to $B_c^+\pi^+\pi^-$ final state [2610]. Few years later the same mass spectrum was investigated by other LHC experiments [2611,2612] and it turned out that the ATLAS structure was very likely the result of a superimposition of two narrower signals (Fig. 202), interpreted as the $B_c(2S)^+$ and $B_c^*(2S)^+$ states. The latter appears in the mass spectrum as a partially reconstructed decay $B_c^*(2S)^+ \to B_c^{*+}\pi^+\pi^-$, where the photon of the $B_c^{*+} \to B_c^+\gamma$ reaction is not reconstructed. Since the B_c^{*+} meson has not been observed yet, the mass of the $B_c^*(2S)^+$ state can not be measured and it is not listed in the PDG. In the next years the upgraded LHC experiments will probe the largely unexplored spectrum of the excited B_c^+ mesons below and above the $B^{(*)}D^{(*)}$ thresholds with the intriguing possibility to observe exotic states as for the other quarkonium systems [2613].

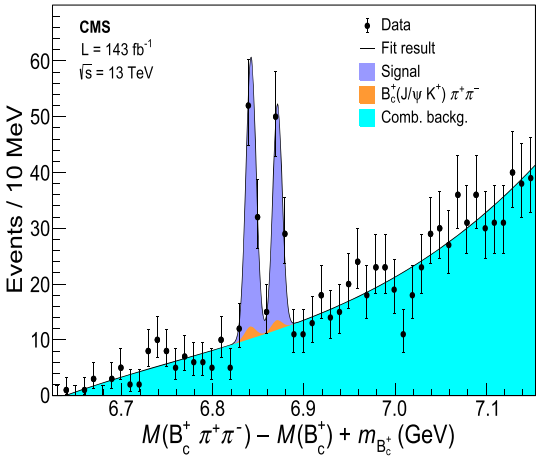


Fig. 202 Observation of the $B_c^*(2S)^+$ (left-most peak) and $B_c(2S)^+$ (right-most peak) states in the $M(B_c^+\pi^+\pi^-) - M(B_c^+) + m_{B_c^+}$ mass distribution [2611]

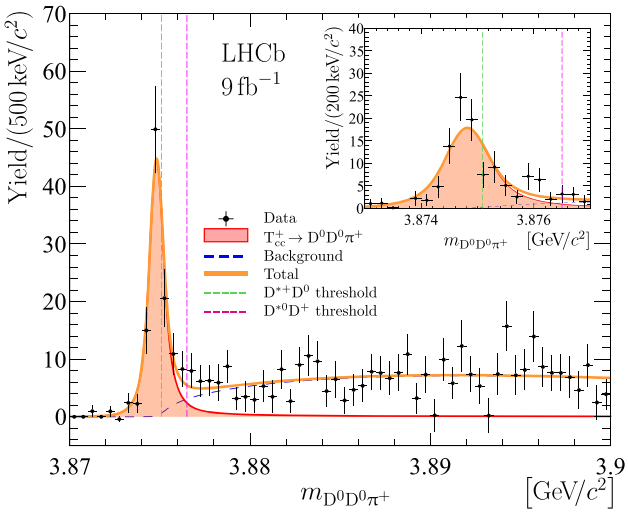


Fig. 203 Distribution of $D^0D^0\pi^+$ mass where the contribution of the non-D^0 background has been statistically subtracted. The $D^{*+}D^0$ and $D^{*0}D^+$ thresholds are indicated with the vertical dashed lines. Inset shows a zoomed signal region with fine binning scheme [1067,2566]

8.5.6 *The doubly charmed $T_{cc}(3875)^+$ state*

All the exotic mesons described so far are featured by a heavy quark–antiquark pair $Q\bar{Q}$ and a light quark–antiquark pair $q\bar{q}$. The observation of several $Q\bar{Q}q\bar{q}$ state has revived the discussion on the existence of of $QQ\bar{q}\bar{q}$ states with two heavy quarks and two light antiquarks. In the limit of a large heavy-quark mass, the two heavy quarks QQ form a heavy point-like color-antitriplet object, that behaves like an antiquark, and the corresponding four-quark state should be bound. The argument that such a state should exist, if the mass of the charm quark is enough, has been discussed extensively, but a consensus was not reached. Even lattice QCD calculations had not provided a definite conclusion [2614]. The $cc\bar{u}\bar{d}$ ground state, hereafter denoted as T_{cc}^+, is predicted with spin-parity quantum numbers $J^P = 1^+$ and isospin $I = 0$. The only known hadron with a similar quark content is the Ξ_{cc}^{++} baryon [2615–2617], a bound state of two c quarks and one u quark. Its measured mass [2618] implies that the mass of the T_{cc}^+ is close to the sum of masses of D^0 and D^{*+} mesons [1066].

The LHCb experiment reported the observation of a narrow state in the $D^0D^0\pi^+$ mass spectrum near the $D^{*+}D^0$ mass threshold compatible with being a T_{cc}^+ tetraquark state [1067,2566]. The $D^0D^0\pi^+$ final state is reconstructed by selecting events with two D^0 mesons and a positively charged pion, all produced at the same proton–proton interaction point. Both D^0 mesons are reconstructed in the $D^0 \to K^-\pi^+$ decay channel. The mass distribution of the selected $D^0D^0\pi^+$ candidates is shown in Fig. 203. A narrow peak near the $D^{*+}D^0$ mass threshold is clearly visible.

An extended unbinned maximum-likelihood fit to the $D^0D^0\pi^+$ mass spectrum is performed by modelling the signal with a Breit–Wigner function $\mathfrak{F}^{BW}$. The measured mass δm and the full width at half maximum (FWHM) of the T_{cc}^+ state are reported in Table 37, where the uncertainties are statistical. The mass parameter δm is defined relative to the $D^{*+}D^0$ mass threshold as $\delta m \equiv m - m_{D^{*+}} - m_{D^0}$, where $m_{D^{*+}}$ and m_{D^0} denote the known masses of the D^{*+} and D^0 mesons. The measured δm value corresponds to a mass of approximately 3875 MeV. Though the use of a standard Breit–Wigner function is sufficient to reveal the existence of a state, it does not take in account the proximity to D^*D thresholds. A more advanced parameterization is needed to probe the physical properties of the resonance. An unitarized Breit–Wigner profile $\mathfrak{F}^U$ is considered as an alternative model for the $T_{cc}(3875)^+$ signal, where the energy-dependent width accounts for the $T_{cc}^+ \to D^0D^0\pi^+$, $T_{cc}^+ \to D^0D^+\pi^0$ and $T_{cc}^+ \to D^0D^+\gamma$ decays. The resulting mass, relative to $D^{*+}D^0$ threshold, and the FWHM of the signal are shown in Table 37 and compared to the results of the $\mathfrak{F}^{BW}$ model. The narrowness of the T_{cc}^+ state varies substantially highlighting the relevance of accounting for the D^*D thresholds. Despite the difference in results, both models can describe the data adequately given the mass resolution of about 400 keV/c^2. The $T_{cc}(3875)^+$ state is the narrowest exotic state observed to date.

The $D^0D^0\pi^+$ events with a mass below the $D^{*+}D^0$ threshold (Fig. 203) are selected to study the $D^0\pi^+$ mass distribution which indicates that the $T_{cc}^+ \to D^0D^0\pi^+$ decay proceeds via an intermediate off-shell D^{*+} meson.

The peak in the $D^0D^0\pi^+$ could be interpreted as the $I_3 = 0$ component of an isotriplet $(\hat{T}_{cc}^0, \hat{T}_{cc}^+, \hat{T}_{cc}^+ +)$ with $cc\bar{u}\bar{u}$, $cc\bar{u}\bar{d}$ and $cc\bar{d}\bar{d}$ quark content, respectively. A search for a $\hat{T}_{cc}^{++}$ state in the $D^+D^0\pi^+$ mass spectrum reports no signal. All the observed properties strongly support the inter-

Table 37 Mass difference $\delta m \equiv m - m_{D^{*+}} - m_{D^0}$ and the full width at half maximum (FWHM) of the $T_{cc}(3875)^+$ state by fitting the $D^0 D^0 \pi^+$ mass spectrum with the $\mathfrak{F}^{BW}$ and $\mathfrak{F}^U$ models. The uncertainties are statistical only. See Refs. [1067,2566] for a complete set of results

	$T_{cc}(3875)^+$	
	δm [keV/c^2]	FWHM [keV/c^2]
$\mathfrak{F}^{BW}$	-279 ± 59	409 ± 163
$\mathfrak{F}^U$	-359 ± 40	47.8 ± 1.9

pretation of the new state as the isoscalar $J^P = 1^+$ $cc\bar{u}\bar{d}$ tetraquark ground state.

Using the $\mathfrak{F}^U$ model, the scattering length a, the effective range r [2622], and the compositeness Z [2623] are determined:

$$a = \left[-(7.16 \pm 0.51) + i\,(1.85 \pm 0.28) \right] \text{fm}, \tag{8.38}$$

$$-r < 11.9\,(16.9)\ \text{fm at 90 (95)\% CL}, \tag{8.39}$$

$$Z < 0.52\,(0.58)\ \text{at 90 (95)\% CL}. \tag{8.40}$$

The real part of the scattering length a can be interpreted as the characteristic size of the state $R_a \equiv -\mathcal{R}e[a] = 7.16 \pm 0.51$ fm which corresponds to a spatial extension as large as expected for molecular states. Within the $\mathfrak{F}^U$ model the resonance pole is found to be located on the second Riemann sheet at $\hat{s} = m_{\text{pole}} - \frac{i}{2}\Gamma_{\text{pole}}$, where

$$\delta m_{\text{pole}} = -360 \pm 40^{+4}_{-0}\ \text{keV/c}^2, \tag{8.41}$$

$$\Gamma_{\text{pole}} = \quad 48 \pm \ 2^{+0}_{-14}\ \text{keV}. \tag{8.42}$$

All exotic hadrons observed so far predominantly decay via the strong interaction; their decay widths vary from a few to a few hundred MeV. The discovery of the $T_{cc}(3875)^+$ meson implies the existence of a $bb\bar{u}\bar{d}$ state that should be stable against strong and electromagnetic interactions: its mass is expected to fall below the $B^{*-}B^0$ and B^-B^0 mass thresholds. The observation of a long-lived exotic state will be an intriguing goal for future experiments.

8.5.7 The fully charmed tetraquark X(6900)

Many QCD-motivated phenomenological models [2624, 2625] have predicted the existence of states consisting of four heavy quarks $T_{QQ\bar{Q}\bar{Q}}$. In 2020 the LHCb collaboration reported the study of the invariant mass spectrum of the J/ψ pairs where both J/ψ mesons are reconstructed via the $\mu^+\mu^-$ decay [2619]. As a result, the reconstruction efficiency is large due to the presence of muons only in the final state. A pair of J/ψ mesons can be produced in proton–proton collisions at LHC via single (SPS) or double (DPS) parton scattering processes, where the two J/ψ are produced in a single or two separated interactions of gluons or quarks, respectively. The SPS process includes both resonant production via intermediate states, such as $T_{c\bar{c}c\bar{c}}$, and nonresonant production.

The $J/\psi J/\psi$ mass distribution (Fig. 204) shows a broad structure just above the kinematic threshold and a narrower peak at about 6.9 GeV, dubbed $X(6900)$. An unusual dip also appears between them. The broad structure can be modelled as a superimposition of two Breit–Wigner structures or as an interference between a Breit–Wigner function and the background. The latter model successes to describe also the dip adequately. The presence of the $X(6900)$ state is established in both models, though the natural width is twice larger in the latter.

Recently the CMS [2620] and ATLAS [2621] collaborations have presented preliminary studies of the $J/\psi J/\psi$ spectrum (Fig. 204). While the $X(6900)$ state is confirmed, there is no consensus on the fit model. Common features are the presence of dips in the spectrum and the need of interference terms to describe it properly. Interestingly the CMS collaboration also claimed the observation and the evidence of two new states $X(6600)$ and $X(7300)$, respectively. A hint of the latter structure was also pointed out by the LHCb collaboration.

Given no single light hadron can mediate the interaction between charmonia to generate a loosely bound molecule, the $X(6900)$ meson seems likely to be a compact tetraquark [2626]. The LHC experiments will profit of larger datasets in a near future which will help to investigate further the resonant nature of the peaks and eventually to measure their spins and parities [2627] in the $J/\psi J/\psi$ and $\psi(2S)J/\psi$ spectra.

Tetraquarks states containing only bottom quarks, $T_{bb\bar{b}\bar{b}}$, have been also searched for by the LHCb and CMS collaborations in the $\Upsilon\mu^+\mu^-$ decay [2628,2629] but no signal have been observed.

8.5.8 Conclusions

The existence of exotic hadronic states with more than minimal quark content ($q\bar{q}$ or qqq) was proposed since the birth of the quark model [17,18]. In the last decades samples of quarkonia larger and larger have been exploited to study their transition and production processes. New and fascinating exotic X, Y, Z states have been observed at a large number of facilities and in different production processes: at tau-charm (BES experiment) and B factories (BaBar and Belle experiments), in hadroproduction at Fermilab Tevatron and the Large Hadron Collider (LHC) at CERN, in photon-gluon fusion at DESY, photoproduction at JLab, and in heavy-ion production and suppression at RHIC, NA60, and LHC. In the upcoming years an unprecedented amount of data will be available from the upgraded experiments CMS, ATLAS, LHCb, ALICE, Belle II and BESIII [1464,2630–2634] and more data will come in future from Panda at FAIR and

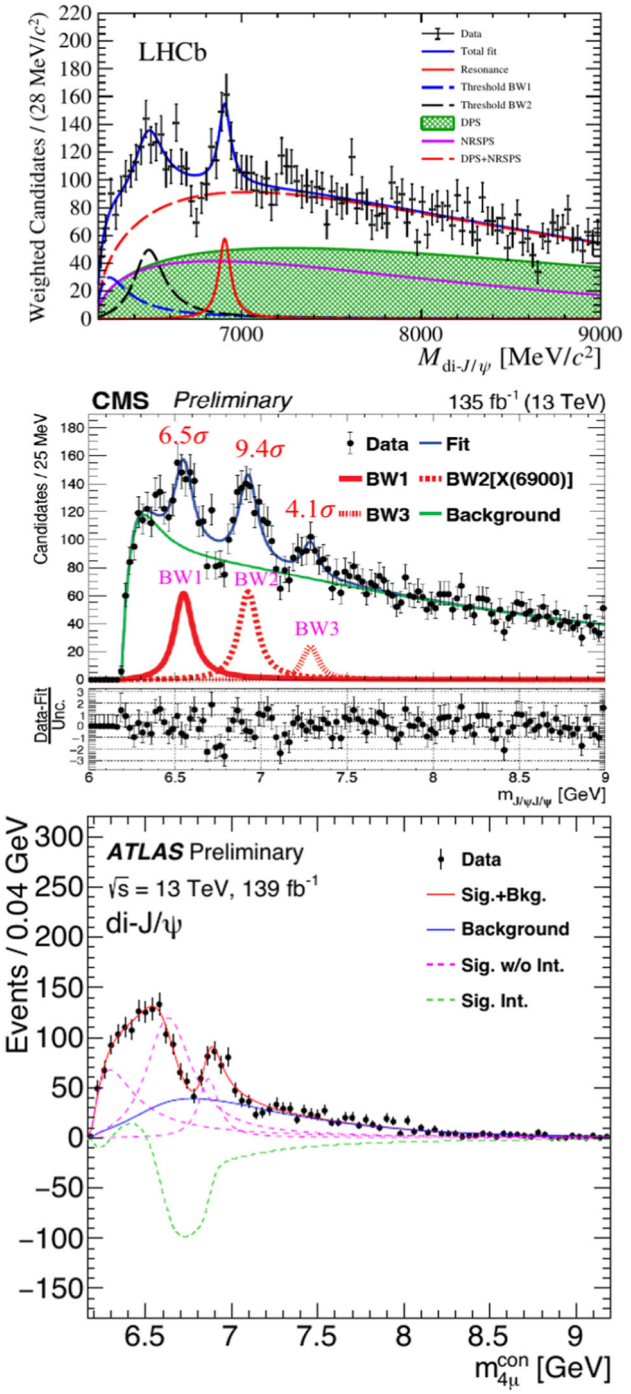


Fig. 204 Invariant mass spectra of J/ψ pair candidates by using data recorded bu the LHCb [2619] (top), CMS [2620] (middle) and ATLAS [2621] (bottom) experiments. The three spectra have been aligned for an easier comparison

the Electron Ion Collider (EIC) [2635,2636]. The measurements of the quantum numbers and classification of the exotic hadrons in $SU(3)$ flavor multiplets will be an important step to understand the nature of the observed exotic states. In addition it will be important to identify observables which can discriminate between different models. For instance the measurement of the effective range has been suggested as a physical quantity able to determine if the $\chi_{c1}(3872)$ is a compact tetraquark or a loosely-bound molecular state [2637].

8.6 Heavy quark–antiquark sector: theory

Nora Brambilla

8.6.1 Introduction

Heavy quarks have been instrumental in accessing the strong interactions as they provide a mass scale m_h which is bigger than $LQCD$: at such scale perturbation theory is valid and scale factorization is useful. Among the systems with heavy quarks, systems with two (or more) heavy quarks are very special, being endowed with a pattern of separated energy scales. Quarkonium in particular, a bound state of a heavy quark and a heavy anti-quark, provides a special tool to study strong interactions.

The 1974 discovery of the J/ψ [91,92], the charmonium ground state, drastically changed and shaped the Standard Model (SM) of particle physics: termed the November revolution, it represented the confirmation of the quark model, the discovery of the charm quark, the confirmation of the GIM mechanism [80] (the mechanism through which flavor-changing neutral currents are suppressed in loop diagrams), and the first discovery of a quark of large mass moving nonrelativistically. It was also the confirmation of QCD in its most peculiar properties of high-energy asymptotic freedom and low-energy confinement [97]. The small width can be explained by the fact that J/ψ is the lowest $c\bar{c}$ energy level and can decay only via annihilation, which makes available in the process a large energy, of order of twice the mass of the charm quark (about 2 GeV). The annihilation width is then proportional to $\alpha_s^2(2m_c)$ which is small due to asymptotic freedom, since m_c is bigger than Λ_{QCD}. Confinement becomes also manifest in the case of quarkonium, where the color-singlet static quark–antiquark interaction potential can be written in terms of a Wilson loop (see e.g. [2638,2639]). Confinement emerges as an area law in the Wilson loop [97], cf. Fig. 205. Correspondingly a linear potential grows with the distance between the quarks [2640] as $V_0 = \lim_{T\to\infty}(i/T)\ln W$, where W is given by $W = \text{Tr P}\exp\{ig_s \oint_{\Gamma_0} dz_\mu \mathcal{A}_\mu(z)\}$, see Fig. 206 and Sect. 6.1.

The energy scales involved in quarkonium span from the hard region, where an expansion in the coupling constant is possible and precision studies may be done, to the low-energy region, dominated by confinement and the many manifestations of nonperturbative dynamics. This property underlies its uniqueness and is the reason for which quarkonium plays a crucial role for a number of problems at the frontier of our

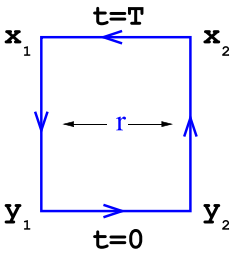


Fig. 205 The static Wilson loop along the circuit Γ_0: it contains the interaction of a static quark–antiquark pair created at a time t = 0 (respectively at space points $\mathbf{y}_1$ and $\mathbf{y}_2$) annihilated at a subsequent large time T (at space points $\mathbf{x}_1$ and $\mathbf{x}_2$) initial and final states are made gauge invariant by the presence of the Schwinger line. The Wilson area law says that the Wilson loop behavior at large distances is exponential in the area of the loop weighted by the string tension σ

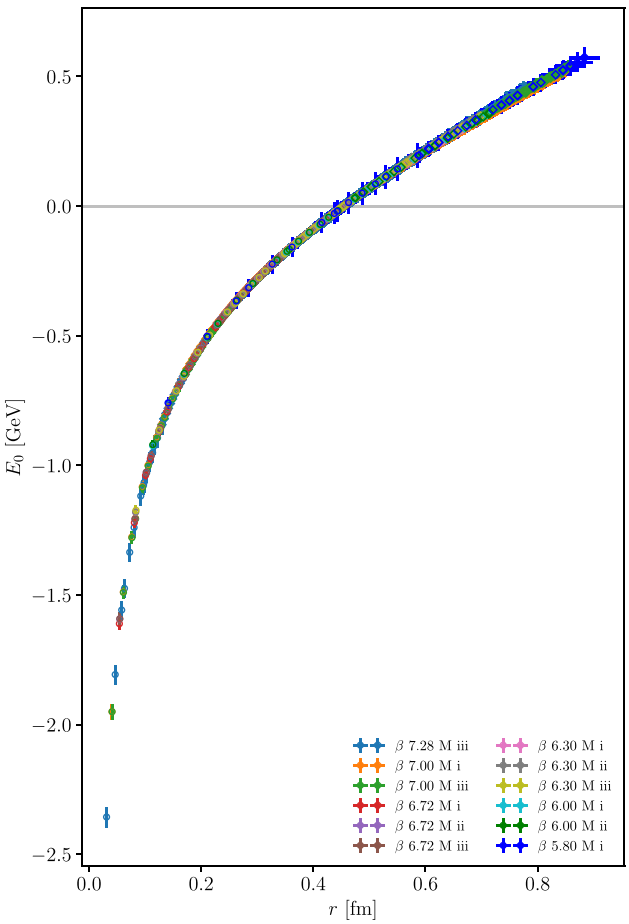


Fig. 206 Results for the static potential in physical units for 2+1+1 dynamical quark flavors. The data are from twelve ensembles of varying lattice spacing (keyed by β) and three choices of light quark mass (denoted "M i", "M ii", "M iii"). Lattice units are eliminated via the r_0/a scale setting, and an unphysical constant is eliminated by setting $V_0(r_0) = 0$. For details see [2641]. This is the first ever determination of the potential with 4 dynamical fermions

research, from the investigation of the confinement dynamics in QCD to the study of deconfinement and the phase diagram of nuclear matter, from the precise determination of Standard Model parameters up to the emergence of exotics X, Y, Z states of an unprecedented nature [1464,1465,2630–2632,2639], as we will summarize in the next sections. It is also the reason for which quarkonium should be addressed with effective field-theory methods to take advantage of the scales separation.

8.6.2 *Scales and effective field theories*

The multiscale nature of quarkonium has made a description within Quantum Field Theory particularly difficult until the advent of non-relativistic effective field theories (NREFTs), cf. Sect. 6.1. When in the eighties of last century, theorists set up to investigate the structure of the energy levels of charmonium and bottomonium, they noticed that it can be reproduced by using a Schrödinger equation with a static potential composed of an attractive Coulomb contribution (with the appropriate $SU(3)$ color factor for a singlet $Q\bar{Q}$) and a term linear in the distance: the famous Cornell potential (see Sect. 2.1 and Refs. [769,2642]:

$$V_0(r) = -\frac{\kappa}{r} + \sigma r + \text{const.} \tag{8.43}$$

This was the quark model description with the potential inspired by QCD. The parameter κ was identified with $\frac{4}{3}\alpha_s$, corresponding to a one-gluon exchange that should dominate at small distances due to asymptotic freedom. The string tension σ corresponds to a constant energy density related to confinement and generates a potential growing with the interquark distance r at large distances. A fit to the states gave $\kappa = 0.52$ and $\sigma = 0.182$ GeV^2. In order to describe the fine and hyperfine characteristics of the spectrum, relativistic corrections to the static potential have been introduced to account for effects of order v^2, i.e. 20% to 30% for the charmonium and up to 10% for the bottomonium spectrum. They appear at the order $1/m_h^2$ in the expansion, involving both spin dependent (spin–spin, tensor and spin–orbit couplings) and purely velocity dependent terms. They were derived in the eighties, either from the semirelativistic reduction of a Bethe–Salpeter equation [816] for the quark–antiquark Green functions (or, equivalently at this level, from the reduction of the quark–antiquark scattering amplitude with an effective exchange) or in some model description like the flux-tube model [2418], for a review see [2638,2639,2643]. The problem of these approaches is the lack of a precise connection to QCD. Taking advantage of NREFTs, quarkonium can be described directly in QCD, and in this way it becomes a probe of strong interactions.

The spectrum of quarkonium, see Fig. 171, clearly states that it is a nonrelativistic system: the difference in the orbital

energy levels is much smaller than the quark mass. Defining v as the heavy quark velocity in the rest frame of the meson in units of c (with $v^2 \sim 0.1$ for the $b\bar{b}$, $v^2 \sim 0.3$ for $c\bar{c}$ systems) the energy levels scales like $m_h v^2$, while fine and hyperfine separations scale like $m_h v^4$. This is the same scaling as for the hydrogen atom (identifying v with the fine structure constant $\alpha_{\rm em}$). This scaling is the signature of a nonrelativistic system. Being nonrelativistic, quarkonia are characterized by a hierarchy of energy scales: the mass m_h of the heavy quark (hard scale), the typical relative momentum $p \sim m_h v$ (in the meson rest frame) corresponding to the inverse Bohr radius $r \sim 1/(m_h v)$ (soft scale), and the typical binding energy $E \sim m_h v^2$ (ultrasoft scale). Of course, for quarkonium there is another scale that can never be switched off in QCD, i.e. $\Lambda_{\rm QCD}$, the scale at which nonperturbative effects become dominant. A similar pattern of scales emerge in the case of baryons composed of two or three heavy quarks [1451,1452] and for the just discovered state $X(6900)$ made by two charm and two anticharm quarks. The pattern of nonrelativistic scales makes all the difference between heavy quarkonia and heavy-light mesons, which are characterized by just two scales: m_h and $\Lambda_{\rm QCD}$.

The correct zero-order problem is thus the Schrödinger equation with potentials. These should, however, be defined and calculated directly in QCD, and nonpotential corrections that should be accounted for. As explained in Sect. 6.1 using the EFT method to integrate out in QCD (in the sector with one heavy quark and one heavy antiquark) the hard scale m_h and the soft scale $m_h v$, give origin to the NREFT called pNRQCD (potential Nonrelativistic QCD) [1462,1490,1492]. The pNRQCD description directly addresses the bound state dynamics, implements the Schrödinger equation as zero-order problem, properly defines the potentials as matching coefficients, and allows to systematically calculate relativistic and retardation corrections. Each correction has a size determined by the power counting in v and in α_s. The EFT allows us to make model-independent predictions and we can use the power counting to attach an error to the theoretical predictions.

When $mv \gg \Lambda_{\rm QCD}$, we speak about weakly-coupled pNRQCD because the soft scale is perturbative and the potentials can be calculated in perturbation theory. The lowest levels of quarkonium, like J/ψ, $\Upsilon(1S)$, $\Upsilon(2S)\ldots$, may be described by weakly coupled pNRQCD, while the radii of the excited states are larger and presumably need to be described by strongly coupled pNRQCD. All this is valid for states away from the strong-decay threshold, i.e. the threshold for a decay into two heavy-light hadrons. In the first case the dynamical degrees of freedom are $Q\bar{Q}$ pairs in color singlet or color octet configuration and ultrasoft gluons, in the second case just $Q\bar{Q}$ pairs in color singlet. The details of the two theories have been presented in Sect. 6.1. The nonperturbative physics in pNRQCD is encoded in a few low-energy correlators that depend only on the gluons and are gauge invariant: these are objects in principle ideal for lattice calculations. Strongly coupled pNRQCD allows us to obtain a definition of the potentials that are given in terms of Wilson loops also in generalized form (i.e. with the insertion of chromoelectric and chromomagnetic field in the static loop). The static potential is given by the static Wilson loop described before that was calculated on the lattice since the inception of QCD [97,351,1568,2640,2644], up to the present state of the art that includes four dynamical quarks in the calculation, see Fig. 206. Some of these potentials have been obtained before the advent of the EFT in the so-called Wilson-loop approach [97,803,1501,1502,2645,2646], but they were missing the contribution of the hard scale. Moreover in the EFT, new (spin-independent) contributions appear at the order $1/m_h$ and at the order $1/m_h^2$ [1556,1557]. The results of strongly coupled pNRQCD – which are valid in the regime in which $m_h v$ is of order $\Lambda_{\rm QCD}$ and where strong decay thresholds are far away – justify the success of the quark model from the QCD perspective. In fact in this regime the only degree of freedom is the $Q\bar{Q}$ singlet, the dynamics is controlled by the Schrödinger equation and ultrasoft corrections are carried only by pions. The potentials, however, are calculated from QCD and they have a structure that is different from what one gets in models, especially for terms related to momentum dependent contributions. This EFT description allows for modifications that could be used to describe X, Y, Z exotics and (combining with finite temperature QCD and open quantum system) the nonequilibrium evolution of quarkonium in medium, as it will summarized later.

8.6.3 The quarkonium potential and confinement

The lattice QCD evaluation of the static Wilson loop clearly displays an area law which is the sign for confinement. Still, it is relevant to investigate the nature of the confinement mechanism. Quarkonium is a golden tool for this aim. Strongly-coupled pNRQCD realizes a scale factorization encoding the low-energy physics in the Wilson loop and its generalized versions, i.e. Wilson loops with insertions of chromoelectric and chromomagnetic fields. All the potentials, static ones and spin and velocity-dependent ones, are given in terms of these gauge invariant nonperturbative objects that no longer depend on the heavy quark degrees of freedom and on the quark flavor. This turns out to be a systematic method to study the QCD confinement properties and put them directly in relation to the quarkonium phenomenology.

The area law emerging in the static Wilson loop at large distance corresponds to the formation of a chromoelectric flux tube between the quark and the antiquark that sweeps the area of the loop: this has been directly observed on the lattice, see Fig. 207. The effects originates from the nonperturbative QCD vacuum that can be imagined as a disordered

medium with whirlpools of color on different scales, thus densely populated by fluctuating fields whose amplitude is so large that they cannot be described by perturbation theory [439]. A QCD vacuum model can be established by making an assumption on the behavior of the Wilson loop in the low energy. The relativistic corrections that involve insertions of gluonic fields in the Wilson loop follow via functional derivative with respect to the quark path see [2638,2646]. One may notice that the part proportional to the square of the angular momentum in the velocity dependent potential at order $1/m_h^2$ takes into account the energy and the angular momentum of the flux tube, which is something that could not be obtained e.g. in any Bethe Salpeter approach with a confining interaction represented by a scalar convolution kernel. The action density or the energy density structure between the static quark and the static antiquark is currently studied both in the lowest energy configuration as well as in the hybrid configurations with excited glue [2647–2649]. The mechanism underlying confinement and flux tube formation has been investigated since long on the lattice [432] using Wilson loops and the't Hooft abelian projection, to identify the roles of magnetic monopoles [2650,2651] and center vortices [438], see e.g. the review [1465].

In the continuum, several models of low-energy QCD have been introduced to explain the flux-tube formation. The models vary from the dual Meissner effect and a dual-abelian Higgs-model picture, from dual QCD [2652] to the stochastic vacuum [2653], to the flux tube model [2418] and an effective low energy string description. Each of these models can be used to obtain analytic estimates of the behavior of the generalized Wilson loops for large distance, which in turn give the static potential and the relativistic corrections as function of r, see e.g. [1571,2654–2657]. Similar nonperturbative configurations leading to confinement can be studied analyzing the Wilson loop in case of baryons with three or two heavy quarks [2658,2659].

8.6.4 States below threshold: quarkonium

On the basis of EFTs and lattice calculations we have reached today a comprehensive understanding of the properties of quarkonium below the strong decay threshold.

Spectra, transition and decays

The power counting of the EFT allows us to attach an error to each prediction. For states with a small radius one can use weakly-coupled pNRQCD with potentials calculated at high order in perturbation theory (see Sect. 6.1) and retardation effects carried by local or non-local condensates. For states with larger radii, the potentials obtained in strongly-coupled pNRQCD (see Sect. 6.1) have been calculated on the lattice [804,1566,1568,2661–2663], and the full quarkonium phenomenology may be obtained using such potentials in the Schrödinger equation. The imaginary parts of the potentials control the decays.

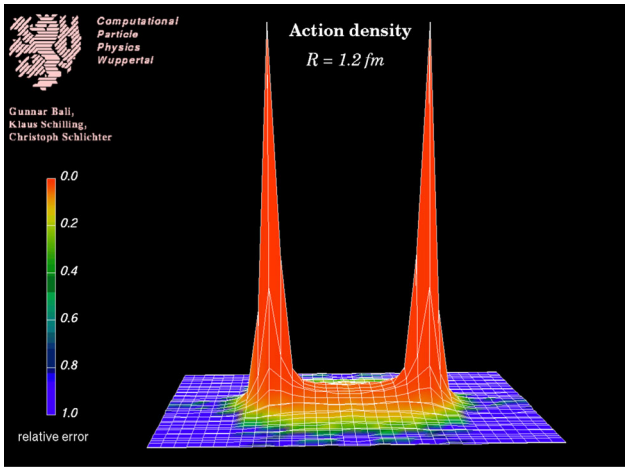


Fig. 207 The origin of the linear potential between the static quark and antiquark may be traced back to a flux tube: a string of gluon energy between the quark pair. Here we present the historical picture of the action density distribution between a static quark antiquark couple in SU(2) at a physical distance of 1.2 fm, from [2660]

On the other hand, direct lattice calculations of quarkonium properties along the years have reached realistic values for the dynamical quark flavors, solid continuum limits and have been extended to the excited states, reaching a comprehensive and precise description [1474,2664–2668]. Electromagnetic M1 and E1 transitions have been calculated in pNRQCD, see e.g. [1551–1554]. There are so many results that it is impossible to discuss all of them here and we refer to some reviews [1462,1464,1465,1518,2630].

Summarizing: today we understand in a precise way, on the basis of QCD, most of the properties of the quarkonium states lying below the strong decay threshold. This has a great impact for the physics of quarkonium, as we explain with examples in the next paragraph.

Precise determination of SM parameters

In the regime in which the soft scale is perturbative, pNRQCD enables precise and systematic higher order calculations of bound states allowing us to extract precisely standard-model parameters like the quark masses and α_s. For example, it has been possible to extract a precise value of α_s at rather low energy by comparing a short-distance lattice calculation with 2+1 flavors of the static energy and a next-to-next-to-next leading order (NNNLO, α_s^4) perturbative pNRQCD calculation of the same quantity, including also ultrasoft logs resummation. When α_s is extrapolated to the mass of the Z, $\alpha_s(M_Z) = 0.11660^{+0.00110}_{-0.00056}$ is obtained. This is a competitive extraction made at a pretty high order of the perturbative expansion [2669,2670]. This method of α_s extraction is now used by several groups, see e.g. [2671,2672]. The QCD static force, defined in terms of a single chromolectric insertion in

the Wilson loop could be used as well to the same scope [2673,2674].

In the same way precise values of the bottom and charm masses can be extracted from measurements of the masses of the lowest states and by comparing them to the formula for the energies in pNRQCD at NNNLO. The renormalon ambiguity between the mass and the static potential cancels and a pretty good determination is possible, see e.g. [1518, 1527,1537] and references therein. Similar methods [1517, 1537] can be used to describe the top anti-top S-wave pair-production cross section near threshold in e^+e^- annihilation and to study the possible achievable accuracy of top-quark mass measurement expected at a future linear collider. A precise determination of the top quark mass is very important for precision tests of the SM, and also due to its crucial role in the vacuum stability of the SM at a very high energy scale. Hence, progress in our understanding of heavy quarkonia leads to an access to key aspects of the SM.

8.6.5 Production

Production of heavy quarkonium has been extensively studied along the years at the Tevatron collider at Fermilab, at Hera at DESY, at B factories and in particular at the LHC where quarkonium production with high statistics at unprecedented values of p_T is measured [1464,1465,1477,1478, 2630–2632]. This is a complex problem encompassing many physical scales still not fully understood, and it can be used to test and extend our understanding of factorization theorems, which are the foundation for all the perturbative calculations in QCD. New theoretical concepts that have been developed here, e.g. arising from kinematic enhancements and from large endpoint logarithms, could have wider applicability in the calculation of high-energy cross sections. Quarkonium production is also relevant to BSM, as certain quarkonium production processes can be used to measure Higgs couplings.

The standard method for calculating quarkonium production rates is the NRQCD [1476] (see Sect. 6.1) factorization approach, where production rates are expressed as perturbatively calculable partonic cross sections multiplied by nonperturbative constants called NRQCD long-distance matrix elements (LDMEs) which are universal. The NRQCD factorization approach is a conjecture that has not been proven to all orders in α_s. Another important theoretical development is the next-to-leading-power (NLP) fragmentation approach [1275,2063], in which quarkonium production rates are expressed as perturbatively calculable partonic cross sections convolved with fragmentation functions, up to corrections suppressed by a factor m_h^4/p_T^4. The NLP fragmentation approach becomes more predictive if NRQCD factorization is used to express the fragmentation functions in terms of NRQCD matrix elements. This organizes the NRQCD factorization expression for the cross section according to powers of m_h^2/p_T^2, which simplifies the calculation of higher-order corrections and the resummation of large logarithms. NRQCD factorization predictions have now been computed at NLO in α_s for many production processes.

The NRQCD approach has brought a great progress into the field even though not all experimental data are understood coherently, and the extraction of the LDMEs remains a complex enterprise [2675–2681]. Recently, it has been possible to factorize the quarkonium production-cross sections at lower energy in pNRQCD [1573–1575], rewriting the octet NRQCD LDMEs, which are the nonperturbative unknowns, in terms of products of wave functions and gauge invariant low energy correlators depending only on the glue and not the on flavor quantum numbers. This allows to reduce by half the number of LDMEs, opens up the possibility of their lattice evaluation and may lead to further progress of the field.

8.6.6 Nonequilibrium evolution in medium

The properties of production and absorption of quarkonium in a nuclear and hot medium are crucial inputs for the study of QCD at high density and temperature (see Sect. 7), reaching out to cosmology.

Heavy ions experiments at the LHC at CERN and at the RHIC at BNL aim at producing the Quark Gluon Plasma (QGP): heavy quarks are good probes of this hot QCD medium. They are produced at the beginning of the collision and remain up to the end. As we discussed, quarkonia are special hard probes as they are multi-scale systems. In the medium besides the energy scales of quarkonium, also the thermal scales of the QGP have to be considered (cf Sect. 6.5): the scale πT related to the temperature, the Debye mass $m_D \sim gT$, with $\alpha_s = g^2/4\pi$, related to the (chromo) electric screening and the scale g^2T related to the (chromo) magnetic screening. In a weakly-coupled plasma, the scales are separated and hierarchically ordered, in a strongly coupled plasma, $m_D \sim T$. To calculate QCD at finite T in real time, Hard Thermal Loop EFT can be used to integrate out the temperature scale. Heavy quarkonium dissociation has been proposed a long time ago as a clear probe of QGP formation through the measurement of the dilepton decay-rate [2152]. The dissociation was related to the screening of the quark–antiquark interaction due to the Debye mass and it was suggested that dissociation would manifest itself in an exponential screening term $\exp(-m_D r)$ in the static potential. One of the key quantities measured in experiments is the nuclear modification factor R_{AA}, a measure for the difference of quarkonium production in pp and in nucleus–nucleus collisions. Since higher excited quarkonium states have larger radii, the expectation was that, as the temperature increases, quarkonium would dissociate first for the higher-mass and

then for lower-mass states giving origin to sequential melting [2152].

In the last decades, using pNRQCD at finite T [1592, 2159], it has been possible to actually define and calculate the $Q\bar{Q}$ potential in medium. In perturbative calculations it was found that the thermal part of the static potential has a real part (roughly described by the free energy) and an imaginary part. The imaginary part comes from two effects: the Landau damping [1592,1593,2153], an effect existing also in QED, and the singlet to octet transition, existing only in QCD [1592]. Which one dominates depends on the ratio between m_D and E. In the EFT one could show that the imaginary part of the potential related to the Landau damping comes from inelastic parton scattering [1596] and the singlet to octet transition from gluon dissociation [1595]. The existence of the imaginary part, first realized in Ref. [2153], changed our paradigm for quarkonium suppression: it has become clear that the state dissociates well before the conventional screening becomes active [1593,2153]. A similar pattern emerges in lattice nonperturbative calculations of the potential [2682,2683].

So far, we have discussed an equilibrium description. However, the evolution of quarkonium in the QGP is an out-of-equilibrium process in which many effects enter: the hydrodynamical evolution of the plasma and the production, dissociation and regeneration of quarkonium in the medium, to quote the most prominent ones. It is necessary therefore to introduce an appropriate framework to describe the real-time nonequilibrium evolution. Recently, using the formalism of open quantum system (see Sect. 6.6) and pNRQCD, it has been possible to describe the nonequilibrium evolution of bottomonium inside a strongly coupled QGP, in a way that incorporates the quantum effects, conserves the number of heavy quarks and considers both color singlet and color octet quarkonium degrees of freedom as well as their recombination [1599,1600]. The results not only describe well the R_{AA}s measured at LHC [1602,2150] (see also [2158,2684]), but they allow also to establish a connection with QCD, since the quarkonium evolution depends only on two transport coefficients given in terms of QCD gluonic correlators characterizing the QGP [1599,1600]. For a review of open quantum system approaches for quarkonium, see [1601,2685].

8.6.7 States at and above threshold: X, Y, Z exotics: intro

As explained in Sect. 8.5, the spectroscopy of charmonium and bottomonium states at or above the open-heavy-flavor thresholds have reserved us several surprises. Experiments at e^+e^- and hadron colliders have discovered many new, unexpected states in the last decades, cf. Sect. 8.5 and Fig. 197. Many of these states are surprisingly narrow, and some have electric charge. The observations of these charged quarkonium states are the first definitive discoveries of manifestly exotic hadrons. These results challenge our understanding of the QCD spectrum. The X, Y, Z offer us a unique opportunity to investigate the dynamical properties of strongly correlated systems in QCD.

As mentioned in Sect. 8.5, These states have been termed X, Y, Z in the discovery publications, without any special criterion. Meanwhile, the Particle Data Group (PDG) has proposed a new naming scheme [2686], that extends the scheme used for ordinary quarkonia, in which the new names carry information on the J^{PC} quantum numbers, see [1427] for more details. Since the situation is still in evolution we will stick to X, Y, Z names. The field is in enormous and very fast development both experimentally and theoretically, with a continuous flux of new papers: we refer to reviews to account for this development [1427,1464,2630,2687–2692]. The X, Y, Z states offer us unique possibilities for the investigation of the dynamical properties of strongly correlated systems: we should develop the tools to gain a solid interpretation from the underlying field theory, QCD. This is a very significant problem with trade off to other fields featuring strong correlations and pretty interesting connections to heavy ion physics, as propagation of these states in medium may help us to scrutinize their structure and composition.

8.6.8 X, Y, Z models and degrees of freedom

Since the $X(3872)$ discovery in 2003, a wealth of theoretical papers appeared to investigate the characteristics of the exotics. Most papers are based on models, which involve a choice of some dominant degrees of freedom and an assumption on their interaction hamiltonian. In the case of states particularly close to their heavy-light threshold, with a very small binding energy and a large scattering length, a more universal picture based on an effective-field-theory molecular description has been put forward [2692–2696] and along the years it has been refined arriving at detailed calculations of the line shapes and the production properties.

A priori the simplest system consisting of only two quarks and two antiquarks (generically called tetraquarks) is already a complicated object and it is unclear whether or not any kind of clustering occurs in it. To simplify the problem, models focus on certain substructures and investigate their implications: in hadroquarkonia the heavy quark and antiquark form a compact core surrounded by a light-quark cloud; in compact tetraquarks the relevant degrees of freedom are compact diquarks and antidiquarks; in the molecular picture two color singlet mesons are interacting at some typical distance: we have no chance here to illustrate all these models and we refer to some recent reviews [1427,2688,2690,2692]. Discussions about exotics therefore often concentrate on the "pictures" of the states, for example the tetraquark interpretation against the molecular one (of which both several different realizations exist). However, as a matter of fact, all the light degrees

of freedom (light quarks, glue, in different configurations) should be there in QCD close or above the strong decay threshold: it is a result of the strong dynamics which one sets in, and when, and which configuration dominates in a given regime.

Even in an ordinary quarkonium, which has a dominant $Q\bar{Q}$, subleading contributions of the Fock space may contribute, which have additional quark–antiquark pairs and active gluons. However, in the most interesting region, close or above the strong decay threshold, where the X, Y, Z have been discovered, the situation is more complicated: there is no mass gap between quarkonium and the creation of a pair of heavy-light mesons, nor to gluon excitations, therefore many additional states appear and are dynamical degrees of freedom to be considered [2687]. Still, m_h is a large scale, and a scale factorization is applicable so that nonrelativistic QCD is still valid. There is still another scale separation that can be used to introduce a description of the bound state similar to what is done in pNRQCD, in which the zero order problem is the Schrödinger equation. Let us consider bound states of two nonrelativistic particles and some light-quark degrees of freedom, e.g. molecules in QED or quarkonium hybrids ($Q\bar{Q}g$ states) or tetraquarks ($Q\bar{Q}q\bar{q}$ states) in QCD: electrons, gluon fields or light quarks fields change adiabatically in the presence of heavy quarks or nuclei. In this situation the interaction between the heavy quarks or the one between nuclei due to the electron cloud may be described at leading order of a nonrelativistic expansion by an effective static energy (or potential) E_κ between the static sources where κ labels different excitations of the light degrees of freedom. A plethora of states can be built on each of the static energies E_κ by solving the corresponding Schrödinger equation, see Figs. 208 and 209. Based on this scale separation one may describe hybrids and tetraquarks using a Born-Oppenheimer (BO) description, similarly to what is done in molecular systems. On the basis of this, the QCD static energies in presence of a static quark and a static antiquark can be classified according to representations of the symmetry group $D_{\infty h}$, typical of diatomic molecules, and labeled by Λ_η^σ (see Fig. 210): Λ is the rotational quantum number $|\vec{\hat{r}} \cdot \vec{K}| = 0, 1, 2, \ldots$, with $\vec{K}$ the angular momentum of the gluons (or in general the nonperturbative collective degrees of freedom), that corresponds to $\Lambda = \Sigma, \Pi, \Delta, \ldots$; η is the CP eigenvalue ($+1 \equiv g$ (gerade) and $-1 \equiv u$ (ungerade)); σ is the eigenvalue of reflection with respect to a plane passing through the $Q\bar{Q}$ axis. The quantum number σ is relevant only for Σ states. In general there can be more than one state for each irreducible representation of $D_{\infty h}$: higher states are denoted by primes, e.g., $\Pi_u, \Pi_u', \Pi_u'', \ldots$. In presence of a light quark that takes part in the binding, isospin quantum numbers should be added. The QCD static energies, E_Γ in Fig. 211, have been calculated on the lattice

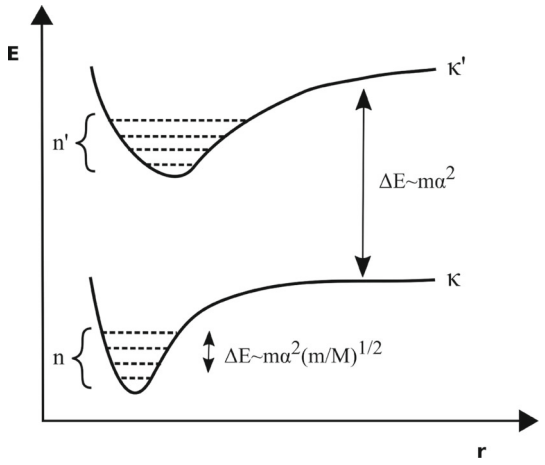


Fig. 208 Pictorial view of electronic static energies in QED, labelled by a collective quantum number κ

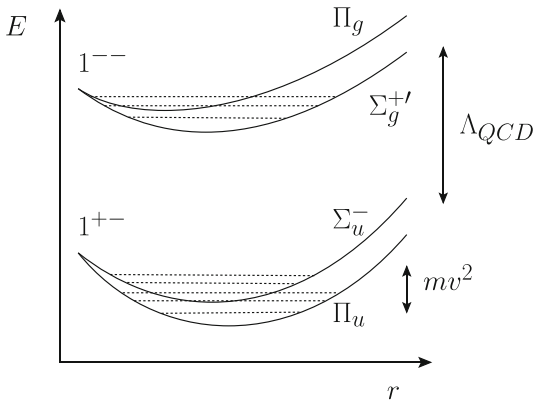


Fig. 209 Pictorial view of the QCD static energies, E_Γ, in QCD. The collective quantum number κ has been detailed in Λ_η^σ as explained in the text. It corresponds to the actual lattice results in Fig. 211

in NRQCD more than 20 years ago [415], Γ representing a choice for Λ_η^σ or in short for the collective quantum number κ. These lattice calculations uses Wilson loops with initial and final states encoding the given quantum numbers, to select a given symmetry. The Born Oppenheimer approximation idea has been first exploited in phenomenological applications by [1586,1588]. This picture may be made precise inside QCD using NREFTs and has the possibility to subdue many different models and pictures. In the next section the content of BOEFT and its implications will be presented.

8.6.9 BO effective field theory

Starting from pNRQED/pNRQCD the BO approximation can be made rigorous and cast into a suitable EFT called Born-Oppenheimer EFT (BOEFT) [1427,1461,1587,1589, 2697–2699] which exploits the hierarchy of scales $\Lambda_{\rm QCD} \gg m_h v^2$.

In Ref. [1589] the BOEFT that describes hybrids has been obtained. In particular, the static potentials and a set of coupled Schrödinger equations were derived and solved to produce the hybrids multiplets for the two first static energies Σ_u^- and Π_u. Such static energies are degenerated at short distance where the cylindrical symmetry gets subdue to a $O(3)$ symmetry and are then labelled by the quantum number of a gluonic operator 1^{+-} called a gluelump. The hybrid static energies are described by a repulsive octet potential plus the gluelump mass in the short distance limit. The $O(3)$ symmetry is broken at order r^2 of the multipole expansion. In the long distance regime the static energies display a behavior linear in r , cf. Fig. 211. The gluelump correlator can be calculated on the lattice to determine the gluelump mass. It depends on the scheme used (the scheme dependence cancels against the analogous dependence in the quark mass and in the octet static potential) but it is of the order of 800 MeV. The hybrid multiplets H_i are constructed from the solution of the Schrödinger equations in correspondence to their J^{PC} quantum numbers. The coupling between the different Schrödinger equations is induced by a non-adiabatic term, known in the Born-Oppenheimer description of diatomic molecules, induced by the non-commutation between the kinetic term and a projector of the cylindrical symmetry in the BOEFT lagrangian. The degeneracy of the static energies at small distance induces a phenomenon called Λ doubling, removing the degeneration between multiplets of opposite parity. This phenomenon is known in molecular physics but with smaller size. This and the structure of the multiplets differ from what is obtained in models for the hybrids, cf. [1589]. The BOEFT hybrid multiplets can be compared to neutral exotic states measured in the bottomonium and charmonium sector [1427]: there are many experimental candidates and to make clear identifications one should study also the decay and production properties in the same framework.

The picture can be generalized to tetraquarks by considering static energies classified by the D_∞ quantum numbers and isospin quantum numbers, extracted from lattice evaluation of the static energies of system of a heavy quark, a heavy antiquark and two light quarks [1461].

Exotic spin structures, decays

In Refs. [2698,2699] the spin-dependent potential of hybrids has been obtained at order $1/m_h$ and $1/m_h^2$ in the quark mass expansion. The potential turned out to be rather different from the spin potential known from standard quarkonium. In fact, a $1/m_h$ contribution appears due to the coupling of the angular momentum of the gluonic excitation (which is not suppressed in m_h) with the total spin of the heavy-quark–antiquark pair. Among the $1/m_h^2$ operators are the standard spin–orbit, total spin squared, and tensor spin operators respectively, which appear for standard quarkonia. But now three novel operators appear in addition. So interestingly and differently from the quarkonium case, the hybrid potential gets a first contribution already at order $\Lambda_{\rm QCD}^2/m$. Hence, spin splittings are remarkably less suppressed in heavy quarkonium hybrids than in heavy quarkonia: this will have a notable impact on the phenomenology of exotics. The nonperturbative low-energy correlators appearing in the factorization can be extracted by fixing them using lattice data on the masses of charmonium hybrids [2667], Then, all bottomonium-hybrid spin-multiplets (more difficult to evaluate on the lattice [1474])

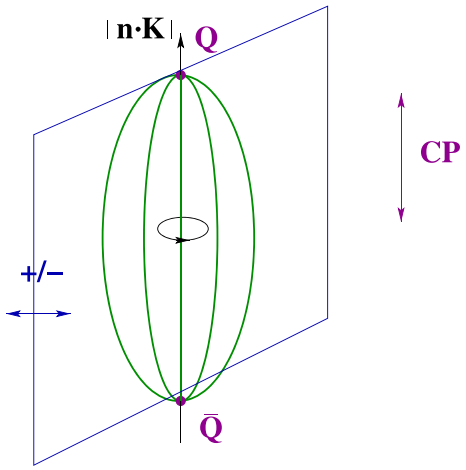


Fig. 210 Symmetries of a system with a static quark and a static antiquark and a nonperturbative cloud (gluonic, light quarks) and respective quantum numbers

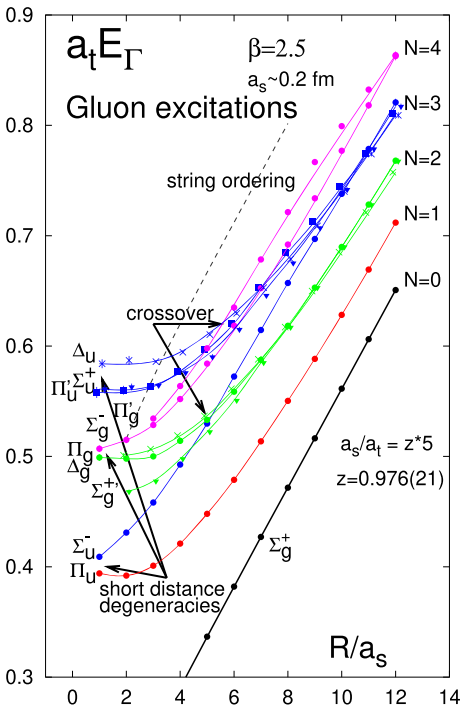


Fig. 211 Hybrid static energies labeled by the D_∞ group quantum number, E_Γ, in lattice units, from the historical Ref. [415]. For updated calculations see [2700]

can be predicted. The BOEFT is therefore able to predict all spin hybrid multiplets, including their decays and transitions [1427].

Avoided level crossing

In the BOEFT the information is carried by the QCD static energies and a few purely gluonic low-energy correlators. The information is relevant, however to describe the static energies in the region close to the threshold of two heavy-light mesons. A phenomenological description has been put forward recently [2701,2702] inspired by lattice calculation in which the avoided-level-crossing-effect is exploited to construct a set of coupled Schrödinger equations and a procedure for the calculation of open-flavor meson–meson scattering cross sections from diabatic potentials. In this framework, the $X(3872)$ naturally emerges [2701]. Such a description can be carried over to BOEFT.

The BOEFT may be used to describe also tetraquarks, double heavy baryons and pentaquarks [1461,1587]. In the case of tetraquarks, a necessary input of the theory is the calculation on the lattice of the generalized Wilson loops with appropriate symmetry and light quark operators. Note that besides the quantum number κ also the isospin quantum numbers $I = 0, 1$ have to be considered. It is interesting to note that the BOEFT approach reconciles the different pictures of exotics based on tetraquarks, molecules, hadroquarkonium…In fact in the plot of the static energy as a function of r for a state with $Q\bar{Q}q\bar{q}$ or $Q\bar{Q}g$ we will have different regions: for short distance a hadroquarkonium picture would emerge, then a tetraquark (or hybrid) one and, when passing the heavy-light mesons line, avoided-cross-level effects should be taken into account and a molecular picture would emerge. QCD would then dictate, through the lattice correlators and the BOEFT characteristics and power counting, which structure prevails and in which precise way. In addition production and suppression in medium may be described in the same approach [1574,2149].

8.6.10 X, Y, Z lattice

Lattice QCD plays a key role for the description of exotics [2666,2703]. For what concerns BOEFT, nonperturbative input from the lattice is needed in the form of static energies, gluelumps correlators, insertions of chromoelectric fields on hybrids states, for a full list see [2687]. Lattice groups have started to calculate some of these crucial quantities [2704–2707].

Direct lattice calculations of the spectrum and properties of exotic states at and above thresholds are extremely challenging. These states are resonances in the pertinent multi-hadron channels and to obtain their properties, scattering amplitudes in the relevant kinematic range should be calculated on the lattice. The approach is based on Lüscher's work. Later generalizations give access to scattering amplitudes of two-hadron elastic channels, of multiple coupled two-hadron inelastic channels, and of three-hadron channels. While the Lüscher method for a single two-hadron elastic channel provides a straightforward mapping between scattering amplitudes and finite-volume energies, this connection is lost for the multi-channel case, and a parameterization of the amplitude is needed. Abundant and precise energy eigenvalues in a given kinematic range should be obtained to constrain these multi-parameter forms, with solid systematic uncertainties. As the calculations move toward physical values of the light-quark masses, the multi-hadron thresholds move towards lower energies and the number of kinematically allowed hadronic channels that need to be included in the determination of scattering amplitudes increases, making everything more challenging. Still interesting information about some exotics mesons could be obtained in these direct lattice calculations, see e.g. [584,2708].

8.6.11 Summary

Quarkonium has been at the origin of QCD. It has been a long way to arrive at describing quarkonium within QCD. It has payed off, making quarkonium a special probe of strong interactions at zero and at finite temperature. We are now in the process to attack the next frontier, i.e. to develop a coherent, field-theory-based description of exotic quarkonium states. Notice, that if new physics involves nonrelativistic bound states, then the techniques that have been developed for understanding quarkonia will be directly applicable. This holds for example for studies of pairs of heavy dark matter in the evolution in the early universe, that well match the studies of the nonequilibrium evolution of quarkonium in medium, or for the production and the spectroscopy of heavy particles of BSM.

9 Baryons

Conveners:

Volker Burkert and Franz Gross

As we are trying to make progress in the complex world of physical sciences, we should not lose sight of what physics is all about: understanding the origin and the history of our universe, and the laws underlying the observations. In this section we also address how excited states of the nucleon fit in to our understanding of the forces and the dynamics of matter in the history of the universe. On the internet we find beautiful representations of the phases through which the universe evolved from the Big Bang (BB) to our times as shown in Fig. 212.

Existing electron accelerators as CEBAF, ELSA, and MAMI, and colliders as BES III have sufficient energy reach

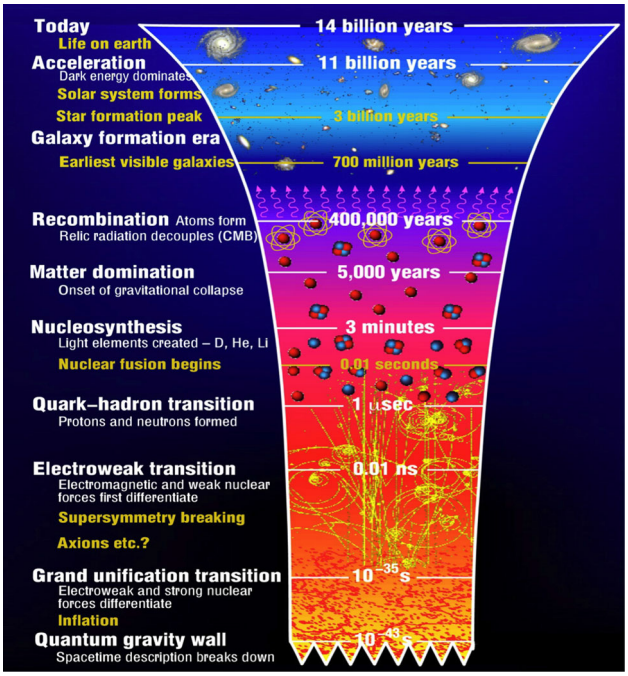


Fig. 212 The evolution of the Universe. The line denoted as *Quark–hadron transition*, is where protons and neutrons are formed

to access this region and study processes in isolation that occurred during this transition in the microsecond old universe and resulted in the freeze out of baryons. There are some marked events that have been of particular significance during the early phases of its history, such as the quark–gluon plasma of non-interacting colored quarks and gluons, and the forming of protons and neutrons. During this transition dramatic events occur – chiral symmetry is broken, quarks acquire mass dynamically, baryon resonances occur abundantly, and colored quarks and gluons are confined. This crossover process is governed by the excited hadrons. During this period strong QCD (sQCD) emerges as the process describing the interaction of colored quarks and gluons.

These are the phenomena that we are exploring with electron and hadron accelerators – the full discovery of the baryon (and meson) spectrum, the role of chiral symmetry breaking and the generation of dynamical quark mass in confinement. While we can not recreate the exact condition in the laboratory, with existing accelerators we can explore these processes in isolation. With electron machines and high energy photon beams in the few GeV energy range we search for undiscovered excitation of nucleons and other baryons.

In this section, Capstick and Crede give an overview over the spectrum of light-quark baryons, followed by a review o the present experimental status by Burkert, Klempt and Thoma. The structure of baryon resonances is explored in electroproduction experiments (Burkert). The section ends with a review of baryons with heavy quarks.

9.1 Theoretical overview of the baryon spectrum

Simon Capstick and Volker Crede

9.1.1 Overview

This contribution examines the constraints on the baryon spectrum imposed by general considerations of flavor, rotational, parity, and particle-exchange symmetries, which lead to a classification scheme for excited baryons. Theoretical approaches to a description of the baryon spectrum based on constituent quark models with various methods for treating the short-range range interactions between quarks are described, and are contrasted to investigations of the spectrum based on lattice and Dyson–Schwinger equation approaches to QCD. Models predict more excited states than are present in the spectrum extracted from data; considerations of how these missing states decay point to alternative ways to produce them, and how to detect their presence once produced. Finally, hybrid baryons with explicit gluonic excitations and the prospect for their discovery are discussed. More detail is given in, for example, reviews of the theoretical approach to the baryon spectrum in Refs. [2709–2711], and reviews of recent experimental developments in Refs. [2712,2713].

9.1.2 Symmetry, group theory, and constraints on the baryon spectrum

Exchange symmetry, baryons, and the color degrees of freedom

The development of $SU(3)_f$ and its isospin subgroup in order to understand the proliferation of what are now known as ground-state baryons led to an understanding that there are states with flavor wave functions that are totally symmetric under exchange of up and down quarks, which are identical in the isospin-symmetric limit. An example is the isospin-3/2 baryon, Δ, with the four charge states Δ^{++}, Δ^{+}, Δ^{0}, and Δ^{-}, three of which were discovered in early πN elastic scattering experiments with charged pion beams, and shown by examining their strong decays to have spin and parity $J^P = 3/2^+$. This led to a paradox: Ground states of few-body systems made up of identical particles usually have spatial wave functions with orbital angular momentum and parity $L^P = 0^+$, and are exchange symmetric. This implies a total quark spin $S = 3/2$, which is also totally symmetric under exchange of the spin-1/2 quarks. However, as fermions, the Pauli principle requires that the wave function of these baryons in the product flavor, spatial, and spin space is totally *antisymmetric*.

The solution is to assign to the quarks an additional degree of freedom, and a wave function in this degree of freedom

which is totally antisymmetric under exchange of the quarks. The simplest way to do this is with a three-valued degree of freedom, now called color. QCD was developed when it was realized that this would result from an $SU(3)_c$ symmetry, where the strong interactions are independent of rotations of the quarks in the color space, with baryons as automatically totally antisymmetric color singlets. This naturally led to a gauge theory which could be the basis for the strong interactions between quarks, and by extension, between all hadrons.

Flavor symmetry in baryons

There is an approximate $SU(3)_f$ symmetry of the strong interactions under exchange of the quarks u, d, and s, which is broken by the higher mass of the strange quark. Gell-Mann [1604] and Okubo [15] were able to write a mass formula for ground-state decuplet ($J^P = 3/2^+$), shown in Fig. 213, and separately for octet ($J^P = 1/2^+$) baryons, shown in Fig. 214, in terms of the eigenvalues of the two generators of the Lie algebra of $SU(3)_f$ that can be simultaneously diagonalized. These generators are the third component of isospin I_3, and the hypercharge $Y = B + S$, where B is baryon number and S is strangeness ($S = -1$ for a strange quark, for historical reasons). Hypercharge is represented by the diagonal matrix $(1, 1, -2)$ in the $\{u, d, s\}$ flavor space. The Gell-Mann-Okubo mass formula ascribes the breaking of symmetry in hadrons to differences in the hypercharge, now understood to be due to the larger mass of the strange quark. It is realized in the ground-state octet baryons as

$$(M_N + M_\Xi)/2 = (3M_\Lambda + M_\Sigma)/4,$$

which holds to a fraction of a percent accuracy, and in ground-state decuplet baryons as the equal-spacing rule

$$M_{\Sigma^*} - M_\Delta = M_{\Xi^*} - M_{\Sigma^*} = M_\Omega - M_{\Xi^*},$$

each approximately 147 MeV, which can be thought of as the difference in the strange and average light (u, d) quark masses. The latter led to the prediction by Gell-Mann [2714] of the existence at around 1680 MeV of the decuplet Ω baryon, made up of three strange quarks. Although the formula is phenomenological, it is now understood in the context of chiral perturbation theory.

De Rujula, Georgi and Glashow [763] were able to explain the above results in the context of a model of hadrons which confined the quarks with a flavor and spin-independent interaction, and used a short-distance interaction between the quarks that results from asymptotic freedom. This is the result of one-gluon exchange, and led to a short-distance potential between two quarks that was Coulomb in nature, and could be interpreted of as due to interactions between two colored spin-1/2 quarks. The mass dependence of the color-magnetic moments of the quarks led naturally to spin- and flavor-dependent interactions between the quarks, which could also explain the mass differences between octet and decuplet baryons of the same flavor, and allow a qualitative understanding of the sign and size of the difference $\Sigma^0 - \Lambda^0$ between the masses of the $I = 1$ and $I = 0$ neutral strange baryons.

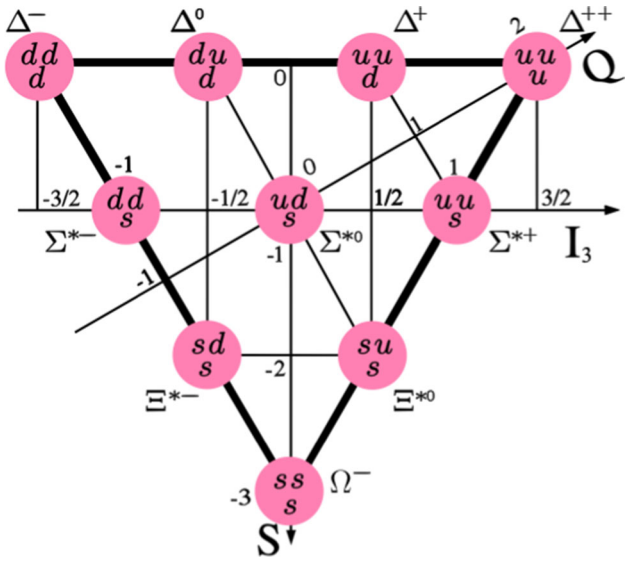


Fig. 213 The ground-state baryon decuplet, with strangeness $(Y - B)$ plotted vs. the third component of isospin I_3

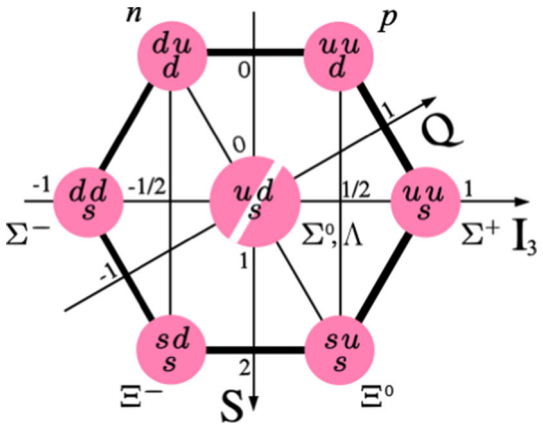


Fig. 214 The ground-state baryon octet, with strangeness plotted vs. the third component of isospin

One consequence of this simple (additive) quark model description of baryons is an understanding of the magnetic moments of the nucleons p, n and other octet and decuplet ground-state baryons. Using the total quark-spin $S = 1/2$ wave function and octet flavor wave functions for the three quarks in nucleons yields

$$\mu_p = \frac{4}{3}\mu_u - \frac{1}{3}\mu_d,$$

and, since the proton can be turned into the neutron by the transformation $u \leftrightarrow d$, we have

$$\mu_n = \frac{4}{3}\mu_d - \frac{1}{3}\mu_u.$$

Fitting this to the measured moments gives quark magnetic moments in the ratio of the quark charges to a good approximation, if we assume the quark masses are identical, and that this light quark mass is around one third of the mass of nucleons. This approach also leads to a qualitative understanding of the magnetic moments of other ground-state octet and decuplet baryons, and the transition moment that affects the rate of the decay $\Sigma^0 \rightarrow \Lambda^0\gamma$. Isgur and Karl [2715] added small contributions to baryon magnetic moments, due to configuration mixing, relativistic corrections, and violations of isospin symmetry, to refine these non-relativistic quark model estimates. The result was better agreement with the moments extracted from experimental data.

Rotational and parity symmetries

Ignoring for now interactions that couple the orbital and spin angular momenta of the quarks, rotational symmetry and the conservation of angular momentum also imply that ground and excited-state baryons should lie in multiplets with a given orbital angular momentum L and total quark spin S, with the overall angular momentum of a baryon given by $\vec{J} = \vec{L} + \vec{S}$. The confining and spin-independent part of the short-range interaction will cause splittings between and within multiplets of states with different orbital angular momentum L, and the short distance interactions between the quarks, for example those in the work of De Rujula, Georgi and Glashow, will further split those multiplets into groups of states with the same total quark spin S.

It is always possible to describe the orbital angular momentum of a basis state used to describe the wave function of a baryon in terms of the angular momentum of the orbital wave functions in the two vectors required to describe the relative positions of the quarks. These can be conveniently chosen to be the Jacobi coordinates

$$\begin{aligned} \vec{\rho} &= \frac{1}{\sqrt{2}}(\vec{r}_1 - \vec{r}_2)\,, \\ \vec{\lambda} &= \frac{1}{\sqrt{6}}(\vec{r}_1 + \vec{r}_2 - 2\vec{r}_3) \end{aligned} \tag{9.1}$$

shown in Fig. 215, where the $\vec{r}_i$ are the vector positions of the three quarks. The total orbital angular momentum is then $\vec{L} = \vec{l}_\rho + \vec{l}_\lambda$, and the parity of the resulting state is $P = (-1)^{l_\rho + l_\lambda}$. It is simple to show that all values of baryon spin and parity can be attained by various choices of the eigenvalues for quarks moving in a static potential; in contrast to the situation in mesons, there are no baryons with 'exotic' quantum numbers. Exotic quantum numbers in mesons require degrees of freedom like those of the glue binding the hadrons together to be in other than their $J^P = 0^+$ ground state.

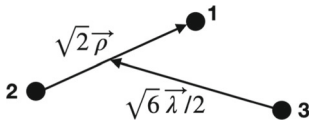


Fig. 215 The three-body Jacobi coordinate vectors $\vec{\rho}$ and $\vec{\lambda}$

This situation is complicated in the presence of spin–orbit (vector in spin and vector in space coupled to a scalar) and tensor ($S = 2$ and $L = 2$ coupled to a scalar) interactions between the quarks. These are present in models which have short-distance interactions between the quarks based on the exchange of a vector boson, such as those due to one-gluon exchange. The evidence for the presence of such interactions in the spectrum of baryons is weaker than that for the presence of interactions which are simultaneous spin and orbital angular momentum scalars; this is discussed in what follows.

The dominance in the baryon spectrum of simultaneous quark spin and orbital angular momentum scalar interactions, when combined with the observation that states assigned quark spin $S = 3/2$ are more massive than those with $S = 1/2$, allowed Klempt [2716] to fit the spectrum of baryons made of $\{u, d, s\}$ quarks with a mass formula. The squares of the masses of baryons are proportional to their orbital angular momenta L, as in Regge theory and, approximately, the spectrum of a linear confining potential. For a given flavor of baryon, more massive recurrences of the same J^P quantum numbers were assigned the same gap in mass-squared as orbital excitations.

Symmetry under particle exchange

The requirement of Pauli symmetry implies, in the isospin symmetric limit, that the wave functions of baryons are totally symmetric under the exchange of light quarks, since the color wave function is totally antisymmetric in the absence of excitation of the gluon fields. For non-strange baryons made up of three light quarks, this means that each component of the wave function must be a basis function for a representation of the exchange group S_3. These basis functions are either totally symmetric (S) under particle exchange, totally antisymmetric (A), or one of a pair with mixed symmetry $\{M_\rho, M_\lambda\}$ that transform into each other under the elements of the permutations of S_3 in a predictable way. Here M_ρ refers to the basis function that is antisymmetric under the exchange $1 \leftrightarrow 2$, and M_λ refers to the basis function that is symmetric under this exchange. The rules for how the mixed-symmetry pair transform into linear combinations of each other under the elements of S_3 can be found, for example, by examining the result of the various permutations on the relative position vectors $\vec{\rho}$ and $\vec{\lambda}$. .

The rules of combining the spin angular momentum of three $S = 1/2$ particles require that the overall spin wave

function of the quarks for all values of the total quark spin projection M_S be either totally symmetric, when $S = 3/2$, or be one of a pair of states of mixed symmetry when $S = 1/2$. The same rules apply for the isospin wave functions for baryons made up of three light quarks, Δ-flavor baryons with $I = 3/2$ and N-flavor baryons with $I = 1/2$.

The lowest lying basis state for the spatial wave functions of ground-state baryons made up only of light quarks have $L^P = 0^+$, and are totally symmetric under quark exchange. Overall exchange symmetry then requires that the flavor and spin wave functions be combined by using the rules for combining two representations of S_3. For Δ baryons this is trivial, since both the spin and flavor wave functions are totally symmetric. For N baryons, the mixed-symmetry spin (χ) and flavor (ϕ) wave functions are combined to the symmetric linear combination

$$\frac{1}{\sqrt{2}}\left(\chi^{M_\rho}\phi^{M_\rho} + \chi^{M_\lambda}\phi^{M_\lambda}\right).$$

The wave functions of baryons with a given flavor and spin-parity J^P can be expanded in a basis of states that satisfy the requirements of antisymmetry under exchange of identical (or nearly identical, for $u \leftrightarrow d$) quarks. A convenient choice is to use a harmonic oscillator (HO) basis, which has the useful feature of being form invariant under Fourier transformation; another is the Sturmian basis, which has improved large momentum behavior useful for calculating decay form factors, but is harder to use in both coordinate and momentum space. Configuration mixing due to the confining potential and the short-range interactions between quarks can be implemented by diagonalization of the Hamiltonian matrix calculated in this basis.

The rules for combining representations of the exchange group S_3 are used to construct this basis from states with given values of radial $\{n_\rho, n_\lambda\}$, orbital $\{L, l_\rho, l_\lambda\}$, and spin S quantum numbers (magnetic quantum numbers have been suppressed, and the sums and Clebsch–Gordan coefficients required to form states of definite $\{L, S, J\}$ are assumed). It is often convenient to expand the wave function up to a given energy, or equivalently polynomial order, which in the HO basis is

$$E = (2n_\rho + l_\rho + 3/2)\hbar\omega_\rho + (2n_\lambda + l_\lambda + 3/2)\hbar\omega_\lambda, \quad (9.2)$$

where ω_ρ and ω_λ are oscillator energies related by $\alpha^2_{\rho,\lambda} = m_{\rho,\lambda}\omega_{\rho,\lambda}$ to the scale α at which the radial wave functions fall with distance and the reduced masses, equal when all three quark masses are the same. Karl and Obryk [2717] give the general procedure up to fourth-order polynomials. Examples of how to construct these bases can be found in [764,2718, 2719]; for a pedagogical overview see, for example, Ref. [2720].

It is not necessary to antisymmetrize the wave functions of baryons under the exchanges $u \leftrightarrow s$ or $d \leftrightarrow s$. It is convenient to use the 'uds' basis [764] for baryons with $S = -1$ or -2, which uses basis states that have either symmetry or antisymmetry under these exchanges.

Baryon resonance classification scheme

The quantum numbers of the total orbital angular momentum $\vec{L} = \vec{l}_\rho + \vec{l}_\lambda$ and spin $\vec{S} = \sum_i \vec{s}_i$ are not good quantum numbers in a relativistic theory. The parity of such states is given by $P = (-1)^{l_\rho + l_\lambda}$. It is always possible to use a basis of states with specific values of L and S which can couple to the total angular momentum J of the baryon being described. These are then mixed in the eigenstates of a Hamiltonian that includes interactions that are not simultaneous scalars in both spin and space, but are overall scalars of the form $\sum_q C_{kq} R_{k,q} S_{k,-q}$, where R_k and S_k are tensor operators of rank k acting on the spatial and spin bases, and the C_{kq} are the coefficients required to make the result an overall scalar. Examples are the tensor ($k = 2$) interactions which occur in models of the short-range interactions between quarks, and spin–orbit ($k = 1$) interactions. To the extent that these interactions are small, a classification scheme based on the $\{L, S\}$ values of the dominant component of a configuration-mixed eigenstate is useful.

It may also be useful to further break down sets of states with the same $\{L, S\}$ values into those with specific spatial symmetries. As an example, consider excitations of N and Δ flavored baryons, which are made up of only $\{u, d\}$ quarks. It is useful to enumerate basis states in a harmonic-oscillator basis. Because of isospin symmetry, this basis has $\omega_\rho = \omega_\lambda$ in Eq. 9.2, so that

$$E = \left[2(n_\rho + n_\lambda) + l_\rho + l_\lambda + 3\right]\hbar\omega = [N + 3]\hbar\omega.$$

Ground states

Baryon states can be classified according to the flavor-spin $SU(6)$ multiplet in which they predominantly lie. This would be an exact symmetry of the Hamiltonian if it was simultaneously invariant under both rotations of quark flavors in the $SU(3)_f$ space, and independent of the spin projections of the quarks. Although useful as a classification scheme, this is clearly only a very approximate symmetry: In addition to the flavor-symmetry breaking effect of the larger strange quark mass, the measured mass difference $M_\Delta - M_N \simeq 300$ MeV shows that interactions between quarks are not independent of their spin. In this scheme the ground-state nucleon lies in an $SU(3)_f$ octet of ground-state baryons with $J = S = 1/2$, which are, in order of increasing strangeness, $\{n, p, \Lambda^0, \Sigma^{+,0,-}, \Xi^{0,-}\}$, giving $(2S + 1) \cdot 8 = 16$ states. The ground-state Δ is a member of an $SU(3)_f$ decuplet of baryons with $J = S = 3/2$, which are $\{\Delta, \Sigma^*, \Xi^*, \Omega\}$, giving $(2S + 1) \cdot 10 = 40$ states. Collectively, these dominantly $L^P = 0^+$ states make up the $SU(6)$ multiplet, labeled $[56, 0^+]$.

Following the notation of Isgur and Karl [2719], we can label harmonic-oscillator basis states by $|X^{2S+1}L_\pi J^P\rangle$, where X is the flavor, L is given in $\{S, P, D \ldots\}$ notation, and π is the spatial exchange symmetry, either totally symmetric (S), mixed symmetry (M), or totally anti-symmetric (A). In this notation, the dominant $N = 0$ components of the ground state non-strange baryons are

$$|N^2 S_S 1/2^+\rangle, \quad |\Delta^4 S_S 3/2^+\rangle.$$

Negative-parity states

The low lying (dominantly $N = 1$) negative-parity non-strange excitations are made up of a triplet of $S = 3/2$ N states and a doublet of $S = 1/2$ N states with mixed flavor symmetry

$$|N^4 P_M(1/2^-, 3/2^-, 5/2^-)\rangle, \quad |N^2 P_M(1/2^-, 3/2^-)\rangle.$$

Their spatial wave functions necessarily have mixed exchange symmetry, since they are proportional to $Y_{1m}(\Omega_\rho) \propto \rho_m$, where ρ_m is a spherical component of the vector $\vec{\rho}$, or $Y_{1m}(\Omega_\lambda) \propto \lambda_m$. There is also a a doublet of $S = 1/2$ Δ states with S flavor symmetry,

$$|\Delta^2 P_M(1/2^-, 3/2^-)\rangle.$$

These negative-parity resonances are members of the $SU(6)$ multiplet $[70, 1^-]$, with two flavor octets of $S = \{1/2, 3/2\}$ states and a decuplet of $S = 1/2$ states [symmetric under $SU(3)_f$], plus a flavor singlet state Λ, with $S = 1/2$. References [764,2717,2720] show how the rules for combining representations of the exchange group S_3 can be applied to yield properly anti-symmetrized basis state wave functions. Basis states with the same flavor, spin, and parity can and will undergo configuration mixing when a model Hamiltonian is diagonalized.

Positive-parity excited states

This scheme can be extended to, for example, positive-parity non-strange excitations, dominantly described by $N = 2$ basis states. This is complicated by the presence of both radial and orbital excitations. There are three radial excitations that have $n_\rho = 1$ or $n_\lambda = 1$, and, in a three-body system, $l_\rho = l_\lambda = 1$ coupled to $L = 0$; linear combinations form $L = 0$ states with definite (S or M) exchange symmetry. There are also orbital excitations with $l_\rho = 2$ or $l_\lambda = 2$, or $l_\rho = l_\lambda = 1$. Since they have positive parity, these baryons must be described by basis states with even $l_\rho + l_\lambda$.

After forming linear combinations of spatial basis states to form anti-symmetrized basis state wave functions, we have a total of 21 basis states that contribute at $N = 2$ to the wave functions of N and Δ states. These are radial excitations

$$|N^2 S_{S'} 1/2^+\rangle, \quad |\Delta^4 S_{S'} 3/2^+\rangle$$

in the $SU(6)$ multiplet $[56', 0^+]$, and

$$|N^4 S_M 3/2^+\rangle, \quad |\Delta^2 S_M 1/2^+\}\rangle, \quad |N^2 S_M 1/2^+\rangle$$

in the $SU(6)$ multiplet $[70, 0^+]$; $L = 2$ orbital excitations that are admixtures of $l_\rho = 2$ and $l_\lambda = 2$,

$$|\Delta^4 D_S(1/2^+, 3/2^+, 5/2^+, 7/2^+)\rangle, \quad |N^2 D_S(3/2^+, 5/2^+)\rangle$$

in the $SU(6)$ multiplet $[56, 2^+]$, and

$$|N^4 D_M(1/2^+, 3/2^+, 5/2^+, 7/2^+)\rangle, \; |\Delta^2 D_M(3/2^+, 5/2^+)\rangle, \; |N^2 D_M(3/2^+, 5/2^+)\rangle,$$

in the $SU(6)$ multiplet $[70, 2^+]$, and $L = 1$ orbital excitations formed from $l_\rho = 1$ and $l_\lambda = 1$

$$|N^2 P_A(1/2^+, 3/2^+)\rangle \tag{9.3}$$

in the $SU(6)$ multiplet $[20, 1^+]$.

The $J^P = 1/2^+$ nucleon and the $J^P = 3/2^+$, isospin $I = 3/2$ Δ ground states have dominant components with $N = 0$, $L = 0$ and $S = 1/2$ and $S = 3/2$, respectively. Spin-independent and spin-scalar (contact) interactions between the quarks (arising from their short-distance interactions and confinement) allow mixing between three basis states: the $N = 0$ ground state $|N^2 S_S 1/2^+\rangle$, and the $L = 0$, $S = 1/2$ states $|N^2 S_{S'} 1/2^+\rangle$ and $|N^2 S_M 1/2^+\rangle$. Tensor (or spin–orbit) interactions cause mixings with the $L = 2$, $S = 3/2$ state $|N^4 D_M 1/2^+\rangle$, and the $L = 1$, $S = 1/2$ state $|N^2 P_A 1/2^+\rangle$. The situation is simpler for the $N = 0$ ground state $|\Delta^4 S_S 3/2^+\rangle$, which mixes with the $L = 0$, $S = 3/2$ radial excitation $|\Delta^4 S_{S'} 3/2^+\rangle$, and the $L = 2$, $S = 3/2$ orbital excitations $|\Delta^4 D_S 3/2^+\rangle$ and $|\Delta^2 D_M 3/2^+\rangle$. The resulting D-wave components in both the N and Δ wave functions can lead to measurable consequences in the photo- and electro-production amplitudes for the transition $\gamma^{(*)} N \to \Delta$. For details, see the review on N and Δ resonance electro-production in the 2022 RPP [616].

Hyperons

If we use a basis of states that imposes $SU(6)$ symmetry despite the larger strange quark mass, this classification scheme extends to the hyperons Λ and Σ, Ξ, and Ω. As an example, the notation of Isgur and Karl [2719] for the ground state $SU(3)_f$ singlet Λ is $|{\Lambda_1}^2 S_M 1/2^+\rangle$. The $SU(3)_f$ singlet wave function is totally antisymmetric under quark exchange, and so is included in the wave function of the radial excitation $|{\Lambda_1}^2 P_A 1/2^+\rangle$, with its antisymmetric spatial wave function; other radial recurrences such as $|{\Lambda_8}^2 S_S 1/2^+\rangle$ necessarily involve the $SU(3)_f$ octet flavor wave functions, so the notation is supplemented by the $SU(3)_f$ multiplet (singlet, octet, or decuplet) in which the state lies. The total number of basis states at each harmonic oscillator level for Ξ baryons, containing two identical strange quarks, is the sum of the number of N and Δ states at that same level. There is a one-to-one

correspondence between basis states for Ω baryons and those of the Δ states.

Constructing the wave functions is made simpler by use of the uds basis [764], which requires overall antisymmetry only under exchange of equal mass quarks. In this basis, the exchange symmetry of the $\vec{\rho}$ and $\vec{\lambda}$ relative coordinate wave functions are specified separately. It is always possible to convert from the uds basis back to a basis with definite $SU(3)_f$ symmetry when that is convenient, for example when calculating strong decays [2721].

9.1.3 Constituent quark models

Constituent quark models treat a baryon as made up of three 'valence' quark degrees of freedom, with the gluon fields providing a static potential in which the quarks move. In flux-tube models [2418,2722,2723] this is treated as the lowest energy state of a system of three strings that meet at a junction, whose energy is proportional to their length. There are several approaches to the treatment of the short-range interactions between the quarks, which are responsible for splitting groups of states which would otherwise be degenerate or have their flavor dependence explained by violations of $SU(3)_f$ symmetry due to the additional mass of the strange quark. These approaches are briefly outlined here.

One-gluon exchange models

The earliest constituent quark models had short-distance interactions based on the exchange of a single gluon [763], which postulate that asymptotic freedom implies that high momentum transfer interactions between quarks are dominated by the exchange of a gluon. The result can be written as the interaction between two color-magnetic dipoles, with a $\vec{\lambda}_i \cdot \vec{\lambda}_j$ dependence on the colors of quarks i and j, and spatial dependence given by the Fourier transform of the vector gluon propagator. Here the λ_i are the generators of $SU(3)_c$ realized in the quark triplet basis. This naturally leads to a spin-independent Coulomb interaction at short range, and, with the assumption of point-like constituent quarks, a 'contact' interaction proportional to

$$\frac{2\alpha_s}{3m_i m_j} \sum_{i<j} \vec{s}_i \cdot \vec{s}_j \delta^3(\vec{r}_{ij}), \tag{9.4}$$

where $\vec{r}_{ij} = \vec{r}_i - \vec{r}_j$ is the relative coordinate of quarks i and j. This approach also results in tensor ($S = 2$ and $L = 2$ coupled to a scalar) and spin–orbit (vector in spin and vector in space coupled to a scalar) interactions between the quarks. There is some evidence for the former in the spectrum of $J^P = 1/2^-, 3/2^-$ nucleon resonances, and from patterns of strong decays of negative-parity excited baryons [2724], for example the $N\eta$ decays of the lightest non-strange $I = 1/2$ (N^*) resonances with $J^P = 1/2^-$, nominally at 1535 and 1650 MeV. Isgur and Karl [764] noted a partial cancellation between spin–orbit interactions resulting from one-gluon exchange and from Thomas precession of the quarks in the confining potential, but the agreement with the spectrum of low-lying negative-parity baryons extracted from data and their one-gluon exchange model was best when they were left out altogether.

The Coulomb, contact and tensor interactions resulting from one gluon exchange were evaluated in low-lying negative parity excited baryons made up of $\{u, d, s\}$ quarks by Isgur and Karl [764]. This was extended to positive-parity excited baryons [2719], where the effects of the difference of the confining potential from that defining the harmonic oscillator basis were also evaluated using perturbation theory. The resulting parameters were fit to the spectrum extracted from data without needing to specify the form of the anharmonicities. In this work and a treatment of ground state baryons [2718], the effects of configuration mixing by the various potentials were taken into account by diagonalization of the Hamiltonian matrix, independently for each sector with $N = 2(n_\rho + n_\lambda) + l_\rho + l_\lambda = 0$ for the ground states, $N = 1$ for the low-lying negative-parity excited states, and $N = 2$ for the positive-parity excited states.

While diagonalization independently by sector has the advantage of simply describing the important physics, the parameters fit to each sector's spectrum may be inconsistent. Systems of light quarks are also relativistic, with $p/m \simeq 1$ when using constituent-quarks, which are effective degrees of freedom with masses that include the effects of sea quark and gluons. These theoretical problems can be solved by simultaneously diagonalizing the Hamiltonian in a large basis, using a relativistic kinetic energy and allowing for other relativistic effects, and using a consistent set of parameters for all baryon excitations [771].

Pseudoscalar-meson exchange models

Glozman and Riska [2725,2726] emphasize the role of chiral symmetry in determining the baryon spectrum by using a short-range interaction between quarks similar to that of Eq. 9.4, but with the exchange of the 'chiral' octet of pseudoscalar mesons between quarks. This leads to a contact interaction between quarks i and j similar in form to that of Eq. 9.4, but proportional to the expectation of the product $\vec{\lambda}_i^f \cdot \vec{\lambda}_j^f$ of $SU(3)_f$ generators. A fit to the spectrum of low-lying negative and positive-parity baryons made up of u, d, and s quarks with harmonic confinement allows first radial recurrence of the nucleon, corresponding to the Roper resonance $N(1440)$, to be lighter than the lightest $J^P = 1/2^-$ orbital excitation, corresponding to $N(1535)$, and the same behavior holds for the Λ baryons, as seen in extractions from experimental data. There are no spin–orbit interactions that arise from pseudoscalar meson exchange; those from other sources are neglected, along with tensor forces that accompany the contact interaction. The calculation of the N and Δ

spectra was refined by performing three-body Faddeev calculations with a Goldstone-boson-exchange interaction plus linear confinement between the constituent quarks in Ref [2727].

Dziembowski, Fabre de la Ripelle, and Miller [2728] put these two approaches together by including the effects of pseudoscalar meson exchange and of one gluon exchange between quarks, neglecting the complexity introduced by tensor and spin–orbit interactions, in a hyper-spherical method calculation that goes beyond wave function perturbation theory. They showed that it is possible to describe the non-strange baryon spectrum using a quark–meson coupling constant that reproduces the measured pion–nucleon coupling constant, and a reasonably small value of the strong-coupling constant, which governs the strength of the one-gluon exchange terms.

Instanton-induced interactions

Instantons are topologically nontrivial gauge-field configurations in 4-dimensional Euclidean space, with field strengths that vanish at large spatial distances. These configurations are localized in both space and (Euclidean) time, and so are instantaneous interactions, which gives rise to their name. They are crucial to understanding the formation of condensates in the QCD vacuum, and how the axial current anomaly gives mass to the η' meson; their presence in QCD also yields short-range interactions between the quarks. Löring, Kretzschmar, Metsch and Petry [2729,2730] investigated the spectrum of baryons in a relativistic model by solving the three-body Bethe–Salpeter equation. This model uses instantaneous pairwise linear confinement, with the Dirac structure required to make the confining potential spin-independent, and an instantaneous two-body interaction based on ′t Hooft's residual interaction, which arises from QCD-instanton effects. This model was able to explain salient features of the non-strange baryon spectrum, such as the low mass of the Roper resonance $|N^2S_{S'}1/2^+\rangle$, and the presence of approximate parity doublets. This study was extended [2731] to the study of excited Λ and Σ hyperons, where the equivalent features of these spectra were also explained, and later to charmed baryons in Ref. [2732].

The Dyson–Schwinger Bethe–Salpeter approach

There has been significant recent progress in understanding the physics of baryons [822,2733] by using the Dyson–Schwinger equations of QCD and Bethe–Salpeter equations [820,2734]. In this approach baryons are relativistic bound states of three quarks, and the treatment of their interactions arising from QCD is non-perturbative, incorporating aspects of confinement and dynamical symmetry breaking. Two paths to solving the three-body problem are taken; direct solution of the three-body Faddeev equation, and decomposition of baryons into quark–diquark systems, with all quark pairs able to constitute the diquark. The latter path requires the calculation of diquark Bethe–Salpeter amplitudes, and diquark propagators. These depend on the quark and gluon propagators and quark–gluon vertex, which are consistent with those used for the Bethe–Salpeter equation for mesons, and with chiral symmetry. Due to the complexity of the three-body system, baryon calculations are performed using the rainbow-ladder approximation, where the $q-q$ kernel has the form of a single gluon exchange with a momentum-dependent vertex strength, summed by the Bethe–Salpeter equation into Feynman diagrams that take the form of a ladder, or rainbow. This construction preserves chiral symmetry.

Using this dynamical quark–diquark approach, the ground state nucleon, $\Delta(1232)\frac{3}{2}^+$, and Roper $N(1440)\frac{1}{2}^+$ resonances are described well [899], as their configurations are dominated by scalar and axial vector diquarks. However, other baryons are sensitive to other diquark channels, which are known to be too strongly bound in this approximation, as are the corresponding scalar and axial-vector mesons. The result is that the other excited baryon masses come out too low. Reducing the strength of the attraction in the pseudoscalar and vector di-quark kernels simulates effects beyond the rainbow-ladder approximation, and the result is good agreement between the calculated spectrum for excited N, Δ, Λ, Σ, Ξ and Ω baryons with $J^P = 1/2^{\pm}, 3/2^{\pm}$, with the exception of the $\Lambda(1405)1/2^-$, $\Lambda(1520)3/2^-$, and to a lesser extent the Roper resonance $N(1440)1/2^+$. The authors of Ref. [899] point out that this is likely due to the lack of a consistent treatment of baryon-meson coupled channel effects.

9.1.4 Missing states in the baryon spectrum

Models of strong decays

For ground-state and low-lying negative-parity excited state baryons made up of $\{u, d\}$ and a single s quark, the spectrum of states extracted from experimental data can be matched to model predictions without ambiguity. (There is little experimental information about the spectrum of the excited Ξ, strangeness $S = -2$, and Ω, strangeness $S = -3$ states.) However, for positive-parity excited states, more states are predicted by models that treat three quarks symmetrically than are present in analyses of the data. This is called the 'missing resonance' problem. One possible explanation is to postulate that they contain static, tightly-bound ud diquarks, which reduces the effective number of degrees of freedom and so the number of excitations in the spectrum [791]. However, lattice QCD calculations of nucleon structure [2735], and of the entire excited baryon spectrum using a broad spread of operators [529,2736] do not show the reduced number of states expected if only 'good' di-quarks prevail, and recent experimental evidence for the existence of states that are ruled out in such models is described below.

A solution to this problem is that, unlike those states seen in partial-wave analyses of elastic πN and $\bar{K}N$ scattering data, these missing positive-parity excited states have weak couplings to the corresponding strong-interaction production channel [2737,2738]. In any case, in order to make a detailed and exhaustive comparison between predictions of any model and the experimental spectrum of excited baryons, a model of the strong decay $B^* \to BM$ of baryons into a ground-state baryon and meson is required. For a detailed, comparative review of such models see Ref. [2709].

One approach is to couple point-like pseudoscalar mesons to the quarks in the decaying baryon, an elementary-meson emission model [2721,2739,2740]. As an example, Koniuk and Isgur [2721] modeled such decays, by coupling point-like pseudoscalar mesons to the quarks in the decaying baryon, and evaluating the transition amplitudes using the configuration-mixed wave functions resulting from the one-gluon exchange model of Isgur and Karl [2719]. They also examined baryon electromagnetic transition amplitudes that can be extracted from meson photo-production experiments. Many states were observed to have small πN or $\bar{K}N$ amplitudes, which would lead to them decoupling from elastic-scattering partial-wave analyses, and the masses and decay amplitudes of those that did not corresponded to those of the observed states.

The internal structure of mesons can be taken into account if strong decays of excited baryons proceed *via* an operator that creates a $q\bar{q}$ pair with vacuum, $^{2S+1}L_J = {}^3P_0$, quantum numbers. The operator is assumed $SU(3)_f$ symmetric; the additional energy required to produce a strange quark pair is taken into account by the kinematics. Strong decay amplitudes are formed by evaluating the required spin and flavor overlaps, forming the expectation value of this operator between wave functions for the final state baryon and meson, and that of the initial excited baryon, and integrating over the relative momentum of the $q\bar{q}$ pair.

9.1.5 Decay-channel couplings

Models for the baryon spectrum that do not take into account decay-channel couplings effectively assume that baryons are infinitely long-lived bound states. In practice, excited baryons decay strongly, with decay widths that are significant fraction of their masses. Excited baryons can have large couplings to continuum states, which can and will affect the positions of the poles in scattering amplitudes that describe these resonances. This can be due to their proximity to decay-channel thresholds, or unusually large couplings to a decay channel, or both. Examples include the low-lying negative-parity resonances $\Lambda(1405)1/2^-$, which has a nominal mass below the $N\bar{K}$ threshold, and $N(1535)1/2^-$, which couples strongly to the $N\eta$ final state, for which it is just above threshold. The authors of Ref. [2741] were among the first to suggest that these states could be dynamically generated resonances. Hyodo and Meißner review the interesting physics of the $\Lambda(1405)$ state in the 2022 RPP [616]. There is also evidence of these effects from lattice QCD [2736] for states like the Roper resonance; its mass changes rapidly as the pion mass in the calculation approaches the physical pion mass, due to strong $N\pi\pi$ channel coupling.

Beyond elastic meson scattering

The observation that baryon resonances could be missing due to weak couplings for both their strong-interaction production and decay in elastic meson-nucleon scattering led to the idea that γN photo-production experiments could excite missing resonances that had appreciable photo-couplings, which could be discovered via their strong decays to final states with more than one pseudoscalar meson. For example, missing N and Δ-flavored baryons could be searched for in proton-target photo-production experiments examining two or three pion final states resulting from the intermediate vector-mesons $\rho(770)$ and $\omega(782)$. In particular, certain missing, positive-parity resonances can be expected to decay to two-pion final states by simultaneous de-excitation of both $l_\rho = 1$ and $l_\lambda = 1$ excitations.

Recent developments from photo-production experiments

The Particle Data Group in its bi-annual updates of the Review of Particle Physics (RPP) lists the known baryon resonances, their properties, and the experimental evidence for their existence in terms of star assignments ranging from one star (poor evidence) to four stars (evidence is strong). Since the 2010 edition of the RPP [2742], much new information about N and Δ-flavored baryons has been added based on recent photo-production experiments. In particular, various polarization observables have played a crucial role in identifying new resonances, or consolidating the existence of those previously poorly known.

The experimental spectroscopy efforts to address the missing resonance problem have concentrated, for the most part, on the $N = 2$ positive-parity non-strange excitations, and on the $N = 1$ negative-parity singly- and doubly-strange states. While only one negative-parity Λ state has yet to be identified to complete the first excitation band, with many more positive-parity candidates unknown for the second excitation band, overall as many Σ and Ξ states are expected as N^* and Δ states combined, and of these, many negative-parity states are still missing. The situation is worse for Ξ and Ω resonances, since their spins and parities have been measured for very few states; speculative J^P assignments based on quark model predictions are listed by the PDG for the majority of the observed states. The potential for new discoveries remains high in the hyperon sector.

Of particular interest in the non-strange sector are those multiplets of the second excitation band where both oscillators have a single orbital excitation, $l_\rho = l_\lambda = 1$, which

combine to either $L = 0$, 1, or 2. Since both relative coordinate vectors must be excited in order to have the necessary exchange symmetry, their presence would rule out tightly-bound, static di-quarks. In fact, a quartet of $S = 3/2$ states

$$|N^4 D_M(1/2^+, 3/2^+, 5/2^+, 7/2^+)\rangle,$$

has been proposed [2743] for the 70-plet $(70, 2^+)$ largely based on the photo-produced double-pion final state.

These states were expected to be seen in double-meson reactions since each oscillator can de-excite via the emission of a meson. There is a new $J^P = 1/2^+$ state, $N(1880)1/2^+$, and the state $N(1900)3/2^+$ has had its likely existence upgraded from 2 to 4 stars in the Review of Particle Properties (RPP) by the PDG [616]. Evidence for two other states in the quartet, $N(2000)5/2^+$ and $N(1990)7/2^+$, is strong in some partial wave analyses but requires additional confirmation, and so remains listed with weak evidence (two stars) in the RPP. Such double-meson reactions had been under-explored until recently, which would explain why these states escaped detection in the past. Another previously one-star resonance, $N(2100)1/2^+$, has been upgraded to three stars, and can be tentatively assigned to the doublet of states with $S = 1/2$ forming the SU(6) multiplet $[20, 1^+]$ of Eq. 9.3, where $l_\rho = l_\lambda = 1$ combine to $L = 1$. There remains at best weak evidence for the second state in this doublet. Although the assignment of experimental N^* candidates to these multiplets is speculative, some optimism persists that the goal of completely mapping the second excitation band for non-strange baryons is within reach once all currently available (polarization) data have been analyzed.

There is also an interesting pattern of parity doublets of N^* baryons with masses around 2 GeV, which might indicate the restoration of the chiral symmetry at higher energies [2744, 2745]. A similar pattern was observed for Δ resonances in the same mass region:

$\Delta(1910)1/2^+$ $\Delta(1900)1/2^-$
$\Delta(1920)3/2^+$ $\Delta(1940)3/2^-$
$\Delta(1905)5/2^+$ $\Delta(1930)5/2^-$
$\Delta(1950)7/2^+$

A detailed study has not yet revealed the missing $7/2^-$ state [2746]. Closest in mass is a previously poorly-known state, $\Delta(2200)7/2^-$, which has since been upgraded from one to three stars based on photo-production data. Interestingly, the corresponding mass difference is observed in the nucleon spectrum between $N(1990)7/2^+$ (two stars) and $N(2190)7/2^-$ (four stars).

The result is that, based on photo-production data, six completely new N^* resonances have been proposed with masses around 2 GeV, and three additional states have been upgraded. No new Δ state has been proposed, but four states have had their status upgraded by the PDG.

9.1.6 Baryons with excited glue

In a strongly-coupled system, *hybrid* baryons with excited gluon degrees of freedom must exist. Unlike in the spectrum of mesons, all J^P quantum numbers are accessible via spatial excitation for a given flavor of baryon, as explained in Sect. 9.1.2. In the absence of exotic quantum numbers, another approach to the discovery of hybrid baryons might be to search for an over-population, relative to the expectations of constituent quark models, of states with a given flavor, spin, and parity quantum numbers. However, it is expected that the lowest-lying states with excited gluon degrees of freedom are positive-parity states that overlap in mass the region in the spectrum where there are already several missing conventional states.

Early approaches to the physics of hybrid baryons include those based on the MIT bag model [2747], large-N_c QCD [2748], and QCD sum rules [2749]. As an example, the calculation of Ref. [2747] confined a constituent gluon and three quarks to an MIT bag, and used $\mathcal{O}(\alpha_s)$ interactions between the constituents. In these studies, the lightest hybrid baryons were found to have N flavor and $J^P = \{1/2^+, 3/2^+\}$, with the lightest of these having $J^P = 1/2^+$ and a mass of approximately 1500 MeV, between those of the two lightest radial excitations of the nucleon, the Roper resonance at 1440 MeV, and the N(1710).

The flux-tube model developed to examine hybrid meson structure and decays by Isgur and Paton [2418] was applied to hybrid baryons in Refs. [2723,2750]. An adiabatic approximation is employed, where a Y-shaped flux tube is allowed to move with the three quark positions fixed, except for center of mass corrections. This defines a potential in which the quarks move, for both conventional (glue in its ground state) and hybrid (glue in its lowest-lying excited state) baryons. The flux-tube dynamical problem can be reduced to the independent motion of the junction and the strings connecting the junction to the quarks. The seven low-lying hybrid baryons are found to be two doublets of $N^2 1/2^+$ and $N^2 3/2^+$ states with quark spin $S = 1/2$, and three states

$$^4\Delta(1/2^+, 3/2^+, 5/2^+)$$

with quark spin $S = 3/2$. Baryon masses are found by using a variational method to solve for the quark energies in these string potentials. Including the hyperfine contact spin–spin term in Eq. 9.4 lowers the mass of the quark–spin 1/2 hybrid states by 110 MeV to 1865 MeV, and raises the mass of the quark–spin 3/2 hybrid states, which coincide with the lightest Δ flavored hybrids, by a similar amount.

Lattice QCD approaches to describing the spectrum of conventional and hybrid baryons assuming isolated bound states [529,2736,2736] are able to determine the spectrum of baryon states up to $J^P = 7/2^\pm$. The results show the same number of states as non-relativistic models based on

three-quark degrees of freedom [2736], with no signs of the reduced number of excitations predicted by di-quark models, or parity doubling. States in this spectrum can be grouped into $SU(6) \times O(3)$ multiplets, with weak mixing. Using many composite QCD interpolating fields, hybrid baryons of N and Δ flavor were identified in Ref. [529] by searching for states with a substantial overlap with operators containing gluonic excitations. This led to doublets of $N1/2^+$ and $N3/2^+$ hybrids, and $N5/2^+$, $\Delta 1/2^+$ and $\Delta 3/2^+$ states at energies above the center of the first band of conventional positive-parity excitations. This suggests that exciting the glue adds a color-octet effective degree of freedom, with roughly the same additional energy in mesons and baryons, that has $J^P = 1^+$, unlike the vector nature of this excitation in the flux-tube model. A $J^P = 1^+$ excitation is expected in the bag model of Ref. [2747], as these are the quantum numbers of the lowest energy, transverse electric mode of a gluon in a spherical bag.

This approach is extended to all baryons made from u, d, and s quarks in Ref. [2736], using operators that lie in irreducible representations of $SU(3)_f$ symmetry, in addition to $SU(4)$ symmetry for the Dirac spins and $O(3)$ symmetry for the orbital state. The spectra that result for non-hybrid states are again consistent with quark model expectations based on weakly broken $SU(6) \otimes O(3)$ symmetry. States with strong hybrid content are usually at about 1 GeV above the corresponding conventional excited states, and the quantum numbers and multiplicity of the positive-parity hybrid states can be roughly predicted by combining a $J^P = 1^+$ gluonic excitation with non-relativistic quark spins, although some of the expected states are not found in the calculation performed at the lowest pion mass. The use of multi-hadron operators will allow the exploration of the energy dependence of and resonances in hadron scattering amplitudes.

A recent proposal prepared by the CLAS12 Collaboration and presented to the Jefferson Lab Physical Advisory Committee aims to experimentally search for hybrid baryon states in electro-produced KY and $p\pi^+\pi^-$ final states by focusing on measurements for $Q^2 < 1.0$ GeV^2. Since the spin and parity of hybrid baryons are expected to be the same as those for conventional states, the experimental signature of hybrid baryons is the distinctively different low-Q^2 evolution of their electro-couplings that originate from the additional gluonic component of their wave function. More details are discussed in the contribution by V. Burkert.

9.2 Light-quark baryons

Volker Burkert, Eberhard Klempt, Ulrike Thoma

9.2.1 Why N's ?*

This was the question with which Nathan Isgur opened his talk at N^*2000 [2751] held at the Thomas Jefferson National Accelerator Facility in Newport News, VA, one year before he passed away, much too early. He gave three answers:

First, nucleons are the stuff of which our world is made. In the Introduction to this section, two of us have outlined the importance of N^*'s and Δ^*'s in the development of the Universe 9, when hadrons materialized from a soup of quarks and gluons at some 10 μs after the big bang. The full spectrum of excited baryon states including those carrying strangeness must be included in hadron gas models that simulate the freeze-out behavior observed in hot-QCD calculations. These simulations aim at finding the underlying processes, to pin-point the "critical point" of the phase transition that is expected to occur between the QGP phase and the hadron phase at a temperature near 155 MeV. Experiments are ongoing at CERN, RHIC and planned at FAIR to study the phase diagram of strongly interacting matter, e.g. by varying the collision energy.

Second, nucleons are the simplest system in which the non-abelian character of QCD is manifest. The proton consists of three (constituent) quarks since the number of colors is three.

Third, baryons are sufficiently complex to reveal physics to us hidden in the mesons. Gell-Mann and Zweig did not develop their quark model along mesons, their simple structure allowed for different interpretations. *Three* quarks resulted in a baryon structure that gave – within SU(3) symmetry – the octet and the decuplet containing the famous Ω^-.

Isgur made many important contributions to the development of the quark model. With Karl he developed the idea that gluon-mediated interactions between quarks bind them into hadrons and constructed a quark model of baryons [2752]. This was a non-relativistic model, hardly justifiable. With Capstick he relativized the model [771], but surprisingly, the pattern of predicted resonances remained rather similar. Isgur always defended the basic principles: hadrons have to be understood in terms of constituent quarks bound in a confining potential and additionally interacting via the exchange of "effective" gluons.

Nearly 20 years later, Meißner ended his contribution [2753] to the N^*2019 conference held in Bonn, Germany, by stating: "Forget the quark model". We need to ask: What has happened in these two decades? What did we know before? What have we learned?

Mapping the excitation spectrum of the nucleon (protons and neutrons) and understanding the effective degrees of freedom are important and most challenging tasks of hadron physics. A quantitative description of the spectrum and properties of excited nucleons must eventually involve solving QCD for a complex strongly interacting multi-particle system. The experimental N^* program currently focuses on the search for new excited states in the mass range just below and above 2 GeV using energy-tagged photon beams in the

few GeV range, and on the study of resonances, their properties, and their internal structure, e.g. in cascade decays and in meson electro-production.

9.2.2 N^*'s: how?

In the previous contribution by Capstick and Crede 9.1 we have seen the complexity of the expected spectrum of nucleon and Δ excitations. Even in the lowest excitation mode with $l_\rho = 1$ or $l_\lambda = 1$, we expect five N^* and two Δ^* states; they are all well established. But already in the second excitation mode, the quark model predicts 13 N^* and 8 Δ^* states The resonances have quantum numbers $J^P = 1/2^+, \ldots, 7/2^+$ and isospin $I = 1/2$ or $3/2$, respectively. All these 21 resonances are expected to fall into a mass range of, let's say, 1600–2100 MeV. This complexity of the light-quark (u and d quarks) baryon excitation spectrum complicates the experimental search for individual states, especially since, as a result of the strong interaction, these states are broad, the typical width being 150–300 MeV. They overlap, interfere, and often several resonances show up in the same partial wave. Grube in his contribution 8.3 has convincingly demonstrated the difficulties of extracting the existence and properties of mesonic resonances from $\pi\pi$ scattering experiments. With nucleon resonances, additional complications due to the nucleon spin emerge: in πN elastic scattering there are two complex amplitudes to be determined, for spin-flip and spin-non-flip scattering.

Pion scattering off nucleons was mostly performed in the pre-QCD era. Nearly all excited nucleon states listed in the Review of Particle Physics (RPP) prior to 2012 have been observed in elastic pion scattering $\pi N \to \pi N$. However there are important limitations in the sensitivity to the higher-mass nucleon states. These may have very small $\Gamma_{\pi N}$ decay widths, and their identification becomes exceedingly difficult in elastic scattering. Three groups extracted the real and imaginary parts of the πN partial-wave amplitude from the data [2754–2756]. Their results are still used as constraints in all modern analyses of photo-induced reactions.

Figure 216a, b shows the real and imaginary part of the S_{11} amplitude for πN scattering. The imaginary part peaks at 1500 MeV and just below 1700 MeV indicating the presence of two resonances, $N(1535)1/2^-$ and $N(1650)1/2^-$. These are known since long and established. Above, there is no clearly visible sign for any additional resonance. Higher-mass resonances – if they exist – must have very small $\Gamma_{\pi N}$ decay widths.

Estimates for alternative decay channels have been made in quark model calculations [2762]. This has led to major experimental efforts at Jefferson Lab, ELSA and MAMI to determine differential cross sections and (double) polarization observables for a variety of meson photoproduction channels. Spring-8 at Sayo in Japan and the ESRF in Grenoble, France, made further contributions to the field.

Figure 216c, d shows an example. In Fig. 216c, the total cross section for η photoproduction off protons and off neutrons is shown [2757,2758]. They are dominated by $N(1535)1/2^- \to N\eta$ interfering with $N(1650)1/2^-$. The opening of important channels is indicated by vertical lines. At the η' threshold, the intensity suddenly drops: significant intensity goes into the $N\eta'$ channel. This is a strong argument in favor of a resonance at or close to the $p\eta'$ threshold. It also clearly demonstrates the advantage of investigating different final states and production mechanisms. In contrast to the πN-S_{11} scattering amplitude, here, already in the total η-photoproduction cross section, a structure relating to $N(1895)1/2^-$ becomes visible. Furthermore, in Fig. 216d, the result of a fit with Legendre moments to the so-called Σ polarization observable for $\gamma p \to \eta p$ is compared to two energy-dependent solutions of the BnGa coupled-channel analysis. Plotted is the coefficient $(a_4)_4^\Sigma$ of the Legendre expansion which receives (among others) a contribution from the interference of the S-wave with the G-wave. Data from different experiments are given with their error bars. The curves represent BnGa fits with (solid curve) and without (dashed curve) inclusion of data on $\gamma p \to \eta' p$. The $N(2190)7/2^-$ (G-wave) was included in both fits. From 1750 MeV to the $p\eta'$-threshold the coefficient is approximately constant, then at the $p\eta'$-threshold, the fit result shows an almost linear rise towards positive values. This change of the coefficient at about 1.9 GeV indicates the presence of a cusp. The strong cusp is an effect of the $p\eta'$ threshold [$E_\gamma = 1447$ MeV (W = 1896 MeV)], the $N\eta'$ amplitude must be strongly rising above threshold. Indeed, the inclusion of the full data set on $\gamma p \to p\eta'$ (cross sections, polarization observables) into the BnGa data base had already confirmed the existence of a new $N(1895)1/2^-$ resonance with a significant coupling to $p\eta$ and $p\eta'$ [2763,2764], first observed in [2765].

This resonance was not seen in classical analyses of πN elastic scattering data.[91] The example shows the importance of inelastic channels and of coupled-channel analyses. Thresholds can be identified by the missing intensity in other channels, cusp effects can show up, all these effects need to be considered and finally contribute to find the correct solution. High-precision and high-statistics data are required as well as a large body of different polarization data.

9.2.3 Photoproduction of exclusive final states

In the photoproduction of a single pseudoscalar meson like $\gamma p \to \eta p$, not only the proton has two spin states but also

[91] Höhler and Manley had claimed a similar state that had been combined with Cutkovsky's result to $N(2090)$.

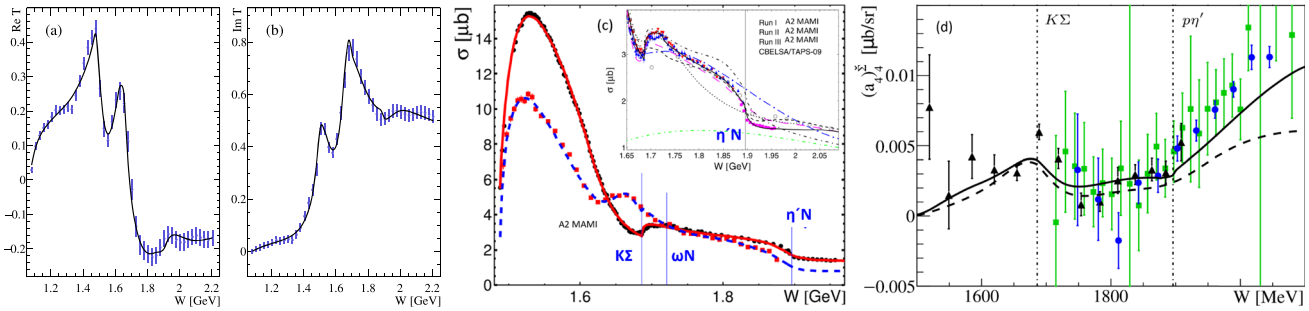


Fig. 216 **a**, **b** Real and imaginary part of the S_{11} πN scattering amplitude. Resonances in this partial wave have quantum numbers $J^P = 1/2^-$. Clearly seen are $N(1535)1/2^-$ and $N(1650)1/2^-$. There is no convincing evidence for any resonance above 1700 MeV. Data points are from [2754], errors are estimates, the curve represents a recent Bonn–Gatchina (BnGa) fit. **c** Total cross sections for $\gamma p \to \eta p$ and $\gamma n \to \eta n$. Important thresholds are marked by lines. The inset shows the η' threshold region for η-photoproduction off the proton (picture adapted from [2757,2758]). **d** The Legendre coefficient of the polarization observable Σ $(a_4)_4^{\Sigma}$ exhibits a cusp at the η' threshold [2759]. The data stems from GRAAL (black), CBELSA/TAPS (blue) and CLAS (green) Picture taken from [2759]. **c**, **d** See publications [2757,2758,2760] for references to the data

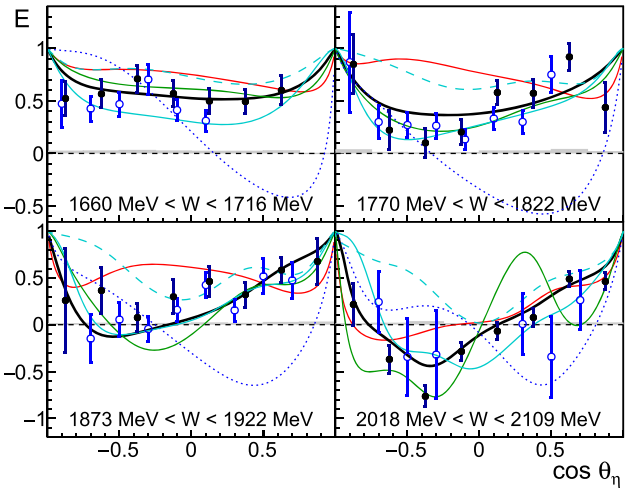


Fig. 217 The double polarization observable E as a function of $\cos\theta_\eta$ in the cms for selected energy bins, black: CBELSA/TAPS [2760], blue: CLAS data [2761] (due to different binning, the energies differ by up to half of the bin size). Colored curves: Predictions from different PWAs (see publication for references), black: BnGa-fit including the data shown here and further new polarization data. Figure adapted from [2760]

the photon has two possible spin orientations. In electroproduction, discussed by Burkert in the subsequent Sect. 9.3, the virtual photon can also be polarized longitudinally. But even for experiments with real photons, there are four complex amplitudes to be determined. There is a large number of observables: the target nucleon can be polarized longitudinally, i.e. in beam direction, or transversely, the photon can carry linear or circular polarization. The final-state nucleon can carry polarization along its flight direction or perpendicular to the scattering plane. There is an intense discussion in the literature on how many independent measurements have to be performed to determine the four complex amplitudes, see Ref. [2713]. In practice, energy-independent analyses in bins of the invariant mass were only done for the very low energy region [2769,2770] or with additional assumptions (see [2771–2773] and references therein).

In most cases, energy-dependent analyses have been performed to extract the information hidden in the photoproduction data. These analyses were pioneered by the Giessen group who made the first coupled-channel analysis of pion and photo-induced reactions to extract properties of nucleon resonances [2774,2775]. Later, the polarization data, in particular those with polarized photon beam and polarized target nucleons, proved to be decisive to reduce ambiguities of the solutions. The double-polarization observable E is one of the beam-target-observables; it requires a circularly polarized photon beam and a longitudinally polarized target. Examples of E for selected W-bins are shown in Fig. 217 for $\gamma p \to p\eta$ [2760]. The data are compared to the predictions of different PWA solutions (colored curves). The curves scatter over a wide range indicating the high sensitivity of the polarization observable on differences in the contributing amplitudes. A new BnGa fit returned masses and widths of N^*-resonances and their $N\eta$-branching fractions [2760], several of them unknown before. Interestingly a $N(1650)1/2^- \to N\eta$-branching fraction of 0.33 ± 0.04 was found supporting the large values reported by BnGa [2765] and the A2 collaborations [2758] while in the RPP'2010, a value of only 0.023 ± 0.022 was given. Recently, also within the Jülich–Bonn dynamical coupled channel approach, a $N\eta$-residue for $N(1650)1/2^-$ was found, larger by almost a factor of two compared to earlier analyses, after inclusion of the new polarization data [2776]. In contrast, the Gießen group [2777] and Hunt and Manley [2778] find very small values for the $N(1650)1/2^- \to N\eta$ branching ratio. Historically, the large $N(1535)1/2^- \to N\eta$ branching fraction and the small one for $N(1650)1/2^- \to N\eta$ has played a significant

role in the development of the quark model [764], of theories based on coupled-channel chiral effective dynamics [2741] and led to several interesting interpretations of the low mass $1/2^-$-resonances (for references see [2760]). The old values from 2010 were obtained without the constraints provided by the new high quality (double) polarization data covering almost the complete solid angle. The impact of polarization observables on the convergence of different PWA-solutions was e.g. also very clearly demonstrated in a common study of pion-photoproduction [2779].

In hyperon decays, the polarization of the Λ or Σ° can be determined by analyzing the parity violating decay $\Lambda \to p\pi^-$. Thus the spin orientation of the final state baryon (recoil polarization) can be determined. Kaon-hyperon production using a spin-polarized photon beam provides access to the beam-, recoil-, target-[92] and to beam-recoil polarization observables. The data had a significant impact on the determination of the resonance amplitudes in the mass range above 1.7 GeV. Precision cross section and polarization data, examples of which are shown in Fig. 218, span the $K^+\Lambda$ and $K^+\Sigma$ invariant mass range from threshold to 2.9 GeV, hence covering the interesting domain where new states could be discovered. Clear resonance-like structures at 1.7 GeV and 1.9 GeV are seen in the $K^+\Lambda$-differential cross section that are particularly prominent and well-separated from other structures at backward angles. At more forward angles (not shown) t-channel processes become prominent and dominate the cross section. The broad enhancement at 2.2 GeV may also indicate resonant behavior although it is less visible at more central angles with larger background contributions. Similar resonance-like structures are observed in the $K\Sigma$ channel (Fig. 218b). Examples for different polarization observables determined for the reaction $\gamma p \to K^+\Lambda$ are shown in the lower row of Fig. 218 for selected bins in the K^+-scattering angle in the γp center-of-mass frame. They are compared to predictions from ANL-Osaka, BnGa-2014 and to a refit from the BnGa-PWA. The large differences between the curves demonstrate the sensitivity of the data to the underlying dynamics. The $K\Lambda$ channel is somewhat easier to understand than the $K\Sigma$ channel, as the iso-scalar nature of the Λ selects isospin-1/2 states to contribute to the $K\Lambda$ final state, while both isospin-1/2 and isospin-3/2 states can contribute to the $K\Sigma$ final state. Of course, here, as well as for other final states, only a full partial wave analysis can determine the underlying resonances, their masses and spin-parity. Polarization data are required to disentangle the different amplitudes.

Energy-dependent analyses have been performed e.g. at Gießen [2774,2775,2780], at GWU [2781] as SAID, in Mainz as MAID [2758], at Kent [2778], at JLab [2782], by the BnGa [2765,2783], the Jülich–Bonn (JüBo) [2776], the ANL-Osaka [2784] and by other groups. A short description of the different methods can be found in Ref. [2713]. Here we emphasize that the energy-dependence of a partial-wave amplitude for one particular channel is influenced by other reaction channels due to unitarity constraints. To fully describe the energy-dependence of a production amplitude, all (or at least the most significant) reaction channels must be included in a coupled-channel approach. Many different final states have been measured with high precision off protons and partly also off neutrons (bound in a deuteron with a quasi-free proton in the final state). Polarization data for meson photoproduction off neutrons are, however, still scarce. A fairly complete list of references can be found in [2713]. Most data are now included in single- and in multi-channel analyses.[93]

The photoproduction data had a strong impact on the discovery of several new baryon states or provided new evidence for candidate states that had been observed previously but lacked confirmation (e.g. [2746,2758,2765]). Many new decay modes were discovered, in particular in the photoproduction of $2\pi^0$ and $\pi^0\eta$, [2783,2785,2786] and references therein. At the NSTAR'2000 workshop, 12 N^* and 8 Δ^* were considered to be established (4*,3*) by the Particle Data Group.[94] These numbers increased to 19 N^* and 10 Δ^* two decades later. Table 38 lists the new resonances below 2300 MeV and those that had not a four-star status in 2010. Resonances which had four stars in 2010 are well established and kept their status. These are:

$N(1440)1/2^+$, $N(1520)3/2^-$, $N(1535)1/2^-$, $N(1650)1/2^-$, $N(1675)5/2^-$, $N(1680)5/2^+$, $N(1720)3/2^+$, $N(2190)7/2^-$, $N(2220)9/2^+$, $N(2250)9/2^-$, $\Delta(1620)1/2^-$, $\Delta(1700)3/2^-$, $\Delta(1905)5/2^+$, $\Delta(1910)1/2^+$, $\Delta(1950)7/2^+$.

A few resonances were removed from the RPP tables. They often had wide-spread mass values, and the old results were redistributed according to their masses and the new findings. Even more impressive is the number of reported decay modes. Our knowledge on N^* and Δ^* decays has at least been doubled.

9.2.4 *Regge trajectories*

Like mesons, baryons fall onto linear Regge trajectories when their squared masses are plotted as a function of their total spin J or their intrinsic orbital angular

[92] The target polarization observable can also be accessed by performing a double-polarization experiment using a linearly polarized photon beam and measuring the baryon polarization in the final state.

[93] A list of data on photoproduction reactions including polarization and double-polarization observables can be found at the BnGa web page: https://pwa.hiskp.uni-bonn.de/.

[94] In PDG notation: 4* Existence certain, 3* almost certain, 2* evidence fair, 1* poor

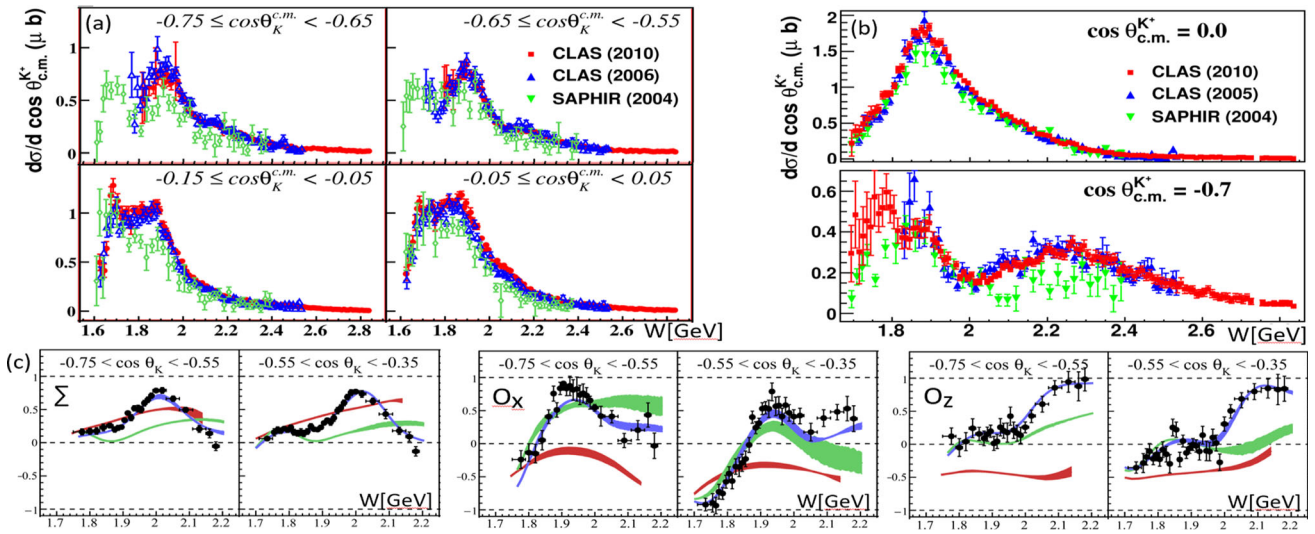


Fig. 218 Invariant mass dependence of the $\gamma p \to K^+\Lambda$ [2766] (**a**) and $\gamma p \to K^+\Sigma$ [2767] (**b**) differential cross sections for selected bins in the polar angle. **c** Examples for polarization observables determined for $\gamma p \to K^+\Lambda$ (only selected bins shown) [2768]. Curves: PWA-predictions from ANL-Osaka (red) and BnGa 2014 (green). Blue: BnGa 2014-refit including the data shown. **a–c** For references to the data and the PWAs see [2766–2768], Picture adapted from [2766–2768]

momentum L. In the case of Δ^*, the leading trajectory consists of $\Delta(1232)3/2^+$, $\Delta(1950)7/2^+$, $\Delta(2420)11/2^+$, $\Delta(2950)15/2^+$. In the quark model, these have intrinsic orbital angular momenta $L = 0, 2, 4, 6$. Figure 219 shows the squared Δ^*-masses as a function of $L + N_{\text{radial}}$, where N_{radial} indicates the intrinsic radial excitation. The resonances $\Delta(1910)1/2^+$, $\Delta(1920)3/2^+$, $\Delta(1905)5/2^+$ have intrinsic $L = 2$ like $\Delta(1950)7/2^+$, and fit onto the trajectory. Also, there are three positive-parity resonances that likely have $L = 4$ with the $5/2^+$ state missing. The two $L = 1$ resonances $\Delta(1620)1/2^-$ and $\Delta(1700)3/2^-$ also have masses close to the linear trajectory. Further, there are resonances in which the ρ or λ oscillator is excited radially to $n_\rho = 1$ or $n_\lambda = 1$ ($N_{\text{radial}} = 1$). Quark models with a harmonic oscillator as confining potential predict that resonances belong to shells. Radial excitations are predicted in the shell $L + 2\,N_{\text{radial}}$. This is not what we find experimentally: the masses are approximately proportional to $L+N_{\text{radial}}$ if $N_{\text{radial}} = 1$ is assigned to $\Delta(1600)3/2^+$, the first radial excitation of $\Delta(1232)3/2^+$, as well as to the $\Delta(1900)1/2^-$, $\Delta(1940)3/2^-$, $\Delta(1930)5/2^-$ triplet, to the two members of a partly unseen quartet $\Delta(2350)5/2^-$ and $\Delta(2400)9/2^-$, and to $\Delta(2750)13/2^-$ (with $L = 5$, $S = 3/2$ and $N_{\text{radial}} = 1$).

Clearly, this is a very simplified picture of the Δ^* spectrum. The picture is that of the non-relativistic quark model – nobody understands why it works.[95] Resonances – assumed to have the same mass if spin orbit-coupling is neglected– have indeed somewhat different masses. But the gross features of the spectrum of Δ^* resonances are well reproduced.

Table 38 Baryon resonances above the $\Delta(1232)$ and below 2300 MeV given in the RPP'2022 in comparison to the resonances considered in the RPP'2010. Resonances with 4* in 2010 are not listed here. See text for further discussion

	RPP 2010	RPP 2022		RPP 2010	RPP 2022
$N(1700)3/2^-$	***	***	$\Delta(1600)3/2^+$	***	****
$N(1710)1/2^+$	***	****	$\Delta(1750)1/2^+$	*	*
$N(1860)5/2^+$	–	**	$\Delta(1900)1/2^-$	**	***
$N(1875)3/2^-$	–	***	$\Delta(1920)3/2^+$	***	***
$N(1880)1/2^+$	–	***	$\Delta(1930)5/2^-$	***	***
$N(1895)1/2^-$	–	****	$\Delta(1940)3/2^-$	*	**
$N(1900)3/2^+$	**	****	$\Delta(2000)5/2^+$	**	**
$N(1990)7/2^+$	**	**	$\Delta(2150)1/2^-$	*	*
$N(2000)5/2^+$	**	**	$\Delta(2200)7/2^-$	*	***
$N(2040)3/2^+$	–	*			
$N(2060)5/2^-$	–	***	N(2080)3/2$^-$	**	–
$N(2100)1/2^+$	*	***	N(2090)1/2$^-$	*	–
$N(2120)3/2^-$	–	***	N(2200)5/2$^-$	**	–

The nucleon spectrum is more complicated. First, there are more resonances, and second, there are two-quark configurations which are antisymmetric in spin and flavor.[96] Due to instanton induced interactions, the relativistic quark model

[95] In addition, we neglect the possible configuration mixing of states in our discussion.

[96] These two-quark configurations are often called *good diquarks*. They may carry orbital-angular momenta, these are not *frozen* diquarks.

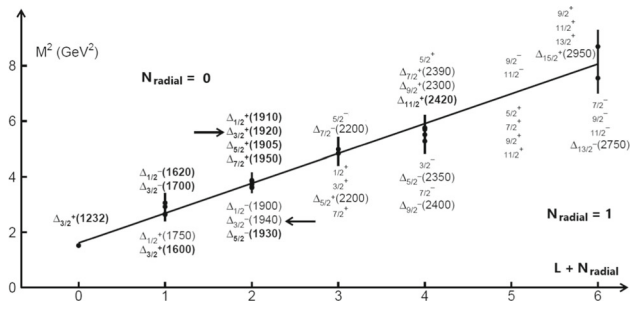


Fig. 219 Regge-like trajectory of Δ^*-resonances. Taken from [2789]

[2730], expects a lowering of states with the respective symmetry. Indeed baryons with two-quark configurations which are antisymmetric in spin and flavor (*good diquarks*) seem to have lower masses than those having *bad diquarks* only. Attempts to include *good-diquark* effects were rather successful [2716,2787]. The χ^2 for the model-data comparison was twice better for the 2-parameter fit than for quark models [2788] when the same mass-uncertainties are assumed.

9.2.5 Hyperons

Nearly no new data on $\bar{K}N$ scattering have become available for several decades except some new data from BNL at very low energy (see Ref. [2790] and references therein). The reaction $\gamma p \to K^+\Sigma\pi$ was studied at JLab and helped to understand the low-energy region [2791]. However, four groups have re-analyzed K^-p reactions using extensive collections of the old data. The new analysis progress was pioneered by the Kent group which performed a comprehensive partial wave analysis [2792,2793]. Energy-independent amplitudes were constructed by starting from an energy-dependent fit and by freezing or releasing sets of amplitudes. The resulting amplitudes were then fit in a coupled-channel approach. The JPAC group performed coupled-channel fits to the partial waves of the Kent group. The fit described the Kent partial waves well while significant discrepancies showed up between data and the observables calculated from their partial-wave amplitudes [2794]. The ANL-Osaka group used the data set collected by the Kent group and derived energy-dependent amplitudes based on a phenomenological SU(3) Lagrangian. Two models were presented which agreed for the leading contributions but which showed strong deviations for weaker contributions [2795,2796]. The BnGa group added further data and tested systematically the inclusion of additional states with any set of quantum numbers. Only small improvements in the fit were found [2797,2798].

The new studies of old data did not change the situation significantly. Some new decay modes were reported, some new but faint signals were found, some were confirmed by one group and missed by others. Several *bumps* were removed from the RPP Tables (for details see [2799]). As a result, our picture of hyperons (with strangeness $S = -1$) remains unclear. Not even all states expected in the first Λ and Σ excitation shell have been seen. In Table 39 all candidates are included.

Very little is known about excited Cascade baryons. A few structures in invariant mass spectra were observed, nearly no spin-parities have been determined. The hope is that at FAIR, JLab and J-PARC (see Sect. 14) new Ξ's and Ω's will be observed and their quantum numbers will be determined.

9.2.6 QCD expectations

The spectrum of excited nucleons has been calculated in different approaches. We list a few here: QCD on a lattice has been used to calculate the spectrum of light-quark baryons including hybrid states (see Sect. 4.5 and [528]). In the Dyson–Schwinger/Bethe–Salpeter approach (see Sect. 5.2 and [2733]) the covariant three-body Fadeev-equation is solved in a rainbow-ladder approximation. The spectra of baryon resonances have been calculated for $J = 1/2^{\pm}$ and $J = 3/2^{\pm}$, reaching for the N^*- and Δ^*-resonances to masses up to about 2000 MeV. AdS/QCD (see Sect. 5.4 and [1003]) predicts a spectrum of N^* and Δ^* that is proportional to $L + N_{\text{radial}}$. Using chiral unitary approaches for the meson–baryon interactions, certain baryon resonances can be generated dynamically (see Sect. 6.2). Various quark models have been developed that treat baryons as bound states of three quarks with constituent masses, a confinement potential and residual quark–quark interactions. The models are discussed in Sect. 9.1. At present, they are still best suited to discuss what has been learned from recent results in the spectroscopy of light baryons.

9.2.7 What did we learn within the quark model?

SU(6)⊗O(3) classification

Table 39 lists the observed N^*-, Δ^*-, Λ^*- and Σ^*-baryons in a SU(6)⊗O(3) classification. This classification assumes non-relativistic constituent quarks. It has been a miracle since the early times of the quark model that this scheme works so well. But baryon resonances often have a leading component in the wave function corresponding to the SU(6)⊗O(3) classification even in relativistic calculations.

The first excitation shell ($N = 1$) is fairly complete. As expected, there are five N^*'s and two Δ^*'s with negative parity. Of the Λ and Σ octet states with negative parity, only the $J^P = 3/2^-$ states are missing.[97] The two states $\Lambda(1800)1/2^-$ and $\Sigma(1750)1/2^-$ are interpreted as

[97] The $N(1700)3/2^-$ is wider than its spin partners and more difficult to identify. This may also be the reason for the absence of the $J^P = 3/2^-$ Λ and Σ states.

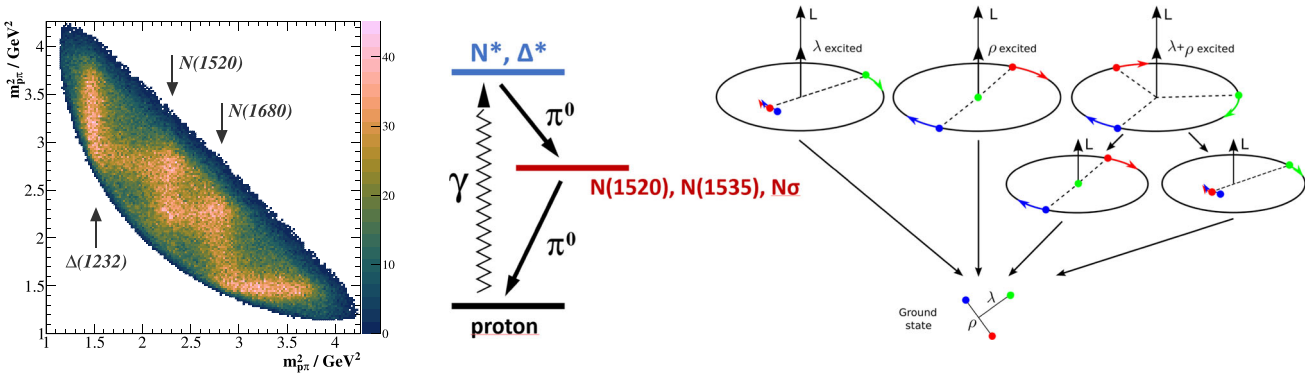


Fig. 220 Left: $\gamma p \to p\pi^0\pi^0$-Dalitz plot for a selected E_γ-bin of 1900–2100 MeV (CBELSA/TAPS) [2800], Middle: Cascade decays of resonances via an immediate state. Right: Classical orbits of nucleon excitations with $L = 2$ (upper row) and $L = 1$ (middle row). Taken from [2786]. The first two pictures in the upper row show excitations of the ρ and λ oscillators, in the third picture both, ρ and λ are excited. When both oscillators are excited, de-excitation leads to an excited intermediate state (middle row)

states with intrinsic spin 3/2: they seem to be spin partners of $\Lambda(1830)5/2^-$ and $\Sigma(1775)5/2^-$. The doublet of negative-parity decuplet Σ states is not uniquely identified. Expected is this doublet at about 1750 MeV, and in the $(56,1_3^-)$-configuration a second doublet at about 2050 MeV and, finally, a triplet at about the same mass. The analysis found (poor) evidence for two doublets, marked [a] in Table 39. The singlet states $\Lambda(1405)1/2^-$ and $\Lambda(1520)3/2^-$ deserve a more detailed discussion.

At higher masses, some choices are a bit arbitrary: Because of its mass, $N(1900)3/2^+$ belongs to the second excitation shell. It may have intrinsic quark spin 1/2 or 3/2, both with $L = 2$. Further, there should be a $3/2^+$ radially excited state with $L = 0$. These three states can mix. Only one of the states is clearly identified. In any case, quark models predict three resonances with $J^P = 3/2^+$ in this mass range while only one is found. Also missing is a doublet of states with $L = 1$ belonging to the 20plet in SU(6)⊗O(3).[98] The production of this doublet is expected to be strongly suppressed for reasons to be discussed below.

Only few hyperons are known that can be assigned to the second excitation shell. The interpretation of some Λ resonances as SU(3) singlet configuration is plausible but not at all compelling.

Missing resonances

In the spectrum of N^* and Δ^*, the first excitation shell is complete, in the second shell, 21 states are expected (two of them likely not observable in πN-elastic scattering or in single/double meson photoproduction), 16 are seen, three are missing. To a large extend, the *missing-resonance* problem is solved for N^* and Δ^*: there are no frozen diquarks. Admittedly, five of the resonances are not yet "established", i.e. have not (yet?) a 3* or 4* status.

In the third shell, only few resonances are known, but the number of expected resonances is quite large and the analysis challenging: 45 N^* and Δ^*, likely with widths often exceeding 300 MeV, are expected to populate an about 400 MeV wide mass range.

Three-quark dynamics in cascade decays

The CBELSA/TAPS collaboration studied cascade decays of high mass resonances via an intermediate resonance down to the ground state nucleon. The analyses were based on a large data base of photoproduction data including final states such as $\gamma p \to p\pi^0\pi^0$ and $p\pi^0\eta$ (see [2783,2785] and Refs. therein). The Dalitz plot of Fig. 220, shows very clearly band-like structures due to the occurrence of baryon resonances in the intermediate state. It was observed that the positive parity N^*- and Δ^*-resonances at a mass of about 1900 MeV show a very different decay pattern. The four N^*-resonances:

$$N(1880)1/2^+, N(1900)3/2^+, N(2000)5/2^+, N(1990)7/2^+,$$

decay with an average branching fraction of $(34 \pm 6)\%$ into $N\pi$ and $\Delta\pi$ and with a branching fraction of $(21 \pm 5)\%$ into the orbitally excited states $N(1520)3/2^-\pi$, $N(1535)1/2^-\pi$, and $N\sigma$. The four Δ^*-states:

$$\Delta(1910)1/2^+, \Delta(1920)3/2^+, \Delta(1905)5/2^+, \Delta(1950)7/2^+,$$

have an average decay branching fraction into $N\pi/\Delta\pi$ of $(44 \pm 7)\%$ while their branching fraction into the excited states mentioned above is almost negligible, only $(5 \pm 2)\%$ [2783]. At the first sight, this is very surprising.

The difference can be traced to the different wave functions. The spin and the flavor wave functions of the four Δ^*-states are both symmetric with respect to the exchange of any two quarks, the spatial wave function needs to be symmetric as well. This means that – having a three-quark-picture

[98] The RPP lists three more N^*/Δ^*-resonances: $N(2040)3/2^+$, $\Delta(2150)1/2^-$, which need confirmation and $N(2100)1/2^+$ which we assign to the 4th shell.

Table 39 The spectrum of N, Δ, Λ and Σ excitations. The first row shows the quantum numbers of the SU(6)⊗O(3) symmetry group. D is the dimensionality of the SU(6) group, L the total internal quark orbital angular momentum, P the parity, N a shell index, S the total quark spin, J the total angular momentum. The assignment of particles to SU(6)⊗O(3) is an educated guess. In the first and second excitation band, all expected states are listed, missing resonances are indicated by a − sign. The third band lists only bands for which at least one candidate exists. The states with an index are special: above 1700 MeV, one pair of Σ states is expected at about 1750 to 1800 MeV, two pairs at about 2000 to 2050 MeV. Two pairs marked[a] are found only. The pairs are shown with the three possible assignments. Likewise, $N(2060)$ and $N(2190)$ marked[b] could form a spin-doublet or be members of a spin-quartet. Likely, the observed pairs of states are mixtures of these allowed configurations (Adapted from [2799])

(D, L_N^P) S J^P	Singlet	Octet			Decuplet	
$(56, 0_0^+)$ $\frac{1}{2}$ $\frac{1}{2}^+$		$N(939)$	$\Lambda(1116)$	$\Sigma(1193)$		
$\frac{3}{2}$ $\frac{3}{2}^+$					$\Delta(1232)$	$\Sigma(1385)$
$(70, 1_1^-)$ $\frac{1}{2}$ $\frac{1}{2}^-$	$\Lambda(1405)$	$N(1535)$	$\Lambda(1670)$	$\Sigma(1620)$	$\Delta(1620)$	$\Sigma(1900)^a$
$\frac{3}{2}^-$	$\Lambda(1520)$	$N(1520)$	$\Lambda(1690)$	$\Sigma(1670)$	$\Delta(1700)$	$\Sigma(1910)^a$
$\frac{3}{2}$ $\frac{1}{2}^-$		$N(1650)$	$\Lambda(1800)$	$\Sigma(1750)$		
$\frac{3}{2}^-$		$N(1700)$	-	-		
$\frac{5}{2}^-$		$N(1675)$	$\Lambda(1830)$	$\Sigma(1775)$		
$(56, 0_2^+)$ $\frac{1}{2}$ $\frac{1}{2}^+$		$N(1440)$	$\Lambda(1600)$	$\Sigma(1660)$		
$\frac{3}{2}$ $\frac{3}{2}^+$					$\Delta(1600)$	$\Sigma(1780)$
$(70, 0_2^+)$ $\frac{1}{2}$ $\frac{1}{2}^+$	$\Lambda(1710)$	$N(1710)$	$\Lambda(1810)$	$\Sigma(1880)$	$\Delta(1750)$	-
$\frac{3}{2}$ $\frac{3}{2}^+$		-	-	-		
$(56, 2_2^+)$ $\frac{1}{2}$ $\frac{3}{2}^+$		$N(1720)$	$\Lambda(1890)$	$\Sigma(1940)$		
$\frac{1}{2}$ $\frac{5}{2}^+$		$N(1680)$	$\Lambda(1820)$	$\Sigma(1915)$		
$\frac{3}{2}$ $\frac{1}{2}^+$					$\Delta(1910)$	-
$\frac{3}{2}$ $\frac{3}{2}^+$					$\Delta(1920)$	$\Sigma(2080)$
$\frac{3}{2}$ $\frac{5}{2}^+$					$\Delta(1905)$	$\Sigma(2070)$
$\frac{3}{2}$ $\frac{7}{2}^+$					$\Delta(1950)$	$\Sigma(2030)$
$(70, 2_2^+)$ $\frac{1}{2}$ $\frac{3}{2}^+$	$\Lambda(2070)$	-	-	-	-	-
$\frac{5}{2}^+$	$\Lambda(2110)$	$N(1860)$	-	-	$\Delta(2000)$	-
$\frac{3}{2}$ $\frac{1}{2}^+$		$N(1880)$	-	-		
$\frac{3}{2}^+$		$N(1900)$	-	-		
$\frac{5}{2}^+$		$N(2000)$	-	-		
$\frac{7}{2}^+$		$N(1990)$	$\Lambda(2085)$	-		
$(20, 1_2^+)$ $\frac{1}{2}$ $\frac{1}{2}^+$	-	-	-	-		
$\frac{3}{2}^+$	-	-	-	-		
$\frac{5}{2}^+$	-					
$(56, 1_3^-)$ $\frac{1}{2}$ $\frac{1}{2}^-$		$N(1895)$	$\Lambda(2000)$	$\Sigma(1900)^a$		
$\frac{3}{2}^-$		$N(1875)$	$\Lambda(2050)$	$\Sigma(1910)^a$		
$\frac{3}{2}$ $\frac{1}{2}^-$					$\Delta(1900)$	$\Sigma(2110)^a$
$\frac{3}{2}^-$					$\Delta(1940)$	$\Sigma(2010)^a$
$\frac{5}{2}^-$					$\Delta(1930)$	-
$(70, 3_3^-)$ $\frac{1}{2}$ $\frac{5}{2}^-$	$\Lambda(2080)$	$N(2060)^b$	-	-	-	-
$\frac{7}{2}^-$	$\Lambda(2100)$	$N(2190)^b$	-	$\Sigma(2100)$	$\Delta(2200)$	-
$(70, 3_3^-)$ $\frac{3}{2}$ $\frac{3}{2}^-$		$N(2120)$	-	-	-	-
$\frac{5}{2}^-$		$N(2060)^b$	-	-	-	-
$\frac{7}{2}^-$		$N(2190)^b$	-	-	-	-
$\frac{9}{2}^-$		$N(2290)$	-	-	-	-

	1. shell	2. shell	3. shell
J^P	$1/2^- - 5/2^-$	$1/2^+ - 7/2^+$	$1/2^- - 9/2^-$
Masses	1500–1750	1700–2100	1900–2300
N	5: 2 2 1	13: 4 5 3 1	30: 7 9 8 5 1
Δ	2: 1 1 –	8: 2 3 2 1	15: 3 5 4 2 1

in mind – that either the ρ- or the λ-oscillator is excited to $\ell = 2$, the other one is not excited. (There is a mixture of the two possibilities $\ell_\rho = 2$, $\ell_\lambda = 0$ or $\ell_\lambda = 2$, $\ell_\rho = 0$). If this state decays, the orbital angular momentum is carried away and the decay products are found preferentially in their ground state.

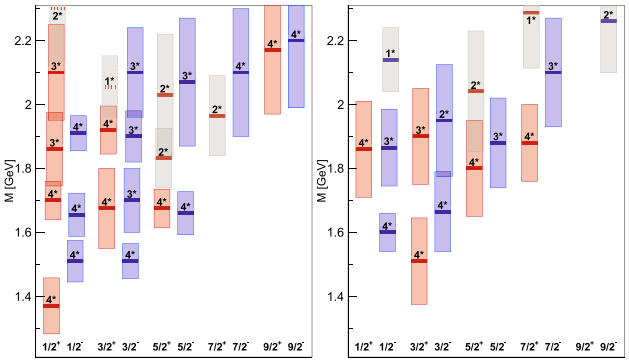


Fig. 221 N*- (left) and Δ^*-resonances (right) above $\Delta(1232)$ for different spin and parities J^P. For each resonance, the real part of the pole position $Re(M_R)$ is given together with a box of length $\pm Im(M_R)$, using the PDG estimates. $2 \cdot Im(M_R)$ corresponds to the total width of the resonance. RPP star ratings are also indicated. If no pole positions are given in the RPP (above the line), the RPP Breit–Wigner estimates for masses and widths are used instead. This is indicated by dashed resonance-mass lines and dashed lines surrounding the boxes. If no RPP-estimates are given, the values above the line have been averaged and the states are shown as gray boxes. This may indicate one measurement above the line only. $\Delta(1750)1/2^+$ is not included, as there is no RPP-value given above the line

The four N^*-states have a spatial wave function with mixed symmetry. Thus the spatial wave function has one part which is mixed-symmetric and one part which is mixed anti-symmetric. In the latter one, both oscillators are excited simultaneously ($\ell_\rho = \ell_\lambda = 1$). If this state decays, one of the excitations remains in the decay product as illustrated in Fig. 220. A similar argument has been used by Hey and Kelly [2801] to explain why the 20'plet in the second excitation shell of Fig. 39 cannot be formed in a πN scattering experiment. For the 20'plet the spacial wave function is entirely antisymmetric, both oscillators are excited simultaneously, and there is no other component in the wave function. A single-step excitation is suppressed.

Parity doublets?

The spontaneous breaking of the chiral symmetry leads to the large mass gap observed between chiral partners: the masses of the $\rho(770)$ meson with spin-parity $J^P = 1^-$ and its chiral partner $a_1(1260)$ with $J^P = 1^+$ differ by about 500 MeV, those of the $J^P = 1/2^+$ nucleon and $N(1535)1/2^-$ by about 600 MeV. In contrast to quark-models expectations and lattice QCD calculations [528] higher-mass baryons are often observed in parity doublets (see Fig. 221), in pairs of resonances having about the same mass, the same total spin J and opposite parities.

This observation and similar observations in meson spectrum has led to the suggestion that chiral symmetry might be effectively restored in highly excited hadrons [2744,2802]. Then, all high-mass resonances should have a parity partner. This is a testable prediction.

In the mass region of 1900 MeV a quartet of well known positive parity Δ^* states exists, consisting of

$\Delta(1910)1/2^+$, $\Delta(1920)3/2^+$, $\Delta(1905)5/2^+$, $\Delta(1950)7/2^+$.

Figure 221 shows the parity partners of the first three states:

$\Delta(1900)1/2^-$, $\Delta(1940)3/2^-$, $\Delta(1930)5/2^-$.

However, the four-star $\Delta(1950)7/2^+$ has no close-by $\Delta(xxx)7/2^-$-state that could serve as parity partner. Where is the closest Δ^* with $J^P = 7/2^-$? Figure 222 shows a resonance scan over the mass region of interest [2746]. There is clear evidence for $\Delta(2200)7/2^-$ (which was upgraded from 1* to 3* based on this result). But its mass difference to $\Delta(1950)7/2^+$ is too large. These two states are no parity partners! Within the quark model and the SU(6)⊗O(3)-systematics, the four positive-parity Δ^*'s have $L = 2$, $S = 3/2$ that couple to $J^P = 1/2^+, \cdots, 7/2^+$. The natural assignment for the three negative-parity Δ^*'s is that they form a triplet with $L = 1$ and $S = 3/2$. Then, they must have one unit of radial excitation. The four positive-parity Δ-states belong to the $2\hbar\omega$ shell and the negative-parity states to the $3\hbar\omega$ shell. With masses considered to be proportional to $L + N_{\text{radial}}$, these seven states are expected to have about the same mass. $\Delta(2200)7/2^-$ has $L = 3, S = 1/2$ and its expected mass is higher. We note that $\Delta(2400)9/2^-$ has $L = 3$, $S = 3/2$, and we assume $N_{\text{radial}} = 1$ for this state (as well as for $\Delta(2750)13/2^-$, see Fig. 219).

9.2.8 Dynamically generated resonances

N^*'s and Δ^*'s

Apart from $\Lambda(1405)1/2^-$ that will be discussed below, the first dynamically generated resonance was the negative-parity $N(1535)1/2^-$ [2741]. At the 1995 International Conference on the Structure of Baryons, Santa Fe, New Mexico, there was a heated discussion between Weise, defending his new approach, and Isgur who argued that $N(1535)1/2^-$ is well understood within the quark model and no new approach is needed. For some time, there was even the idea that there could be two overlapping states but this is excluded by data. Later, in Refs. [2803,2804], $N(1535)1/2^-$ and $N(1650)1/2^-$, were both shown to be generated dynamically. However, $\Delta(1620)1/2^-$ was not.[99] An important question remains: Are (qqq)-resonance poles and dynamically generated poles different descriptions of the same object or do they present different (orthogonal) states?

[99] It should be mentioned that not only the SU(6)⊗O(3)-systematics in the spectrum seems to indicate a 3-quark-nature of $N(1535)1/2^-$ and $N(1650)1/2^-$ but also the electroproduction results discussed in the following Sect. 9.3 indicate that $N(1535)1/2^-$ is a 3-quark state with little meson–baryon contribution only (Q^2 dependence of the transition form factor $A_{1/2}$).

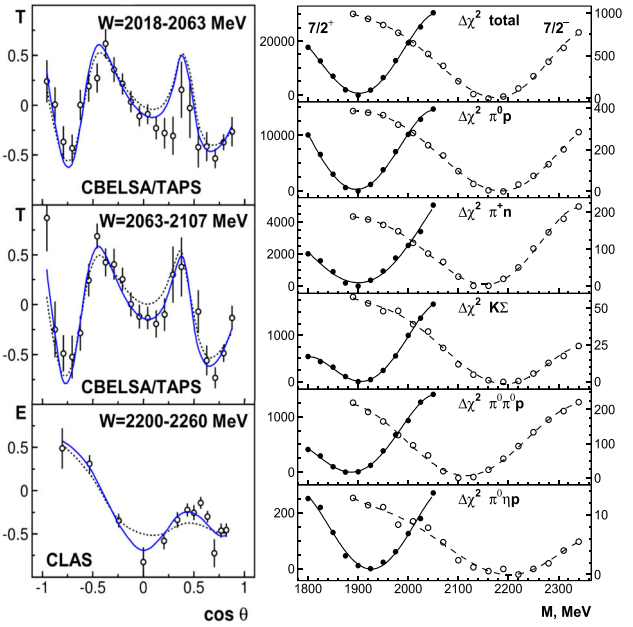


Fig. 222 Left: The new polarization observables T and E shown for selected mass bins (see [2746] for Refs. to the data). The fit curves represent the best fits with (solid) and without (dashed) inclusion of $\Delta(2200)7/2^-$. Right: The increase in pseudo-χ^2 of the fit to a large body of pion- and photo-produced reactions when the mass of $\Delta(1950)7/2^+$ (solid points) or $\Delta(2200)7/2^-$ (open circles) is scanned. The scale on the left (right) abscissa refers to the $7/2^+$ ($7/2^-$) partial wave. The curves are to guide the eye. Adapted/taken from [2746]

The $\Lambda(1405)1/2^-$

The $\Lambda(1405)1/2^-$ mass is very close to the $N\bar{K}$ threshold. Kaiser, Waas and Weise [2805] proved that the resonance can be generated dynamically from $N\bar{K} - \Sigma\pi$ coupled-channel dynamics. Oller and Meissner [2806] studied the S-wave $N\bar{K}$ interactions in a relativistic chiral unitary approach based on a chiral Lagrangian obtained from the interaction of the octet of pseudoscalar mesons and the ground state baryon octet and found two isoscalar resonances in the $\Lambda(1405)1/2^-$ mass region and one isovector state. In a subsequent paper [2807], Jido et al. studied the the effects of SU(3) breaking on the results in detail. These two papers had an immense impact on the further development. It is the only result in light-baryon spectroscopy that is in clear contradiction to the quark model. It introduces a new state $\Lambda(1380)1/2^-$, that has no role in a quark model, it enforces an interpretation of $\Lambda(1405)1/2^-$ as mainly SU(3) octet resonance, and it interprets $\Lambda(1670)1/2^-$ as high-mass partner of $\Lambda(1405)1/2^-$. The $\Lambda(1405)1/2^-$ and $\Lambda(1670)1/2^-$ would then be the strange partners of the $N(1535)1/2^-$ and the $N(1650)1/2^-$. In quark models, $\Lambda(1405)$ is a mainly SU(3) singlet resonance and the octet states $\Lambda(1670)1/2^-$ and $\Lambda(1800)1/2^-$ are the strange partners of $N(1535)1/2^-$ and $N(1650)1/2^-$ (see Table 39). In the quark-model interpretation, the hyperon states $\Lambda(1405)1/2^-$ and $\Lambda(1670)1/2^-$ have close-by $J^P =$

$3/2^-$ partners (the $J^P = 3/2^-$-partner of $\Lambda(1800)1/2^-$ is missing but there is $\Lambda(1830)5/2^-$). The masses of the mainly octet states are about 130 MeV above their non-strange partners.

This conflict initiated an attempt to fit (nearly) all existing data relevant for $\Lambda(1405)1/2^-$ in the BnGa approach [2808]. The data could be fit with one single resonance in the $\Lambda(1405)1/2^-$ region but were also compatible, with a slightly worsened χ^2, with a description using two resonances with properties as obtained in the chiral unitary approach.

9.2.9 Outlook

There is not yet a unified picture of baryons. Regge-like trajectories ($M^2 \propto L + N_{\text{radial}}$) are best described by AdS/QCD. Unitary effective field theories describe consistently meson–baryon interactions, and some resonances can be generated dynamically from their interaction. The quark model is useful to understand cascade decays of highly excited states and is indispensable to discuss the full spectrum including missing resonances. The symmetry of quark pairs, symmetric or anti-symmetric with respect to their exchange, has a significant impact on baryon masses. This effect could be due to an effective gluon exchange. More likely seems an interpretation by quark and gluon condensates, e.g. by instanton-induced interactions. Based on the new high quality (polarized) photoproduction data, new baryon resonances were discovered and our knowledge of properties of existing resonances has increased considerably. Yet, our understanding is still unsatisfactory mirroring the complexity of QCD in the non-perturbative regime. New results from lattice QCD are eagerly awaited and new experiments are needed to understand the spectrum and the properties of baryon resonances in further detail. Those include further precise photoproduction experiments measuring polarization observables not only off the proton but also off the neutron. Multi-meson final states have to be studied further. Strange baryon resonances need to be addressed. Other production processes such as electroproduction, $\bar{p}p$-annihilation, experiments with π- or K-beams and baryon resonances produced in J/ψ or ψ'-decays will also contribute to improve our understanding of the bound states of the strong interaction.

9.3 Nucleon resonances and transition form factors

Volker D. Burkert

Meson photoproduction has become an essential tool in the search for new excited light-quark baryon states. As discussed in the previous section, many new excited states have been discovered thanks to high precision photoproduction data in different final states [2771], and are now included in recent editions of the Review of Particle Physics (RPP) [616]. The exploration of the internal structure of excited states and the effective degrees of freedom contributing to s-channel resonance excitation requires the use of electron beams, which is the subject of this contribution, where the virtuality (Q^2) of the exchanged photon can be varied to pierce through the peripheral meson cloud and probe the quark core and its spatial structure. Electroproduction can thus say something about if a resonance is generated through short distance photon interaction with the small quark core, or through interaction with a more extended hadronic system.

The experimental exploration of resonance transition form factors reaches over 60 years with many review articles describing this history. Here we refer to a few recent ones [2809–2812]. A review of recent electroproduction experiments in hadron physics and their interpretation within modern approaches of strong interaction physics can be found in Ref. [2813].

Electroproduction of final states with pseudoscalar mesons (e.g. $N\pi$, $p\eta$, $K\Lambda$) have been employed at Jefferson Laboratory mostly with the CEBAF Large Acceptance Spectrometer (CLAS) operating at an instantaneous luminosity of $10^{34}\,\text{sec}^{-1}\,\text{cm}^{-2}$. In Hall A and Hall C, pairs of individual well-shielded focusing magnetic spectrometers are employed with more specialized aims and limited acceptance, but operating at much higher luminosity. This experimental program led to new insights into the scale dependence of effective degrees of freedom, e.g. meson–baryon, constituent quarks, and dressed quark contributions. Several excited states, shown in Fig. 223 assigned to their primary $SU(6) \otimes O(3)$ supermultiplets, have been studied this way, mostly with CLAS in Hall B. Most of the resonance couplings have been extracted from single pseudoscalar meson production. In electroproduction, there are 6 complex helicity amplitudes, requiring a minimum of 11 independent measurements for a complete[100] model-independent determination of the amplitudes. In addition, measurements of isospin amplitudes require additional measurements. Following this, the complex amplitudes would need to be subjected to analyses of their phase motions to determine resonance masses on the (real) energy axis, or poles in the (complex) energy plane. Fortunately, in the lower mass range a variety of constraints can be applied to limit the number of unknowns when fitting the cross section data. These include the masses of states quite well known from hadronic processes or from meson photoproduction. Also, the number of possible angular momenta is limited to $l_\pi \leq 2$ in the examples discussed in the following. Additional constraints come from the Watson theorem [2815] that relates the electromagnetic phases to the hadronic ones, and the use of dispersion relations, assuming the imaginary

[100] With the exception of an overall phase that cannot be determined.

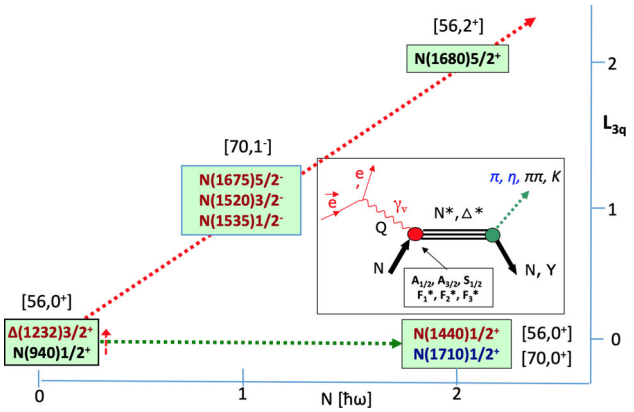


Fig. 223 Excited states of the proton that have been studied in electroproduction to determine their resonance transition amplitudes or form factors. States highlighted in red are discussed in this subsection. Graphics from Ref. [2814]

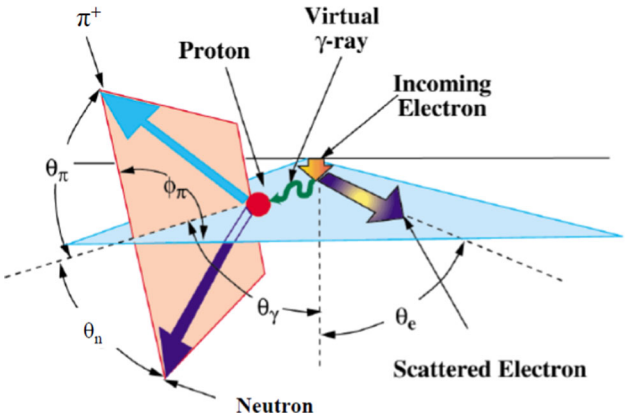


Fig. 224 The kinematics of single π^+ electro-production off protons in the laboratory system

parts of the amplitude are given by the resonance contribution, and the real parts determined through dispersion integrals and additional pole terms. Other approaches use unitary isobar models that parameterize all known resonances and background terms, and unitarize the full amplitudes in a K-matrix procedure. In the following, we show results based on both approaches, where additional systematic uncertainties have been derived from the differences in the two procedures.

The availability of electron accelerators with the possibility of generating high beam currents at CEBAF at Jefferson Lab in the US and MAMI at Mainz in Germany, has enabled precise studies of the internal structure of excited states in the N^* and the Δ^* sectors employing s-channel resonance excitations in large ranges of photon virtuality Q^2. This has enabled probing the degrees of freedom relevant in the resonance excitation as a function of the distance scale probed. This is the topic we will elucidate in the following sections and the relevance to (approximations to) QCD. First we briefly discuss the formalism needed for a quantitative analysis of the single pseudoscalar meson electroproduction.

9.3.1 Formalism in the analysis of electroproduction of single pseudoscalar mesons

The simplest process used in the study of resonance transition amplitudes is single pion or kaon production, e.g. $ep \to e\pi^+ n$. Single π^+ and π^0 production are most suitable for the study of the lower-mass excited states as they couple dominantly to the excited states with masses up to ≈ 1.7 GeV. It may then be useful to describe in more detail the analysis techniques and the formalism used. The unpolarized differential cross section for single pseudoscalar meson production can be written in the one-photon exchange approximation as:

$$\frac{d\sigma}{dE_f d\Omega_e d\Omega_\pi} = \Gamma \frac{d\sigma}{d\Omega_\pi}, \tag{9.5}$$

where Γ is the virtual photon flux,

$$\Gamma = \frac{\alpha_{em}}{2\pi^2 Q^2} \frac{(W^2 - M^2)E_f}{2ME_i} \frac{1}{1-\epsilon}, \tag{9.6}$$

where M is the proton mass, W the mass of the hadronic final state, ϵ is the photon polarization parameter, Q^2 the photon virtuality, E_i and E_f represent the initial and the final electron energies, respectively. Moreover,

$$\epsilon = \left[1 + 2\left(1 + \frac{\nu^2}{Q^2}\right)\tan^2\frac{\theta_e}{2}\right]^{-1} \tag{9.7}$$

and

$$\frac{d\sigma}{d\Omega_\pi} = \sigma_T + \epsilon\sigma_L + \epsilon\sigma_{TT}\cos 2\phi_\pi + \sqrt{2\epsilon(1+\epsilon)}\sigma_{LT}\cos\phi_\pi .$$

The kinematics for single π^+ production is shown in Fig. 224. The observables of the process $\gamma_v p \to \pi N'$ can be expressed in terms of six parity-conserving helicity amplitudes [2811, 2816, 2817] :

$$H_i = \langle \lambda_\pi; \lambda_N | T | \lambda_{\gamma_v}; \lambda_p \rangle, \tag{9.8}$$

where λ denotes the helicity of the respective particle, $\lambda_\pi = 0$, $\lambda_N = \pm\frac{1}{2}$, $\lambda_{\gamma_v} = \pm 1, 0$, and $\lambda_p = \pm\frac{1}{2}$, and H_i are complex functions of Q^2, W, and θ_π^*.

9.3.2 Multipoles and partial wave decompositions

The response functions in (1) are given by:

$$\sigma_T = \frac{\vec{p}_\pi W}{2KM}(|H_1|^2 + |H_2|^2 + |H_3|^2 + |H_4|^2), \tag{9.9}$$

$$\sigma_L = \frac{\vec{p}_\pi W}{2KM}(|H_5|^2 + |H_6|^2), \tag{9.10}$$

$$\sigma_{TT} = \frac{\vec{p}_\pi W}{2KM} Re(H_2 H_3^* - H_1 H_4^*), \tag{9.11}$$

$$\sigma_{LT} = \frac{\vec{p}_\pi W}{2KM} Re[H_5^*(H_1 - H_4) + H_6^*(H_2 + H_3)], \tag{9.12}$$

where $\vec{p}_\pi$ is the pion 3-momentum in the hadronic center-of-mass system, and K is the equivalent real photon lab energy needed to generate a state with mass W:

$$K = \frac{W^2 - M^2}{2M}. \tag{9.13}$$

The helicity amplitudes $H_i, i = 1$–6, can be expanded into Legendre polynomials:

$$\begin{aligned}
H_1 &= \frac{1}{\sqrt{2}} \sin\theta \cos\frac{\theta}{2} \sum_{l=1}^{\infty} (B_{l+} - B_{(l+1)-})(P_l'' - P_{l+1}'') \\
H_2 &= \sqrt{2} \cos\frac{\theta}{2} \sum_{l=1}^{\infty} (A_{l+} - A_{(l+1)-})(P_l' - P_{l+1}') \\
H_3 &= \frac{1}{\sqrt{2}} \sin\theta \sin\frac{\theta}{2} \sum_{l=1}^{\infty} (B_{l+} + B_{(l+1)-})(P_l'' + P_{l+1}'') \\
H_4 &= \sqrt{2} \sin\frac{\theta}{2} \sum_{l=1}^{\infty} (A_{l+} + A_{(l+1)-})(P_l' + P_{l+1}') \\
H_5 &= \sqrt{2} \cos\frac{\theta}{2} \sum_{l=1}^{\infty} (C_{l+} - C_{(l+1)-})(P_l' - P_{l+1}') \\
H_6 &= \sqrt{2} \sin\frac{\theta}{2} \sum_{l=1}^{\infty} (C_{l+} + C_{(l+1)-})(P_l' + P_{l+1}'),
\end{aligned} \tag{9.14}$$

where the A_{l+} and B_{l+} etc., are the transverse partial wave helicity elements for $\lambda_{\gamma p} = \frac{1}{2}$ and $\lambda_{\gamma p} = \frac{3}{2}$, and $C_\pm$ the longitudinal partial wave helicity elements. In the subscript, $l+$ and $(l+1)-$ define the π orbital angular momenta, and the sign $\pm$ is related to the total angular momentum $J = l_\pi \pm \frac{1}{2}$. In the analysis of data on the $N\Delta(1232)$ transition, linear combinations of partial wave helicity elements are expressed in terms of electromagnetic multipoles:

$$M_{l+} = \frac{1}{2(l+1)}[2A_{l+} - (l+2)B_{l+}] \tag{9.15}$$

$$E_{l+} = \frac{1}{2(l+1)}(2A_{l+} + lB_{l+}) \tag{9.16}$$

$$M_{l+1,-} = \frac{1}{2(l+1)}(2A_{l+1,-} - lB_{l+1,-}) \tag{9.17}$$

$$E_{l+1,-} = \frac{1}{2(l+1)}[-2A_{l+1,-} + (l+2)B_{l+1,-}] \tag{9.18}$$

$$S_{l+} = \frac{1}{l+1}\sqrt{\frac{\vec{Q}^{*2}}{Q^2}} C_{l+} \tag{9.19}$$

$$S_{l+1,-} = \frac{1}{l+1}\sqrt{\frac{\vec{Q}^{*2}}{Q^2}} C_{l+1,-}, \tag{9.20}$$

where $\vec{Q}^*$ is the photon 3-momentum in the hadronic rest frame. The electromagnetic multipoles are often used to describe the transition from the nucleon ground state to the $\Delta(1232)$, which is dominantly described as a magnetic dipole transition M_{1+}. The electromagnetic multipoles as well as the partial wave helicity elements are complex quantities and contain both non-resonant and resonant contributions. In order to compare the results to model predictions and LQCD, an additional analysis must be performed to separate the resonant parts $\hat{A}_\pm$, $\hat{B}_\pm$, etc., from the non-resonant parts of the amplitudes. In a final step, the known hadronic properties of a given resonance can be used to determine photocoupling helicity amplitudes that characterize the electromagnetic vertex:

$$\hat{A}_{l\pm} = \mp F C^I_{\pi N} A_{1/2}, \tag{9.21}$$

$$\hat{B}_{l\pm} = \pm F \sqrt{\frac{16}{(2j-1)(2j+3)}} C^I_{\pi N} A_{3/2}, \tag{9.22}$$

$$\hat{S}_{l\pm} = -F \frac{2\sqrt{2}}{2J+1} C^I_{\pi N} S_{1/2}, \tag{9.23}$$

$$F = \sqrt{\frac{1}{(2j+1)}\pi} \frac{K}{p_\pi} \frac{\Gamma_\pi}{\Gamma^2}$$

where the $C^I_{\pi N}$ are isospin coefficients. The total transverse absorption cross section for the transition into a specific resonance is given by:

$$\sigma_T = \frac{2M}{W_R \Gamma}(A_{1/2}^2 + A_{3/2}^2). \tag{9.24}$$

Experiments in the region of the $\Delta(1232)\frac{3}{2}^+$ resonance often determine the electric quadrupole ratio R_{EM}

$$R_{EM} = \frac{Im(E_{1+})}{Im(M_{1+})} \tag{9.25}$$

and the scalar quadrupole ratio R_{SM}

$$R_{SM} = \frac{Im(S_{1+})}{Im(M_{1+})} \tag{9.26}$$

where E_{1+}, S_{1+}, and M_{1+} are the electromagnetic transition multipoles at the mass of the $\Delta(1232)\frac{3}{2}^+$ resonance.

9.3.3 Resonance analysis tools

A model-independent determination of the amplitudes contributing to the electro-excitation of resonances in single pseudoscalar pion production $ep \to e'N\pi$ (see kinematics of single pion production in Fig. 224) requires a large number of independent measurements at each value of the electron kinematics W, Q^2, the hadronic cms angle $\cos\theta^\pi$, and the azimuthal angle ϕ^π describing the angle between the electron scattering plane and the hadronic decay plane. Such a measurement requires full exclusivity of the final state and

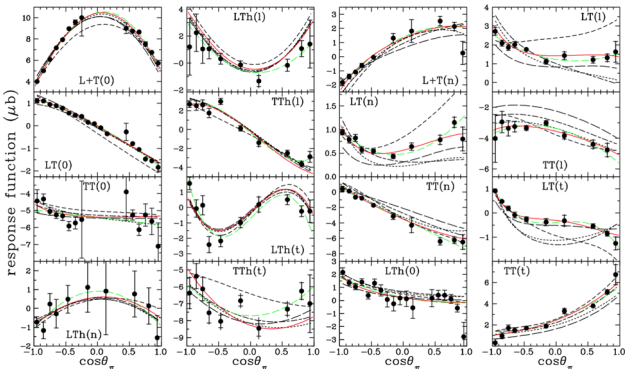


Fig. 225 JLab/Hall A data for $\vec{e}p \to e\vec{p}\pi^0$ response functions at W=1.232 GeV and $Q^2 = 1.0$ [2818]. Notations refer to transverse (t), normal (n) and longitudinal (l) components of the proton recoil polarization. The curves correspond to results obtained using SAID (short dashed), MAID (dashed-dotted), and the dynamical models DMT [2819] (dotted), and SL [2820] (long-dashed/green). The other curves correspond to Legendre and multipole fits performed by the authors

employing both polarized electron beams and the measurements of the nucleon recoil polarization.

Such measurements would in general require full 4π coverage for the hadronic final state. The only measurement that could claim to be complete was carried out at Jefferson Lab in Hall A [2818] employing a limited kinematics centered at resonance for $\vec{e}p \to e'\vec{p}\pi^0$ at $W = 1.232$ GeV, and $Q^2 \approx 1$ GeV^2. Figure 225 shows the 16 response functions extracted from this measurement. The results of this measurement in terms of the magnetic $N\Delta$ transition form factor and the quadrupole ratios are included in Fig. 226 among other data. They coincide very well with results of other experiments [2821–2824] using different analysis techniques that may be also applied to broader kinematic conditions, especially higher mass resonances. Details of the latter are discussed in [2811,2825]. We briefly summarize them here:

- *Dispersion Relations* have been employed in two ways: One is based on fixed-t dispersion relations for the invariant amplitudes and was successfully used throughout the nucleon resonance region. Another way is based on DR for the multipole amplitudes of the $\Delta(1232)$ resonance, and allows getting functional forms of these amplitudes with one free parameter for each of them. It was employed for the analysis of the more recent data.
- *The Unitary Isobar Model (UIM)* was developed in [2831] from the effective Lagrangian approach for pion photoproduction [2832]. Background contributions from t-channel ρ and ω exchanges are introduced and the overall amplitude is unitarized in a K-matrix approximation (Fig. 227).
- *Dynamical Models* have been developed, as SAID from pion photoproduction data [2833], the Sato-Lee model was developed in [2834]. Its essential feature is the consistent description of πN scattering and the pion electroproduction from nucleons. It was utilized in the study of $\Delta(1232)$ excitations in the $ep \to ep\pi^0$ channel [2820]. The Dubna–Mainz–Taipei model [2835] builds unitarity via direct inclusion of the πN final state in the T-matrix of photo- and electroproduction.

9.3.4 *Models for light-quark resonance electroproduction*

In order to learn from the meson electroproduction data about the internal spin and spatial electromagnetic structure, it is essential to have advanced models available with links to the fundamentals of QCD.

While most of the analyses have focused on single pseudoscalar meson production, such as

$$\gamma_v p \to N\pi,\ p\eta,\ K\Lambda,\ K\Sigma,$$

more recent work included the $p\pi^+\pi^-$ final state both in real photoproduction [2837] as well as in electroproduction [2838]. The 2-pion final state has more sensitivity to excited N^* and Δ^* states in the mass range above 1.6 GeV, with several states dominantly coupling to $N\pi\pi$ final states, enabling the study of their electromagnetic transition form factors in the future.

9.3.5 *The $N\Delta(1232)\frac{3}{2}^+$ transition*

The Δ^{++} isobar was first observed 70 years ago in Enrico Fermi's experiment that used a π^+ meson beam scattered off the protons in a hydrogen target [2839]. The cross section showed a sharp rise above threshold towards a mass near 1200 MeV. While the energy of the pion beam was not high enough to see the maximum and the fall-off following the peak, a strong indication of the first baryon resonance was observed. It took 12 more years and the development of the underlying symmetry in the quark model before a microscopic explanation of this observation could emerge. There was, however, a problem; while the existence of the $\Delta^{+,0,-}$ could be explained within the model, the existence of the $\Delta(1232)^{++}$, which within the quark model would correspond to a state $|u\uparrow u\uparrow u\uparrow\rangle$, was forbidden as it would have an overall symmetric wave function. It took the introduction of para Fermi statistics [31] what later became "color" (see Sect. 1.2), to make the overall wave function anti-symmetric. In this way the $\Delta^{++}(1232)$ resonance may be considered a harbinger of the development of QCD.

The nucleon to $\Delta(1232)\frac{3}{2}^+$ transition is now well measured in a large range of Q^2 [2822–2824]. At the real photon point, it is explained by a dominant magnetic transition from the nucleon ground state to the $\Delta(1232)$ excited state. Additional contributions are related to small D-wave components

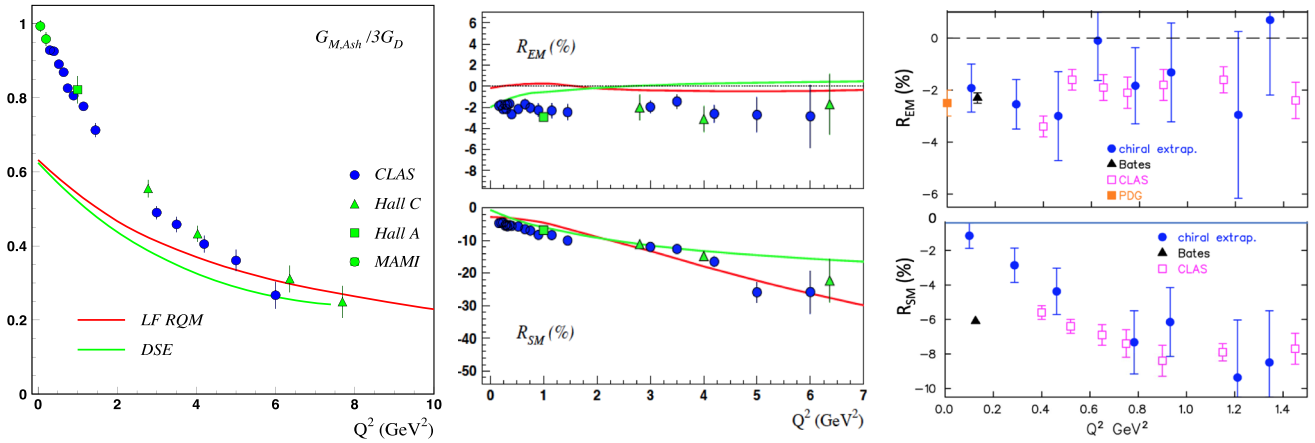


Fig. 226 The $N\Delta(1232)$ transition amplitudes. Left: The magnetic NΔ transition form factor normalized to the dipole form factor and compared with the Light-Front Relativistic Quark Model (LFRQM) [2826,2827] with running quark mass, and with results using the Dyson–Schwinger Equation [2828]. Both predictions are close to the data at high Q^2. At $Q^2 < 3\,\text{GeV}^2$ meson–baryon contributions are significant. Middle: The electric (top) and scalar (bottom) quadrupole/magnetic-dipole ratios R_{EM} and R_{SM}. Right: R_{EM} and R_{SM} from Lattice QCD [2829,2830] compared to data in the low Q^2 domain

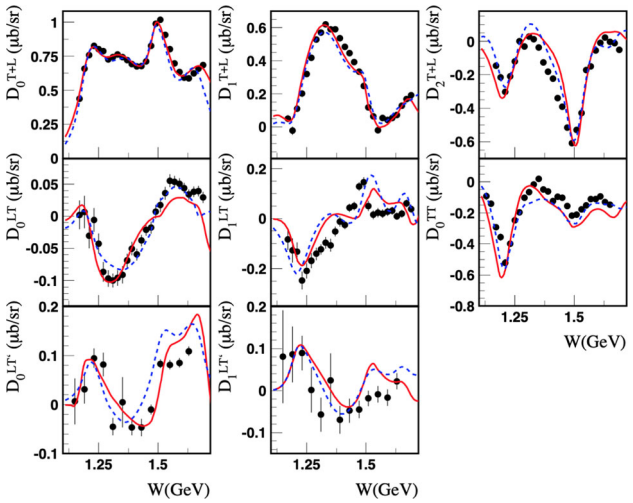


Fig. 227 Sample of results of an analysis by the JLab group of the Legendre moments of $\vec{e}p \to e\pi^+ n$ structure functions in comparison with experimental data [2836] at $Q^2 = 2.44\,\text{GeV}^2$. The solid (dashed) curves correspond to results obtained using the DR (UIM) approach

in both the nucleon and the $\Delta(1232)$ wave functions leading to electric quadrupole and scalar quadrupole transitions. These remain in the few % ranges at small Q^2. The magnetic transition is to $\approx 65\%$ given by a simple spin flip of one of the valence quarks as seen in Fig. 226. The remaining 35% of the magnetic dipole strength is attributed to meson–baryon contributions.

The electric quadrupole ratio R_{EM} was found as:

$$R_{EM} \approx -0.02. \tag{9.27}$$

There has been a longstanding prediction of asymptotic pQCD, that $R_{EM} \to +1$ at $Q^2 \to \infty$. Results on the magnetic transition form factor $G_{Mn,Ash}$, defined in the Ash convention [2840], and on the quadrupole transition ratios are shown in Fig. 226. $G_{Mn,Ash}$ is shown normalized to the dipole form factor, but shows a much faster Q^2 fall-off compared to that. In comparison to the advanced LF RQM with momentum-dependent constituent quark mass, and with the Dyson–Schwinger Equation (DSE-QCD) results, there is good agreement at the high-Q^2 end of the data. The discrepancy at small $Q^2 = 0$ is likely due to the meson–baryon contributions at small Q^2, which are not modeled in either of the calculations.

The quadrupole ratio R_{EM} shows no sign of departing significantly from its value at $Q^2 = 0$, even at the highest $Q^2 \approx 6.5\,\text{GeV}^2$. Both calculations barely depart from $R_{EM} = 0$, and remain near zero at all $Q^2 > 2\,\text{GeV}^2$. This indicates that the negative constant value shown by the data is likely due to meson–baryon contributions that are not included in the theoretical models. For the scalar quadrupole ratio R_{SM} the asymptotic prediction in holographic QCD (hQCD) [2845] is:

$$R_{SM} = \frac{ImS_{1+}}{ImM_{1+}} \to -1, \quad \text{at } Q^2 \to \infty, \tag{9.28}$$

while R_{EM} in hQCD is predicted to approach +1 asymptotically. The R_{SM} data show indeed a strong trend towards increasing negative values at larger Q^2, semi-quantitatively described by both calculations at $Q^2 < 4\,\text{GeV}^2$. The Dyson–Schwinger equation approach predicts a flattening of R_{SM} at $Q^2 > 4\,\text{GeV}^2$, while the Light Front relativistic Quark Model predicts a near constant negative slope of $R_{SM}(Q^2)$ also at higher Q^2.

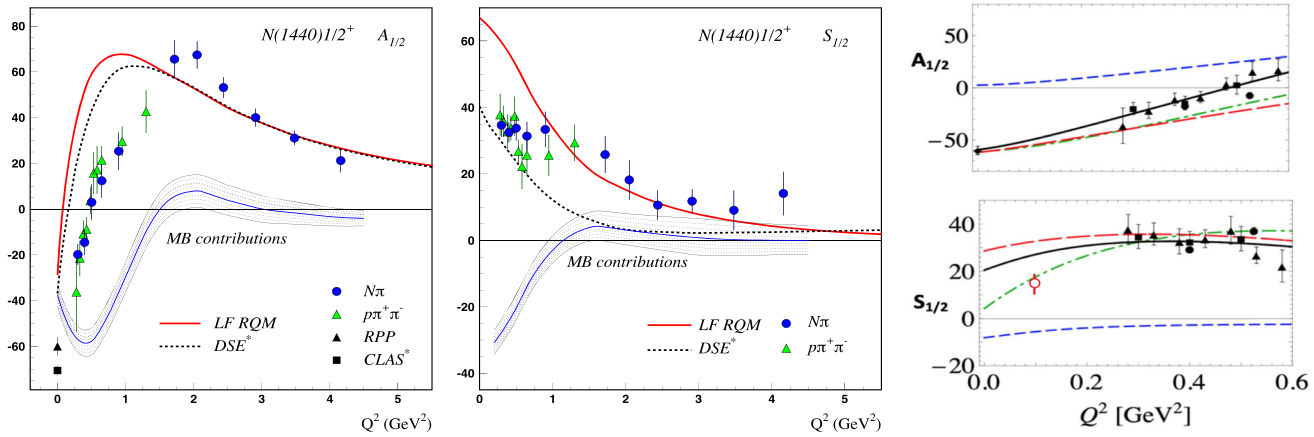


Fig. 228 Helicity transition amplitudes for the proton to Roper $N(1440)\frac{1}{2}^+$ excitation compared to model calculations in LFRQM, DSE, and EFT; see text. Left: Transverse $A_{1/2}$ amplitude. Middle: Scalar $S_{1/2}$ amplitude. Right: Helicity amplitudes of the Roper resonance $N(1440)\frac{1}{2}^+$ at low Q^2. Data are compared to calculations within Effective Field Theory [2841], shown in solid black lines. The other broken lines are parts of the full calculations. The data are from [2821,2842,2843]. The open red circle at $Q^2 \approx 0.1$ GeV^2 is the result of an analysis of $ep \to ep\pi^0$ data from MAMI [2844]

9.3.6 *The Roper resonance* $N(1440)\frac{1}{2}^+$

The Roper resonance, discovered in 1964 [2846] in a phase shift analysis of elastic πN scattering data, has been differently interpreted for half a century. In the non-relativistic quark model (nrQM), the state is the first radial excitation of the nucleon ground state with a mass expected around 1750 MeV, much higher than the measured Breit–Wigner mass of $\approx$ 1440 MeV. This discrepancy is now understood as the consequence of a dynamical coupled channel effect that shifts the mass below the mass of the $N(1535)1/2^-$ state, the negative-parity partner of the nucleon [2847]. Another problem with the quark model was the sign of the transition form factor $A_{1/2}(Q^2 = 0)$, predicted in the nrQM as large and positive, while experimental analyses showed a negative value.

These discrepancies resulted in different interpretations of the state that could only be resolved with electroproduction data from CLAS at Jefferson Lab, the development of continuous QCD approximations in the Dyson–Schwinger equation approach [2848] and Light Front Relativistic QM with momentum-dependent quark masses [2826] shown in Fig. 228, and Lattice data [2849,2850]. A recent review of the history and current status of the Roper resonance, is presented in a colloquium-style article published in Review of Modern Physics [2851].

Descriptions of the baryon resonance transitions form factors, including the Roper resonance $N(1440)\frac{1}{2}^+$, have also been carried out within holographic models [2852,2853]. In the range $Q^2 < 0.6$ GeV^2, calculations based on meson–baryon degrees of freedom and effective field theory [2841] have been successfully performed, as may be seen in Fig. 228. Earlier model descriptions, such as the Isgur-Karl model that describe the nucleon as a system of 3 constituent quarks in a confining potential and a one-gluon exchange contribution leading to a magnetic hyperfine splitting of states [764,2752], and the relativized version of Capstick [771] have popularized the model that became the basis for many further developments and variations, e.g. the light front relativistic quark model, and the hypercentral quark model [2854]. Other models were developed in parallel. The cloudy bag model [777] describes the nucleon as a bag of 3 constituent quarks surrounded by a cloud of pions. It has been mostly applied to nucleon resonance excitations in real photoproduction, $Q^2 = 0$ [777,2855], with some success in the description of the $\Delta(1232)\frac{3}{2}^+$ and the Roper resonance transitions.

There is agreement with the data at $Q^2 > 1.5$ GeV^2 for these two states, while the meson–baryon contributions for the $\Delta(1232)$ are more extended, and agreement with the quark based calculations is reached at $Q^2 > 4$ GeV^2. The calculations deviate significantly from the data at lower Q^2, which indicates the presence of non-quark core effects. For the Roper resonance such contributions have been described successfully in dynamical meson–baryon models [2856] and in effective field theory [2841]. Calculations on the Lattice for the N-Roper transition form factors F_1^{pR} and F_2^{pR}, which are combinations of the transition amplitudes $A_{1/2}$ and $S_{1/2}$, have been carried out with dynamical quarks [2850]. The results agree well with the data in the range $Q^2 < 1.0$ GeV^2, where data and calculations overlap Fig. 229.

New electroproduction data on the Roper [2844] and on several higher mass states have been obtained in the 2-pion channel, specifically in $ep \to e'p\pi^+\pi^-$ [2857].

The mass of the Roper state has been computed on the Lattice and extrapolated to the physical pion mass, show-

ing good agreement with the physical value measured with a Breit–Wigner parametrization. It should be noted that the Roper mass measured at the pole in the complex plane is significantly different from the value obtained using the BW ansatz. Supported by an extensive amount of single pion electroproduction data, covering the full phase space in the pion polar and azimuthal center-of-mass angles, and accompanied by several theoretical modeling, we can summarize our current understanding of the $N(1440)\frac{1}{2}^+$ state as follows:

- The Roper resonance is, at heart, the first radial excitation of the nucleon.
- It consists of a well-defined dressed-quark core, which plays a role in determining the system's properties at all length scales, but exerts a dominant influence on probes with $Q^2 > m_N^2$, where m_N is the nucleon mass.
- The core is augmented by a meson cloud, which both reduces the Roper's core mass by $\approx 20\%$, thereby solving the mass problem that was such a puzzle in constituent quark model treatments, and, at low Q^2, contributes an amount to the electroproduction transition form factors that is comparable in magnitude with that of the dressed quark core, but vanishes rapidly as Q^2 is increased beyond m_N^2.

As stated in the conclusions of [2851]: "The fifty years of experience with the Roper resonance have delivered lessons that cannot be emphasized too strongly. Namely, in attempting to predict and explain the QCD spectrum, one must fully consider the impact of meson–baryon final state interactions and the coupling between channels and states that they generate, and look beyond merely locating the poles in the S-matrix, which themselves reveal little structural information, to also consider the Q^2 dependencies of the residues, which serve as a penetrating scale-dependent probe of resonance composition."

9.3.7 *Transition form factors of $N(1535)\frac{1}{2}^-$ – a state with a hard quark core.*

This state is the parity partner state to the ground state nucleon, with the same spin 1/2 but with opposite parity, its quark content requires an orbital L=1 excitation in the transition from the proton. In the $SU(6) \otimes O(3)$ symmetry scheme, the state is a member of the $[70, 1^-]$ super multiplet. This state couples equally to $N\pi$ and to $N\eta$ final state. It has therefore be probed using both decay channels $ep \to ep\eta$ and $ep \to eN\pi^{+,0}$. Because of isospin $I = 1/2$ for nucleon states, the coupling to the charged $\pi^+ n$ channel is preferred over $\pi^0 p$ owing to the Clebsch–Gordon coefficients.

The $A_{1/2}$ helicity amplitude for the $\gamma pN(1535)\frac{1}{2}^-$ resonance excitation shown in Fig. 229 represents the largest range in Q^2 of all nucleon states for which resonance transition form factors have been measured as part of the broad experimental program at JLab.

For this state, as well as for the $N(1440)\frac{1}{2}^+$ state, advanced relativistic quark model calculations [2860], DSE-QCD calculations [2848] and Light Cone sum rule results [2861] are available, employing QCD-based modeling of the excitation of the quark core for the first time.

The transverse transition amplitude $A_{1/2}$ of $N(1535)\frac{1}{2}^-$ is a prime example of the power of meson electroproduction to unravel the internal structure of the resonance transition. In the previous Sect. 9.2, the nature of this state is discussed as a possible example of a dynamically generated resonance. The electroproduction data shown here reveal structural aspects of the state and its nature that require a different interpretation. The transition form factor $A_{1/2}$ of the state, shown in Fig. 229, is quantitatively reproduced over a large range in Q^2 by two alternative approaches, the LFRQM and the LCSR. Both calculations are based on the assumptions of the presence of a 3-quark core. Note that there is a deviation from the quark calculations at $Q^2 < 1{-}2\,\mathrm{GeV}^2$, highlighted as the shaded area in Fig. 229, which may be assigned to the presence of non-quark contributions. Attempts to compute the transition form factors within strictly dynamical models have not succeeded in explaining the data [2862]. The discrepancy could be resolved if the character of the probe, meson (pion) in the case of hadron interaction and short range photon interaction in the case of electroproduction, probe different parts of the resonance's spatial structure: peripheral in case of meson scattering and short distance behavior in electroproduction. The peripheral meson scattering and low Q^2 meson photoproduction reveal the dynamical features of the state, whereas high Q^2 electroproduction reveals the structure of the quark core.

9.3.8 *The helicity structure of the $N(1520)\frac{3}{2}^-$*

The $N(1520)\frac{3}{2}^-$ state corresponds to the lowest excited nucleon resonance with $J^P = \frac{3}{2}^-$. Its helicity structure is defined by the Q^2 dependence of the two transverse transition amplitudes $A_{1/2}$ and $A_{3/2}$. They are both shown in Fig. 230. A particularly interesting feature of this state is that at the real photon point, $A_{3/2}$ is strongly dominant, while $A_{1/2}$ is very small. However, at high Q^2, $A_{1/2}$ is becoming dominant, while $A_{3/2}$ drops rapidly. This behavior is qualitatively consistent with the expectation of asymptotic QCD, which predicts the transition helicity amplitudes to behave like:

$$A_{1/2} \propto \frac{a}{Q^3}, \quad A_{3/2} \propto \frac{b}{Q^5}. \tag{9.29}$$

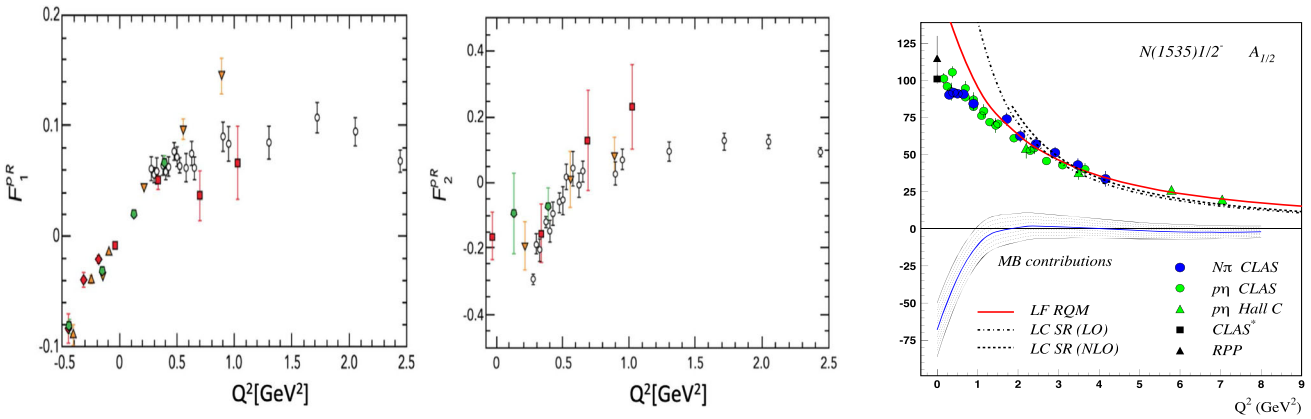


Fig. 229 Left and middle: Dirac and Pauli transition form factors F_1 and F_2 for the proton to $N(1440)1/2^+$ transition compared to Lattice QCD calculations [2850] with pion masses (in GeV): 0.39 (red squares), 0.45 (orange triangles), and 0.875 (green circles) on the $N_f = 2 + 1$ anisotropic lattices, compared to CLAS results (black circles). The F_1 and F_2 form factors are linear combinations of the $A_{1/2}$ and $S_{1/2}$ amplitudes. Right: The transverse transition helicity amplitude $A_{1/2}$ versus Q^2. At $Q^2 > 2\,\text{GeV}^2$ the data are well described by the light-cone sum rules LCSR [2858]. The light front relativistic quark model (LFRQM) [2859] describes that data at $Q^2 > 1\,\text{GeV}^2$

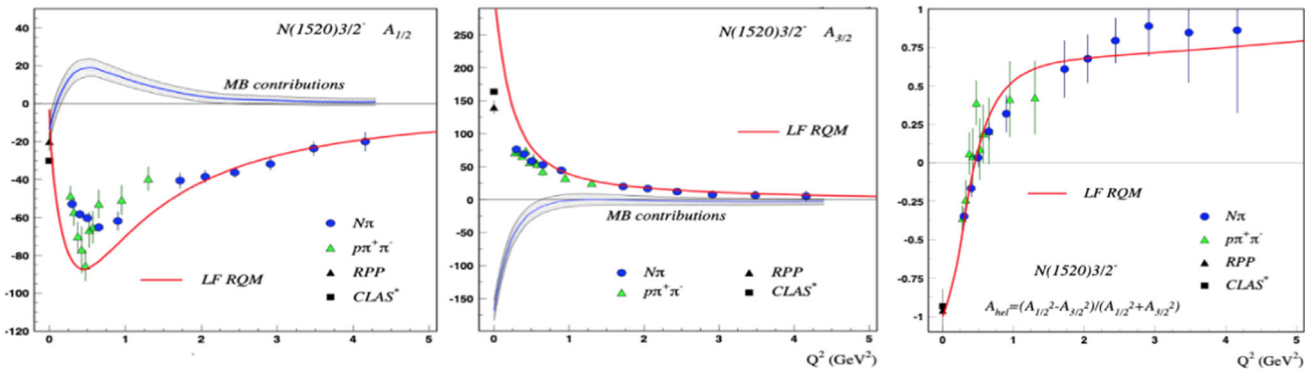


Fig. 230 The transverse helicity transition amplitudes of $N(1520)\frac{3}{2}^-$ versus Q^2, compared to the LFRQM, $A_{1/2}$ (left), $A_{3/2}$ (middle). The shaded area indicates the contribution from non-quark contributions as estimated from the difference of the measured data points and the LF RQM contribution, likely due to hadronic contributions. Right: Helicity asymmetry A_{hel}, as defined in Eq. (9.30). Graphics from Ref. [2811]

The helicity asymmetry

$$A_{hel} = \frac{A_{1/2}^2 - A_{3/2}^2}{A_{1/2}^2 + A_{3/2}^2}, \tag{9.30}$$

shown in Fig. 230, illustrates this rapid change in the helicity structure of the $\gamma_v pN(1520)3/2^-$ transition. At $Q^2 > 2$ GeV^2, $A_{1/2}$ fully dominates the process.

9.3.9 *The helicity transition amplitudes to the* $N(1535)\frac{1}{2}^-$ *resonance*

The Roper $N(1440)\frac{1}{2}^+$ resonance, at the core, is a radial excitation. Its parity partner, the $N(1535)\frac{1}{2}^-$, in the quark model, is an orbitally excited quark state of the nucleon. It is then interesting to compare the transition amplitude to the $N(1535)\frac{1}{2}^-$ with the amplitude to the Roper resonance. The $N(1535)\frac{1}{2}^-$ is, together with the $\Delta(1232)\frac{3}{2}^+$, the best measured state, and both its transverse and longitudinal (scalar) amplitudes are well measured [2811]. Figure 231 shows the transverse amplitude $A_{1/2}$ versus Q^2. They reveal a very different behavior at low Q^2, where $N(1535)\frac{1}{2}^-$ indicates only small effects from meson–baryon contributions below $Q^2 \approx 1\,\text{GeV}^2$, while the $N(1440)\frac{1}{2}^+$ changes sign at small Q^2 and reveals a much more prominent impact of meson–baryon contributions. The Q^2 dependence of the $N(1535)\frac{1}{2}^-$ is well reproduced by LC SR in LO and NLO. There have been attempts to explain the transition form factor of the $N(1535)\frac{1}{2}^-$ as a dynamically generated resonance [2862] that does not achieve quantitative agreement with experiment and concludes that admixture with a genuine three-quark state is demanded that could help to better reproduce

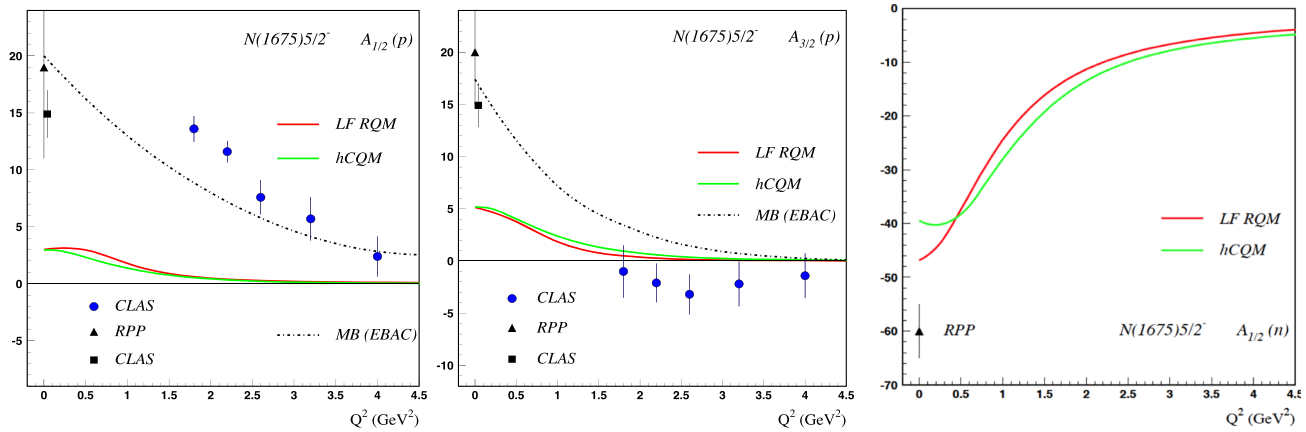


Fig. 231 The transverse amplitudes of the proton to $N(1675)\frac{5}{2}^-$ transition compared to the LF RQM [2863], hypercentral QM [2864], and contributions from meson–baryon (MB) coupled channel dynamics [2865]. Left: $A_{1/2}$, Middle:$A_{3/2}$. Both quark models predict very small amplitudes for the proton, while the meson–baryon contributions estimate is large and is close to the data. Right: $A_{1/2}$ for neutron target (only photoproduction data available) compared to the LFRQM and hCQM. Both quark models predict large amplitudes for neutrons, more than factor 10 compared to protons at $Q^2 = 0$. Assuming similar meson–baryon contributions as in the proton case with opposite sign could quantitatively explain the single measured value at the photon point

the magnitude or the Q^2 falloff of the $A_{1/2}$ helicity amplitude.

9.3.10 *The $N(1675)\frac{5}{2}^-$ state – revealing the meson–baryon contributions*

In previous discussions we have concluded that meson–baryon degrees of freedom provide significant strength to the resonance excitation in the low-Q^2 domain where quark based approaches LF RQM, DSE/QCD, and LC SR calculations fail to reproduce the transition amplitudes quantitatively. Our conclusion rests, in part, with this assumption. But, how can we be certain of the validity of this conclusion?

The $N(1675)\frac{5}{2}^-$ resonance allows us to test this assumption, quantitatively. Figure 231 shows our current knowledge of the transverse helicity amplitudes $A_{1/2}(Q^2)$ and $A_{3/2}(Q^2)$, for proton target compared to RQM [2859] and hypercentral CQM [2864] calculations. The specific quark transition for a $J^P = \frac{5}{2}^-$ state belonging to the $[SU(6) \otimes O(3)] = [70, 1^-]$ supermultiplet configuration, in nonrelativistic approximation prohibits the transition from the proton in a single quark transition. This suppression, known as the Moorhouse selection rule [760], is valid for the transverse transition amplitudes $A_{1/2}$ and $A_{3/2}$ at all Q^2. It should be noted that this selection rule does apply to the transition from a proton target, it does not apply to the transition from the neutron, which is consistent with the data. Modern quark models that go beyond single quark transitions, confirm quantitatively the suppression resulting in very small amplitudes from protons but large ones from neutrons. Furthermore, a direct computation of the hadronic contribution to the transition from protons confirms this (Fig. 231). The measured helicity amplitudes off the protons are almost exclusively due to meson–baryon contributions as the dynamical coupled channel (DCC) calculation indicates (dashed line). The close correlation of the DCC calculation and the measured data for the case when quark contributions are nearly absent, supports the phenomenological description of the helicity amplitudes in terms of a 3-quark core that dominate at high Q^2 and meson–baryon contributions that can make important contributions at lower Q^2.

9.3.11 *Resonance lightfront transition charge densities*

Knowledge of the helicity amplitudes in a large Q^2 allows for the determination of the transition charge densities on the light cone in transverse impact parameter space (b_x, b_y) [2866]. The relations between the helicity transition amplitudes and the Dirac and Pauli resonance transition form factors are given by:

$$A_{1/2} = e\frac{Q_-}{\sqrt{K}(4M_N M^*)^{1/2}}\{F_1^{NN^*} + F_2^{NN^*}\} \tag{9.31}$$

$$S_{1/2} = e\frac{Q_-}{\sqrt{K}(4M_N M^*)^{1/2}}\left(\frac{Q_+Q_-}{2M^*}\right)\frac{(M^* + M_N)}{Q^2} \times\left\{F_1^{NN^*} - \frac{Q^2}{(M^* + M_N)^2}F_2^{NN^*}\right\}, \tag{9.32}$$

where M^* is the mass of the excited state N^*, $K = \frac{M^{*2}-M_N^2}{2M^*}$ is the equivalent photon energy, Q_+ and Q_- are short hands for $Q_\pm \equiv \sqrt{M^* \pm M_N)^2 + Q^2}$. The charge and magnetic lightfront transition densities $\rho_0^{NN^*}$ and $\rho_T^{NN^*}$, respectively,

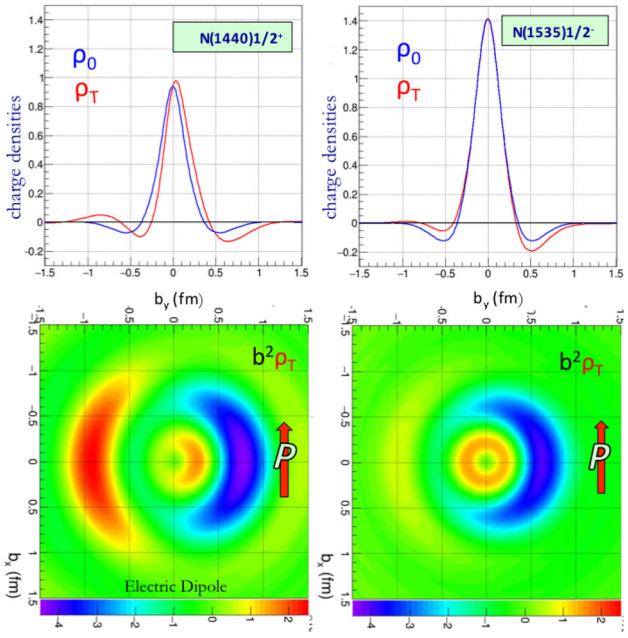


Fig. 232 Left panels: $N(1440)$, top: projection of charge densities on b_y, bottom: transition charge densities when the proton is spin polarized along b_x. Right panels: same for $N(1535)$. Note that the densities are scaled with b^2 to emphasize the outer wings. Color code:negative charge is blue, positive charge is red. Note that all scales are the same for ease of comparison [2868]. Graphics credit: F.X. Girod

are given as:

$$\rho_0^{NN^*}(\vec{b}) = \int_0^\infty \frac{dQ}{2\pi} J_0(bQ) F_1^{NN^*}(Q^2) \tag{9.33}$$

$$\rho_T^{NN^*}(\vec{b}) = \rho_0^{NN^*}(\vec{b}) + \sin(\phi_b - \phi_s) \times \int_0^\infty \frac{dQ}{2\pi} \frac{Q^2}{(M^* + M_N)} J_1(bQ) F_2^{NN^*}(Q^2). \tag{9.34}$$

A comparison of $N(1440)\frac{1}{2}^+$ and $N(1535)\frac{1}{2}^-$ is shown in Fig. 232. There are clear differences in the charge transition densities between the two states. The Roper state has a softer positive core and a wider negative outer cloud than $N(1535)\frac{1}{2}^-$ and develops a larger shift in b_y when the proton is polarized along the b_x axis.

Similar transverse charge transition densities can be defined for $J^P = \frac{3}{2}^+$ states such as the $\Delta(1232)\frac{3}{2}^+$. This has been studied in [2867] and is shown in Fig. 233.

9.3.12 Single quark transition model

Many of the exited states for which there is information about the transition form factors available have been assigned as members of the $[SU(6), L^P] = [70, 1^-]$ super multiplet of the $[SU(6) \otimes O(3)]$ symmetry group. In a model, where only single quark transitions to the excited states are considered [2869–2871], only 3 of the amplitudes need to be known to determine the remaining 16 transverse helicity amplitudes for all states in $[70, 1^-]$ including on neutrons. However, the picture is now more complicated due to the strong admixture of meson–baryon components to the single quark transition especially in the lower Q^2 range. This requires a model to separate the single quark contributions from the hadronic part before projections for other states can be made [2872].

9.3.13 Higher mass baryons and hybrid baryons

The existence of baryons containing significant active gluonic components in the wave function has been predicted some decade ago [529] employing Lattice QCD simulations. The lowest such "hybrid" state is expected to be a $J^P = \frac{1}{2}^+$ nucleon state. LQCD projects a mass of 1.3 GeV above the nucleon mass, i.e. approximately 2.2–2.3 GeV, and several other states should appear close by in $J^P = \frac{1}{2}^+$ and $J^P = \frac{3}{2}^+$, as seen in Fig. 234.

How do we identify these states? Hybrid baryons have same spin-parity as other ordinary baryons. In contrast to hybrid mesons, there are no hybrid baryons with "exotic" quantum numbers. One possibility is to search for more states than the quark model predicts in some mass range. The other possibility is to study the transition form factors of excited states. Hybrid states may be identified as states with a different Q^2 behavior than what is expected from a 3-quark state. The sensitivity [2873] is demonstrated for the Roper resonance that projected a very rapid drop of the $A_{1/2}(Q^2)$ with Q^2, and $S_{1/2}(Q^2) \sim 0$ prediction. Both are incompatible with what we know today about the Roper resonance. Precision electroproduction data in the mass range above 2 GeV will be needed to test high mass states for their potential hybrid character, e.g. from experiments at CLAS12 [2874].

9.3.14 Conclusions and outlook

In this contribution we have focused on more recent results of nucleon resonance transition amplitudes and their interpretation within LQCD and within most advanced approaches, e.g. in light front relativistic quark models and approaches with traceable links to first principle QCD such as Dyson–Schwinger Equations [2875] and light cone sum rules [2858]. These calculations describe the transition form factors at $Q^2 \geq 2$ GeV^2, while at lower Q^2 values hadronic degrees of freedom must be included and could even dominate contributions of the quark core.

For the lowest mass states, $\Delta(1232)\frac{3}{2}^+$ and the Roper $N(1440)\frac{1}{2}^+$, LQCD calculations have been carried out that are consistent with the data within large uncertainties. These calculations are about one decade old, and new data, with higher precision in more extended kinematic range have been

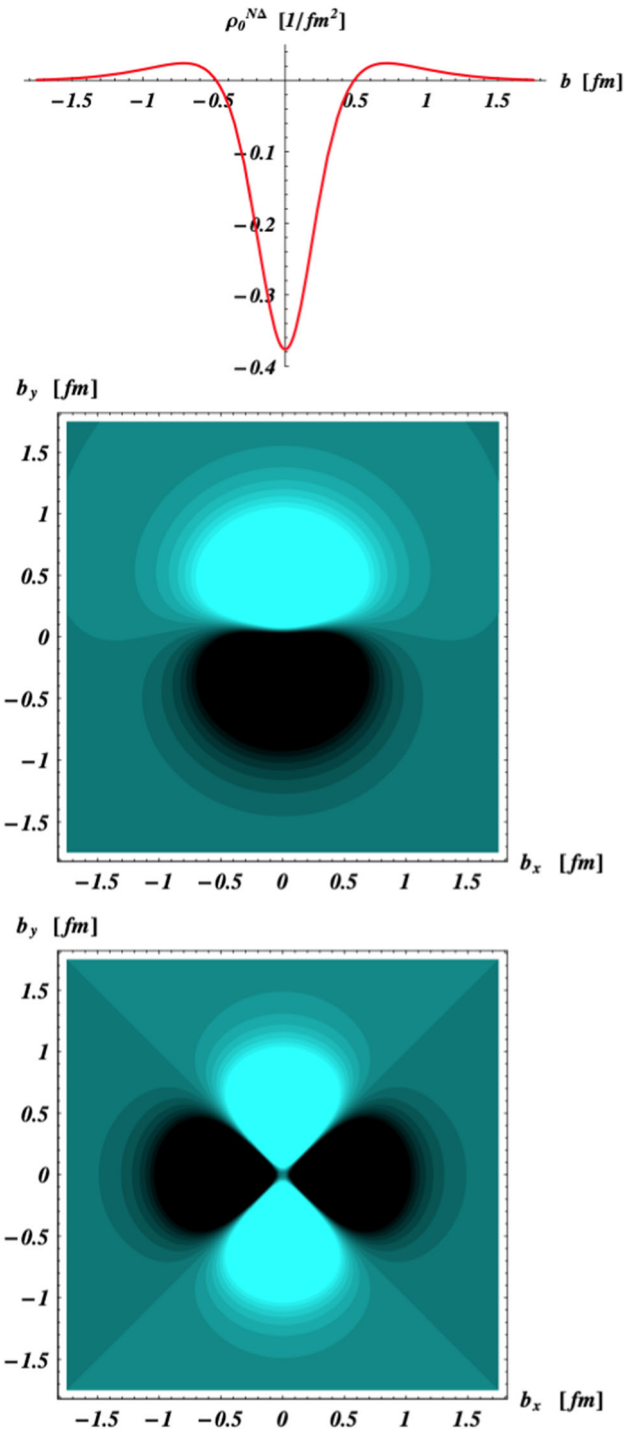


Fig. 233 Quark transverse transition charge density corresponding to the $p \to \Delta^+$ transition. Light color indicates positive charge, dark color indicates negative charge. Top: p and Δ are unpolarized. Middle: p and Δ are polarized along b_x axis generating an electric dipole along the b_y axis. Bottom: Quadrupole contribution to transition density. Graphics adapted from [2867]

added to the database that warrant new Lattice calculations at the physical pion mass to be carried out.

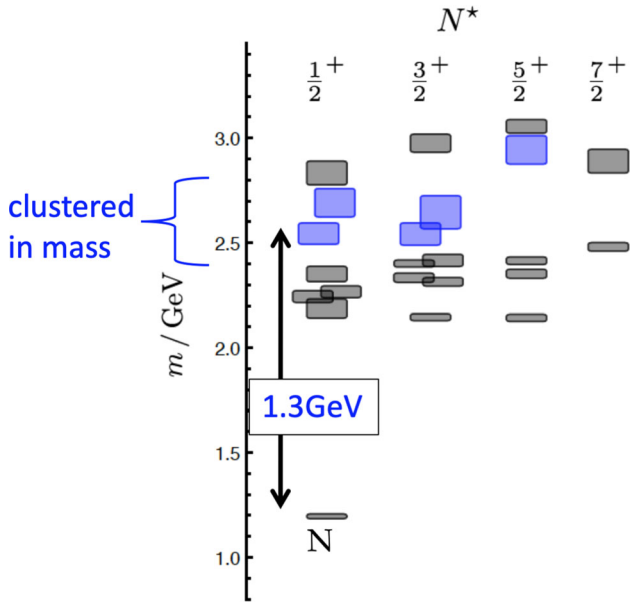


Fig. 234 Projections of excited baryons with dominant gluonic components (marked in blue shades) in LQCD with 400 MeV pions. The lowest hybrid baryon is projected with mass 1.3 GeV above the nucleon mass. The $1/2^+$ and $3/2^+$ states are clustered in a narrow mass range of about 200 MeV

Over the past decade, eight baryon states in the mass range from 1.85 to 2.15 GeV have been either discovered or evidence for the existence of states has been significantly strengthened. Some of these states are in the mass range and have J^{PC} quantum numbers that could have significant contributions of gluonic components. Such "hybrid" states are in fact predicted in LQCD [529]. These states appear with the same quantum numbers as ordinary quark excitations, and can only be isolated from ordinary states due to the Q^2 dependence of their helicity amplitudes [2873], which is expected to be quite different from ordinary 3-quark excitation. The study of hybrid baryon excitations then requires new electroproduction data especially at low Q^2 [2874] with different final states and with masses above 2 GeV. Despite the very significant progress made in recent years to further establish the light-quark baryon spectrum and explore the internal structure of excited states and the relationship to QCD [2813,2876], much remains to be done. A vast amount of precision data already collected needs to be included in the multi-channel analysis frameworks, and polarization data are still to be analyzed. There are approved proposals to study resonance excitation at much higher Q^2 and with higher precision at Jefferson Lab with CLAS12 [2877,2878], which may begin to reveal the transition to the bare quark core contributions at short distances.

A new avenue of experimental research has recently been opened up with the first data-based extraction of a gravitational property of the proton, its internal pressure distribution,

which is represented by the gravitational form factor $D^q(t)$. It is one of the form factors of the QCD matrix element of the energy–momentum tensor, its internal pressure and shear stress distribution in space [2879,2880]. These properties, as well as the distribution of mass and angular momentum, are accessible directly in gravitational interaction, which is highly impractical. However the relevant gravitational form factor $D^q(t)$ for the ground state nucleon can be accessed indirectly through the process of deeply virtual Compton scattering and in time-like Compton scattering [1286,2881]. Both processes, having a $J = 1$ photon in the initial state as well as in the final state, contain components of $J = 2$ that couple to the proton through a tensor interaction, as gravity does [2882].

Mechanical properties of resonance transitions have recently been explored for the $N(1535)\frac{1}{2}^- \to N(938)$ gravitational transition form factors calculations in [2883] and in [2884]. To access these novel gravitational transition form factors experimentally, the generalized parton distributions for nucleon-to-resonance transitions must be studied. The framework for studying the $N \to N(1535)$ transition GPDs, which would enable experimental access to these mechanical properties, remains to be developed. The required effort is quite worthwhile as a new avenue of hadron physics has opened up that remains to be fully explored.

9.4 Heavy-flavor baryons

Eberhard Klempt and Sebastian Neubert

9.4.1 Introduction

Baryons with one heavy quark Q and a light diquark qq provide an ideal place to study diquark correlations and the dynamics of the light quarks in the environment of a heavy quark. The heavy quark is almost static and provides the color source to the light quarks. Here, we attempt to understand the dynamics leading to the spectrum of baryons with one heavy quark.

The Review of Particle Physics [616] lists 28 charmed baryons (16 with known spin-parity) and 19 bottom baryons (11 with known spin-parity). One doubly charmed state has been detected, the ground state Ξ_{cc}^{++}. (Its isospin partner Ξ_{cc}^{+} is known as well, with poor evidence and one star in RPP, but we do not count isospin partners separately.) In the decays of the lightest bottom baryon, exotic $J/\psi p$ states, incompatible with a three-quark configuration, have been have been reported in studies of the reaction $\Lambda_b \to J/\psi p K^-$ [2885,2886]. The search for further states and attempts to understand the underlying dynamics of heavy baryons are active fields in particle physics. New information can be expected from the upgrades of LHC, BELLE and J-PARC, and from the new FAIR facility at GSI (see Sect. 14).

Table 40 Masses and lifetimes of baryon ground states with one b-quark. The second line gives the mass in MeV, the third line the life time in fs

Λ_b^0	Ξ_b^-	Ξ_b^0	Ω_b^-
5619.60 ± 0.17	5797.0 ± 0.6	5791.9 ± 0.5	6045.2 ± 1.2
1464 ± 11	1572 ± 40	1480 ± 0.030	1640^{+180}_{-170}

Table 41 Masses and lifetimes of baryon ground states with one c-quark. The second line gives the mass in MeV, the third line the life time in fs

Λ_c^0	Ξ_c^+	Ξ_c^0	Ω_c^-
2286.46 ± 0.14	2467.71 ± 0.23	2470.44 ± 0.2	2695.2 ± 1.7
201.5 ± 2.7	453 ± 5	151.9 ± 2.4	268 ± 26

9.4.2 Ground states of heavy baryons

Masses and lifetimes

Table 40 presents masses and life times of the ground states of heavy baryons containing a b-quark. Naively, one could expect all these life times to represent the life time of the b quark, that they all agree with the life time of the B^0 meson. This life time is $\tau_{B^0} = (1519 \pm 4)$ fs. Indeed, all life times agree within $\sim 10\%$ percent.

This is not at all the case when the b-quark is replaced by a c-quark (see Table 41). The D^0 has a life time $\tau_{D^0} = (410.3 \pm 1.0)$ fs, the D^+ has $\tau_{D^+} = (1033 \pm 5)$ fs. The life times of charmed baryons are spread over a wide range and do not agree with the life times of D mesons. In addition to the decay of the c-quark, the $c\bar{d}$ pair in a D^0 meson can annihilate into a W^+, a process forbidden for the D^+. In B decays, the corresponding CKM matrix element is small, and this effect is suppressed. Further significant corrections are required to arrive at a consistent picture for the decays of charmed mesons and baryons. The authors of Ref. [1243] have performed an extensive study of the lifetimes within the heavy quark expansion, and have included all known corrections. The impact of the charmed-quark mass and of the wavefunctions of charmed hadrons were carefully studied. Then, qualitative agreement between their calculations and the experimental data was achieved. For a more detailed discussion, see Sect. 5.7.

The first state with two charmed quarks, the Ξ_{cc}^+ was reported by the SELEX collaboration in two decay modes at a mass of (3518.9 ± 0.9) MeV and with 5–6σ [2887,2888]. In later searches, this state was never confirmed. The LHCb collaboration found its doubly charged partner Ξ_{cc}^{++} [2618]. Its mass is (3621.6 ± 0.4) MeV, its life time (25.6 ± 2.7) fs. Later, the LHCb collaboration reported evidence for a Ξ_{cc}^+ baryon at (3623.0 ± 1.4) MeV [2889]. It is seen with 3–4σ only but its mass is better compatible with an interpretation

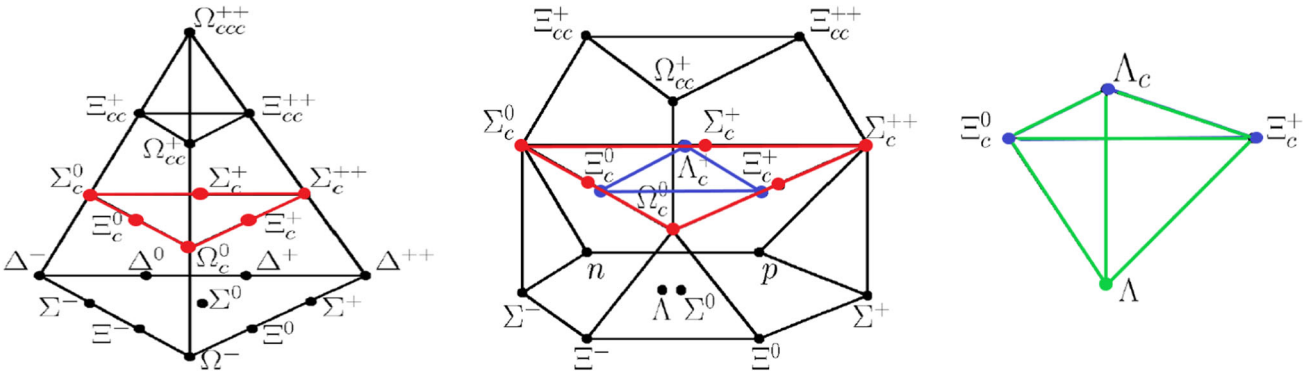


Fig. 235 Ground-state heavy baryons in SU(4). Baryons with one charm quark are represented by colored dots. Left: The symmetric 20-plet. Center: Baryons in the mixed-symmetry 20-plet. The mixed-symmetry 20-plet contains a sextet with a symmetric light-quark pair ($SU_F(3)$ multiplicity 6) and a triplet with an anti-symmetric light-quark pair ($SU_F(3)$ multiplicity $\bar{3}$). Right: The fully antisymmetric 4-plet

of Ξ_{cc}^+ and Ξ_{cc}^{++} as isospin partners. A search for the Ξ_{bc}^+ remained unsuccessful [2890].

The flavor wave function: SU(4)

In this contribution we discuss baryons with one heavy-quark flavor, with either a charm or a bottom quark. Overall, we consider five quarks, u, d, s, c, b, but we will not discuss baryons with one light ($q = u, d, s$) and two different heavy quarks like $\Xi_{cb}^+ = (ucb)$. Thus we can restrict ourselves to SU(4). The four quarks have very different masses, and the SU(4) symmetry is heavily broken, nevertheless it provides a guide to classify heavy-quark baryons. Three-quark baryons can classified according to

$$4 \otimes 4 \otimes 4 = 20_s \oplus 20_m \oplus 20_m \oplus 4_a \tag{9.35}$$

into a fully symmetric 20-plet, two 20-plets of mixed symmetry and a fully antisymmetric 4-plet. In states with one heavy quark only, there is one light quark pair. The light diquark can be decomposed

$$3 \otimes 3 = \bar{3}_a \oplus 6_s \tag{9.36}$$

The light diquark in the 6-plet is symmetric, in the $\bar{3}$-plet antisymmetric.

Figure 235a shows the symmetric 20-plet, which contains the well-known baryon decuplet and a sextet of charmed baryons. In addition to Ξ_{cc}^+ and Ξ_{cc}^{++}, a Ω_{cc}^+ (with two charmed and one strange quarks) and a Ω_{ccc}^{++} are expected but not yet observed. All baryons in the symmetric 20-plet in the ground state have a total spin $J = 3/2$. The three quark pairs are symmetric with respect to (w.r.t.) their exchange, in particular the pair of light quarks is symmetric w.r.t. their exchange, they have $SU_F(3)$ multiplicity 6. Baryons with three charmed quarks have not yet been discovered.

Figure 235b shows the mixed symmetry 20-plet of heavy baryons. In the ground state they have $J = 1/2$. Baryons with one heavy quark occupy the second layer. The 6-plet and the $\bar{3}$-plet are indicated. The sextet in the first floor has a symmetric light-quark pair, the two light-heavy quark pairs are then antisymmetric in flavor. The 3-plet in the first floor has an antisymmetric light-quark pair, the light-heavy quark pairs are then symmetric in flavor.

Finally, there is a fully anti-symmetric 4-plet. It is shown in Fig. 235c. Ground-state baryons have a symmetric spatial wave function. The spin of three fermions coupling to $J/1/2$ has mixed symmetry. A fully symmetric (space), a fully antisymmetric (flavor) and a mixed-symmetry (spin) wave function cannot be coupled to a fully symmetric wave function. Hence baryons with no orbital excitations cannot be in the 4-plet. Only excited baryons can have a fully antisymmetric flavor wave function. Below, in Sect. 9.4.5, the wave functions and their symmetries are discussed in more detail.

9.4.3 *Excited baryons: selected experimental results*

BaBar, BELLE and LHCb:

Most information on heavy baryons stems from three experiments, BaBar, BELLE and LHCb even though many discoveries had already been made before with the Split-Field-Magnet, by the SELEX, UA and LEP experiments at CERN, and by the CDF experiment at FERMILAB. BaBar at SLAC (US) and BELLE at KEK (Japan) study the decays of B mesons produced in asymmetric e^+e^- storage rings with beam energies of 9 (KEK: 7) GeV for electrons and 3.1 (KEK: 4) GeV for positrons resulting in a center-of-mass energy equal to the $\Upsilon(4S)$ mass of 10.58 GeV. The LHCb experiment is placed at the Large Hadron Collider at CERN operating at $\sqrt{s} = 13.6$ GeV. The experiment is a single-arm forward spectrometer covering the pseudorapidity range $2 \leq \eta \leq 5$. It is designed for the study of particles containing b or c quarks. All three detectors have vertex reconstruction capabilities; BaBar and BELLE track charged particles in tracking chambers placed in the 1.5 T magnetic field of a supercon-

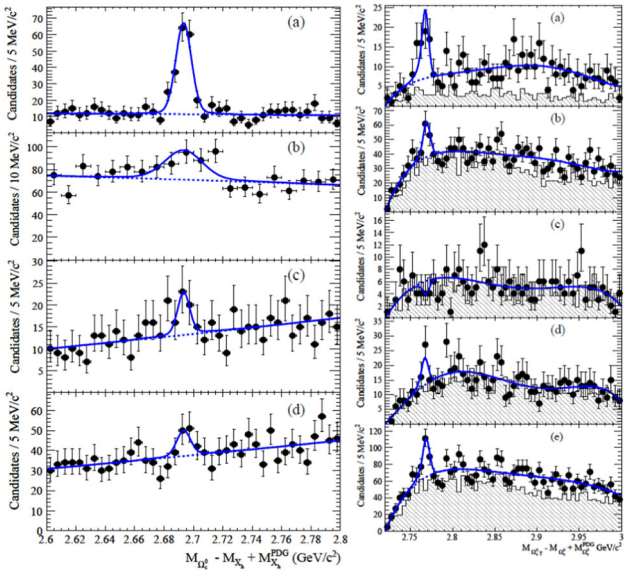


Fig. 236 Left: The invariant mass distributions of Ω_c^0 candidates in their decay to $\Omega^-\pi^+$ (**a**), $\Omega^-\pi^+\pi^0$ (**b**), $\Omega^-\pi^+\pi^-\pi^+$ (**c**), $\Xi^-K^-\pi^+\pi^+$ (**d**). $M_{\Omega_c^0}$ is the reconstructed mass of Ω_c^0 candidates, X_h denotes the daughter hyperon. Right: Invariant mass distribution of $\Omega_c^* \to \Omega_c\gamma$ for the individual Ω_c^0 decay modes (**a**–**d**) and for the sum (**e**). (Adapted from [2891])

ducting solenoid. Particle identification is provided by a measurement of the specific ionization and by detection of the Cherenkov radiation in reflecting ring imaging Cherenkov detectors. CsI(Tl)-crystal electromagnetic calorimeters allow for energy measurements of electrons and photons. LHCb is equipped with silicon-strip detector located upstream and downstream of a dipole magnet with a bending power of about 4 Tm. Photons, electrons and hadrons are identified by a calorimeter system consisting of scintillating counters and pre-shower detectors, and an electromagnetic and a hadronic calorimeter. Muons are identified by a system composed of alternating layers of iron and multiwire proportional chambers.

In the following we discuss three important results from these experiments that demonstrate the capabilities of the detectors.

Observation of $\Omega_c^{*0}(2770)$ decaying to $\Omega_c^0\gamma$ by BaBar

The Babar experiment studied the inclusive reaction $e^+e^- \to \Omega_c^{*0} X$ where X denote the recoiling particles [2891]. Ω_c^0 baryons are identified via different decay modes and reconstructed with a mass resolution $\sigma_{\rm RMS} = 13\,{\rm MeV}$. The γ is reconstructed in the Ω_c^0 CsI(Tl) calorimeter. Figure 236 shows the reconstructed Ω_c^0 and the Ω_c^{*0} in its $\Omega_c^{*0} \to \Omega_c^0\gamma$ decay. Obviously, the $\Omega_c^{*0}(2770)$ is equivalent to $\Delta^0(1232)$ with the u, d, d quarks exchanged by c, s, s, and the transition corresponds to the $\Delta(1232) \to N\gamma$ decay.

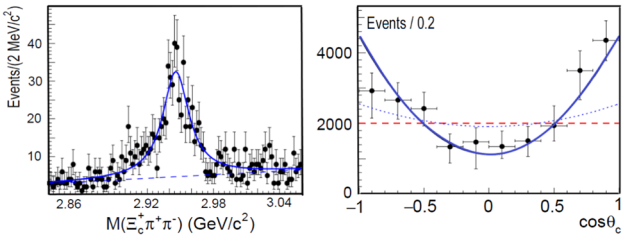


Fig. 237 Left: The $\Xi_c^+\pi^-\pi^+$ invariant mass distribution for events in which the $\Xi_c^+\pi^-$ invariant mass is compatible with the $\Xi_c^0(2645)$ mass. Right: The helicity angle θ_c between the direction of the π^- relative to the opposite direction of the $\Xi_c^+(2970)$ in the rest frame of the $\Xi_c^0(2645)$. (Adapted from Ref. [2892])

First determination of the spin and parity of the charmed-strange baryon $\Xi_c^+(2970)$ by BELLE

The BELLE collaboration identified $\Xi_c^+(2970)$ in the decay chain $\Xi_c^+(2970) \to \Xi_c^0(2645)\pi^+ \to \Xi_c^+\pi^-\pi^+$; Ξ_c^+ is reconstructed from its decay into $\Xi^-\pi^+\pi^+$ [2892]. Due to its mass, $\Xi_c^0(2645)$ is likely the spin excitation with $J^P = 3/2^+$ of the $J^P = 1/2^+$ ground state Ξ_c^0. The helicity angle in the primary decay, i.e. the angle between the π^+ and the opposite of the boost direction in the c.m. frame both calculated in the $\Xi_c^+(2970)$ rest frame, proved to be insensitive to some likely J^P combinations. However, the predictions for different J^P's vary significantly for the angular distributions in the secondary decay (see Fig. 237).

The analysis shows that quantum numbers $J^P = 1/2^+$ are preferred for $\Xi_c^+(2970)$. These are the quantum numbers of the Roper resonance. The BELLE collaboration noted that its mass difference to the Ξ_c ground state is about 500 MeV. The same excitation energy is required to excite the Roper resonance $N(1440)$, the $\Lambda(1600)$ and the $\Sigma(1660)$, all with $J^P = 1/2^+$.

First observation of excited Ω_b states by LHCb

The LHCb collaboration searched for narrow resonances in the $\Xi_b^0 K^-$ invariant mass distribution [2893]. The Ξ_b^0 has a lifetime of $(1.48 \pm 0.03)10^{-12}$ s, $c\tau \approx 500\,\mu$m, which is sufficiently long to separate the interaction and the decay vertices. Four peaks can be seen (Fig. 238), which correspond to excited states of Ω_c. With the given statistics, quantum numbers can not yet be determined.

9.4.4 The mass spectrum of excited heavy baryons

Figure 239 shows the mass spectrum of heavy baryons with a single charm or bottom quark. Established light baryons with strangeness are shown for comparison. The quantum numbers of low-mass heavy baryons are mostly known, for higher-mass states this information is often missing. The masses are given as excitation energies above the Λ (Λ_c, Λ_b) mass.

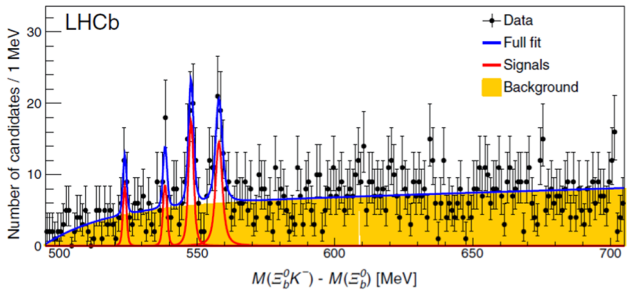


Fig. 238 Distribution of the mass difference $M_{\Xi_b^0 K^-} - M_{\Xi_b^0}$ for $\Xi_b^0 K^-$ candidates. The background is given by the wrong-sign candidates $\Xi_b^0 K^+$. (From Ref. [2893].)

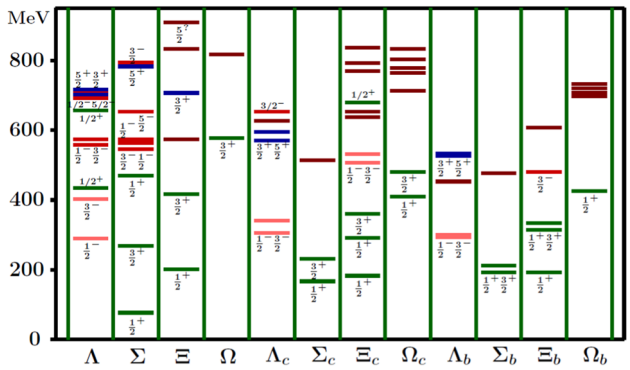


Fig. 239 Heavy baryons with charm or bottomness and a comparison with light baryons with strangeness. All heavy baryons are shown, light baryons are shown at the pole mass and are only included with 3* or 4* rating. When two quantum numbers are given, the first one refers to the lower-mass state, the second one to the state above. The states with $L = l_\rho = l_\lambda = 0$ are shown in green, states with $L = 1$ in red (orange for members of $\bar{4}_F$), states $L = 2$ in blue, states with unknown spin-parity in brown

At the first glance, the spectrum looks confusing. The Λ spectrum is crowded, there is a low-mass negative-parity spin doublet, a second doublet – at about the same mass as a Σ spin doublet – a pair with $J^P = 1/2^-$ and $5/2^-$ where a $3/2^-$ state seems to be missing, and then a positive-parity doublet with $J^P = 3/2^+, 5/2^+$. In the Λ_c spectrum, the higher-mass negative-parity states and the positive-parity doublet are inverted in mass.[101] The $3/2^+ - 1/2^+$ hyperfine splitting decreases rapidly when going from Σ and Ξ to Σ_c and Ξ_c and from Σ_b and Ξ_b. It is interesting to note that a similar pattern is observed in mesons: the hyperfine splitting decreases when going from $\rho - \pi$ to $D^* - D$ and to $B^* - B$. Also, there is one Ξ $1/2^+$ ground state but two states for Ξ_c and Ξ_b. The lowest-mass Ω has $J^P = 3/2^+$, in the charm sector, two low-mass Ω_c states are known with $J^P = 1/2^+$ and $3/2^+$, the Ω_b spectrum has just one low-mass state with $J^P = 1/2^+$.

[101] This inversion was predicted by Capstick and Isgur long before the states were discovered [771].

9.4.5 *Heavy baryons as three-quark systems*

The spatial wave function

The orbital wave functions of excited states are classified into two kinds of orbital excitations, the λ-mode and the ρ-mode (see Eq. (9.1)). In heavy baryons with one heavy quark, the λ-mode is the excitation of the coordinate between the heavy quark and the light diquark, and the ρ-mode is the excitation of the diquark cluster. In light-baryon excitations, the λ and ρ oscillators are mostly both excited, e.g. to $l_\lambda = 1, l_\rho = 0$ and $l_\lambda = 0, l_\rho = 1$, the two components of the wave function having a relative $+$ or $-$ sign. In heavy baryons with one heavy quark, the mixing between these two configurations is small.

The two oscillators have different reduced masses, m_ρ and m_λ:

$$m_\rho = \frac{m_q}{2} \quad , \quad m_\lambda = \frac{2m_q m_Q}{2m_q + m_Q}. \tag{9.37}$$

The ratio of harmonic oscillator frequencies is then given by

$$\frac{\omega_\lambda}{\omega_\rho} = \sqrt{\frac{1}{3}(1 + 2m_q/m_Q)} \leq 1. \tag{9.38}$$

In the heavy-quark limit ($m_Q \to \infty$), the excitation energies in the λ oscillator are reduced by a factor $\sqrt{3}$.

Diquarks

We first consider the light diquark. The two light quarks can have either the symmetric flavor structure $\mathbf{6}_F$ or the antisymmetric flavor structure $\bar{\mathbf{3}}_F$. The spin of the light diquark can be $s_{qq} = s_l = 1$ or $s_l = 0$ leading to a symmetric or an antisymmetric spin wave function. The color part of the wave function is totally antisymmetric. Hence flavor and spin wave functions are linked. In an S-wave, scalar ("good" or g) and axial-vector ("bad" or b) diquarks can be formed. The intrinsic quark spins couple to the internal orbital angular momentum l_ρ, leading to excited diquarks with orbital excitations.

$$(l_\rho = 0,\ \mathbf{S}) \begin{cases} s_l = 0\ (\mathbf{A}),\ \bar{\mathbf{3}}_F\ (\mathbf{A}),\ j_{qq} = 0, & \text{(g)} \\ s_l = 1\ (\mathbf{S}),\ \mathbf{6}_F\ (\mathbf{S}),\ j_{qq} = 1, & \text{(b)} \end{cases}$$

$$(l_\rho = 1,\ \mathbf{A}) \begin{cases} s_l = 0\ (\mathbf{A}),\ \mathbf{6}_F\ (\mathbf{S}),\ j_{qq} = 1, & \text{(g)} \\ s_l = 1\ (\mathbf{S}),\ \bar{\mathbf{3}}_F\ (\mathbf{A}),\ j_{qq} = 0/1/2, & \text{(b)} \end{cases}$$

$$(l_\rho = 2,\ \mathbf{S}) \begin{cases} s_l = 0\ (\mathbf{A}),\ \bar{\mathbf{3}}_F\ (\mathbf{A}),\ j_{qq} = 2, & \text{(g)} \\ s_l = 1\ (\mathbf{S}),\ \mathbf{6}_F\ (\mathbf{S}),\ j_{qq} = 1/2/3, & \text{(b)} \end{cases}$$

$$\ldots$$

where we have denoted the total angular momentum of the light diquark as j_{qq}.

Coupling of angular momenta

Figure 240 shows how the orbital angular momentum and the diquark spin couple to the total diquark angular momentum j_l. This in turn couples to the heavy-quark spin s_Q giving rise to spin doublets (or just spin-1/2 states for $j_l = 0$). Note

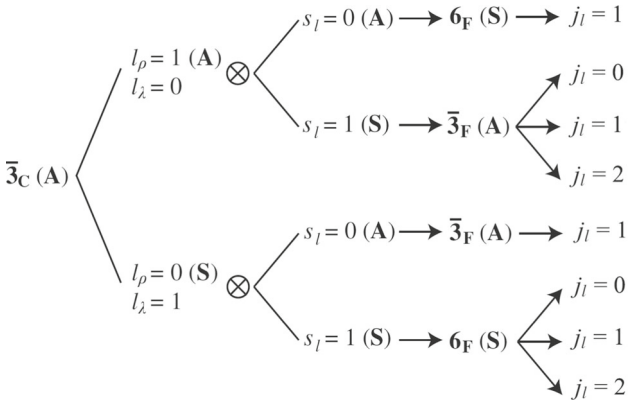


Fig. 240 Heavy baryons in P-wave. The light diquark couples to the spin of the heavy quark. The light diquark of Λ and Ξ heavy baryons are in the antisymmetric flavor $\bar{3}_F$ and in the symmetric 6_F in the case of Σ, Ξ' and Ω (Adapted from [2894])

the Λ and Ξ spin doublet with $s_l = 0$ and $\bar{3}_F$. In this case the wave function is antisymmetric in spin and flavor, this is a "good" diquark.

Only a few heavy baryons are known with $L = 2$: Λ_c and Λ_b with spin-parity $3/2^+$ and $5/2^+$. The other expected states seem to show up only in the higher-mass, less-explored region. The two observed doublets can be assigned to a configuration in which $l_\rho = 2$, $L_\lambda = 0$, and the diquark is in $\bar{3}_F$ and $s_l = 0$.

$$\bar{\mathbf{3}}_\mathbf{F} : L = 2 \otimes s_l = 0(\mathbf{A}) – j_l = 2$$

The diquark is a "good" diquark. Note that the states $\Lambda_c(2860)$, $\Lambda_c(2880)$ with spin-parity $3/2^+$ and $5/2^+$ ($L = 2$) are below $\Lambda_c(2940)$ with $3/2^-$. The latter state has a "bad" diquark and is excited to $L = 1$ in the ρ oscillator. In this competition, the "good" diquark and λ excitation with $L = 2$ wins over "bad" diquark and ρ oscillator even though the orbital angular momentum of $\Lambda_c(2940)$ is $L = 1$!

Table 42 gives a survey of the coupling scheme of Qqq baryons. The spin and orbital angular momentum of the two light quarks couple to j_q, and when combined with the heavy-quark spin s_Q, the final J^P results. There are also states with mixed excitations like $l_\rho = 1$, $l_\lambda = 1$. These are unlikely to be produced (see Sect. 9.2) and are not included here. Λ and Ξ with $s_q = 0$ and $l_\rho = 0$ have a "good" light diquark. For the Λ_c we denote the light diquark by $[u, d]$. Note that also one light and the heavy quark can be antisymmetric in their spin and flavor wave function. We write $\Sigma_b = [ub]s$.

Heavy quark limit

When $m_Q \to \infty$, the heavy quark spin s_Q is conserved. Due to the conservation of the total angular momentum J, also the angular momentum carried by the light quarks is conserved. Hence all interactions which depend on the spin of the heavy quark disappear. Thus, the mass difference within a spin doublet with, e.g., $J^P = 3/2^+$ and $1/2^+$, will disappear

Table 42 The λ- and ρ-mode assignments of the P and D-wave excitations of singly-heavy baryons. l_ρ, l_λ are orbital angular momenta of the two oscillators, L the total orbital angular momentum, s_q is the spin, j_q the total angular momentum of the diquark, and J the total spin

l_ρ	l_λ	L	s_q	j_q	Λ, Ξ	Σ, Ξ', Ω	J^P
0	1	1	0	1	λ	ρ	$1/2^-, 3/2^-$
0	1	1	1	0	ρ	λ	$1/2^-$
1	0	1	1	1	ρ	λ	$1/2^-, 3/2^-$
1	0	1	1	2	ρ	λ	$3/2^-, 5/2^-$
0	2	2	0	2	λ	–	$3/2^+, 5/2^+$
2	0	2	0	2	ρ	–	$3/2^+, 5/2^+$
0	2	2	1	1	–	λ	$1/2^+, 3/2^+$
0	2	2	1	2	–	λ	$3/2^+, 5/2^+$
0	2	2	1	3	–	λ	$5/2^+, 7/2^+$
2	2	2	1	1	–	ρ	$1/2^+, 3/2^+$
2	0	2	1	2	–	ρ	$3/2^+, 5/2^+$
2	0	2	1	3	–	ρ	$5/2^+, 7/2^+$

in the heavy-quark limit. Indeed, the mass differences

$$M_{\Sigma(1520)3/2^+} - M_{\Sigma(1190)} = 230\,\text{MeV}$$
$$M_{\Sigma_c(2520)3/2^+} - M_{\Sigma_c(2455)} = 65\,\text{MeV}$$
$$M_{\Sigma_b(5830)3/2^-} - M_{\Sigma_b(5820)} = 20\,\text{MeV}$$

decrease as m_Q becomes large.

9.4.6 A guide to the literature

The first prediction of the full spectrum of baryons including charmed and bottom baryons was presented by Capstick and Isgur [771], 3 years before the first baryon with bottomness was discovered. The publication remained a guideline for experimenters for now 36 years! Capstick and Isgur used a relativized quark model with a confining potential and effective one-gluon exchange. Based on the quark model, further studies of the mass spectra of heavy baryons were performed. They are numerous, and only a selection of papers can be mentioned here.

Ebert, Faustov and Galkin calculated the mass spectra for orbital and radial excitations and constructed Regge trajectories [2895]. Yu, Li, Wang, Lu, and Ya [2896] calculated the mass spectra and decays of heavy baryons excited in the λ-mode. Li, Yu, Wang, Lu, and Gu [2897] restricted the calculation – again based on the relativized quark model – to the Ξ_c and Ξ_b families. In their model, all excitations are in the λ-mode.

Migura, Merten, Metsch, and Petry [2898] calculated excitations of charmed baryons within a relativistically covariant quark model based on the Bethe–Salpeter-equation in instantaneous approximation. Interactions are given by a linearly rising three-body confinement potential and a fla-

Table 43 Increase of baryon masses with the number of strange quarks

	$3n \to 2ns$	$2ns \to n2s$	$n2s \to 3s$
$\Delta^-(1232)3/2^+$	$\Sigma^-(1385)3/2^+$	$\Xi^-(1530)3/2^+$	$\Omega^-3/2^+$
	+155 MeV	+148 MeV	+ 137 MeV
$\Sigma_c^0(2520)3/2^+$	$\Xi_c^0(2645)3/2^+$	$\Omega_c^0(2770)3/2^+$	
	+128 MeV	+ 120 MeV	
$\Sigma_c^0(2455)1/2^+$	$\Xi_c'^0\ 1/2^+$	$\Omega_c^0\ 1/2^+$	
	+121 MeV	+116 MeV	
$\Sigma_b^-(5816)1/2^+$	$\Xi_b'^0\ 1/2^+$	$\Omega_b^-1/2^+$	
	+120 MeV	+111 MeV	

vor dependent two-body force derived from QCD instanton effects. Valcarce, Garcilazo and Vijande [2899] performed a comparative Faddeev study of heavy baryons with nonrelativistic and relativistic kinematics and different interacting potentials that differ in the description of the hyperfine splitting. The authors conclude that the mass difference between members of the same $SU_F(3)$ configuration, either $\bar{3}_F$ or 6_F, is determined by the interaction in the light-heavy quark subsystem, and the mass difference between members of different representations is mainly determined by the dynamics of the light diquark.

Chen, Wei and Zhang [2900] derive a mass formula in a relativistic flux tube model to calculate mass spectra for Λ and Ξ heavy baryons and assign quantum numbers to states whose quantum numbers were not known. Faustov and Galkin [2901] assigned flavor- and symmetry dependent masses and form factors to diquarks and calculated the masses of heavy baryons within a relativistic quark–diquark picture. Quantum numbers are suggested for the Ω_c excitations [2902,2903] and other states with unknown spin-parities. A further diquark model, again with adjusted diquark masses, is presented by Kim, Liu, Oka, and Suzuki [2904] exploiting a chiral effective theory of scalar and vector diquarks according to the linear sigma model.

QCD sum rules have been exploited to study P-wave heavy baryons and their decays within the heavy quark effective theory (see [2905] and Refs. therein). The low-lying spectrum of charmed baryons has also been calculated in lattice QCD with a pion mass of 156 MeV [2906]. The results – comparing favorably with the data – are compared to earlier lattice studies that are not discussed here.

All calculations reproduce the observed spectrum with good success, with a large number of parameters. For the reader, it is often not easily seen what are the main driving forces that generate the mass spectrum. Clearly, a confinement potential is mandatory, spin dependent forces are necessary. In the following phenomenological part we try to identify the leading effects driving the resonance spectrum.

9.4.7 Phenomenology of heavy baryons

We start with a simple observation: masses of baryons increase when a u or d quark is replaced by an s quark (see Table 43). For light baryons, this is known as U-spin rule. The constituent s-quark mass decreases in heavy baryons. Note that the difference of current quark masses is $m_s - m_n \approx 124$ MeV (see Table 43).

In Table 44 we show the mass difference of the lowest-mass $J^P = 3/2^-$ states with (u, d, s, c) or (u, d, s, b) quarks and the $J^P = 1/2^+$ ground states: The mass differences are surprisingly small. The $N(1520) - N$ mass difference is 580 MeV, much larger than the mass differences seen here. In the table, [ud] represents wave functions with a u, d quark pair that is anti-symmetric in spin and flavor. These diquarks are often called *good diquarks*. The presence of good diquarks leads to a stronger binding. In the 4-plet, all three quark pairs have such a component w.r.t. their exchange. We denote this by [ud,us,ds]. Thus there are three good diquarks in the wave function. This fact leads to the low masses of the 4-plet members. The similarity of the mass splittings supports similar interpretations of the four resonances from $\Lambda(1520)$ to $\Xi_{b\ 3/2^-}^0$.

In most publications, both resonances, $\Lambda_c(2595)1/2^-$ and $\Lambda_c(2625)3/2^-$, are discussed as 3-quark baryons. However, Nieves and Pavao [2907] have studied these two resonances in an effective field theory that incorporates the interplay between $\Sigma_c^{(*)}\pi - ND^{(*)}$ baryon-meson dynamics and bare P-wave cud quark-model state and suggest that these two resonances are not heavy quark symmetry spin partners. Instead, they see

$$\Lambda_c(2625)3/2^-$$

as a dressed three-quark state while $\Lambda_c(2595)1/2^-$ is reported to have a predominant molecular structure. Nevertheless, the two states $\Lambda_c(2625)3/2^-$ and $\Lambda_c(2595)1/2^-$ obviously form a spin doublet.

The mass shift in H atoms between the two ground states with electron and proton spins parallel or antiparallel is called hyperfine splitting. We borrow this expression to discuss the

Table 44 Mass splitting between baryon ground states belonging to the symmetric 20plet (with $J^P = 3/2^+$) and to the mixed-symmetry 20plet (with $J^P = 1/2^+$)

$\Xi_{b\,3/2^-}^0$	[us,ub,sb]	$\Xi_{b\,1/2^+}^0$	[us]	$\delta M = 310\,\text{MeV}$
$\Lambda_{b\,3/2^-}^0$	[ud,ub,db]	$\Lambda_{b\,1/2^+}^0$	[ud]	$\delta M = 300\,\text{MeV}$
$\Xi_{c\,3/2^-}^+$	[us,uc,sc]	$\Xi_{c\,1/2^+}^+$	[us]	$\delta M = 350\,\text{MeV}$
$\Lambda_{c\,3/2^-}^+$	[ud,uc,dc]	$\Lambda_{c\,1/2^+}^+$	[ud]	$\delta M = 400\,\text{MeV}$
$\Lambda(1520)$	[ud,us,ds]	$\Lambda_{1/2^+}$	[ud]	$\delta M = 400\,\text{MeV}$

Table 45 Mass splitting between baryons with fully symmetric wave functions and baryons with antisymmetric quark pairs. The [us] indicates an antisymmetric quark pair

			δM [MeV]	m_q [GeV]	$\delta M \cdot m_q$
$\Sigma_{b\,3/2^+}$	Λ_b	$[ud]b$	0.211	~ 0.3	0.063
$\Sigma_{c\,3/2^+}$	Λ_c	$[ud]c$	0.232	~ 0.3	0.070
$\Sigma_{3/2^+}$	Λ	$[ud]s$	0.268	~ 0.3	0.080
$\Delta_{3/2^+}$	N	$[ud]u$	0.292	~ 0.3	0.088
$\Xi_{b\,3/2^+}$	Ξ_b	$[us]b$	0.163	~ 0.45	0.073
$\Xi_{c\,3/2^+}$	Ξ_c	$[us]c$	0.177	~ 0.45	0.080
$\Xi_{3/2^+}$	Ξ	$[us]s$	0.217	~ 0.45	0.098
$\Sigma_{3/2^+}$	Σ	$[us]$u	0.191	~ 0.45	0.086
$\Xi_{c\,3/2^+}$	$\Xi_c^{'}$	$[uc]s$	0.067	~ 1.4	0.093
$\Sigma_{c\,3/2^+}$	Σ_c	$[uc]u$	0.065	~ 1.4	0.090
$\Xi_{b\,3/2^+}$	$\Xi_b^{'}$	$[ub]s$	0.020	~ 4.25	0.085
$\Sigma_{b\,3/2^+}$	Σ_b	$[ub]u$	0.021	~ 4.25	0.089

difference between the ground states with all three quark spins adding to $J = 3/2$ (and belonging to the symmetric 20-plet) and with those having $J = 1/2$ (that belong to the mixed-symmetry 20-plet). We thus compare masses of the fully symmetric 20_s-plet with those from the $\bar{3}$-plet or 6-plet within the 20_m-plet (see Table 45). The two configurations differ by the orientation of the heavy-quark spin relative to the spin of the light diquark. According to the heavy-quark-spin symmetry, this mass difference has to vanish with $m_Q \to \infty$. In the Table we assume constituent quark masses of 0.15 GeV (u, d), 0.3 GeV (s), 1.25 GeV (c) and 4.1 GeV (b).

The $J^P = 3/2^+$ states have a fully symmetric flavor wave function, the $J^P = 1/2^+$ states have an antisymmetric quark pair (a good diquark) that is indicated in the list. Their effect scales with $1/m_q$. The mass shift due to the presence of good diquarks is expected for instanton-induced interactions.

Heavy baryons at higher mass:
Next we discuss the higher-mass negative-parity states. In light-baryon spectroscopy, there are seven negative-parity Λ states expected in the first excitation level: two singlet states with $J^P = 1/2^-, 3/2^-$, two octet states with intrinsic total quark spin $s = 1/2$ and $J^P = 1/2^-, 3/2^-$, and a $J^P = 1/2^-, 3/2^-, 5/2^-$ triplet with $s = 3/2$. In light baryons, both λ and ρ oscillator are coherently excited. In heavy-quark baryons, the two oscillators decouple, and the λ and ρ modes are well separated. The low-lying spin-doublet of P-wave Λ_Q states is dominated by a λ-mode excitation, the other five expected states are excited in the ρ mode.

Unfortunately, only one negative-parity state at a higher mass has been reported, the $\Lambda_c(2940)3/2^-$. Its mass is 653 MeV above the Λ_c^+. We interpret this state as l_ρ excitation with a diquark spin $s = 1$. The $\Lambda(1690)$ $3/2^-$ is only 570 MeV above the Λ, it is excited in both the λ and the ρ mode.

The mass of $\Lambda_c(2940)3/2^-$ (with intrinsic orbital angular momentum $L = 1$) is above the masses of the positive-parity states $\Lambda_c(2860)3/2^+$ and $\Lambda_c(2880)5/2^+$ (having $L = 2$). Yet, the mass of $\Lambda(1690)3/2^-$ falls well below the masses of $\Lambda(1890)3/2^+$ and $\Lambda(1820)5/2^+$ for reasons discussed above.

9.4.8 Pentaquarks

In 2015, the LHCb collaboration reported the observation of two exotic structures in the $J/\psi p$ system, a broad resonant structure with a Breit–Wigner width of about 200 MeV called $P_c(4380)^+$ and a narrow state called $P_c(4450)^+$ [2885]. The exotic structures were observed in the reaction $\Lambda_b^0 \to J/\psi K^- p$. An excited three-quark nucleon cannot decay into J/ψ p, this would violate the OZI rule. Hence the minimal quark content is $(c\bar{c}uud)$. The findings met with great interest; the publication is quoted more than 1600 times (2023, October). Indeed narrow baryonic resonances with hidden charm had been predicted several years before as dynamically generated states [2908–2910].

A multitude of different interpretations of the observed structures is offered in the literature, but none is accepted anonymously. There are numerous reviews on tetra- and pentaquarks and their possible interpretations [1427,2692,2911, 2911–2914].

With increased statistics, $P_c(4312)^+$ was confirmed and the higher-mass $P_c(4450)^+$ was shown to be split into two narrow overlapping structures, $P_c(4440)^+$ and $P_c(4457)^+$ [2886]. The existence of the broad resonance was not confirmed. The data and a fit are shown in Fig. 241 which also displays some relevant thresholds. In addition, a further smaller structure can be seen at 4380 MeV, close to the $\Sigma_c^{+*}\bar{D}^0$ threshold. A narrow structure here is expected in molecular models (see e. g. [2915]), but due to limited statistics there was no attempt to describe it in the recent LHCb analysis [2886]. The resonant parameters – including the broad structure at 4380 MeV – are reproduced in Table 46.

Quantum numbers $J^P = 3/2^-$ and $5/2^+$ were preferred for $P_c(4380)^+$ and $P_c(4450)^+$. In the later publication [2886], no quantum numbers are determined.

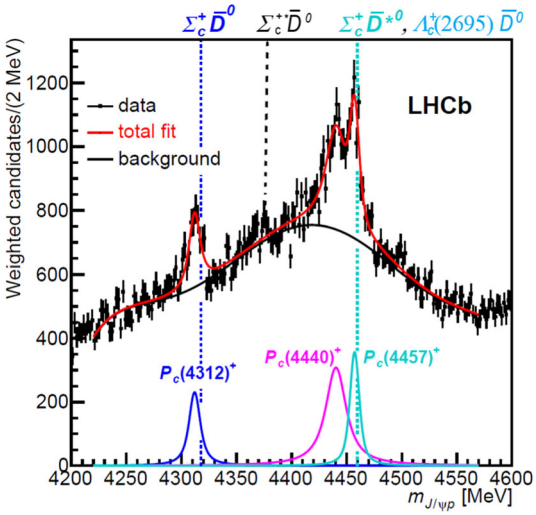


Fig. 241 The $J/\psi\, p$ mass distribution fitted with three BW amplitudes and a sixth-order polynomial background. The thresholds for the $\Sigma_c^+\bar{D}^0$. (Adapted from [2886].)

Table 46 $J/\psi p$ and $J/\psi\,\Lambda$ pentaquarks found by the LHCb collaboration

$P_c(4312)^+$:	M =	$(4311.9 \pm 0.7\,^{+6.8}_{-0.6})$	MeV
[2886]	Γ =	$(9.8 \pm 2.7\,^{+3.7}_{-4.5})$	MeV
$P_c(4380)^+$:	M =	(4380 ± 30)	MeV
[2885]	Γ =	(205 ± 90)	MeV
$P_c(4440)^+$:	M =	$(4440.3 \pm 1.3\,^{+4.1}_{-4.7})$	MeV
[2886]	Γ =	$(20.6 \pm 4.9\,^{+8.7}_{-10.1})$	MeV
$P_c(4457)^+$:	M =	$(4457.3 \pm 0.6\,^{+4.1}_{-1.7})$	MeV
[2886]	Γ =	$(6.4 \pm 2.0\,^{+5.7}_{-1.9})$	MeV
$P_c(4337)^+$:	M =	$(4337\,^{+7}_{-4}\,^{+2}_{-2})$	MeV
[2916]	Γ =	$(29\,^{+26}_{-12}\,^{+14}_{-14})$	MeV
$P^0_{cs}(4459)$:	M =	$(4458.8 \pm 2.9\,^{+4.7}_{-1.1})$	MeV
[2918]	Γ =	$(17.3 \pm 6.5\,^{+8.0}_{-5.7})$	MeV
$P^0_{cs}(4338)$:	M =	$(4338.2 \pm 0.7 \pm 0.4)$	MeV
[2919]	Γ =	$(7.0 \pm 1.2 \pm 1.3)$	MeV

In the reaction $B_s^0 \to J/\psi\,\bar{p}\,p$ a pentaquark-like structure, named $P_c(4337)^+$, was observed in the $J/\psi\,\bar{p}$ and $J/\psi\,p$ mass distributions [2916]. The significance, as determined from a 3-body amplitude analysis, is between 3.1 and $3.7\,\sigma$. Its Breit–Wigner parameters are incompatible with the structures observed in Λ_b decays. The lighter state at 4312 MeV was not found in this reaction, highlighting the importance of the production mechanism for the formation of these resonances. However, it has been pointed out in [2917] that in a region with many close-by thresholds, the Breit–Wigner parameters measured in a particular channel may differ significantly from the pole location.

Strange counterparts to these pentaquark states, e.g. resonances in the $J/\psi\,\Lambda$ system, are denoted by P_{cs} and have $(c\bar{c}uds)$ as minimal quark content. A peak has been reported by LHCb in the reaction $\Xi_b^- \to J/\psi\,\Lambda K^-$ [2918]. Close to the $\Xi_c^0\,D^{*0}$ threshold a further peak was found, with a mass and width given in Table 46, too.

The $J/\psi\,\Lambda$ system was also investigated in 2019 by CMS [2920], exploiting the small phase space available in the B-meson decay $B^- \to J/\psi\,\Lambda\bar{p}$. The analysis showed that the observed spectrum was incompatible with a pure phase space distribution. Very recently, the LHCb collaboration reported a new analysis of this process [2921]. Now, a signal in the $J/\psi\,\Lambda$ subsystem, with preferred quantum numbers $J^P = 1/2^-$, was established at high significance, named $P^0_{cs}(4338)$. Due to the presence of the second (anti)baryon, the phase space in the B-meson decay is too small to access the heavier pentaquark state found in the Ξ_b decay.

These structures have stimulated an intense discussion of the nature of these structures. Do they originate from threshold singularities due to rescattering in the final state leading to a logarithmic branching point in the amplitude? Are they hadronic molecules like the deuteron? Are they compact or triple-quark–diquark systems or states where a $c\bar{c}$ center is surrounded by light quarks?

The peaks are mostly seen very close to important thresholds. Thus they could originate from threshold singularities. We refer to a few publications [2689,2922–2924]. The LHCb collaboration studied this hypothesis and found it incompatible with the data, but the attempts continued [2925–2928].

Very popular are interpretations as bound states composed of charmed baryons and anti-charmed mesons or of charmonium states binding light-quark baryons. The pentaquark states are then seen to be of molecular nature and be bound by coupled-channel dynamics [2915,2929–2939]. Diquark–triquark models were studied [2940–2943], and sum rules are exploited in Refs. [2944,2945].

9.4.9 *Concluding remarks*

The study of hadrons with heavy quarks has developed into a fascinating new field of particle physics. Particular excitement is due to the discovery of unconventional structures that are hotly debated. But also the "regular" heavy hadrons yield very useful information on the interactions of quarks in the confinement region.

10 Structure of the nucleon

Conveners:
Volker Burkert and Franz Gross

After discussion of the baryon spectrum in the previous section, this section focuses on the nucleon, the most studied of

all hadrons. Soon after the proton and neutron were established as the constituents of atomic nuclei, experiments measuring their magnetic moments μ_N found that these spin-1/2 particles are not point-like elementary fermions with expected $\mu_p = 1.0\mu_N$ for the proton, and $\mu_n = 0$ for the neutron. Instead $\mu_p \approx 2.5\mu_N$ and $\mu_n \approx -1.5\mu_N$, showing that the nucleons have significant structure. The discovery that the proton and the neutron are not point-like objects gave birth to the field of hadron structure explorations discussed in this section. Beginning with the Nobel prize winning measurement of the finite size of the proton in elastic electron–proton scattering experiments (Hofstadter, 1956) there have been generations of electron scattering measurements studying the proton and neutron form factors, reviewed by Andrew Puckett

In 1968 experiments employing high-energy electrons scattering from proton targets at SLAC found surprisingly large inelastic cross sections, or structure functions, which rather than falling rapidly with the exchanged four-momentum squared Q^2 (as would elastic cross sections) were observed to "scale" with Q^2. The observation of scaling suggested scattering from point-like quarks in the proton, which could most naturally be described in terms of parton distribution functions (PDFs). These PDF measurements have shed light on the momentum distributions of the different quark species (Wally Melnitchouk), and with the use of spin-polarized electrons and polarized nucleon targets the quark contributions to the nucleon spin have been precisely measured (Xiangdong Ji), putting significant challenges on the theory of QCD to reproduce or predict the results of these measurements.

As these studies continue, both in experiment with high precision measurements, and in theory, new challenges have arisen with the discovery of the generalized parton distributions that lead to the assembly of 3-dimensional tomographic images of the quark (and gluon) transverse spatial and longitudinal momentum distributions employing deeply virtual exclusive processes (Andreas Schafer and Feng Yuan). The challenges here will be on the experiments to access these generalized parton distributions (GPDs) and transverse momentum distributions (TMDs) from experiments like deeply virtual Compton scattering and deeply virtual meson production, and on the phenomenology aiding the analysis. Some of the measurements are underway at Jefferson Lab in several experiment halls. The EIC will vastly extend the kinematic reach of the measurements into the gluon dominated regime.

10.1 Form factors

Andrew Puckett

10.1.1 Introduction

Elastic scattering of nucleons by point-like, leptonic probes is among the simplest observable processes sensitive to the nucleon's internal structure. The study of elastic electron–nucleon scattering started in the 1950s with the pioneering measurements by Robert Hofstadter and collaborators in HEPL (the High Energy Physics Lab) at Stanford [599] at incident electron energies of up to 550 MeV. Among the highlights of this work were the first conclusive demonstration of a deviation of the elastic electron–proton scattering cross section from point-like behavior, and the first direct measurement of the proton's finite size, leading to the awarding of the 1961 Nobel Prize in Physics to Hofstadter *"for his pioneering studies of electron scattering in atomic nuclei and for his thereby achieved discoveries concerning the structure of the nucleons"*.

In the Standard Model, the lepton–nucleon interaction is purely electroweak. Due to the nucleon's finite size and complicated structure, the elastic scattering cross section falls much more rapidly as a function of the squared four-momentum transfer Q^2 than the point-like scattering cross section. Given the limitations of past, present, and planned lepton–hadron scattering facilities, *elastic* scattering of leptons by nucleons only occurs with sufficient probability to be practically measurable at energy scales where electromagnetic interactions are dominant; i.e., at four-momentum transfers $Q^2 \ll M^2_{W,Z}$, where $M_W(M_Z) \approx 80(91)$ GeV is the $W(Z)$ boson mass. As such, for most practical purposes this process can be interpreted in the framework of low-order perturbation theory in quantum electrodynamics (QED). However, the elastic form factors of the nucleon for charged- and neutral-current weak interactions are interesting in their own right and accessible even at relatively low energies in neutrino scattering [2946] and through parity-violating asymmetries in polarized electron scattering that are sensitive at leading order to the interference between photon and Z exchange amplitudes [2947–2949].

The use of elastic lepton–nucleon scattering as a precision probe of nucleon structure and dynamics remains a highly active area of investigation at low and high energies. The improvements in energy reach and precision of these measurements over decades have led to many important discoveries and surprises that have dramatically reshaped our understanding of the nucleon. This section will present a brief summary of the status of the nucleon's elastic scattering form factors, their definition and physical interpretation, outstanding challenges and problems, and the near-future outlook for further advancements.

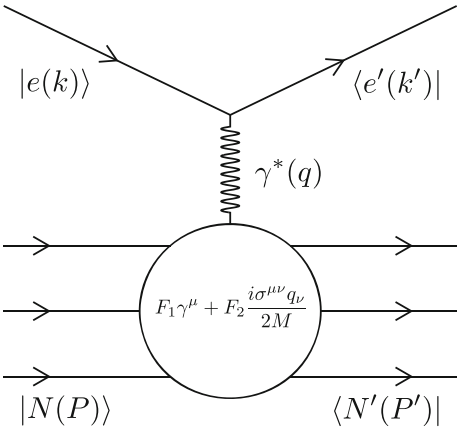


Fig. 242 Feynman diagram for elastic eN scattering in the one-photon-exchange approximation. Initial and final nucleons are represented by three lines each to indicate the nucleon's three-quark valence structure, while the circle represents the modification of the photon-nucleon vertex function by the nucleon's internal structure. See text for details

10.1.2 Theoretical formalism

The starting point for the interpretation of elastic electron–nucleon (eN) scattering is the one-photon-exchange (OPE) approximation, which is roughly analogous to the first Born approximation and/or the plane wave impulse approximation in non-relativistic quantum scattering theory. In the following discussion the terms OPE and Born approximation will be used interchangeably. The tree-level Feynman diagram for $eN \to eN$ is depicted schematically in Fig. 242. An incident electron of four-momentum $k \equiv (E_e, \mathbf{k})$ scatters from a nucleon of mass M, assumed to be initially at rest in the lab frame, with initial four-momentum $P \equiv (E_N, \mathbf{p}) = (M, \mathbf{0})$. The electron recoils with four-momentum $k' \equiv (E'_e, \mathbf{k}')$ and the nucleon recoils with four-momentum $P' \equiv (E'_N, \mathbf{p'})$ after absorbing the four-momentum transfer $q \equiv k - k'$. Energy and momentum conservation in this two-body scattering process dictate $P' = P + q = (M + E_e - E'_e, \mathbf{k} - \mathbf{k}')$. Together with the requirement that the final-state particles be "on mass shell"; i.e., that they satisfy the relativistic energy–momentum relation for a free particle ($E^2 = p^2 + m^2$), the kinematics of the elastic eN scattering process are entirely specified by just two independent variables, commonly chosen to be the incident electron energy E_e and the electron scattering angle θ_e that are directly observed experimentally. Another main variable of interest is the squared four-momentum transfer $Q^2 \equiv -q^2 = -(k - k')^2 > 0$. In the nucleon rest frame (or in the center-of-momentum frame, or any other frame in which the initial momenta of the colliding particles are collinear), the scattering process is independent of the azimuthal scattering angle of the electron.

In most modern electron–nucleon scattering experiments, it is safe to use the ultrarelativistic approximation for the electron ($|\mathbf{k}| = E_e, |\mathbf{k}'| = E'_e, k^2 = k'^2 = 0$), as the incident beam energies required for sensitivity to the non-trivial details of nucleon structure are generally quite large compared to the electron mass. Moreover, the vast majority of elastic electron–nucleon scattering data come from fixed-target experiments, in which the target nucleus is at rest in the lab frame. Unless otherwise noted, all of the following expressions apply in the initial nucleon's rest frame.

To develop intuition for the physical interpretation of elastic eN scattering, it is useful to consider the closely related process of ultrarelativistic electron scattering from a static charge distribution $\rho(\mathbf{r})$ with total charge Ze, given in the OPE approximation by:

$$\left(\frac{d\sigma}{d\Omega_e}\right) = \frac{Z\alpha^2}{4E_e^2}\frac{\cos^2\left(\frac{\theta_e}{2}\right)}{\sin^4\left(\frac{\theta_e}{2}\right)}|F(\mathbf{q})|^2 \tag{10.1}$$

$$\equiv \left(\frac{d\sigma}{d\Omega_e}\right)_{\text{Mott}}|F(\mathbf{q})|^2, \tag{10.2}$$

where E_e is the incident electron energy, θ_e is the electron scattering angle, α is the fine structure constant, and $F(\mathbf{q})$ is the electron scattering *form factor* given by the Fourier transform of the charge distribution:

$$F(\mathbf{q}) \equiv \int \rho(\mathbf{r})e^{i\mathbf{q}\cdot\mathbf{r}}d^3\mathbf{r}, \tag{10.3}$$

with $\mathbf{q} \equiv \mathbf{k} - \mathbf{k}'$ the three-momentum transfer in the scattering process. The Mott cross section as defined in Eq. (10.2) describes the scattering of ultrarelativistic electrons from a point-like target of charge Ze with zero spin and zero magnetic moment, in the limit where target recoil is negligible. In the electron–nucleon scattering case, this corresponds to the requirement $Q^2 \ll 2ME_e$. When target recoil is not negligible, the electron loses energy in the collision, and the Mott cross section is modified by the factor E'_e/E_e:

$$\left(\frac{d\sigma}{d\Omega_e}\right)_{\text{Mott}} = \frac{Z\alpha^2}{4E_e^2}\frac{\cos^2\left(\frac{\theta_e}{2}\right)}{\sin^4\left(\frac{\theta_e}{2}\right)}\frac{E'_e}{E_e}. \tag{10.4}$$

In much of the modern literature, Eq. (10.4) is taken as the definition of the Mott cross section, whereas in Mott's original paper, the target recoil factor E'_e/E_e is not included. Hereafter, we will use the definition (10.4) unless otherwise noted.

The most general form of the single-photon-exchange amplitude $\mathcal{M}$ for elastic eN scattering consistent with Lorentz invariance, gauge invariance, and parity conservation as required by QED, and under the assumption that the

nucleon is a spin-1/2 fermion obeying the Dirac equation, can be expressed using the Feynman rules of QED (see, e.g., [2140]) as follows (in "natural units" $\hbar = c = 1$):

$$\mathcal{M} = 4\pi\alpha\bar{u}(k')\gamma^{\mu}u(k)\left(\frac{g_{\mu\nu}}{q^2}\right)\bar{u}(P')\Gamma^{\nu}u(P). \tag{10.5}$$

Here $\mathcal{M}$ is the Lorentz-invariant single-photon-exchange amplitude, $\bar{u}$ and u represent free-particle Dirac spinors for the incoming and outgoing particles, evaluated at the relevant four-momenta, γ^{μ} is a Dirac γ matrix, $g_{\mu\nu}$ is the Minkowski metric tensor, and Γ^{ν} represents the photon-nucleon vertex function, given by:

$$\Gamma^{\mu} = F_1(q^2)\gamma^{\mu} + \frac{i\sigma^{\mu\nu}q_{\nu}}{2M}F_2(q^2), \tag{10.6}$$

with $\sigma^{\mu\nu} \equiv \frac{i}{2}\left[\gamma^{\mu}, \gamma^{\nu}\right]$ the antisymmetric tensor formed from $\gamma^{\mu}, \gamma^{\nu}$. The form factors $F_1(q^2)$ (Dirac) and $F_2(q^2)$ (Pauli) can be regarded as matrix elements of the electromagnetic current operator between final and initial nucleon states. They are real-valued functions of q^2, which is the only independent Lorentz scalar variable on which the photon-nucleon vertex function Γ^{μ} can depend. The convention (10.6) for the $\gamma^{*}N$ vertex function is the most commonly used one in the literature, and is constructed such that the amplitude is real (assuming real-valued form factors).[102] F_1 and F_2 represent the (electron) helicity-conserving and (electron) helicity-flip amplitudes, respectively. The nucleon's charge and Dirac ("non-anomalous") magnetic moment distributions determine the behavior of $F_1(q^2)$, while $F_2(q^2)$ measures the contribution of the "anomalous" magnetic moment distribution to the scattering.

Experimentally, the following linearly independent combinations of F_1 and F_2, known as the Sachs electric (G_E) and magnetic (G_M) form factors [2950], are more convenient:

$$G_E = F_1 - \tau F_2 \tag{10.7}$$

$$G_M = F_1 + F_2. \tag{10.8}$$

The differential cross section in OPE is given in terms of the Sachs form factors by [599,2950–2952]

$$\frac{d\sigma}{d\Omega_e} = \left(\frac{d\sigma}{d\Omega_e}\right)_{\text{Mott}} \frac{\epsilon G_E^2 + \tau G_M^2}{\epsilon(1+\tau)}, \tag{10.9}$$

where τ and ϵ are kinematic parameters defined as

$$\tau \equiv \frac{Q^2}{4M^2} \tag{10.10}$$

[102] Since no other diagrams interfere with the OPE at the same order in α, we are of course free to choose the phase of the OPE amplitude arbitrarily without affecting physical observables.

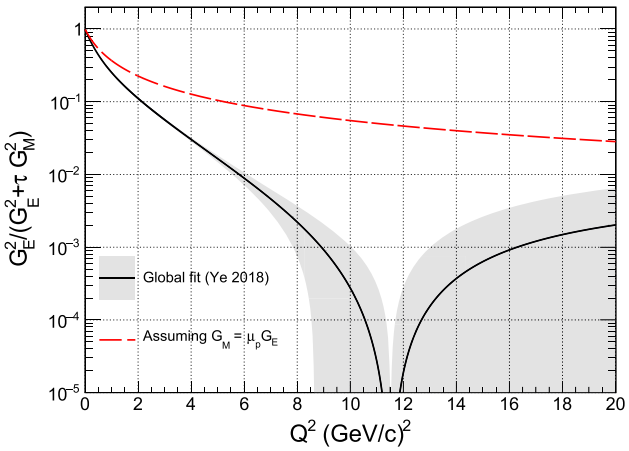


Fig. 243 Q^2 dependence of the ratio $G_E^2/(G_E^2+\tau G_M^2)$ for the proton, representing the maximum fraction of the reduced cross section carried by the electric term (at $\epsilon = 1$). The central value and uncertainty band of the curve are calculated from the global fit of Ref. [1076]. The dashed line shows the ratio that would be obtained under the assumption of form factor scaling ($G_M^p = \mu_p G_E^p$)

$$\epsilon \equiv \left[1 + 2(1+\tau)\tan^2\left(\frac{\theta_e}{2}\right)\right]^{-1}. \tag{10.11}$$

In the OPE approximation, ϵ can be interpreted as the longitudinal polarization of the virtual photon [2952]. The electric and magnetic contributions to the scattering can be separated by measuring the cross section while varying the beam energy and the scattering angle in such a way as to hold Q^2 constant while varying ϵ, a technique known as Longitudinal/Transverse (L/T) separation or Rosenbluth separation. The "reduced" cross section

$$\sigma_R \equiv \epsilon(1+\tau)\frac{(d\sigma/d\Omega_e)_{\text{Measured}}}{(d\sigma/d\Omega_e)_{\text{Mott}}},$$

is linear in ϵ, with slope (intercept) equal to G_E^2 (τG_M^2).

In the limit of very small Q^2, corresponding to long-wavelength virtual photons, the cross section behaves as if the nucleon were a point particle of charge ze ($z = +1(0)$ for proton (neutron)) and magnetic moment $\mu = (z+\kappa)$ (in units of the nuclear magneton), with κ the anomalous magnetic moment. In this limit, the form factors thus become $G_E(0) = z$ and $G_M(0) = z + \kappa$. For small but finite Q^2 such that $\tau \ll \epsilon G_E^2/G_M^2$, the electric term dominates the cross section, and if target recoil is neglected, Eq. (10.9) takes the same form as Eq. (10.2), with $G_E \equiv F(\mathbf{q})$. Thus, in the low-energy limit, the electric form factor can be identified with the Fourier transform of the charge density. Similar reasoning leads to an interpretation of G_M as a Fourier transform of the nucleon's magnetization density.

The Rosenbluth formula (10.9) describes *unpolarized* electron–nucleon scattering. At large values of Q^2, the magnetic term dominates the OPE cross section, and the sensitivity of the Rosenbluth method to G_E vanishes (see Fig. 243).

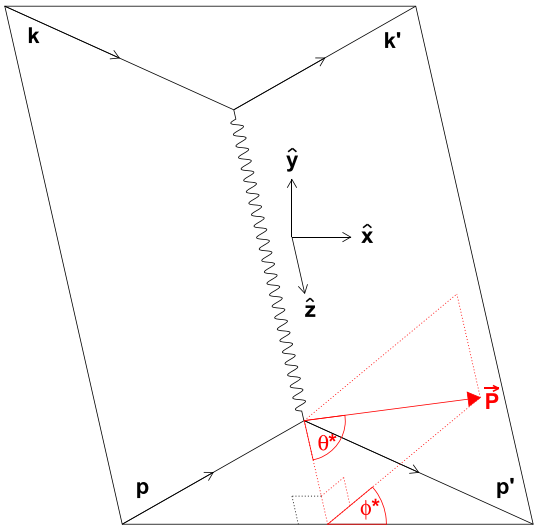


Fig. 244 Standard coordinate system for nucleon polarization components in elastic eN scattering. The arrow labeled $\vec{P}$ indicates the nucleon polarization direction and illustrates the definitions of the angles θ^* and ϕ^* between $\vec{P}$ and the momentum transfer $\mathbf{q}$. The x or "t" (transverse) axis is parallel to the reaction plane but perpendicular to the momentum transfer. The y or "n" (normal) axis is perpendicular to the reaction plane defined by $\hat{n} \equiv \hat{q} \times \hat{k}$. The z or "ℓ" (longitudinal) axis is along the momentum transfer direction, which coincides with the outgoing nucleon direction in the lab frame. The direction of the x axis is chosen so that the Cartesian basis $(\hat{x}, \hat{y}, \hat{z})$ is right-handed

As the use of electron scattering to investigate nuclear structure expanded during the 1960s and 1970s, and as the technology to produce spin-polarized electron beams and nuclear targets was being developed and improved, several authors independently developed the theory of spin-polarized elastic eN scattering in the OPE approximation and examined the implications for future measurements of polarization observables [2953–2956]. Nonzero asymmetries arise when the incident electron beam is longitudinally polarized and either the target nucleon is also polarized, or the recoil nucleon polarization is measured, or both. Asymmetries involving transverse electron beam polarization are generally suppressed by factors of m_e/E_e relative to longitudinal asymmetries, and while such asymmetries have been measured and are interesting in their own right, they are not ideal observables for measuring electromagnetic form factors, and they will not be considered further in this section.

Figure 244 illustrates the "standard" coordinate system used in most of the literature on polarized elastic eN scattering. In the case where the target nucleon is polarized, the asymmetry in the scattering cross section between positive and negative electron beam helicities is given by

$$A_{eN} \equiv \frac{\sigma_+ - \sigma_-}{\sigma_+ + \sigma_-} \tag{10.12}$$

$$= P_{\text{beam}} P_{\text{targ}} \left[A_t \sin\theta^* \cos\phi^* + A_\ell \cos\theta^* \right], \tag{10.13}$$

where P_{beam} is the longitudinal electron beam polarization, P_{targ} is the magnitude of the target nucleon polarization, and the angles θ^*, ϕ^* are defined in Fig. 244. The asymmetries A_t and A_ℓ are given in terms of τ, ϵ, and the form factor ratio $r \equiv G_E/G_M$ by:

$$A_t = -\sqrt{\frac{2\epsilon(1-\epsilon)}{\tau}} \frac{r}{1 + \frac{\epsilon}{\tau} r^2}$$

$$A_\ell = -\frac{\sqrt{1-\epsilon^2}}{1 + \frac{\epsilon}{\tau} r^2}. \tag{10.14}$$

Equations (10.14) show that the sensitivity of the double-spin asymmetry A_{eN} to the form factor ratio is generally highest when the target is polarized perpendicular to the momentum transfer but parallel to the scattering plane; i.e., along the x direction in Fig. 244. Note also that the asymmetries are sensitive to the *ratio* G_E/G_M, but not G_E or G_M separately. When the target is unpolarized, the longitudinally polarized electron transfers polarization to the outgoing nucleon. The nonvanishing components of the transferred polarization in OPE are

$$P_t = P_{\text{beam}} A_t$$

$$P_\ell = -P_{\text{beam}} A_\ell. \tag{10.15}$$

Here P_t and P_ℓ are the in-plane transverse and longitudinal components of the recoil nucleon's polarization, respectively. The sign change of P_ℓ relative to A_ℓ reflects the spin flip required to conserve angular momentum when the nucleon absorbs a transversely polarized virtual photon. The ratio P_t/P_ℓ is directly proportional to the form factor ratio G_E/G_M:

$$\frac{G_E}{G_M} = -\frac{P_t}{P_\ell} \sqrt{\frac{\tau(1+\epsilon)}{2\epsilon}} = -\frac{P_t}{P_\ell} \frac{E_e + E'_e}{2M} \tan\left(\frac{\theta_e}{2}\right) \tag{10.16}$$

Measurements of the differential cross sections, Eq. (10.9), and polarization observables, Eqs. (10.14) and (10.16), in elastic eN scattering are the main source of knowledge of the nucleon's electromagnetic form factors, which are among the most important precision benchmarks for testing theoretical models of the nucleon. Moreover, precise knowledge of these form factors is required for the interpretation of many different experiments in nuclear and particle physics. In the next section, we summarize the existing data on nucleon form factors.

10.1.3 Experimental data

Figures 245, 246, 247, 248 summarize the state of empirical knowledge of the nucleon electromagnetic form fac-

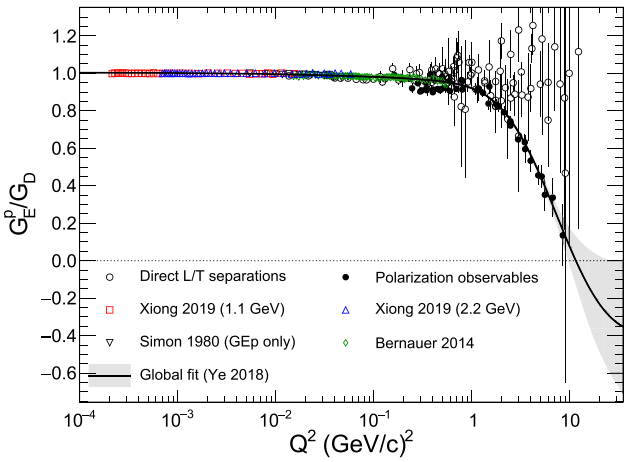


Fig. 245 (Approximate) World data for G_E^p/G_D. "Direct L/T separations" are published point extractions of G_E^p from Rosenbluth plots. The points labeled "Bernauer 2014" are the direct Rosenbluth extractions from the Mainz A1 dataset [618,2957]. The data labeled "Xiong 2019" are from the PRad experiment [2958]. The global fit is from [1076]. See text for details

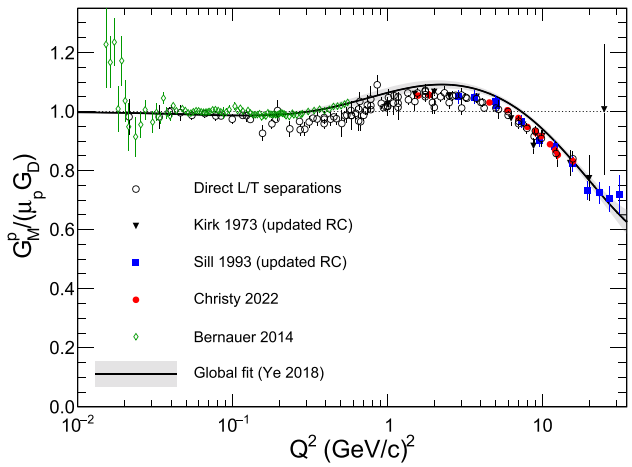


Fig. 246 (approximate) World data for $G_M^p/(\mu_p G_D)$. The "Direct L/T separations" are published point extractions of G_M^p from Rosenbluth plots. The Kirk 1973 data [2959] and the Sill 1993 data [2960] are point G_M^p extractions from single cross section measurements, with updated radiative corrections as detailed in Ref. [2961]. The data labeled "Christy 2022" are the point G_M^p extractions from the individual cross section measurements published in Ref. [2961]. The points labeled "Bernauer 2014" are the direct Rosenbluth extractions from the Mainz A1 dataset [618,2957]. The global fit curve is that of Ref. [1076]. See text for details

tors, as of this writing. The proton form factors G_E^p and G_M^p extracted from cross section measurements, as well as the neutron magnetic form factor G_M^n, can be described to within $\approx$ 10% over most of the measured Q^2 range by $G_E^p \approx G_M^p/\mu_p \approx G_M^n/\mu_n \approx G_D$, where G_D is the "dipole" form factor defined as

$$G_D = \left(1 + \frac{Q^2}{\Lambda^2}\right)^{-2}, \tag{10.17}$$

with the scale parameter $\Lambda^2 = 0.71$ (GeV/c)^2 defining the so-called "standard dipole". The neutron electric form factor G_E^n has a very different Q^2 dependence; since the neutron has zero *net* charge, $G_E^n(0) = 0$. Nevertheless, the neutron *rms* charge radius has been determined with good precision via neutron–electron scattering length measurements (see Ref. [616] and references therein). Existing measurements of G_E^n in quasi-elastic electron scattering on bound neutrons in light nuclear targets, shown in Fig. 247, exhibit a rapid rise with Q^2 to an appreciable fraction of G_D (nearly $\approx$ 50% at the highest Q^2 for which we have reliable G_E^n data). Precise high-Q^2 measurements of G_E^p/G_M^p using the polarization transfer method revealed that G_E^p starts falling much faster than G_D above 1 (GeV/c)^2, while G_M^p/μ_p falls to about 70% of G_D at the highest measured Q^2 values. Reliable neutron form factor data only reach $Q^2 \approx 3.4(4.5)$ (GeV/c)^2 for $G_E^n(G_M^n)$, but significant expansions in the Q^2 reach of the neutron data are anticipated in the near future.

The three-dimensional Fourier transform of G_D gives an exponentially decreasing charge density as a function of the radial distance from the center of the nucleon, assuming a spherically symmetric density. The mean square radius of the nucleon charge density is related to the slope of the electric form factor in the limit $Q^2 \to 0$:

$$\left\langle r_E^2 \right\rangle = -6 \left.\frac{dG_E}{dQ^2}\right|_{Q^2=0}. \tag{10.18}$$

For the standard dipole form factor, the implied charge radius is $\sqrt{\langle r_E^2 \rangle_D} = 0.81$ fm, which is in rough agreement with modern, precise determinations of the proton charge radius from electron scattering and the spectroscopy of electronic and muonic hydrogen. See Ref. [2962] for a very recent, in-depth review of the experimental and theoretical status of the proton charge radius.

Proton data and discussion

Figures 245 and 246 show most of the existing data for the proton electric and magnetic form factors G_E^p and G_M^p/μ_p, respectively, normalized to G_D, over the entire measured Q^2 range. While not comprehensive, the data shown are sufficiently representative of the Q^2 coverage and precision of the entire world data. The points shown as empty circles in Figs. 245 and 246 are published point extractions of G_E^p and G_M^p based on direct L/T separations from Rosenbluth plots, and are taken from Refs. [2961,2963–2971]. These extractions are not entirely independent of each other in terms of cross section input, as several of the analyses combined data from multiple experiments at similar Q^2 values.

The points shown as filled circles in Fig. 245 are based on direct measurements of the ratio G_E^p/G_M^p using polarization

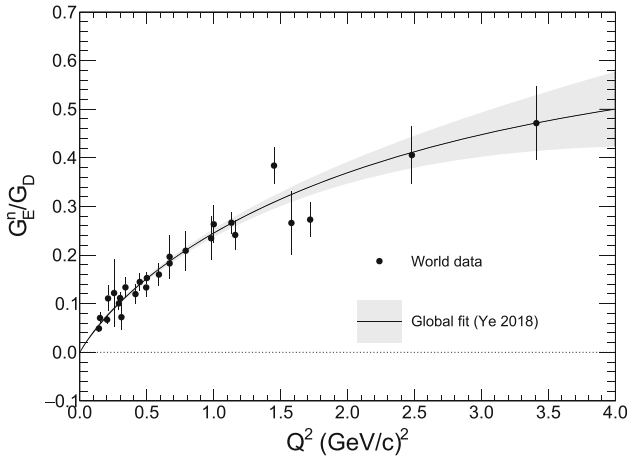


Fig. 247 World data for neutron electric form factor G_E^n/G_D. See text for references, details

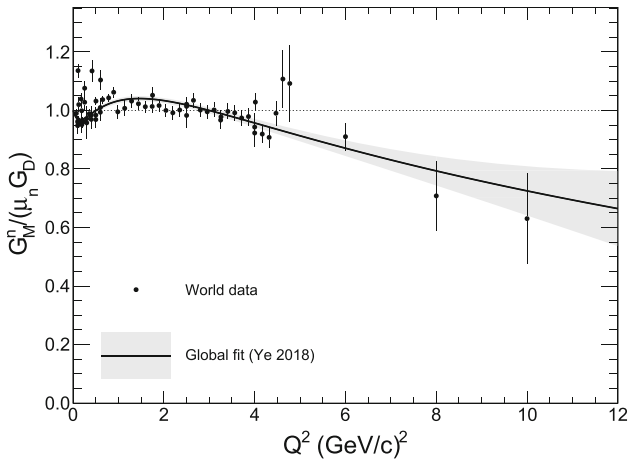


Fig. 248 World data for neutron magnetic form factor G_M^n/G_D. See text for references, details

observables, converted to G_E^p/G_D values using the global fit to G_M^p from Ref. [1076]. The polarization data for G_E^p include measurements based on the polarization transfer technique of Eq. (10.16) (Refs. [2972–2985]), and the beam-target double-spin asymmetry method, Eq. (10.14) [2986–2988]. The points labeled "Bernauer 2014" in Figs. 245 and 246 are the direct Rosenbluth separations from the Mainz A1 dataset [618,2957]. The points at very low Q^2 labeled "Simon 1980" and "Xiong 2019" in Fig. 245 are direct extractions of G_E^p from individual cross section measurements based on the assumption of form factor scaling ($G_M^p = \mu_p G_E^p$) in the case of Ref. [2989], or using the Kelly fit to G_M^p [610] in the case of Ref. [2958]. In Fig. 246, the G_M^p values extracted from the cross sections published in Refs. [2959,2960] are based on the updated analysis in Ref. [2961], which used the "state-of-the-art" radiative corrections described in Ref. [2990]. It must also be noted that the global fits shown in Figs. 245 and 246 include phenomenological two-photon-exchange corrections that have *not* been applied to the published form factor extractions. These corrections tend to increase the value of G_M^p by roughly 2–3% in the Q^2 range where the discrepancy between Rosenbluth and polarization results is largest.

The extraction of nucleon form factors from cross section measurements generally requires corrections to account for the effects of higher-order QED radiative processes in order to isolate the OPE term from which G_E^2 and/or G_M^2 can be determined. While each of these higher-order terms is at least $\mathcal{O}(\alpha)$ relative to the Born term, their combined effect on the observed cross sections can be significant; typically as much as 10–30% at modest-to-large Q^2 [2991]. As a general rule, the magnitude of the radiative correction (RC) to the elastic cross section tends to increase at large Q^2 values and/or large θ_e/small ϵ, and also depends on experiment-specific parameters including detector acceptance and resolution, electron beam properties, and target geometry, material, and density. Additionally, the calculation of the RC depends strongly on whether the experiment detects the scattered electron only (most common), the recoil proton only (see, e.g., Ref. [2970]), or both final-state particles. For many experiments, the RC calculation is an important source of uncertainty in the extraction of the Born cross section, which is not directly observable, and can dramatically change the slope of the Rosenbluth plot in converting measured cross sections to Born cross sections [2991].

At next-to-leading order in α, the "standard", model-independent RCs to $ep \to ep$ scattering include vacuum polarization, vertex, and self-energy terms that are purely virtual and depend only on Q^2, Bremsstrahlung (real photon emission), which depends strongly on both Q^2 and ϵ and modifies the reaction kinematics, and two-photon-exchange (TPE), in the limit where one of the two exchanged photons is "soft". The contribution of "hard" TPE, in which both exchanged photons carry a "large" momentum, cannot presently be calculated model-independently, and is neglected in the standard radiative correction procedures. It is thought to be largely responsible for the discrepancy between cross sections and polarization observables [2992] in high-Q^2 extractions of G_E^p, and is presently the subject of vigorous worldwide experimental and theoretical investigation. For a recent review of the subject, see Ref. [2993].

For conventional RC, most of the earlier published extractions of the proton form factors relied on the work of Tsai [2994] or Mo and Tsai [2995]. Following the initial discovery of the rapid fall-off of G_E^p/G_M^p at large Q^2 using polarization transfer [2973], and the resulting large discrepancy between two different observables sensitive (in principle) to the same fundamental property of the proton, Maximon and Tjon [2996] refined the mathematical treatment of these corrections and removed many of the approximations made in the expressions of Mo and Tsai, including an exact calculation of the soft Bremsstrahlung contributions.

Several authors [2961,2990,2997] have recently examined the quantitative differences between the calculations of Ref. [2995] and the more accurate approach of Ref. [2996], and studied the impact of these differences on previously published extractions of the form factors. Updating the published cross sections to use the more modern RC prescriptions is a non-trivial undertaking, especially for the older experiments, since the required modifications depend on details of the experiments and the associated data analyses that in some cases were not thoroughly documented in the final publications.

The most recent and comprehensive effort thus far to update published elastic ep cross sections to use "state-of-the-art" RC in the high-Q^2 region was described in Ref. [2961]. The reanalysis focused on a subset of high-Q^2 experiments from Jefferson Lab and SLAC for which the original publications provided sufficient details of the experimental parameters and the RC prescriptions and cutoffs used that they could be corrected in a self-consistent way [2961]. As noted by the authors of [2961] and earlier by [2990], the effect of updating the RC to the older SLAC data is to reduce, but not eliminate, the magnitude and significance of the discrepancy in the high-Q^2 region. The new, precise cross sections from Jefferson Lab's Hall A [2961] extend the Q^2 range for which a statistically significant discrepancy between cross sections and polarization observables is established.

In the polarization transfer method, the simultaneous measurement of the recoil nucleon polarization components P_t and P_ℓ offers many advantages in the control of experimental uncertainties. In particular, the form factor ratio can be determined in a single measurement, eliminating uncertainties resulting from changes in experimental parameters such as the beam energy, detector angles, spectrometer magnetic field settings, target polarization and spin direction, and others. Moreover, the beam polarization and many other sources of systematic uncertainty associated with recoil nucleon polarimetry cancel in the ratio P_t/P_ℓ, see, e.g., Ref. [2982], and reversal of the electron beam helicity reverses the direction of the recoil nucleon polarization while leaving all other experimental parameters unchanged, providing for robust cancellation of systematic effects associated with polarimeter acceptance and/or detection efficiency [2998]. The dramatically different behavior of G_E^p implied by the polarization data has profound implications for theoretical modeling of nucleon structure, as discussed below.

While polarization measurements of G_E/G_M are generally thought to have small systematic uncertainties, it must be noted that the published data exhibit significant internal tension in the region 0.1–1 GeV^2 where several high-precision experiments give somewhat conflicting results [2975,2983–2985,2987]. Despite this unresolved tension, polarization observables are generally regarded as giving the most reliable determination of G_E^p at large Q^2 values, due to their superior sensitivity to G_E as compared to the Rosenbluth method, and their relative insensitivity to radiative corrections [2999,2999–3001] and higher-order QED corrections neglected by the standard RC procedures, such as hard Two-Photon-Exchange (TPE) [2981,2982]. This property derives from the fact that polarization asymmetries are ratios of polarized and unpolarized cross sections, that tend to be affected similarly by radiative processes. The P_t/P_ℓ ratio in the polarization transfer method is a ratio of such ratios, and the model-independent RC to this ratio tend to be utterly negligible compared to the uncertainties in the presently measured range of Q^2 [2982]. Moreover, a precise search for evidence of hard TPE contributions in this observable found no significant effect [2981] at 2.5 GeV^2, with the ratio $\mu_p G_E^p/G_M^p$ showing no variation with ϵ in the range 0.15–0.8 with $\approx 1\%$ total uncertainties.

Assuming that polarization measurements give the "true" value of G_E^p, the fractional contribution of the ϵG_E^2 term to the OPE cross section falls rapidly with Q^2, as shown in Fig. 243. Based on the global fit of Ref. [1076], the electric term contributes at most 10% of σ_R at 2 (GeV/c)^2, 2% at 5 (GeV/c)^2, and even less at higher Q^2, basically wiping out any meaningful sensitivity to G_E, since its contribution to σ_R becomes comparable to the limits of experimental accuracy and to the expected magnitude of higher-order QED corrections that are theoretically and experimentally uncertain.

In addition to efforts to resolve the difficulties with G_E^p at large Q^2, there has been a renewed effort to improve the precision of elastic ep scattering data at very low Q^2, since the CREMA collaboration first published an extremely precise extraction of the proton radius from Lamb shift measurements in muonic hydrogen [3002], yielding a radius of about 0.84 fm, smaller by roughly seven standard deviations than the previous consensus value (at the time) of 0.88 fm from electron–proton scattering and spectroscopy of ordinary hydrogen. The Mainz A1 collaboration [618,2957] carried out a systematic program of over 1,400 precision cross section measurements spanning the Q^2 range 0.003–1 GeV^2 using the "traditional" method based on magnetic spectrometers. They published several direct fits of G_E^p and G_M^p to their cross section data, testing various functional forms to accurately quantify the uncertainties. They also published direct L/T separations for $Q^2 \gtrsim 0.02$ GeV^2. While the Mainz G_E^p extraction is in good agreement with the rest of the world data, their G_M^p results, whether from global fits or direct L/T separations, are in significant tension with the other world data,[103] as is evident from Fig. 246. The slower fall-off with Q^2 of the Mainz G_M^p implies a smaller magnetic radius; indeed, the published Mainz extraction of the proton mag-

[103] Note, however, that the Mainz dataset implies a G_E^p/G_M^p ratio that is consistent with the high-precision polarization measurements by Zhan et al., Ref. [2984].

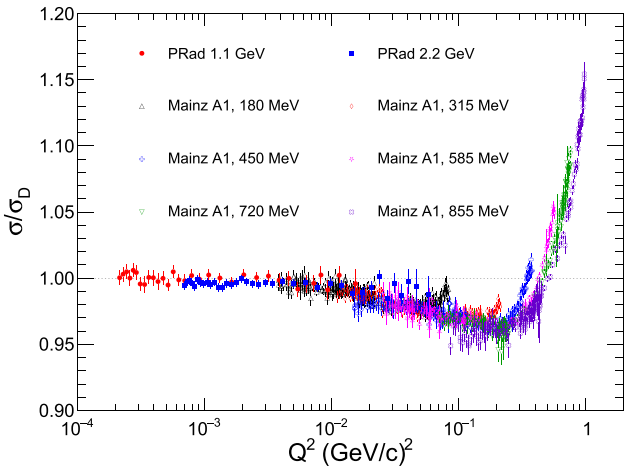


Fig. 249 Comparison of PRad [2958] and Mainz A1 [618] elastic $ep \to ep$ cross sections, normalized to the "standard dipole" cross section σ_D, calculated from Eq. (10.9) assuming $G_E^p = G_M^p/\mu_p = G_D$; i.e., $\epsilon(1+\tau)\sigma_D/\sigma_{\text{Mott}} = G_D^2\left(\epsilon + \mu_p^2\tau\right)$

netic radius r_M^p is about three standard deviations below the consensus of extractions based on other world data.

More recently, the PRad collaboration [2958] performed new $ep \to ep$ cross section measurements using a novel, magnetic-spectrometer-free method involving precision calorimetry, a windowless gas target, and a simultaneous measurement of the pure electroweak process of Möller scattering ($e^-e^- \to e^-e^-$) to constrain the absolute cross section normalization. Their measurements reached a minimum Q^2 of 0.0002 GeV^2 with small statistical and systematic uncertainties, achieving a proton radius measurement of $r_p \approx 0.83 \pm 0.01$ fm, consistent with the muonic hydrogen value.

Figure 249 shows the PRad and Mainz cross section data, normalized to the "standard dipole" cross section, calculated from Eq. (10.9) under the assumption $G_E^p = G_M^p/\mu_p = G_D$. The low end of the PRad Q^2 range is in a regime where the cross section is indistinguishable from point-like behavior within experimental precision; at the lowest Q^2 of the PRad dataset, $G_D^2 \approx 0.999$. This is unsurprising given that the de Broglie wavelength of the virtual photon $\lambda = \hbar c/\sqrt{Q^2} \approx$ 13 fm is large compared to r_p at this Q^2.

Neutron data and discussion

The neutron electromagnetic form factors are much more difficult to measure accurately than those of the proton, due primarily to the absence of free neutron targets of sufficient density for electron scattering experiments at large Q^2. The small cross sections for high-energy electromagnetic interactions can generally only be measured accurately in high-luminosity experiments, and the neutron's instability and zero charge severely limit the number of free neutrons that can be collected in a suitably small volume for a suitable duration for such experiments. As such, essentially all knowledge of neutron electromagnetic form factors at meaningfully large Q^2 values comes from measurements of electron scattering on bound neutrons in light nuclear targets such as deuterium and ^{3}He.

Since $G_E^n(0) = 0$, the cross section for elastic en scattering is dominated by the magnetic term over essentially the entire measured Q^2 range, even at relatively low Q^2. The neutron form factors are accessible experimentally through a number of scattering observables on light nuclear targets, including cross sections and spin asymmetries. Model-dependent extractions of neutron elastic form factors from measurements of elastic electron–deuteron scattering have also been attempted at relatively low Q^2 values (see, e.g., [3003–3006]), but are subject to large theoretical and experimental systematic uncertainties, and are generally considered less reliable than extractions from measurements of quasi-elastic scattering on bound nucleons in deuterium and/or Helium-3, although they are qualitatively consistent.

Figures 247 and 248 show most of the existing data for G_E^n and G_M^n, respectively, excluding extractions based on elastic ed cross section measurements. For G_E^n, essentially all reliable data of reasonable precision come from measurements of polarization observables, since the (quasi)-elastic $(e, e'n)$ cross section has relatively low sensitivity to G_E^n over the entire accessible Q^2 range. The data shown in Fig. 247 include extractions from asymmetry measurements on polarized deuterium targets (Refs. [3007–3010]), polarized ^{3}He targets (Refs. [3011–3015]), and via recoil neutron polarization on unpolarized deuterium (Refs. [3016–3018]).

The most reliable known method to determine the neutron magnetic form factor G_M^n is the so-called "ratio" or "Durand" technique [3019], in which "neutron-tagged" and "proton-tagged" quasi-elastic electron scattering on a deuterium target are measured simultaneously, and the ratio of cross sections

$$^2\text{H}(e, e'n)p/^2\text{H}(e, e'p)n$$

is measured. The simultaneous measurement of quasi-free scattering on bound protons and neutrons in deuterium, combined with the precise knowledge of the free proton cross section, allows a determination of the free neutron cross section with very small uncertainties. In particular, the electron acceptance and detection efficiency, the data acquisition deadtime, and the luminosity cancel exactly in the n/p ratio, and nuclear effects such as Fermi motion and binding, final-state interactions, meson-exchange currents, and others, as well as QED radiative corrections, tend to affect the $d(e, e'n)p$ and $d(e, e'p)n$ cross sections nearly identically [3020], for sufficiently tight cuts on the photon-nucleon invariant mass W^2, and the angle θ_{pq} between the detected nucleon's momentum and the momentum transfer direction, determined from the scattered electron's kinematics, to

ensure exclusivity of the reaction. The main source of experimental uncertainty with the ratio method is in the knowledge of the acceptance/detection efficiency for protons and neutrons. Of the data shown in Fig. 248, Refs. [3020–3024] used the ratio method, Refs. [3025–3028] extracted G_M^n from the beam-target double-spin asymmetry in inclusive quasielastic electron scattering on polarized Helium-3, and Refs. [3029–3031] extracted G_M^n from absolute cross section measurements in either inclusive scattering on deuterium or coincidence $d(e, e'n)p$ scattering. The low-Q^2 data for G_M^n show some inconsistencies, suggesting underestimated theoretical or experimental systematic uncertainties in some of the older measurements. The Super BigBite Spectrometer (SBS) Collaboration in Jefferson Lab's Hall A recently collected data using the ratio method to extend the knowledge of G_M^n to $Q^2 = 13.5$ GeV^2 with very small statistical and systematic uncertainties. The CEBAF Large Acceptance Spectrometer (CLAS) collaboration in Jefferson Lab's Hall B has also collected data for G_M^n up to $Q^2 \approx 10$ GeV^2, with qualitatively different sources of systematic uncertainty. Both datasets are currently under analysis.

10.1.4 Theoretical interpretation of nucleon form factors

As the spacelike electromagnetic form factors are among the simplest, most clearly interpretable, and best-known measurable dynamical properties of the nucleon, they constitute important benchmarks for testing theoretical models. Figure 250 shows the world data for the nucleon's spacelike EMFFs together with selected theoretical models and the expected results from the ongoing high-Q^2 form factor program in Hall A at Jefferson Lab by the Super BigBite Spectrometer (SBS) collaboration. The SBS measurements of the neutron magnetic form factor were completed during the Oct. 2021–Feb. 2022 running period in Hall A, and the data are currently under analysis. The SBS measurement of G_E^n/G_M^n using a polarized ^{3}He target is underway as of October 2022 and will run through March of 2023, and the polarization transfer measurements of G_E^n/G_M^n and G_E^p/G_M^p are expected to take data in 2023–2024. The expansion of the Q^2 range and precision of the proton and neutron data will severely test theoretical models of nucleon structure.

The calculation of nucleon structure from first principles in QCD is presently only possible using the methods of lattice gauge theory. The accuracy of lattice QCD calculations is rapidly improving with increases in computing power and improvements in the control of systematic errors, and the range of measurable quantities lattice QCD can predict continues to expand. Nevertheless, the prediction of nucleon form factors and other observables of hard exclusive processes from lattice QCD (see Refs. [609,611,3036,3037] and references therein for recent efforts at low and high Q^2) has not yet reached a level of precision and accuracy commensurate with that of the experimental data, particularly at high energies. As such, its predictions cannot yet be conclusively "tested" by the form factor data. Instead, the existing data serve to guide the improvement of the calculations. Meanwhile, the continued use of QCD-inspired phenomenological models, approximations, effective theories, and continuum methods provides valuable insight and improved understanding of the relevant degrees of freedom and dynamical effects at different energy scales when compared to the data.

For asymptotically large Q^2 values, pQCD (perturbative QCD) predicts the scaling behavior of the nucleon form factors based on simple constituent counting rules and helicity conservation [226]. For the nucleon, its three-quark valence structure predicts $F_1 \propto Q^{-4}$ and $F_2 \propto Q^{-6}$ (see the discussion in Sect. 5.9). While the proton data at the highest measured Q^2 values are in superficial qualitative agreement with the pQCD scaling predictions, it has been argued [3038,3039] that the pQCD mechanism of multiple hard gluon exchange is not applicable to exclusive processes in the presently accessible range of Q^2. More recently, Belitsky et al. considered the effects of both leading and subleading twist contributions to the nucleon's light-cone wavefunctions in a pQCD analysis of the Pauli form factor F_2, deriving the modified logarithmic scaling expression $Q^2F_2/F_1 \propto \ln^2\left(Q^2/\Lambda^2\right)$, with a range of values of $\Lambda \approx 200-300$ MeV describing the proton data rather well [3040]. However, in an analysis of the quark flavor decomposition of the spacelike FFs [3041] shortly following the publication of data for G_E^n/G_M^n up to 3.4 (GeV/c)^2 [3011] and G_M^n up to 4.8 (GeV/c)^2 [3020], it was noted that the neutron F_2^n/F_1^n data do not follow this logarithmic scaling, at least not for values of Λ similar to those fitting the proton data.

Dispersion theoretical analysis, including models based on the assumption of VMD (Vector Meson Dominance) [3042], provide a coherent, self-consistent framework for the joint interpretation of spacelike and timelike nucleon form factors over the entire physical range of Q^2. VMD-based models were among the earliest to describe the global features of the nucleon form factors and predicted the high-Q^2 falloff of G_E^p/G_M^p decades before the polarization transfer experiments. A key assumption of VMD and VMD-based models is that the virtual photon-nucleon interaction at low to intermediate Q^2 is dominated by vector meson pole terms, which contribute significantly to the dispersion integrals connecting the spacelike and timelike regions through the requirements of unitarity and analyticity of the form factors considered as functions of q^2 in the complex plane. For a recent review of the dispersion theoretical analysis of nucleon EMFFs, see Ref. [3043].

As mentioned above, in the very low-energy limit, when target recoil can be neglected, the form factors can be interpreted as three-dimensional Fourier transforms of the spa-

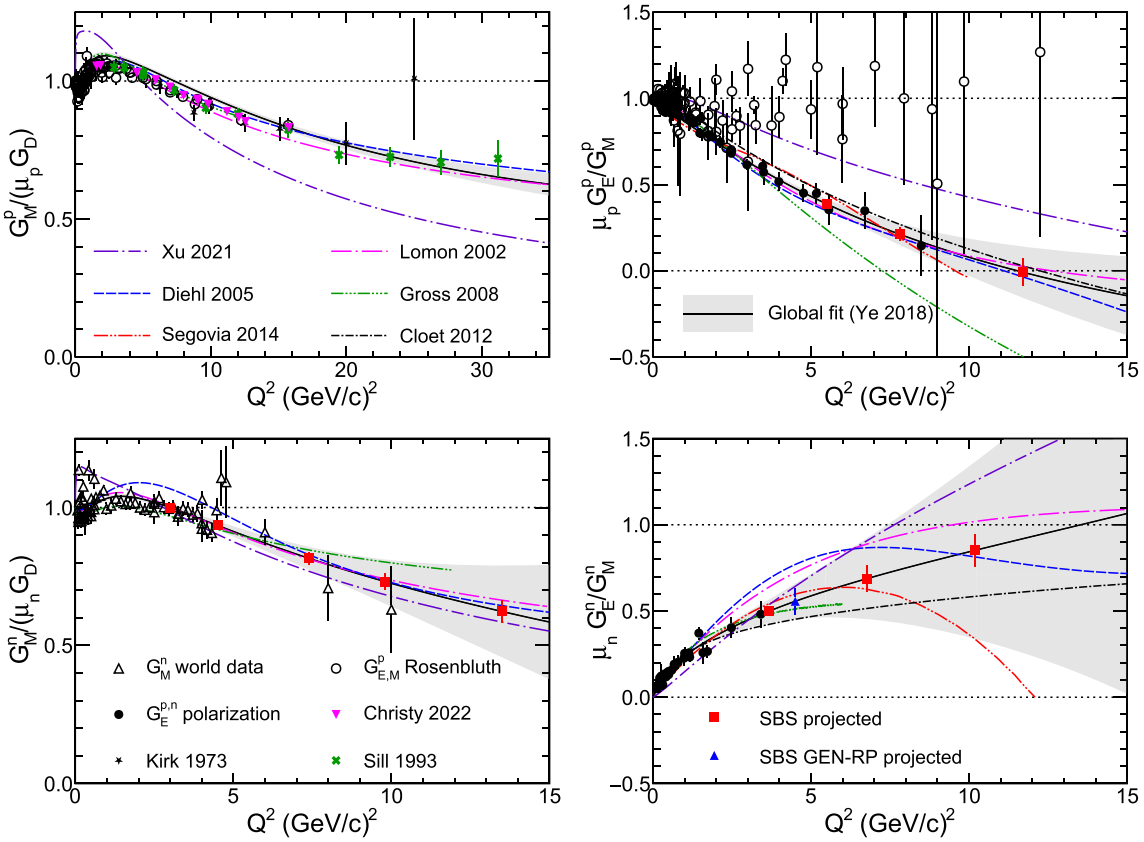


Fig. 250 Data for all four nucleon electromagnetic form factors at spacelike Q^2 with selected theoretical models, the global fit from [1076], and the projected results from the ongoing SBS program at Jefferson Lab, plotted arbitrarily at the values from the global fit. Theoretical curves shown are the BLFQ calculations of Ref. [941] (Xu 2021), the VMD-based model of Ref. [3032] (Lomon 2002), the GPD-based model of Ref. [3033] (Diehl 2005), the covariant spectator model of Ref. [3034] (Gross 2008), the DSE-based calculation of Ref. [2828] (Segovia 2014), and the quark–diquark model of Ref. [3035] (Cloet 2012). Data references are the same as those given in the text and described in the captions of Figs. 245, 246, 247, 248. See text for details

tial distributions of charge (G_E) and current (G_M) in the nucleon. While this naive density interpretation is invalidated by relativity for finite momentum transfers, several authors have extracted three-dimensional rest-frame densities from the form factors using model-dependent relativistic prescriptions to relate the Sachs form factors measured at a four-momentum transfer Q^2 to the static rest frame densities. A common feature of such extractions is the identification of the Sachs form factors G_E and G_M with Fourier transforms of the Breit frame[104] charge and current densities. The Breit frame densities are then modified by a boost factor $k^2 = Q^2/(1+\tau)$ relating Q^2 to the wave number k in the nucleon rest frame, and another model-dependent factor relating the Sachs FF to the so-called "intrinsic" form factors $\tilde{\rho}(k)$, defined as Fourier–Bessel transforms of the rest frame densities. The latter correction attempts to account for the Q^2-dependent boost of the nucleon wavefunction itself from the rest frame to the Breit frame. Kelly [3044] used expansions in a complete set of radial basis functions and a relativistic boost prescription consistent with the pQCD asymptotic behavior to minimize model-dependence and estimate the uncertainties in the radial densities due to the finite Q^2 range of the data. Among his key findings were a broader charge density for the proton compared to its magnetization density, consistent with the fall-off of the polarization data for G_E^p/G_M^p, and a neutron charge density described by a positive core surrounded by a negative exterior, consistent with pion-cloud models.

While the three-dimensional radial densities extracted from the form factors are necessarily model-dependent, a model-independent density interpretation of the form factors exists through sum rules relating the form factors to moments of Generalized Parton Distributions (GPDs) [1081]. Miller [3045,3046] showed that in the infinite momentum frame,

[104] The Breit or "brick-wall" frame in elastic eN scattering is the frame in which there is no energy transfer in the collision. It is related to the nucleon rest frame by a boost along the momentum transfer direction, with a boost factor $\gamma = \sqrt{1+\tau}$.

the impact-parameter-space densities of charge and magnetization in the nucleon are two-dimensional Fourier–Bessel transforms of the Dirac (F_1) and Pauli (F_2) form factors, respectively. Examples of empirical extractions of the transverse densities from the form factor data can be found in Refs. [2867,3047]. In apparent contrast to model-dependent extractions of 3D rest frame densities such as Kelly's, the neutron's transverse charge density exhibits a negative core surrounded by a positive exterior, contradicting the qualitative predictions of several models.

The form factors also play an important role in efforts to extract the GPDs from measurements of Deeply Virtual Compton Scattering (DVCS) and other hard exclusive processes. Through the aforementioned sum rules, the form factors F_1 and F_2 impose fairly powerful constraints on, respectively, the vector (H) and tensor (E) GPDs that enter the Ji sum rule for the nucleon spin decomposition [1081]. If good measurements and/or models of the GPDs exist, they can be used to predict the form factors [3048]. Alternatively, when combined with the forward parton distributions measured in deep inelastic scattering, the form factors can be used to constrain the GPDs [3033,3049], particularly at high Bjorken x and/or large $-t$. Apart from the direct constraints, precise knowledge of the form factors is also required for analysis of experiments attempting to measure GPDs, to separate the contributions of the DVCS and Bethe–Heitler processes, which interfere at the same order in α and are experimentally indistinguishable.

Constituent quark models (CQMs) have a long history in nuclear physics and predate the emergence of QCD as the accepted theory of strong interactions within the Standard Model. For a review and modern perspective on the role of the quark model in nuclear physics, see [513]. The early non-relativistic constituent quark model was successful in explaining the observed spectra of baryons and mesons as qqq (fermionic) and $q\bar{q}$ (bosonic) bound states, and making qualitative predictions of meson and baryon masses and magnetic moments. Indeed, one of the original motivations for the introduction of the color quantum number prior to the development of QCD was to preserve the Pauli exclusion principle for low-lying baryon states, whose combined spin/flavor/orbital quantum numbers are symmetric under the exchange of any two quarks. This issue was particularly acute for the spin-3/2 baryon decuplet. To explain dynamical properties of hadrons in terms of constituent quarks, a model for the confining quark–quark interaction and the resulting quark wavefunctions is needed. The "bare" u and d valence quark constituents of nucleons appearing in the QCD Lagrangian are almost massless compared to the nucleon mass. As such, the nucleon, considered as the ground state of a bound system of three light quarks, is characterized by a large ratio of binding energy to constituent mass, making a fully relativistic treatment mandatory to obtain realistic phenomenology and accurate descriptions of the data. A common feature of CQM calculations of nucleon structure is the "dressing" of the bare, almost-massless valence quarks by gluons and quark–antiquark pairs, leading to massive constituent quarks and/or diquarks as effective degrees of freedom, often carrying their own internal structure. While a full review of relativistic constituent quark model calculations of nucleon form factors is beyond the scope of this section, a fairly comprehensive overview is given in Ref. [3050] (see also Sect. 5).

In recent years, Hamiltonian light-front field theory has emerged as a useful framework for the nonperturbative solution of invariant masses and correlated parton amplitudes of self-bound systems [922]. Xu et al. recently applied this framework to calculate the structure of the nucleon using the method of Basis Light Front Quantization (BLFQ) [941]; see also Sect. 5.3. Their calculation used an effective light-front Hamiltonian with quarks as the only effective degrees of freedom, a transverse confining potential from light-front holography supplemented by a longitudinal confinement, and a one-gluon-exchange interaction with a fixed coupling. The light-front wave functions resulting from the solution of this Hamiltonian were then used to calculate the nucleon form factors, parton distributions, and other dynamical properties. The first form factor results from BLFQ [941], solved in the valence space of three quarks, are compared to the data and a selection of other theoretical models in Fig. 250. Such comparisons indicate the need for improvements to the magnetic form factors within BLFQ, particularly in the low-Q^2 region. Augmenting the BLFQ basis with dynamical gluons may provide such improvements [953].

In recent years, significant progress has occurred in the explanation and prediction of a wide range of measurable dynamical properties of hadrons in continuum nonperturbative QCD [821], within the framework of Dyson–Schwinger Equations (DSE). In this framework, the high-Q^2 behavior of proton and neutron form factors is very sensitive to the behavior of the momentum-dependent dressed quark mass function that governs the transition from massive, constituent-quark-like behavior at low energies to light, parton-like behavior at high energies [2875]. Moreover, the flavor decomposition of the form factors enabled by combined proton and neutron measurements, soon to be extended to Q^2 values up to 10 GeV^2, has the potential to elucidate the importance of diquark correlations in nucleon structure [793,3035]. Over the longer term, looking past the ongoing SBS program, major efforts are underway to establish intense polarized and unpolarized positron beams at Jefferson Lab, which will facilitate precise e^+p/e^-p comparisons over a much larger range of Q^2 and ϵ than presently available, hopefully leading to a decisive resolution of the Rosenbluth/polarization discrepancy for the proton, as part of a larger physics program using positron beams [3051]. The planned Electron-Ion Collider at Brookhaven National Lab-

oratory should be capable of measuring the elastic *ep* cross sections to a Q^2 of up to 40–50 (GeV/c)^2 [3052]. A proposed "low-cost" upgrade [3053] of Jefferson Lab's Continuous Electron Beam Accelerator Facility (CEBAF) to a maximum energy of 20+ GeV using fixed-field alternating gradient magnets to achieve 6–7 additional passes through the CEBAF linear accelerators would enable further expansions of the Q^2 reach for G_E^p, G_E^n, and G_M^n to at least 20 GeV^2.

10.2 Parton distributions

Wally Melnitchouk

10.2.1 Theoretical foundations

Parton distribution functions are the prototypical examples of QCD quantum correlation functions, which allow high-energy lepton and/or hadron scattering processes to be described in terms of quarks and gluons (or partons) (for reviews see Refs. [3054–3058]). The PDF for a quark of flavor i in a nucleon (moving with momentum p) is defined by the Fourier transform of a forward matrix element of quark bilinear operators, which in the $A^+ = 0$ gauge can be written as

$$f_{i/N}(x,\mu^2) = \frac{1}{4\pi}\int dz^-\, e^{-ixp^+z^-} \times \langle N(p)|\bar{\psi}_i(z^-)\,\gamma^+\,\psi_i(0)|N(p)\rangle, \quad (10.19)$$

where ψ_i is the quark field operator, x is the light-cone momentum fraction of the proton carried by the parton, and μ is the renormalization scale. Analogous expressions can be written for antiquark and gluon PDFs, the latter in terms of the gluon field strength tensor, $F^A_{\mu\nu}$.

The utility of PDFs is that they allow one to relate various high-energy scattering reactions, which would otherwise not be easily related to one another, and make predictions for new reactions in terms of the same set of PDFs obtained from previous experiments. The key to this is the ability to factorize the scattering process into a process-dependent, perturbatively calculable hard scattering cross section and the process-independent, nonperturbative function parametrized by the PDF. An important virtue of PDFs is that in the infinite momentum frame (or on the light-front) they can be simply interpreted as probability densities describing how the proton's momentum is shared amongst the different parton constituents, as a function of the fraction x of the proton's momentum carried by the parton [1341].

Since quarks and gluons have nonzero spin, the fundamental distributions are the PDFs for a specific helicity (spin projection along the direction of motion), $f_i^{\uparrow}$ and $f_i^{\downarrow}$, corresponding to parton spins aligned and antialigned with the proton spin, respectively. Unpolarized scattering experiments are therefore only sensitive to sums of the helicity PDFs, $f_i = f_i^{\uparrow} + f_i^{\downarrow}$, while measurements involving polarized beams and/or targets are required to obtain information on differences, $\Delta f_i = f_i^{\uparrow} - f_i^{\downarrow}$.

Traditionally, PDFs have been determined in global QCD analyses by simultaneously fitting a wide variety of data for large momentum transfer processes. Typically, the PDFs are parametrized in terms of some functional form, the parameters of which are determined by fitting the calculated cross sections to data. Once the PDFs are determined at some initial momentum transfer scale, the DGLAP Q^2 evolution equations (see Sect. 2.3) are used to compute them at all other scales needed for the calculations. The standard data sets used in global analyses include deep-inelastic scattering (DIS) of charged leptons from proton or nuclear targets (or neutrinos from heavy nuclei), Drell–Yan (DY) inclusive lepton-pair production in hadron–hadron scattering, and the production of photons, $W^{\pm}$ or Z bosons, or jets at large transverse momentum in hadronic collisions (see Sect. 10.2.2). We discuss the specific reactions and relevant data sets in more detail in the following.

10.2.2 Physical processes and experimental observables

Historically, the main source of information on proton PDFs has been the DIS of leptons from protons or nuclei, starting from the pioneering experiments at SLAC in the late 1960s. In the one-boson exchange approximation, the differential DIS cross section can be written as a product of leptonic and hadronic tensors,

$$\frac{d^2\sigma}{d\Omega dE'} \sim \alpha\, L^{\mu\nu}\, W_{\mu\nu},$$

where α is the fine structure constant, $\Omega = \Omega(\theta,\phi)$ is the laboratory solid angle of the scattered lepton, and E' is the scattered lepton energy. Using constraints from Lorentz and gauge invariance, the hadronic tensor $W_{\mu\nu}$ can be decomposed into several independent terms,

$$W_{\mu\nu} = -\widetilde{g}_{\mu\nu}\,F_1 + \frac{\widetilde{p}_\mu \widetilde{p}_\nu}{p\cdot q}\,F_2 + i\epsilon_{\mu\nu\alpha\beta}\,p^\alpha q^\beta\,F_3 + i\epsilon_{\mu\nu\alpha\beta}\frac{q^\alpha}{p\cdot q}\Big[s^\beta\,g_1 + \Big(s^\beta - \frac{s\cdot q}{p\cdot q}\,p^\beta\Big)g_2\Big], \quad (10.20)$$

where p_μ and q_μ are the nucleon and exchanged boson four-momenta, $\widetilde{g}_{\mu\nu} = g_{\mu\nu} - q_\mu q_\nu/q^2$, and $\widetilde{p}_\mu = p_\mu - (p\cdot q/q^2)q_\mu$. The nucleon polarization four-vector s^β satisfies $s^2 = -1$ and $p\cdot s = 0$. The structure functions $F_{1,2,3}$ and $g_{1,2}$ contain the complete information about the structure of the nucleon in DIS, and are generally functions of two variables, conventionally chosen to be the Bjorken scaling variable $x = Q^2/2p\cdot q$ and the exchanged boson virtuality Q^2. In the Bjorken limit, in which both Q^2 and $p\cdot q \to \infty$ (or invariant final state hadron mass $W^2 = (p+q)^2 =$

$M^2 + Q^2(1-x)/x \to \infty$), but x is fixed, the structure functions become simple functions of x only.

Unpolarized scattering

For spin-averaged scattering, the nucleon structure is parametrized in terms of the vector F_1 and F_2 structure functions, and the vector-axial vector interference F_3 structure function, which requires weak currents. According to QCD factorization theorems, the structure functions F_j ($j = 1, 2, 3$), can be written in factorized form as convolutions of hard coefficient functions and PDFs, weighted by respective electroweak charges,

$$F_j(x, Q^2) = \sum_{i=q,\bar{q},g} e_i^2 \left[C_i^j \otimes f_i\right](x, Q^2), \tag{10.21}$$

where the convolution symbol is defined by $[A \otimes B](x) = \int_x^1 (dy/y) A(x) B(x/y)$. The coefficient functions C_j^i can be computed perturbatively in a series in α_s. At leading order (LO) in α_s, C_j^i is a δ function, and the structure functions reduce to linear combinations of the PDFs,

$$F_1(x) = \frac{1}{2} \sum_q e_q^2 \, q^+(x), \tag{10.22a}$$

$$F_2(x) = 2x F_1(x), \tag{10.22b}$$

$$F_3(x) = 2 \sum_q g_V^q \, g_A^q \, q^-(x), \tag{10.22c}$$

where $q^\pm = q \pm \bar{q}$ denote the C-even (odd) flavor combinations, and we use the short-hand notation $q(x) \equiv f_q(x)$ or $\bar{q}(x) \equiv f_{\bar{q}}(x)$ for a quark or antiquark PDF of flavor q in the proton, and $g(x) \equiv f_g(x)$ for the gluon PDF. The F_3 structure function vanishes for photon exchange, but is nonzero for the exchange of weak bosons, with $g_{V(A)}^q$ representing the vector (axial vector) coupling of the boson to the quark q. Equation (10.22) correspond to the simple parton model of inclusive DIS, in which the structure functions are interpreted as parton densities. At finite energies, the logarithmic Q^2 dependence from the evolution equations described in Sect. 2.3, as well as residual Q^2 dependence associated with power corrections (see below), give corrections to the simple parton model expectations. (Note also that at LO the Bjorken x variable coincides with the parton momentum fraction; however, at higher orders these are different.)

Many DIS experiments have been performed with charged lepton beams on proton targets, which for neutral currents in the one-photon exchange approximation constrain the flavor combination $4u^+ + d^+ + s^+$ (Z boson exchange would involve a different linear combination of PDFs, involving the weak mixing angle, $\sin^2\theta_W$). For a neutron, the corresponding linear combination would be $4d^+ + u^+ + s^+$. In practice, free neutron targets do not exist, so deuterium is often used as a proxy, which then requires nuclear corrections be made to extract the free neutron structure information (see Sect. 10.2.3).

Charged current neutrino and antineutrino interactions constrain different combinations of q^+ or q^- PDFs for the $F_{1,2}$ or F_3 structure functions, respectively, depending on the type of target used, so that by combining data on different targets and with different beams one can in principle isolate specific combinations of q or $\bar{q}$. A special case is provided by charm production in ν and $\bar{\nu}$ DIS, which is sensitive to the s and $\bar{s}$ PDFs, respectively (although in practice this involves heavy targets for which model-dependent nuclear corrections must be made).

The gluon PDF plays a lesser role in inclusive DIS, as it enters the cross sections at higher order, $\mathcal{O}(\alpha_s)$. In practice, it is mainly constrained through the Q^2 dependence of the structure functions, and the longitudinal structure function F_L, which depends on differences at higher order between the left- and right-hand sides of Eq. (10.22b). The strongest constraints on $g(x)$ in DIS have come from the HERA ep collider data at very small x values [3059].

Since the PDFs are universal, the functions appearing in the DIS structure functions are the same as those that describe the structure of the incoming hadrons in hadronic collisions. In analogy with the QCD factorization for DIS, the cross section for the high energy scattering of hadron A (momentum p_A) and hadron B (momentum p_B) to an inclusive state in which a particle C is identified (such as a vector boson, photon, or jet) can generally be written as

$$\sigma_{AB\to CX}(p_A, p_B) = \sum_{a,b} \int dx_a \, dx_b \, f_{a/A}(x_a) \, f_{b/B}(x_b) \times \hat{\sigma}_{ab\to CX}(x_a p_A, x_b p_B), \tag{10.23}$$

where x_a and x_b are the corresponding parton momentum fractions, and $\hat{\sigma}_{ab\to CX}$ is the partonic cross section. For vector boson production ($W^\pm$, Z^0, or lepton pairs produced from virtual photons), the process proceeds at LO through $q\bar{q}$ annihilation. In particular, the Drell–Yan lepton-pair production cross sections in pp and pn collisions depend on the combinations

$$\sigma^{pp} \sim 4u(x_a)\,\bar{u}(x_b) + d(x_a)\,\bar{d}(x_b) + (x_a \leftrightarrow x_b) + \cdots$$
$$\sigma^{pn} \sim 4d(x_a)\,\bar{d}(x_b) + u(x_a)\,\bar{u}(x_b) + (x_a \leftrightarrow x_b) + \cdots$$

where the ellipses indicate contributions from heavier quarks. (The pn cross section is again obtained from deuterium data.) As we discuss below, ratios of these cross sections at kinematics such that $x_a \gg x_b$, where the hadron A can be approximated by its valence structure, can be used to constrain the $\bar{d}/\bar{u}$ ratio. In contrast, the inclusive production of $W^\pm$ bosons constrains products of the form $q(x_a)\,\bar{q}'(x_b)$ with specific weights given by the appropriate CKM matrix elements.

For $p\bar{p}$ collisions at the Tevatron, for example, at large values of rapidity (very asymmetric values of x_a and x_b) at

LO one has

$$\sigma^{W^+} \sim u(x_a)\, d(x_b) + \bar{d}(x_a)\, \bar{u}(x_b) + \cdots$$
$$\sigma^{W^-} \sim d(x_a)\, u(x_b) + \bar{u}(x_a)\, \bar{d}(x_b) + \cdots$$

where the PDFs in the antiproton have been related to those in the proton. For large and positive rapidity, $x_a > x_b$ and the antiquark PDFs can be neglected, so that these cross sections depend only on the u and d PDFs. Because of the missing neutrino resulting from the W decays, one cannot directly reconstruct the rapidity distributions, and typically the charged lepton rapidity asymmetry for $W^{\pm}$ production is presented. The decay process means that the constraints on the PDFs are less direct, but such measurements still provide useful constraints on the d/u ratio at moderate values of x.

Recent data from the ATLAS Collaboration at the LHC [3060,3061] on $W^{\pm}$ and Z production and decay suggested a rather larger strange quark sea than traditionally obtained from neutrino scattering, with the ratio $R_s = (s + \bar{s})/(\bar{u} + \bar{d}) \approx 1.13$ at parton momentum fraction $x \approx 0.02$, compared with the traditionally accepted value of $R_s \approx 0.4$ from neutrino scattering analyses. In contrast, a simultaneous analysis of PDFs and fragmentation functions including semi-inclusive $\pi^{\pm}$ and $K^{\pm}$ production data, along with single-inclusive e^+e^- annihilation cross sections into hadrons [665,3062], favored a strong suppression of the strange PDF at intermediate x values, correlating with an enhancement of the $s \to K^-$ fragmentation function. The question of the magnitude and shape of the strange (and antistrange) PDF remains a topic of considerable phenomenological interest.

Other observables that can constrain PDFs are inclusive jet or photon, dijet, and photon + jet production cross sections. Generally, these have greater sensitivity to the gluon PDF at large x than DIS reactions. Dijet production triple differential cross sections yield more information than single jet cross sections because the rapidity of the second jet is also constrained, thereby helping to constrain the momentum fractions of the PDFs. Direct or isolated photon production can also constrain the gluon PDF through the subprocess $qg \to \gamma q$ [3063]. Photon + jet production offers similar constraints, but now the subprocesses are weighted by the squared charge of the parton to which the photon couples. A summary of the kinematic coverage of the existing datasets used to constraint unpolarized PDFs is shown in Fig. 251.

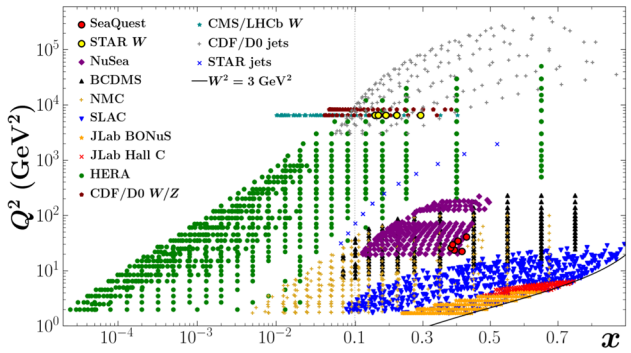


Fig. 251 Kinematic coverage of datasets used in global QCD analyses. The variable x represents Bjorken-x for DIS and Feynman-x for vector boson and jet production, while the scale Q^2 represents the four-momentum transfer squared for DIS, the mass squared of the intermediate boson for vector boson production, and the transverse momentum squared for jet production. A DIS cut of $W^2 = 3$ GeV^2 is indicated in the bottom right hand corner (solid back line)

Polarized scattering

For spin-dependent reactions, the structure functions g_1 and g_2 are extracted from DIS measurements with longitudinally polarized leptons scattered from a nucleon or nucleus that is polarized either longitudinally or transversely relative to the beam. For longitudinal beam and target polarization, the difference between the cross sections for spins aligned and antialigned is dominated by the g_1 structure function, while the g_2 structure function requires measurements with the target polarized transversely to the beam polarization. In practice one often measures the polarization asymmetry A_1, which is given as a ratio of spin-dependent and spin-averaged structure functions,

$$A_1 = \frac{1}{F_1(x)}\left[g_1(x) - \frac{4M^2x^2}{Q^2} g_2(x)\right], \tag{10.24}$$

where M is the nucleon mass. At small values of x^2/Q^2, the asymmetry simplifies to $A_1 \approx g_1/F_1$.

In analogy with the unpolarized F_1 structure function, the structure function g_1 can be expressed at LO in terms of differences between quark distributions with spins aligned and antialigned with that of the nucleon,

$$g_1(x) = \frac{1}{2}\sum_q e_q^2\, \Delta q^+(x). \tag{10.25}$$

The g_2 structure function, on the other hand, does not have a simple partonic interpretation. However, its measurement provides information on the subleading, higher-twist contributions which parametrize long-range multi-parton correlations in the nucleon. The dependence on both spin-dependent and spin-averaged structure functions in A_1 illustrates the need to consistently analyze both unpolarized and polarized PDFs simultaneously, as will be discussed below.

As with unpolarized measurements, historically most constraints on spin-dependent PDFs have come from polarized charged-lepton DIS experiments. For charged lepton scattering from polarized proton targets, the g_1 structure function depends on the combination $4\Delta u^+ + \Delta d^+ + \Delta s^+$, while for the neutron the combination would be $4\Delta d^+ + \Delta u^+ + \Delta s^+$. In practice, polarized ^{3}He targets are usually used as effective sources of polarized neutron, since the neutron carries almost 90% of the ^{3}He spin, while polarized deuterons, which have

equal proton and neutron spin contributions, are sensitive to the isoscalar combination $5(\Delta u^+ + \Delta d^+) + 2\Delta s^+$.

At next-to-leading order (NLO), the polarized gluon distribution Δg also enters in the g_1 structure function. The mixing with the quark flavor singlet contribution to g_1 under Q^2 evolution can then be used to provide constraints on Δg.

Semi-inclusive DIS (SIDIS) provides additional independent combinations of spin-dependent PDFs that can be used to separate individual quark and antiquark flavors. At high energies, production of hadrons h in the current fragmentation region, primarily pions or kaons, is proportional to products of PDFs and quark $\rightarrow$ hadron fragmentation functions. Typically, such experiments measure the semi-inclusive polarization asymmetry, which at LO can be written as a ratio of spin-dependent to spin-averaged SIDIS cross sections,

$$A_1^h(x,z) = \frac{\sum_q e_q^2 \left(\Delta q(x)\, D_q^h(z) + \Delta\bar{q}(x)\, D_{\bar{q}}^h(z)\right)}{\sum_q e_q^2 \left(q(x)\, D_q^h(z) + \bar{q}(x)\, D_{\bar{q}}^h(z)\right)}, \tag{10.26}$$

where $D_q^h(z)$ is the fragmentation function for the scattered quark to produce a hadron h with a fraction z of the quark's energy. For large z, the produced hadron has a high probability of containing the scattered parton, providing a tag on the initial parton distribution.

The fragmentation functions D_q^h can be determined from other reactions, such as inclusive single hadron production in e^+e^- annihilation or pp collisions. One can then weight particular quark or antiquark flavors by selecting favored (such as $D_u^{\pi^+}$ or $D_{\bar{d}}^{\pi^+}$) or unfavored ($D_{\bar{d}}^{\pi^+}$ or $D_{\bar{u}}^{\pi^+}$) fragmentation functions for specific hadrons. The polarized strange quark PDF Δs, in particular, can be constrained from SIDIS K production data. The polarized gluon distribution Δg can also be constrained from SIDIS data on charmed or high-p_T hadron production through the photon–gluon fusion process.

Inclusive particle production in polarized proton–proton collisions provides an additional method of determining spin-dependent sea quark and gluon PDFs. The cross sections for the production of $W^\pm$ bosons in scattering longitudinally polarized protons from unpolarized protons, $\vec{p}\, p \rightarrow W^\pm X$, depend on products of spin-dependent and spin-averaged PDFs,

$$\Delta\sigma^{W^+} \sim \Delta\bar{d}(x_a)\, u(x_b) - \Delta u(x_a)\, \bar{d}(x_b),$$
$$\Delta\sigma^{W^-} \sim \Delta\bar{u}(x_a)\, d(x_b) - \Delta d(x_a)\, \bar{u}(x_b).$$

At large positive (negative) rapidities, $x_a \gg x_b$ ($x_a \ll x_b$), the cross sections are dominated by a single flavor, while at mid-rapidities both u and d flavors contribute.

Inclusive jet (or π^0) production in double-polarized proton–proton scattering, $\vec{p}\,\vec{p} \rightarrow$ jet (or π^0) $+ X$, is sensitive to the polarized gluon PDF. The first evidence for a small, but nonzero Δg was observed by the STAR Collaboration at RHIC in jet data at $\sqrt{s} = 200$ GeV, although recent Monte Carlo analysis [976] suggests that the sign of Δg is not unambiguously determined by these data. A summary of the kinematic coverage of the existing datasets used to constraint helicity PDFs is shown in Fig. 252.

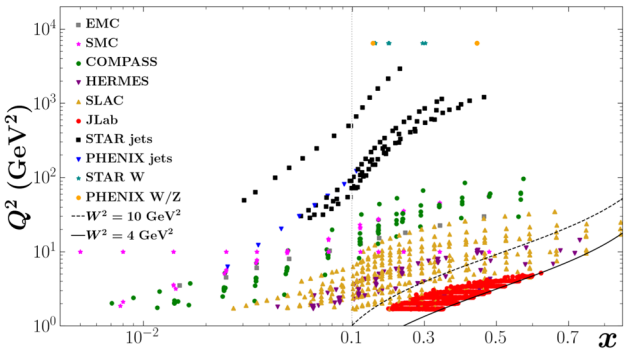


Fig. 252 As in Fig. 251 but for spin-dependent observables

10.2.3 Global QCD analysis

With the growing number of high energy scattering experiments in the 1970s and 1980s came the need to systematically and uniformly analyze the data with the tools that were being developed in perturbative QCD. The concept of fitting datasets from various experiments globally with a single set of quark, antiquark and gluon PDFs dates back to the early analyses of Duke and Owens [3064] and Morfin and Tung [3065]. Since then, a number of dedicated efforts have been made worldwide to fit both unpolarized and polarized scattering experiments in terms of spin-averaged and spin-dependent PDFs.

The standard paradigm has been to parametrize the PDFs at some input scale Q_0 and then evolve using the appropriate evolution equations to the scales needed for the calculation of each experimental observable. The parameters of the PDFs are estimated by comparing each calculated observable with the data using χ^2 minimization techniques. All of the global PDF analysis groups use some variation of this approach, although the details of the implementation differ between different groups.

PDF parametrizations and constraints

A typical parametrization at the input scale Q_0 for a generic (unpolarized or polarized) PDF f is

$$xf(x, Q_0^2) = a_0\, x^{a_1} (1-x)^{a_2}\, P(x), \tag{10.27}$$

where $P(x)$ represents a smoothly varying function, such as a polynomial in x or $\sqrt{x}$, or more elaborate forms based on neural networks [3066] or self-organizing maps [3067]. Some of the parameters in the input distributions can be determined from physical constraints. For example, in the unpolarized case the conservation of valence quark number gives for the

first moments

$$\int_0^1 dx\, u^-(x, Q_0^2) = 2, \quad \int_0^1 dx\, d^-(x, Q_0^2) = 1, \qquad (10.28)$$

and zero for all other flavors, while the momentum sum rule requires

$$\int_0^1 dx\, x\Big[\sum_q^{n_f} q^+(x, Q_0^2) + g(x, Q_0^2)\Big] = 1, \qquad (10.29)$$

where the number of flavors at the input scale Q_0^2 is usually taken to be $n_f = 3$.

In the polarized case the first moments of the C-even distributions can be related to octet baryon weak decay constants. For the isovector combination, corresponding to the Bjorken sum rule,

$$\int_0^1 dx\, (\Delta u^+ - \Delta d^+)(x, Q_0^2) = g_A, \qquad (10.30)$$

where $g_A = 1.270(3)$ is the nucleon axial charge, while for the SU(3) octet one has

$$\int_0^1 dx\, (\Delta u^+ + \Delta d^+ - 2\Delta s^+)(x, Q_0^2) = a_8, \qquad (10.31)$$

where the octet axial charge $a_8 = 0.58(3)$ is extracted from hyperon β-decays assuming SU(3) flavor symmetry [3068]. Note that the sum rules (10.28)–(10.31) are preserved under Q^2 evolution.

Power corrections

We should note that the theoretical results summarized above have been obtained within the framework of perturbative QCD in the limit when both Q^2 and W are much larger than all hadron mass scales, $Q^2, W^2 \gg M^2$, where the cross sections are dominated by their leading twist contributions. In actual experiments performed at finite beam energy E, the maximum values of Q^2 and W are limited, which restricts the available coverage in Bjorken x. This is especially relevant at large x in DIS, where for fixed Q^2, as $x \to 1$ the final state hadron mass W decreases as one descends into the region dominated by nucleon resonances at $W \lesssim 2$ GeV. The resonance region may be treated using the concept of quark–hadron duality [3069], although this goes beyond the scope of the usual perturbative QCD analysis.

In the low-Q^2 region, power corrections to the Bjorken limit results that scale as powers of $\Lambda^2_{\rm QCD}/Q^2$ become increasingly important. In the operator product expansion, these are associated with matrix elements of higher twist operators, associated with multi-parton correlations which characterize the long-range nonperturbative interactions between quarks and gluons. While providing glimpses into the dynamics of quark confinement, the power corrections are viewed as unwelcome backgrounds in efforts aimed solely at extracting leading twist PDFs. Other subleading corrections are associated with target mass corrections (TMCs), which are of kinematical origin and arise from nonzero values of hadron masses [3070–3074].

Regardless of their origin, the various power suppressed corrections to the leading twist results can be absorbed into phenomenological functions, such as

$$F_i(x, Q^2) = F_i^{\rm LT}(x, Q^2) + \frac{h_i(x)}{Q^2} + \cdots, \qquad (10.32)$$

for an unpolarized structure function F_i, for example, where $F_i^{\rm LT}$ denotes the leading twist contribution. The higher twist corrections are sometimes assumed to be multiplicative, with the functions h_i proportional to the leading twist contribution. Possible additional Q^2 dependence of the higher twist contributions, such as from radiative $\alpha_s(Q^2)$ corrections, is usually neglected.

Nuclear corrections

Since nucleons bound in a nucleus are not free, the parton distributions $f_{i/A}$ in a nucleus A deviate from a simple sum of PDFs in the free proton and neutron, $f_{i/A} \neq Z f_{i/p} + (A-Z) f_{i/n}$, where Z is the number of protons. This is especially relevant at small values of x, where nuclear shadowing effects suppress the nuclear to free isoscalar nucleon (N) ratio, $f_{i/A}/(A f_{i/N}) < 1$, and at large x, where the effects of Fermi motion, nuclear binding, and nucleon off-shellness give rise to the "nuclear EMC effect" [3075–3077]. For spin-dependent PDFs, the different polarizations of the bound nucleons and nuclei also need to be taken into account.

In the nuclear impulse approximation, where scattering takes place incoherently from partons inside individual nucleons, the PDF in a nucleus can be expressed as a convolution of the PDF in a bound nucleon and a momentum distribution function $f_{N/A}$ of nucleons in the nucleus [3078–3080]. The momentum distribution, or "smearing function", can be computed from nuclear wave functions, incorporating nuclear binding and Fermi motion effects. Coherent rescattering effects involving partons in two or more nucleons give rise to nuclear shadowing corrections to the impulse approximation, and such effects are typically important only in the small-x region. In general, the relation between PDFs in a nucleus and in a nucleon can be written as

$$f_{i/A} = \sum_{N=p,n} \big[f_{N/A} \otimes f_{i/N}\big] + \delta^{(\rm off)} f_{i/A} + \delta^{(\rm shad)} f_{i/A}, \qquad (10.33)$$

where the term $\delta^{(\rm off)} f_{i/A}$ represents nucleon off-shell or relativistic corrections that account for modification of the parton structure of the nucleon in the nuclear medium. A similar expression can be written for spin-dependent PDFs.

At large Q^2 the smearing function has a probabilistic interpretation in terms of the light-cone momentum fraction y of

the nucleus carried by the struck nucleon. Typically, the function $f_{N/A}$ is steeply peaked around $y \approx 1$, becoming broader with increasing mass number A as the effects of binding and Fermi motion become more important. In the limit of zero binding, $f_{N/A}(y) \to \delta(1-y)$, and one recovers the free nucleon case. This assumption has often been made in global PDF analyses. More recently, however, the important role of nuclear corrections has been more widely appreciated, especially in connection with extractions of the free neutron structure function data from measurements involving deuterons and other light nuclei [3080–3086].

For neutrino scattering, to increase the relatively low rates and obtain sufficient statistics for analyses such as strange PDF extraction [3087,3088], experiments have usually resorted to using heavier nuclear targets, such as iron or lead. Extractions from such data are complicated by the presence of nuclear corrections in neutrino structure functions [3089,3090], as well as effects of the nuclear medium on the charm quark propagation in the final state [3091].

For spin-dependent scattering, the scarcity of data and larger uncertainties at small x and at high x, where nuclear corrections are most prominent, has meant that most global analyses have relied exclusively on the effective polarization *ansatz*, in which the polarized PDF in the nucleus $\Delta f_{i/A}$ is related to the polarized PDFs in the proton and neutron as $\Delta f_{i/A} \approx \langle\sigma\rangle^p \, \Delta f_{i/p} + \langle\sigma\rangle^n \, \Delta f_{i/n}$, where $\langle\sigma\rangle^{p(n)}$ is the average polarization of the proton (neutron) in the nucleus. In practice, along with polarized protons, only polarized deuterium and ^{3}He nuclei have been used in DIS experiments. As experiments at high luminosity facilities such as Jefferson Lab at 12 GeV push to explore the higher-x region, nuclear corrections from smearing and off-shell effects will become more relevant.

Uncertainty quantification

There are several sources of PDF uncertainties that enter in global QCD analyses. These include uncertainties on the experimental data, the approximations used in computing the partonic cross sections, and the parametrizations used to describe the PDFs. The experimental errors on the data can be directly propagated to the fitted PDFs. The most common method for implementing this is the Hessian method, described in Ref. [3092]. The elements of the Hessian matrix are given by partial derivatives of the χ^2 function,

$$H_{ij} = \frac{1}{2}\frac{\partial^2\chi^2}{\partial a_i \, \partial a_j}, \tag{10.34}$$

where a_i denotes the i^{th} PDF parameter. The Hessian matrix is generated during the minimization procedure and its inverse gives the error matrix. The eigenvectors of the error matrix can then be used to define eigenvector parameter sets, from which the error bands for the PDFs or for specific processes are calculated. An important point to note is that the error bands generally depend on a χ^2 tolerance. Mathematically, the expectation is that the 1σ parameter errors correspond to an increase of χ^2 by one unit from the minimum value, $\Delta\chi^2 = 1$. However, it has been suggested [3093] that inconsistencies between different data sets should be handled by introducing a larger value to be used, $\Delta\chi^2 > 1$. This "χ^2 tolerance" varies between groups ($\Delta\chi^2 \sim 10-100$), and allowance must be made for this when comparing the resulting error bands.

On the other hand, it has been argued [3094] that the tolerance criterion effectively changes the likelihood function, which is usually defined in terms of the χ^2 function. In contrast, neural network based approaches suggest that the use of a tolerance criterion is not necessary [3066,3095–3097]. In practice, the similar size of the uncertainties obtained in such different approaches may be coincidental and due to the likelihood deformation and resulting uncertainty inflation, as observed in a recent comparative study using toy data [3094]. Furthermore, concern has also been expressed [3098] that a meta-analysis, such as PDF4LHC [3099], that combines existing PDFs from different groups may obscure the fundamental connection between experimental data and theory and hide the true meaning of the uncertainties, if these ultimately originate from different choices of the likelihood function.

An alternative to the usual linear propagation of errors in the Hessian method which avoids ambiguities associated with tolerance criteria, and which is useful for minima that are not well behaved or defined, is the Monte Carlo method. To propagate the experimental errors a number of replica data sets are randomly generated within the original errors, and these replica sets are then fitted with the resulting replica PDF sets treated using standard statistics [3096]. The central values are computed as the averages over replicas, while the uncertainties are given by the envelope of predictions.

In practice, the data resampling method has been used by the NNPDF [975,3100] and JAM [665,976,3086] collaborations, although these groups differ in their approach to PDF parametrizations. While the JAM collaboration uses a traditional polynomial functional form for the function $P(x)$ in Eq. (10.27), the NNPDF group implements a similar basic parametric form that is supplemented by a series of trained neural network weights. The dependence on the functional form for the PDF can be minimized by choosing a flexible parametrization with parameters that are well-constrained by data. Outside of kinematic regions covered by data, the PDFs are not constrained, and care must be taken when using them in extrapolated regions at small or large x.

The approximations made in computing partonic cross sections naturally introduce uncertainties in PDFs, although these can be rather difficult to quantify reliably. One of these is the uncertainty arising from the truncation of the perturbative series. These can be estimated to some extent by compar-

ing LO, NLO, NNLO, and recently even approximate N^{3}LO [3101] fits, although not all processes are known to the same accuracy. The topic of "missing higher order uncertainties" and how to estimate them has in fact attracted some attention recently in global PDF fitting efforts [3057].

Perturbative QCD calculations also depend to some degree on the choices made for the renormalization and factorization scales for each physical process. The choices will change the results for different processes, and the fitted PDFs must compensate these changes. A closely related issue is the choice of the strong running coupling $\alpha_s(M_Z)$, which is fitted together with the PDF parameters in some analyses, and fixed to the global average in others. Finally, the choice of data sets and kinematic cuts can of course affect the extracted PDFs, and these choices and the reasons for them need to be assessed when drawing conclusions from PDF comparisons.

10.2.4 Spin-averaged PDFs

Using the technology outlined in the previous sections, a number of global QCD analyses efforts have produced sets of unpolarized proton PDFs, with groups in Europe and the the US at the forefront of the data analyses. The European groups include the UK-based MSHT [1107] group and the ABM [3102] group, which use standard global fitting methodology; the HERAPDF [3059] analysis, which includes only data from the H1 and ZEUS experiments at HERA; and previously the Dortmund [3103] group, which pioneered the approach of dynamically generating PDFs through Q^2 evolution from a low input scale. More recently, the NNPDF [3100] collaboration introduced an approach based on neural networks.

US-based efforts have centered around the CTEQ collaboration, which involves two derivative analyses of nucleon PDFs, by the CT (CTEQ-Tung et al.) [663] and CJ (CTEQ-Jefferson Lab) [3084] groups. The former focuses more on LHC-related phenomenology, while the latter has developed methodologies needed for describing data over a broad energy range including the low-Q^2 and W domain. The Jefferson Lab-based JAM [3086,3104] collaboration uses a Monte Carlo approach with simultaneous determination of PDFs and other types of distributions, such as fragmentation functions and spin-dependent PDFs. In the following we illustrate the current state of knowledge of the spin-averaged proton PDFs, including the u and d valence quark distributions and the flavor structure of the proton sea.

Valence quark distributions

Valence quarks give the global properties of the nucleon, such as its baryon number and charge. Knowledge of their momentum distributions is important for many reasons, especially at high values of x, where a single quark carries most of the nucleon's momentum. The large-x region is a unique testing ground, for example, for various nonperturbative models of the nucleon [1088,1341,3081,3105,3106]. Reliable determination of PDFs at large x is also important for searches for new physics beyond the Standard Model in collider experiments at the LHC [3098,3107].

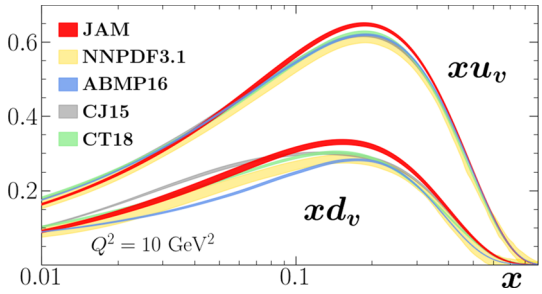


Fig. 253 Valence u and d quark PDFs versus x from several global QCD analyses: JAM21 [3086], NNPDF [664], ABMP [3102], CJ15 [3084], and CT18 [663] at a scale $Q^2 = 10$ GeV^2

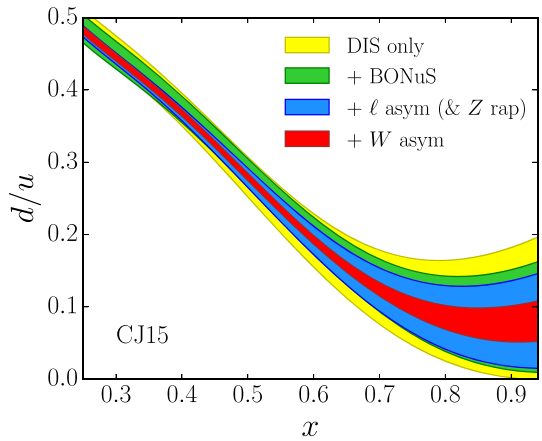


Fig. 254 Impact of various data sets on the d/u ratio at $Q^2 = 10\,\text{GeV}^2$, using the CJ15 PDFs set [3084]

The valence u and d PDFs are illustrated in Fig. 253 from several PDF groups. The u quark PDF is fairly well constrained (due to its larger charge) by the relatively abundant proton DIS data that have been collected over several decades at SLAC, CERN, DESY and Jefferson Lab. The d quark distribution, on the other hand, relies in addition on neutron structure functions, whose determination requires both proton and deuteron DIS data. Studies of nuclear effects in the deuteron suggest that the uncertainties related to nucleon interactions increase significantly at large x [3081], leading to large uncertainties in the d/u PDF ratio for $x \gtrsim 0.6$, as Fig. 254 illustrates. Inclusion of tagged deuteron data from the BONuS experiment at Jefferson Lab [3108,3109], and in particular the lepton and W boson asymmetry data from $p\bar{p}$ collisions at the Tevatron [3110–3112], reduces the uncertainty considerably in the experimentally constrained region up to $x \sim 0.8$.

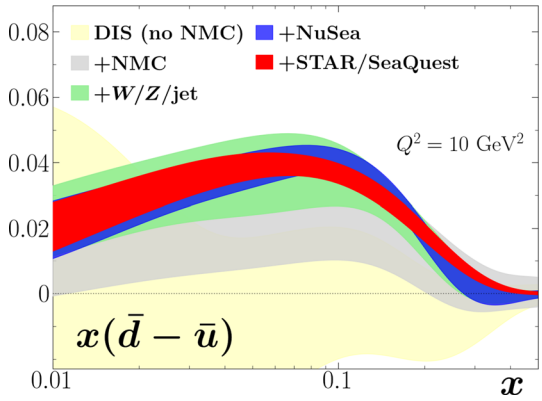


Fig. 255 Comparison of $x(\bar{d} - \bar{u})$ with different combinations of datasets [3104]: DIS only, excluding NMC (gold band); with NMC (gray); with W, Z, and jet production from RHIC, Tevatron, and the LHC (green); with NuSea (blue); and finally with the SeaQuest DY and STAR W-lepton ratio (red)

Light quark sea

Because inclusive DIS measures only C-even combinations of PDFs, q^+, to disentangle quark from antiquark contributions requires other types of observables, such as the DY cross sections, where the q and $\bar{q}$ PDFs are weighted differently. As discussed in Sect. 10.2.2, ratios of pd to pp cross sections at $x_a \gg x_b$ are directly sensitive to the ratio $\bar{d}/\bar{u}$. The flavor asymmetry $\bar{d} - \bar{u}$ is illustrated in Fig. 255, which shows the impact of various data sets. Starting with inclusive DIS data only and excluding data from the NMC experiment, the asymmetry is consistent with zero within large uncertainties. Including the NMC data [3113,3114], the asymmetry gives an indication of deviation from zero in the range $0.01 < x < 0.2$. When W-lepton, reconstructed W and Z boson, and jet production data from RHIC, Tevatron, and LHC are further included (but not the new STAR data [3115]), the asymmetry becomes significantly larger, and more distinguishable from zero below $x = 0.3$.

The new constraints come primarily from the high-precision W asymmetry measurements from the Tevatron and LHC, which are sensitive to $\bar{u}$ and $\bar{d}$. The further addition of the NuSea DY data [3116] greatly decreases the uncertainty, showing that these data provide a strong constraint on the asymmetry even when compared to the Tevatron and LHC W-lepton asymmetries. Finally, the inclusion of the new SeaQuest [3117] and STAR [3115] data reduces the uncertainty on the asymmetry even further, while increasing the magnitude at $x \gtrsim 0.2$. The behavior of the asymmetry seen in Fig. 255 is consistent with expectations from nonperturbative models of the nucleon in which the excess of $\bar{d}$ over $\bar{u}$ in the proton sea has been that associated with chiral symmetry breaking, and the consequent prevalence of the virtual $p \to n\pi^+$ dissociation [3118–3120].

Strange quarks

The strange quark distribution has generally been more difficult to determine experimentally than the nonstrange sea. While the size of the strange to nonstrange ratio R_s has been controversial, with values ranging from $R_s \approx 0.4$ from neutrino DIS at $x \approx 0.02$ to $R_s \approx 1$ from analyses that included ATLAS data on W/Z production [3060,3061], an independent and underutilized source of information at lower energies is semi-inclusive production of pions or kaons. Analysis of SIDIS data has often been complicated by the need to know both the PDFs of the initial state and the fragmentation functions describing hadronization to the final state, as assumptions about the latter can lead to significant differences in the extracted PDFs [3121,3122]. For any definitive conclusion a combined analysis of PDFs and fragmentation functions is necessary, which was first performed by the JAM group [665,3062].

Including data from the standard datasets used for unpolarized PDFs, along with SIDIS multiplicities and e^+e^- annihilation data to constrain the fragmentation functions [3123], the most striking result of the simultaneous JAM fit was a significantly reduced strange quark PDF compared with that reported by ATLAS. In particular, the strange to nonstrange ratio was found in the JAM global QCD analysis to be $R_s \approx 0.2-0.3$ at $x \sim 0.02$ [3123], as Fig. 256 illustrates, in contrast to values of $R_s \sim 1$ inferred from the ATLAS data [3060,3061]. The most significant source of the strange suppression is the SIDIS and SIA K production data. Without these data, the s^+ PDF is poorly constrained, in contrast to the light flavor sea, which is not strongly affected by the SIDIS multiplicities. Consequently, while the ratio R_s varies over a large range without SIDIS (and SIA) data, and at low x is compatible with $R_s \approx 1$, once those data are included its spread becomes dramatically reduced.

Note that while the original ATLAS PDF fit [3060,3061] was consistent with $R_s \sim 1$ for all $x \gtrsim 10^{-3}$, the more recent ATLAS analysis [3124] including new W/Z + jet data gives a strongly suppressed R_s at $x \gtrsim 0.01$ with significantly smaller uncertainties, and more in line with other determinations. The ratio at smaller x ($x \approx 0.023$) remains unsuppressed, however, consistent with the earlier ATLAS results and higher than most PDF parametrizations [3124].

The SIDIS $K^\pm$ production data could also in principle discriminate between the s and $\bar{s}$ PDFs, which could have different x dependence [3125–3130]. As shown in Fig. 256, however, the current data do not indicate any significant $s - \bar{s}$ asymmetry within uncertainties. Future high-precision SIDIS data from Jefferson Lab or the Electron-Ion Collider may allow more stringent determinations of the s and $\bar{s}$ PDFs [3131], as would inclusion of W + charm production data from the LHC [3132,3133].

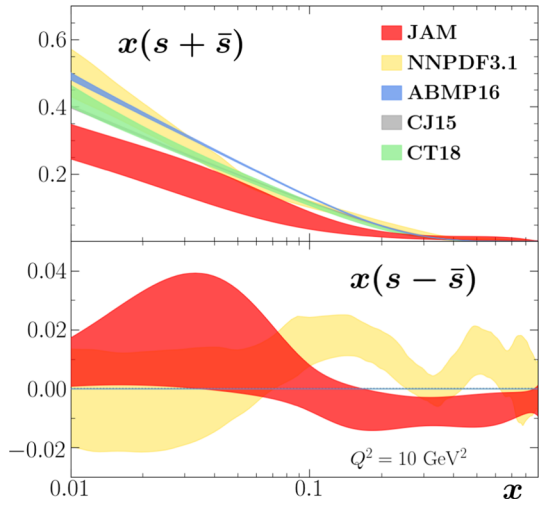


Fig. 256 Sum and difference of the s and $\bar{s}$ PDFs from several global QCD analyses, as in Fig. 253

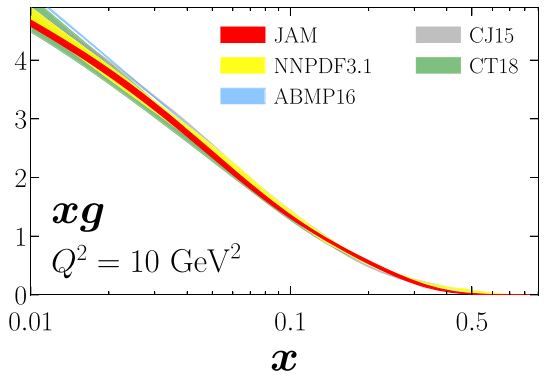


Fig. 257 Unpolarized gluon PDF xg from various QCD global analyses at a scale of $Q^2 = 10$ GeV^2 from several global QCD analyses, as in Fig. 253

Gluons and heavy quarks

Gluons play an important role in the study of nucleon structure, contributing some 50% of the nucleon's overall (linear) momentum, and indirectly provide some constraints on quark PDFs through the momentum sum rule, Eq. (10.29). Since photons do not couple directly to gluons, the constraints on the gluon PDF $g(x)$ from DIS come via the Q^2 evolution of the F_2 structure function at low values of x. In addition, measurements at HERA of the longitudinal structure function, which has a leading contribution at $\mathcal{O}(\alpha_s)$ through the $\gamma^* g \to q\bar{q}$ process, have allowed $g(x)$ to be relatively well determined at low x. More directly, inclusive jet and photon production cross sections at hadron colliders have constrained $g(x)$ at moderate x values, although there is somewhat more uncertainty in the behavior at high x. A survey of various determinations of the gluon PDF at $Q^2 = 10$ GeV^2 is illustrated in Fig. 257 for the same set of PDF parametrizations as in Fig. 253.

Since that the gluon PDF is accompanied by α_s in DIS structure functions, in practice there is a correlation between the value of α_s obtained in global PDF analyses and the shape of gluon distribution, with larger α_s leading to a smaller $g(x)$ at small x and (via the momentum sum rule correlation) a larger $g(x)$ at large x. An interesting question is whether α_s should be fitted as a parameter in global analyses or, since it is a parameter of the QCD Lagrangian and should be the same for all processes, fixed to the world average value for $\alpha_s(M_Z)$. Comparisons of results with $\alpha_s(M_Z)$ fitted or fixed may indicate which processes are responsible for any differences [3103].

A related question is the shape of heavy quark PDFs, such as the charm distribution, which is known to contribute $\sim 30\%$ of the total F_2 measured at HERA at small x values. Here the main production mechanism is photon-gluon fusion, so that data on inclusive charm production could also provide valuable constraints on the gluon PDF in the nucleon. The question of whether there is a sizable nonperturbative charm component at a low energy input scale [3134–3138] also remains controversial [3139–3141], with recent analyses claiming both positive [3142] and negative evidence [3143].

10.2.5 Spin-dependent PDFs

Considerable progress has been made in understanding the spin structure of the nucleon since the first precision polarized DIS experiments at CERN in the late 1980s indicated an anomalously small fraction of the proton spin carried by quarks. A rich program of spin-dependent inclusive and semi-inclusive DIS, as well as polarized proton–proton scattering experiments has followed, vastly improving our knowledge of spin-dependent PDFs of the nucleon over the last two decades. While the spin-dependent data have not been as abundant as those available for constraining spin-averaged PDFs, several dedicated global QCD analyses of spin-dependent PDFs to be performed. The main current global efforts include the DSSV group [1294,3144,3145], the NNPDF collaboration [975,3146], and the JAM collaboration [974,3147], extending earlier efforts by the LSS [3121], BB [3148], KATAO [3149] and AAC [3150] groups.

Polarized valence quarks

As for the unpolarized PDFs, the spin-dependent Δu^+ distribution is the most strongly constrained helicity PDF, largely by the proton g_1 structure function data. The corresponding Δd^+ distribution, which has a negative sign, is smaller in magnitude compared with Δu^+ and has larger relative uncertainties, especially at intermediate and large values of x. The size of the uncertainties depends somewhat on the theoretical assumptions made for the distributions. For example, if one assumes only the SU(2) symmetry constraint (10.30) for the difference $\Delta u^+ - \Delta d^+$, the uncertainties on the individual

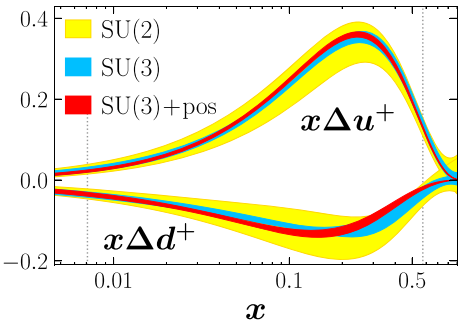


Fig. 258 Polarized $x\Delta u^+$ and $x\Delta d^+$ PDFs from the JAM analysis [976] for various scenarios: assuming SU(2) symmetry (10.30) (yellow bands), SU(3) symmetry (10.31) (blue bands), and in addition the PDF positivity constraint (red bands)

Δu^+ and Δd^+ PDFs are significantly larger than assuming in addition the SU(3) symmetry relation (10.31) involving also the strange polarization Δs^+.

This is illustrated in Fig. 258 for the JAM parametrization [976], which also shows the result of a fit that enforces in addition positivity constraints on the unpolarized PDFs. Whether spin-averaged PDFs need to be positive beyond LO in α_s has been debated recently in the literature [3151], and generally it is understood that the positivity constraint should hold only at LO [3152]. The general features of the Δu^+ and Δd^+ PDFs in Fig. 258 are similar to those found by other global QCD analysis groups [975,3145], which reflects the common origin in the constraints on these PDFs from proton and neutron DIS data. In contrast, without the additional assumption of SU(3) symmetry [974,3153], the strange helicity PDF remains largely unconstrained [976,1295].

Polarized sea quarks

Since inclusive polarized DIS experiments measure C-even combinations of PDFs, Δq^+, additional constraints, either from theory or experiment, are needed to separate the individual quark and antiquark distributions. Additional experimental constraints come from the semi-inclusive production of hadrons, in which spin-dependent PDFs are weighted by fragmentation functions, as well as particle production in polarized hadron collisions, which involve products of spin-dependent (and spin-averaged) PDFs.

The strongest constraints on the polarization of the sea have come from recent W-lepton production data from polarized protons collisions at RHIC [3154–3156]. The effect of the polarized W data is a clear nonzero antiquark asymmetry $\Delta\bar{u} - \Delta\bar{d}$ for $0.01 \lesssim x \lesssim 0.3$, as Fig. 259 illustrates for the recent JAM analysis [3147]. Qualitatively similar, although not as pronounced, behavior was also observed in the earlier DSSV [3145] and NNPDF [975] fits, although these made stronger theoretical assumptions about PDF positivity and SU(3) symmetry. The observed polarized sea asymmetry is

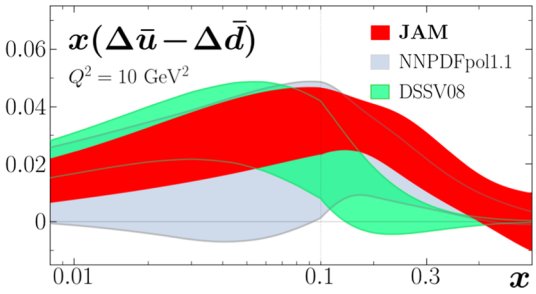


Fig. 259 Polarized sea quark asymmetry $x(\Delta\bar{u} - \Delta\bar{d})$ from the JAM [3147], NNPDF [975] and DSSV [3145] analyses

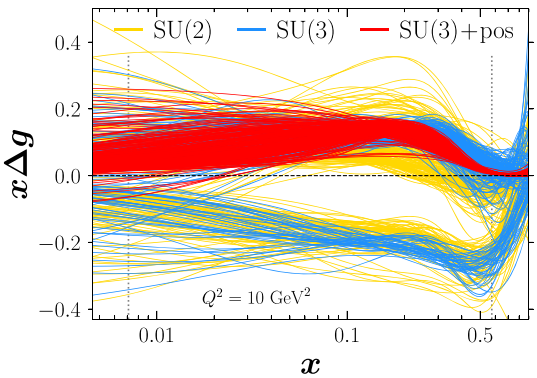


Fig. 260 Monte Carlo replicas for the gluon helicity PDF $x\Delta g$ fitted under various theory assumptions according to the SU(2) (yellow lines), SU(3) (blue lines) and SU(3)+positivity (red lines) scenarios [976]

also similar to expectations from some nonperturbative models of the nucleon [3157–3160].

Polarized gluons

The sign and magnitude of the gluon polarization is a critical component to understanding the decomposition of the proton's spin amongst its quark and gluon constituents. The first clear indication of a positive Δg came from analysis of RHIC jet production data in polarized proton–proton collisions, which were used by the DSSV group to extract a nonzero signal for gluon momentum fractions between $x \approx 0.05$ and ≈ 0.2. More recently, the JAM collaboration [976] performed a simultaneous global fit to unpolarized and polarized data, testing in particular the sensitivity to theoretical assumptions about axial charges and PDF positivity.

The results of the simultaneous analysis, illustrated in Fig. 260, show that indeed the gluon helicity can depend strongly on the constraints imposed. Interestingly, without restricting PDFs to be positive and assuming SU(3) flavor symmetry for the axial vector charges, existing polarized data allow solutions containing negative gluon polarization, in addition to the standard positive gluon solutions found in earlier analyses, giving equally acceptable descriptions of the data. A negative gluon polarization would imply rather large quark or gluon orbital angular momentum contributions, in order

to satisfy the proton spin sum rule. It will be important to verify the sign and magnitude of the gluon polarization in future experiments [3161], as well as explore possible insight gained from lattice QCD calculations [3162].

10.2.6 Outlook

Our knowledge of the detailed partonic structure of the nucleon has improved tremendously in recent years, thanks to high precision experiments and advances in computational and data analysis methods. With planned measurements at facilities such as Jefferson Lab, the LHC, and the future Electron-Ion Collider (EIC) [3163], we can look forward to further breakthroughs in addressing long-standing questions about the momentum and spin distributions of quarks and gluons in the nucleon.

The new experiments will probe hitherto unexplored corners of kinematics in which PDFs have been difficult to determine. An example is the behavior of PDFs and PDF ratios such as $\Delta q^+/q^+$ in the limit as $x \to 1$, which are particularly sensitive to the details of nonperturbative quark–gluon dynamics [3081,3164]. The new data will allow one to test basic theoretical assumptions such as SU(2) and SU(3) symmetry, PDF positivity, and charge symmetry in PDFs. The latter, which is expected to be broken by light quark mass differences, $m_u \neq m_d$, and by electromagnetic corrections, will need to be taken into account if one hopes for PDF accuracy at the few-percent level. Further inroads into solving the proton spin puzzle, through the determination of the total spin contributions from quarks, antiquarks and gluons, will require measurements of spin structure functions down to smaller values of x [3165,3166], which will be one of the focuses of the EIC program [3163].

The aim of few-percent precision in PDFs will also require a more systematic treatment of radiative effects, which in the past have been treated using approximate prescriptions. Recently, a combined QED+QCD approach to factorization has been developed [3167], and while the differences with the traditional methods are not large for inclusive processes, for more exclusive reactions, such as semi-inclusive DIS [3168], the simultaneous paradigm of self-consistently incorporating QED and QCD effects and determining different types of distributions within the same analysis will be necessary.

Along with the new measurements, it is likely that complementary information will be needed from lattice QCD simulations, especially for quantities that will be difficult to access from experiment. Indeed, the first exploratory simultaneous analyses of experimental and lattice data have already been made recently [1090,3169]. Future success in mapping out and understanding the quark and gluon structure of the proton will thus require a coordinated effort on the multiple fronts of experiment, theory, lattice simulation, and data analysis.

10.3 Spin structure

Xiangdong Ji

The nucleon (proton and neutron) is a spin-1/2 composite particle made from three valence quarks. Every model of the nucleon gives an explanation for its spin structure [3170–3174], from the Skyrme model [3170], to Gell-Mann and Zweig's quark model [3171,3172], and to many other models popular in 1970s and 1980s [3173,3174]. The simplest and most successful one is the quark model which, among others, inspired the discovery of QCD [55], predicted that the entire nucleon spin is carried out by the three valence quarks [31, 2719,3175]. The non-relativistic quark model has indeed a simple explanation for the nucleon spin and the associate magnetic moments [31], also for their excited states [2719]: Three constituent quarks are all in the s-wave orbit in the nucleon, and their spins couple to $1/2$ in a way consistent with the $\mathrm{SU}(2_{\mathrm{spin}} \times 3_{\mathrm{flavor}})$, a combined spin-flavor symmetry group [3175].

The quark-model picture for the spin was put under a test through polarized deep-inelastic scattering (DIS) on a polarized proton [3176]. The EMC collaboration made the first definitive measurement for the fraction of the proton spin carried by quarks in 1987 [3177,3178], and the result

$$\Delta\Sigma(Q^2 = 10.7\mathrm{GeV}^2) = 0.060 \pm 0.047 \pm 0.069, \quad (10.35)$$

is consistent with zero. The discrepancy has inspired large amount of experimental and theoretical studies which have been summarized in a number of excellent reviews [3179–3183]. Perhaps the most important lesson we have learned is that the QCD quarks probed in polarized DIS are very different from those in the constituent quark models, and that QCD has a much more sophisticated way to build up the proton spin.

Understanding the nucleon spin in QCD remains an important challenge in hadron structure physics, particularly, in experiment. In the following, we will briefly review the current status and future perspective for this topic, focusing on the questions such as: does it make sense to talk about the different parts of the proton spin? What will be an interesting decomposition for the spin? To what extent do we believe that we can measure these parts experimentally? How can we calculate these contributions in fundamental theory and put them to experimental tests?

10.3.1 Spin sum rules in QCD

Angular momentum (AM) or spin structure of a composite system can be studied through various contributions to the total. In quantum field theories, the individual parts are renormalization scale and scheme dependent, although the total is not. The most popular convention in the literature

is to use dimensional regularization and modified minimal subtraction, indicated by the dependence on the scale μ. To understand the proton spin, we can start from QCD AM operator expressed in terms of individual sources,

$$\vec{J}_{\rm QCD} = \sum_{\alpha} \vec{J}_{\alpha}(\mu) \ . \tag{10.36}$$

Through the above, one can express the total spin 1/2 as contributions from different parts. This has been one of the main methods to explore the origins of the proton spin in the literature. Since the individual contributions are the expectation values of the AM sources in the entire wave function, they are neither integers nor half-integers: they are the quantum mechanical average of probability amplitudes.

There exists more than one way to split the AM and derive spin sum rules for the proton. However, a physically-interesting spin sum rule shall have the following properties:

Experimental measurability
The overwhelming interest in the proton spin began with the EMC data. Much of the followup experiments, including polarized RHIC [3184], Jefferson Lab 12 GeV upgrade [3185] and Electron-Ion Collider (EIC) [1293,3186], have been partially motivated to search a full understanding of the proton spin.

Frame Dependence
Since spin is an intrinsic property of a particle, one naturally searches for a description of its structure independent of a reference frame. How the individual contributions depend on the proton momentum or reference frame requires understanding of the Lorentz transformation properties of $\vec{J}_{\alpha}$. Moreover, the longitudinal and transverse spins behave differently under frame transformation and therefore have very different experimental implications. Since the proton structure probed in high-energy scattering is best described in the infinite momentum frame (IMF), a partonic picture of the spin is phenomenologically interesting to explore.

In the rest frame, the proton state $|\vec{P}=0,\vec{s}\rangle$ can be defined with the angular momentum quantized along $\vec{s}$,

$$\vec{s}\cdot\vec{J}\left|\vec{P}=0,\vec{s}\right\rangle = 1/2\left|\vec{P}=0,\vec{s}\right\rangle \ , \tag{10.37}$$

where we have dropped the "QCD" subscript on $\vec{J}$. Boosting the above to an arbitrary Lorentz frame, one has ($\hbar=1$)

$$(-W^{\mu}S_{\mu})|PS\rangle = 1/2|PS\rangle \ , \tag{10.38}$$

where $|PS\rangle$ have definite four-momentum P^{μ} and spin polarization four-vector S^{μ}, $S^{\mu}=(\gamma\vec{s}\cdot\vec{\beta},\ \vec{s}+(\gamma-1)\vec{s}\cdot\hat{\beta}\hat{\beta})$ with $S^{\mu}S_{\mu}=-1$, $P^{\mu}S_{\mu}=0$, $\hat{\beta}$ the direction of $\vec{\beta}=\vec{v}/c$, $\gamma=(1-\beta^2)^{-1/2}$ the boost factor, and W^{μ} is the relativistic spin (or Pauli–Lubanski) four-vector ($\epsilon^{0123}=1$) [3187]

$$W^{\mu} = -\frac{1}{2}\epsilon^{\mu\alpha\lambda\sigma}J_{\alpha\lambda}P_{\sigma}/M, \tag{10.39}$$

$$= \gamma(\vec{J}\cdot\vec{\beta},\ \ \vec{J}+\vec{K}\times\vec{\beta}) \tag{10.40}$$

where $\vec{K}$ is the Lorentz-boost operator defined in terms of the $0\,i$ components of the Lorentz generator $J^{\alpha\beta}$. In the second line of the equation, we have replaced the four-momentum operator P_{σ} by its eigenvalue specifying a Lorentz frame $\vec{\beta}$. One can use Eq. (10.38) to develop spin sum rules in any frame,

$$\langle PS|(-W^{\mu}S_{\mu})|PS\rangle = 1/2 \ , \tag{10.41}$$

by expressing the left-hand side as the sums of expectation values. Thus the covariant spin is not only related to the AM operator but also to the boost $\vec{K}$. However, it is desirable to develop a spin picture in terms of the AM operator alone in a general Lorentz frame.

Without loss of generality, one can assume the proton momentum is along the z-direction $\vec{P}^z=(0,0,P^z)$. In the case of longitudinal polarization, one has $\vec{s}_z=(0,0,1)$, $-W^{\mu}S_{\mu}=J^z$, and Eq. (10.41) becomes the total helicity,

$$\langle PS_z|J^z|PS_z\rangle = 1/2 \ , \tag{10.42}$$

which is boost-invariant along the z-direction. This is a starting point to construct helicity sum rules. Since the helicity is independent of momentum, the individual contributions are generally sub-leading order in high-energy scattering.

For transverse polarization along the x-direction, $\vec{s}_x=(1,0,0)$, and Eq. (10.41) becomes

$$\langle PS_x|\gamma(J^x-\beta K^y)|PS_x\rangle = 1/2 \ , \tag{10.43}$$

which contains the boost operator K^y from the transformation of J^x under the Lorentz boost along z. Since $\vec{K}$ and $\vec{J}$ transform under Lorentz transformation as $(1,0)+(0,1)$, we can deduce separate relations:

$$\langle PS_x|J^x|PS_x\rangle = \gamma/2$$
$$\langle PS_x|K^y|PS_x\rangle = \gamma\beta/2,$$

true as expectation values. Therefore a transverse polarization sum rule from the AM operator starts from

$$\langle PS_x|J^x|PS_x\rangle = \gamma/2. \tag{10.44}$$

Because the transverse angular momentum J^x depends on the longitudinal momentum of the proton, its expectation value grows under boost, a fact less appreciated in the literature.

To obtain a spin sum rule, we need an expression for the QCD AM operator. It can be derived through Noether's theorem based on space-time symmetry of the QCD lagrangian density. Straightforward calculation yields the *canonical* AM expression [3188]

$$\vec{J}_{\rm QCD} = \int d^3\vec{x}\left[\psi_f^{\dagger}\tfrac{1}{2}\vec{\Sigma}\psi_f + \psi_f^{\dagger}\vec{x}\times(-i\vec{\partial})\psi_f + \vec{E}_a\times\vec{A}_a + E_a^i(\vec{x}\times\vec{\partial})A_a^i\right] \ , \tag{10.45}$$

where ψ_f is a quark field of flavor f, $\vec{\Sigma} = \text{diag}(\vec{\sigma}, \vec{\sigma})$ with $\vec{\sigma}$ the Pauli matrices, A^i_a vector potentials of gauge fields with color $a = 1, \ldots 8$, E^i_a color electric fields, and the contraction of flavor and color indices is implied. The above expression contains four different terms, each of which has clear physical meaning in free-field theory. The first term corresponds to the quark spin, the second to the quark orbital angular momentum (OAM), the third to the gluon spin, and the last one to the gluon OAM. Apart from the first term, the rest are not manifestly gauge-invariant under the general gauge transformation $A^\mu \to U(x)\,(A^\mu + (i/g)\partial^\mu)\,U^\dagger(x)$. However, the total is invariant under the gauge transformation up to a surface term at infinity which can be ignored in physical state matrix elements.

On the other hand, using the Belinfante improvement procedure (Belinfante, 1939) one can obtain a gauge-invariant form [1081],

$$\vec{J}_{\rm QCD} = \int d^3x \left[\psi_f^\dagger \frac{1}{2} \vec{\Sigma} \psi_f + \psi_f^\dagger \vec{x} \times (-i\vec{\nabla} - g\vec{A})\psi_f + \vec{x} \times (\vec{E} \times \vec{B}) \right] . \tag{10.46}$$

All terms are manifestly gauge invariant, with the second term as mechanical or kinetic OAM, and the third term gluon AM.

Helicity sum rule

Using Eq. (10.42) and the gauge-invariant QCD AM in Eq. (10.46), one can write down a helicity sum rule [1081],

$$\frac{1}{2}\Delta\Sigma(\mu) + L^z_q(\mu) + J_g(\mu) = 1/2 \tag{10.47}$$

where $\Delta\Sigma/2$ is the quark helicity contribution, and L^z_q is quark OAM contribution. Together, they give the total quark AM contribution J_q. The last term, J_g, is the gluon contribution. Both contributions can be obtained from the twist-two form factors of the energy–momentum tensor $T^{\mu\nu}$ [1081] (see below). One important feature of the above sum rule is that it is independent of the proton's momentum [3189]. This is an important feature because the sources of the proton spin does not depend on observer's reference frame so long as helicity is a good quantum number.

On the other hand, the canonical form of the AM operator in Eq. (10.45) allows deriving an infinite number of helicity sum rules with choices of gauges and/or frames of reference [3182,3190]. The usefulness of such sum rules are questionable as they are not relevant to experiment. However, the gluon spin contribution in the IMF and light-cone gauge $A^+ = 0$ is measurable. Jaffe and Manohar proposed a canonical spin sum rule in a nucleon state with $P^z = \infty$ [3188],

$$\frac{1}{2}\Delta\Sigma + \Delta G + \ell_q + \ell_g = \frac{1}{2} \tag{10.48}$$

where ΔG is the gluon helicity and $\ell_{q,g}$ are quark and gluon OAM, respectively. Considerable attention has been given to the above sum rule because of its relevance to high-energy scattering. For example, the total quark helicity contribution can be written in terms of parton sum rule,

$$\Delta\Sigma = \int_{-1}^{1} dx(\Delta u(x) + \Delta d(x) + \cdots) , \tag{10.49}$$

where $\Delta q(x)$ is the quark helicity distribution function. Moreover, ΔG has been defined and measured experimentally as the first moment of the gauge-invariant polarized gluon distribution [3191]

$$\begin{aligned} \Delta G(Q^2) &= \int_0^1 dx\ \Delta g(x, Q^2) , \\ \Delta g(x) &= \frac{i}{2x(P^+)^2} \int \frac{d\lambda}{2\pi} e^{i\lambda x} \\ &\quad \times \langle PS|F^{+\alpha}(0)W(0,\lambda n)\tilde{F}_\alpha{}^+(\lambda n)|PS\rangle , \end{aligned} \tag{10.50}$$

where $\tilde{F}^{\alpha\beta} = \frac{1}{2}\epsilon^{\alpha\beta\mu\nu}F_{\mu\nu}$, and the light-cone gauge link $W(\lambda n, 0)$ is defined in the adjoint representation of SU(3). In the light-cone gauge $A^+ = 0$, the nonlocal operator in Eq. (10.50) reduces to the free-field form in the Jaffe-Manohar sum rule. Additionally, one can write a parton sum rule for each of the OAM contributions

$$\ell_q = \int_{-1}^{1} dx \ell_q(x) , \tag{10.51}$$

$$\ell_g = \int_{-1}^{1} dx \ell_g(x) , \tag{10.52}$$

which give a more detailed picture of AM distributions in partons compared with the frame-independent sum rule above.

It appears that one can define a gauge-variant quantity which can be measured in experiment! This has inspired much debate about the gauge symmetry properties of the gluon spin operator and myriads of experimentally-unaccessible spin sum rules [3182]. It turns out, however, that the key is not about generalizing the concept of gauge invariance, it is about the proton state in the IMF [3192]. In particular, $A^+ = 0$ is a physical gauge as it leaves the transverse polarizations of the radiation field intact. This justifies the physical meaning of $\vec{E} \times \vec{A} = \vec{E}_\perp \times \vec{A}_\perp$ as the gluon spin (helicity) operator in the Jaffe-Manohar sum rule.

Comparing the two helicity sum rules Eqs. (10.47) and (10.48) above, they must be related in some way in the IMF. In fact, their relation is [3193,3194]

$$J_g = \Delta G + \ell_g + \ell_{\rm int} \tag{10.53}$$

$$L_q = \ell_q - \ell_{\rm int} \tag{10.54}$$

where $\ell_{\rm int}$ represents the interaction AM and does not have a simple parton interpretation.

Transverse spin sum rules

For transverse polarization, a spin sum rule is less straightforward and much controversy exists in the literature [3182, 3195]. First of all, the transversely-polarized proton is not an eigenstate of the transverse AM operator. Second, the expectation value of the transverse AM has a intriguing frame dependence due to the center-of-mass contribution, which must be properly subtracted. Finally, there are two contributions to the transverse AM which transform differently under Lorentz boost and must combine properly to generate the total result. The delicate balance of two contributions entails two separate transverse spin sum rules.

The transverse spin has a simple frame-independent sum rule [3196],

$$J_q + J_g = 1/2 \,, \tag{10.55}$$

which is the same as the helicity sum rule due to Lorentz symmetry. One can separate the contributions to the quark into spin and orbit ones, however, such a separation is frame-dependent and therefore less interesting.

In the IMF, the above sum rule becomes partonic sum rules [3195,3197],

$$J_q = \int_{-1}^{1} dx J_q(x) \,, \tag{10.56}$$

$$J_g = \int_{-1}^{1} dx J_g(x) \,, \tag{10.57}$$

where $J_q(x)$ and $J_g(x)$ are twist-2 transverse angular momentum densities of the quarks and gluons. They are related to quark and gluon unpolarized densities and generalized parton distributions through $J_q(x) = (1/2)x(q(x) + E_q(x))$ and $J_g(x) = (1/2)x(g(x) + E_g(x))$.

The second transverse spin sum rule can best be discussed in the IMF, where there is a sub-leading partonic sum rule for the transverse spin, corresponding to the twist-three part of the canonical angular momentum density $J^\perp$ in Eq. (10.45). In a simple form, one can write [3198]

$$\frac{1}{2}\Delta\Sigma_T + \Delta G_T + \ell_{qT} + \ell_{gT} = \frac{1}{2} \,. \tag{10.58}$$

The various terms have partonic interpretations in the IMF,

$$\Delta\Sigma_T = \int_{-1}^{1} dx g_T(x) \,, \tag{10.59}$$

$$\Delta G_T = \int_{-1}^{1} dx \Delta G_T(x) \,, \tag{10.60}$$

$$\ell_{qT} = \int_{-1}^{1} dx \ell_{qT}(x) \,, \tag{10.61}$$

$$\ell_{gT} = \int_{-1}^{1} dx \ell_{gT}(x) \,, \tag{10.62}$$

where $g_T(x) = g_1(x) + g_2(x)$ and $G_T(x)$ are transverse spin densities of quarks and gluons, respectively, and ℓ_{qT} and ℓ_{gT} are the corresponding twist-three transverse OAM densities. Because of Lorentz symmetry, the values of these integrated quantities with T are exactly the same as the ones without T in Jaffe-Manohar sum rule. However, the parton densities for the transversely polarized proton are different from those in the longitudinally polarized one. For instance, for the quark spin, the difference is the well-known $g_2(x)$ structure function.

10.3.2 Lattice calculations

At present, the only systematic approach to solve the QCD proton structure is the lattice field theory [97], see, Sect. 4. There are less systematic approaches such as Schwinger–Dyson (Bethe–Salpeter) equations [800] and instanton liquid models [1410] in which a certain truncation is needed to find a solution, see, Sect. 5. Although much progress has been made in these other directions, we focus on the lattice QCD method.

A complete physical calculation on the lattice faces a number of obstacles. First the angular momentum is flavor-singlet quantity, and as such, one needs to compute the disconnected diagrams for the quarks. Since up and down quarks are light, computational demands at the physical pion mass are very high. Moreover, one also has to compute gluon observables to complete the picture, which is known to be very noisy. At the same time, one needs to keep the lattice space sufficiently small and the physical volume large enough. All of these add up to an extremely challenging task. However, a computation with all these issues considered has become possible recently, see for example Ref. [3204]. An additional challenge is present in computing light-cone correlations with a real time variable. The recent development of large-momentum effective theory (LaMET) has opened the door for such computations [637,638,646].

The matrix elements of local operators, $\Delta\Sigma$, J_q and J_g are relatively simple to calculate using the standard lattice QCD technology. Much progress has been made in understanding the content of manifestly gauge-invariant helicity sum rule in Eq. (10.47), and also the transverse spin sum rule in Eq. (10.55).

The first calculations have been about the $\Delta\Sigma$ from different quark flavors. A large amount of work has been summarized in a recent review [3205]. Three most recent calculations are in Refs. [3199–3201], with some at the physical quark mass. Table 47 is taken from Ref. [3203] and shows a summary of the recent lattice results on the quark helicity. The strange quark contribution was also calculated in

Table 47 Results of quark spin for the u, d and s flavors from three recent lattice calculations by the Cyprus group [3199], χQCD[3200], PNDME [3201] in the $\overline{\text{MS}}$ scheme at 2 GeV are listed. $\Delta(u+d)$(CI) and $\Delta(u+d)$(DI) are the spins of the u and d quarks in the connected insertion (CI) and disconnected insertion (DI). Three analyses of experiments from de Florian et al. [1294], NNPDF [975] and COMPASS [3202] are also listed for comparison. Source: Ref. [3203]

	Δu	Δd	Δs	$g_A^3 = \Delta u - \Delta d$	$\Delta(u+d)$(CI)	$\Delta(u/d)$(DI)	$\Delta\Sigma$
Cyprus	0.830(26)(4)	− 0.386(16)(6)	− 0.042(10)(2)	1.216(31)(7)	0.598(24)(6)	− 0.077(15)(5)	0.402(34)(10)
χQCD	0.846(18)(32)	− 0.410(16)(18)	− 0.035(8)(7)	1.256(16)(30)	0.580(16)(30)	− 0.072(12)(15)	0.401(25)(37)
PNDME	0.777(25)(30)	− 0.438(18)(30)	− 0.053(8)	1.218(25)(30)			0.286(62)(72)
de Florian et al. ($Q^2 = 10$ GeV^2)	$0.793^{+0.011}_{-0.012}$	$-0.416^{+0.011}_{-0.009}$	$-0.012^{+0.020}_{-0.024}$				$0.366^{+0.015}_{-0.018}$
NNPDFpol1.1 ($Q^2 = 10$ GeV^2)	0.76(4)	− 0.41(4)	− 0.10(8)				0.25(10)
COMPASS ($Q^2 = 3$ GeV^2)	[0.82, 0.85]	[−0.45, −0.42]	[−0.11, −0.08]	1.22(5)(10)			[0.26, 0.36]

Refs. [3206,3207] through the anomalous Ward identity, and $\Delta s + \Delta\bar{s} = -0.0403(44)(78)$. The total quark spin contribution to the proton helicity is about 40%.

To calculate the total quark orbital and gluon AM contributions, one can start with the AM density, $M^{\mu\nu\lambda}$, of QCD, from which the AM operator is defined. It is well-known that the AM density is related to the energy–momentum tensor (EMT) $T^{\mu\nu}$ through [3188],

$$M^{\mu\nu\lambda}(x) = x^\nu T^{\mu\lambda} - x^\lambda T^{\mu\nu}. \tag{10.63}$$

The individual contributions to the EMT, hence AM density, can be written as the sum of quark and gluon parts,

$$T^{\mu\nu} = T_q^{\mu\nu} + T_g^{\mu\nu}\ , \tag{10.64}$$

where

$$T_q^{\mu\nu} = \frac{1}{2}\left[\bar{\psi}\gamma^{(\mu} i\overrightarrow{D}^{\nu)}\psi + \bar{\psi}\gamma^{(\mu} i\overleftarrow{D}^{\nu)}\psi\right], \tag{10.65}$$

$$T_g^{\mu\nu} = \frac{1}{4}F^2 g^{\mu\nu} - F^{\mu\alpha}F^{\nu}{}_{\alpha}\ , \tag{10.66}$$

where T_q includes quarks of all flavor. The expectation values of the AM densities can be derived from the off-forward matrix elements of EMT [1081],

$$\begin{aligned}\langle P'S|T_{q/g}^{\mu\nu}(0)|PS\rangle = \bar{U}(P'S)\Big[&A_{q/g}(\Delta^2)\gamma^{(\mu}\bar{P}^{\nu)} \\ &+ B_{q/g}(\Delta^2)\frac{\bar{P}^{(\mu}i\sigma^{\nu)\alpha}\Delta_\alpha}{2M} \\ &+ C_{q/g}(\Delta^2)\frac{\Delta^\mu\Delta^\nu - g^{\mu\nu}\Delta^2}{M} \\ &+\bar{C}_{q/g}(\Delta^2)Mg^{\mu\nu}\Big]U(PS)\ ,\end{aligned} \tag{10.67}$$

where $\bar{P} = (P + P')/2$, $\Delta = P' - P$, and A, B, C and $\bar{C}$ are four independent form factors. It has been shown that

$$J_q = 1/2(A_q(0) + B_q(0)) \tag{10.68}$$

and similarly for the gluon.

The calculation of the total quark and gluon angular momenta started from Ref. [3208] in which the quark part including the disconnected diagrams was calculated without dynamical quarks. The result is the total quark contribution is $J_q = 0.30 \pm 0.07$, i.e. 60%; therefore about 40% of the proton spin must be carried by the gluon. Following other quenched studies [3209,3210], dynamical simulations took over [3211–3215]. A complete study of the angular momentum decomposition was made in Ref. [3206] in quenched formalism, and later in Ref. [3199]. It was found that the quark orbital angular momentum contributes about 47% and gluon angular momentum contributes 28%.

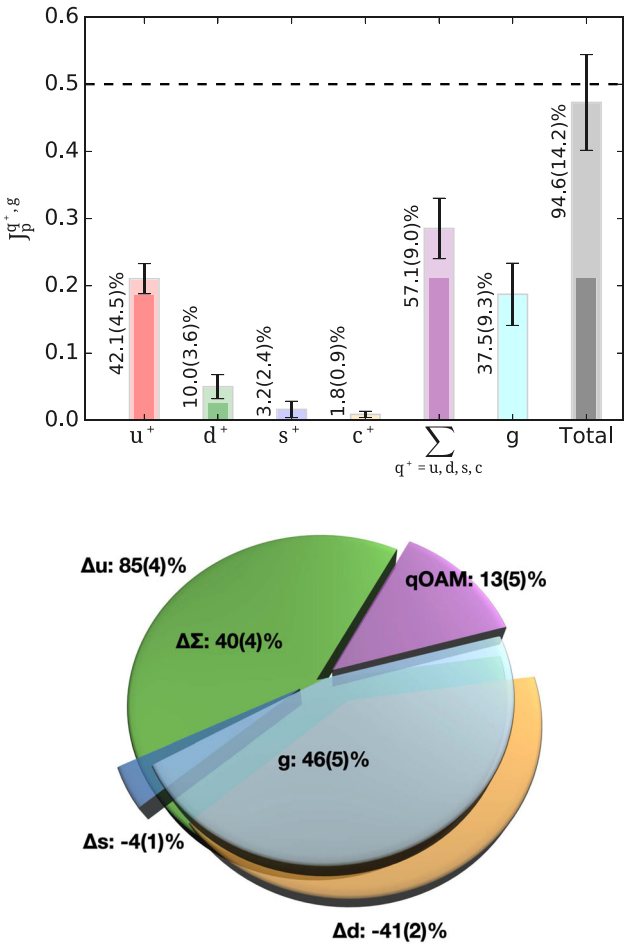


Fig. 261 (upper) Proton spin decomposition in terms of different quark flavors and gluon from Ref. [3204]. (lower) Spin decomposition in terms of quark helicity, OAM and gluon contributions from Ref. [3216]

A complete dynamical simulation at the physical pion mass has been finished recently [3204]. It was found that the total quark spin contribution is about 38.2%, and the orbital angular momentum contribution of the quarks is about 18.8%, much reduced compared with quenched simulations. The total gluon contribution is 37.5%. The resulting picture is shown in Fig. 261. The total spin is 94.6% with an error bar of 14.2%. The spin decomposition in terms of the total quark helicity $\Delta\Sigma = \Delta u + \Delta d + \Delta s$, and quark OAM, and the gluon J_g for $n_f = 2 + 1$ has been calculated in Ref. [3216].

Calculation of the gluon helicity has not been possible for many years because it is intrinsically a light-cone quantity. However, a progress in 2013 was made by studying the frame dependence of non-local matrix elements. One can match the large-momentum matrix element of a static "gluon spin" operator, which is calculable in lattice QCD, to ΔG in the IMF [3192]. This idea was a prototype of LaMET, which was soon put forward as a general approach to calculate all parton physics [637,638]. Using LaMET, one can also calculate the polarized gluon helicity distribution $\Delta g(x)$ in a region of $x \sim 0.2-0.8$. However, the approach does not allow one to calculate the integrated ΔG starting from spatial correlation functions of gluon field strength.

The computation of parton OAM on lattice has been suggested in terms of lattice phase-space Wigner distribution, in which a quark bilinear non-local operator form factor is calculated [3217,3218]. The non-local operator contains a Wilson line to make it gauge invariant. The canonical OAM can be constructed with Wilson lines along the main direction of the proton momentum going to infinity. One can in principle obtain the local gauge invariant OAM with a Wilson line connecting the two quark fields with a straight line. The result seems to be consistent with the calculation discussed above. The result in Ref. [3218] suggests that the isovector canonical OAM has a different sign from the mechanical one, and with a magnitude about 40% larger. One issue with this type of calculation is the renormalization, which can be done with LaMET matching.

One can also calculate the total parton OAM using local operators in a fixed gauge [3219] following the similar approach for the gluon helicity. Matching coefficients between IMF and finite momentum frame have been calculated. One particular feature of the calculation is fixed-gauge which is challenging both on lattice and QCD perturbation theory. On lattice, local gauge condition can lead to the Gribov copies; on the other hand, perturbation theory in a physical gauge requires better understanding at large orders.

Finally, the spin structure of the nucleon in the IMF requires calculations of various light-cone distributions, which include the quark and gluon helicity distributions $\Delta q(x)$, and $\Delta G(x)$, OAM distributions $J_q(x)$ and $J_g(x)$ through GPDs, and OAM distributions $\ell_q(x)$, $\ell_g(x)$, $g_2(x)$, $\Delta G_T(x)$, $\ell_{qT}(x)$, and $\ell_{gT}(x)$.

Shown in Fig. 262 are twist-2 angular momentum densities of up and down quarks in a transversely-polarized nucleon, obtained from phenomenological fit to lattice form factors and generalized parton distributions (GPDs) [3220]. They can be compared with direct lattice calculations and experimental data to be discussed below.

10.3.3 Experiments and phenomenology

Since the EMC experiments, there have been extensive experimental efforts around the globe to investigate the quark and gluon spin contributions to the proton spin, with two important improvements: higher precision and wider kinematic coverage. Majority of these efforts continued in line of the EMC experiment, measuring the polarized structure functions in inclusive DIS with polarized lepton on polarized target (proton, neutron, deuteron). Two important new initiatives have also emerged. First, the DIS experiment facilities extended their capabilities to measure the spin asymme-

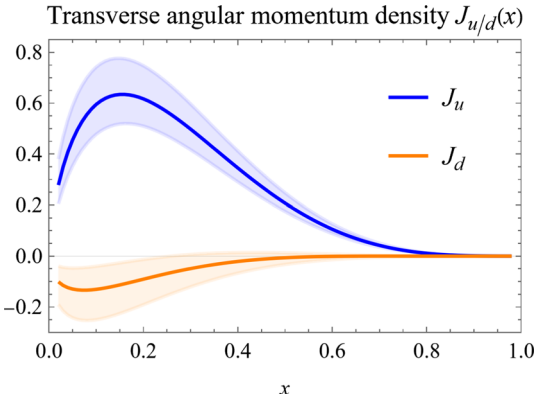


Fig. 262 Angular momentum density distributions of up and down quarks in a transversely polarized proton, fitted to lattice form factors and GPDs [3220]

tries in the semi-inclusive hadron production in DIS (SIDIS), which can help to identify the flavor structure in the polarized quark distributions. Second, the Relativistic Heavy Ion Collider (RHIC) at the Brookhaven National Laboratory (BNL) started the polarized proton–proton experiments. This facility opened new opportunities to explore the proton spin, in particular, for the helicity contributions from gluon and sea quarks (see the previous subsection for experimental data and analysis).

To take into account the constraints from all experiments, it is important to perform a global analysis of the polarized parton distributions from the world-wide data. In these analyses, one has to make some generic assumptions about the functional form (in terms of the unpolarized parton distributions) with a few parameters to fit to data, see, e.g., Refs. [975,3144,3202], where perturbative corrections have been included up to next-to-leading order. Very interesting results, in particular, for the double spin asymmetries in inclusive jet production from the RHIC experiments have provided more strong constraint on the gluon spin [3222], see Fig. 263. This promises great potential for future RHIC experiments to further reduce the uncertainties due to greater statistics [1276,3223].

The total quark spin contribution to the proton spin $\Delta\Sigma$ has been well determined from the DIS measurements. For this quantity, all of the global fits agree well with each other, which essentially gives $\Sigma_q \approx 0.30$ with uncertainties around 0.05. However, for sea quark polarizations including $\bar{u}$, $\bar{d}$ and s ($\bar{s}$), there exist great uncertainties, in particular, for the strange quark polarization [975,1295,3144], which mainly comes from SIDIS measurements from HERMES and COMPASS. Recently, it was also found that the W boson spin asymmetries at $\sqrt{s} = 500$ GeV RHIC have also improved the constraints on $\bar{u}$ and $\bar{d}$ polarization [3224].

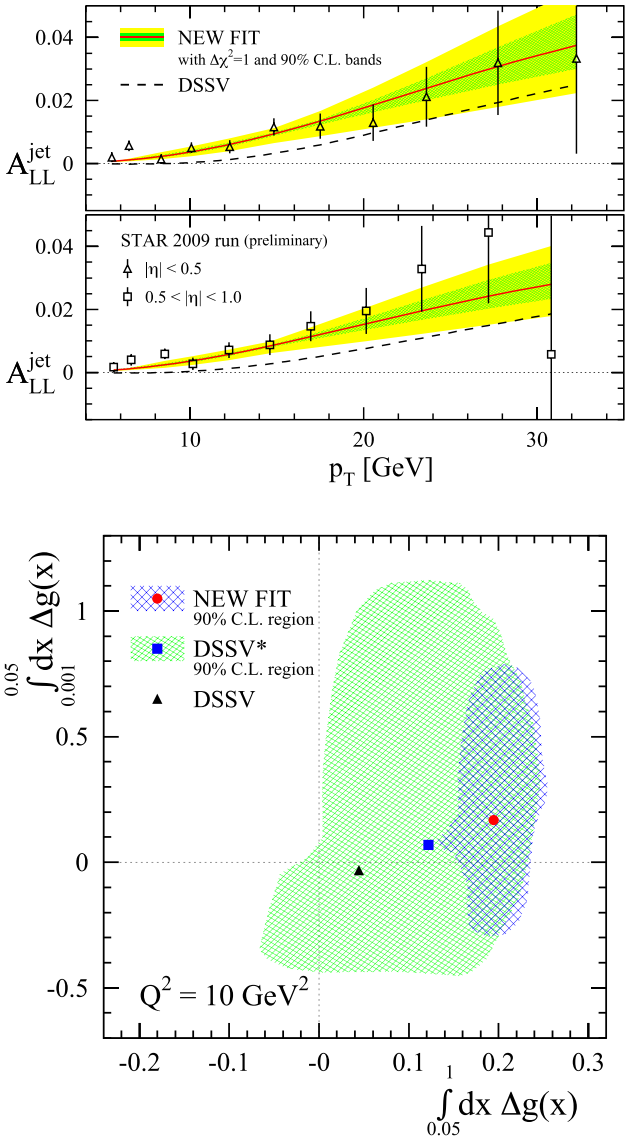


Fig. 263 (upper) Double spin asymmetry in inclusive jet production at RHIC and (lower) constraints on the gluon helicity contribution to the proton spin. Source: Ref. [3144]

The OAM of the quarks may be extracted from measurement of GPD [1081],

$$J_q = \frac{1}{2}\Sigma_q + L_q = \lim_{t\to 0}\frac{1}{2}\int dx x\left[H^q(x,\xi,t) + E^q(x,\xi,t)\right], \tag{10.69}$$

where J_q is the total quark contribution to the proton spin, H and E are GPDs. After subtracting the helicity contribution $\Delta\Sigma$ from various experiments, the above equation will provide the quark OAM contribution to the proton spin. The GPDs can be measured in many different experiments, for example, deeply virtual compton scattering (DVCS) and hard exclusive meson production. Experimental efforts have

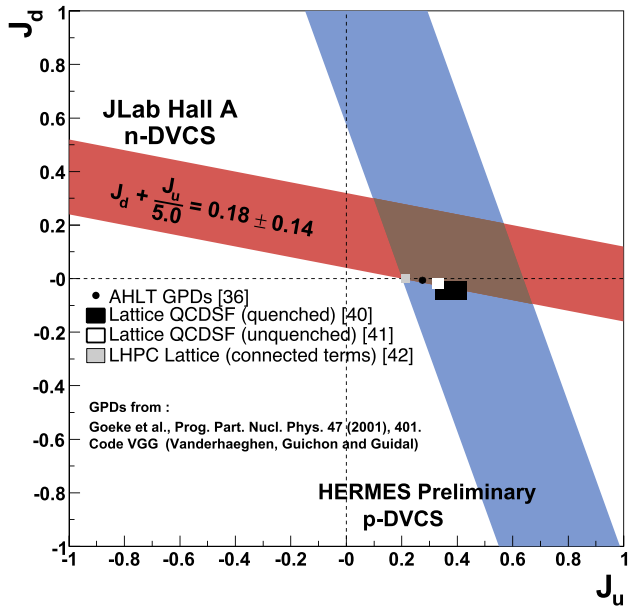


Fig. 264 Model-dependent constraints on the up and down quark total angular momentum from DVCS measurement at JLab. Source: Ref. [3221]

been made at various facilities, including HERMES at DESY, Jefferson Lab, and COMPASS at CERN.

In real photon exclusive production in DIS process, the DVCS amplitude interferes with the Bethe–Heitler (BH) amplitude. This will, on one hand, complicate the analysis of the cross section, on the other hand, provide unique opportunities to direct access the DVCS amplitude through the interference. To obtain the constraints on the quark OAMs from these experiments, we need to find the observables which are sensitive to the GPD Es. Experiments on the DVCS from JLab 6 GeV Hall A [3221] and HERMES at DESY [3225] have shown strong sensitivity to the quark OAM in nucleon, see, e.g., Fig. 264. In these experiments, the single spin asymmetries associated with beam or target in DVCS processes are measured, including the beam (lepton) single spin asymmetry and (target) nucleon single spin (transverse or longitudinal) asymmetries.

A less model-dependent approach to extract the AM information from DVCS or similar experiments is to perform a global analysis. Several theory groups have been working on global analyses of the DVCS and DVEM processes [3226–3228]. Recently, a framework to make general analysis of GPDs similar to CTEQ program [1082], called GPDs through universal momentum parametrization (GUMP) [3220], has been proposed based on the previous work on conformal moments expansion [3229,3230]. The framework, once including the ξ dependence, can be used to fit experimental cross sections and asymmetries. In this way, the quark AM extracted will have less systematic error. In addition, this approach allows us to get the twist-2 quark AM densities, $J_q(x)$, with constraints from experimental data. A number of important AM densities in the spin sum rules depend on information from twist-3 GPDs, such as canonical OAM densities in both longitudinally and transversely polarized proton. Extracting the relevant GPDs from experimental data will be very challenging due to the kinematic suppression.

For the gluon GPDs and AM density $J_g(x)$, one of the most interesting processes is heavy quarkonium production in hard exclusive DIS. This is in particular important at the EIC machine. In early 2020, DOE announced that the next major facility for nuclear physics in US will be a high-energy high-luminosity polarized EIC to be built at BNL. The primary goal of the EIC is to precisely image gluon distributions in nucleons and nuclei, revealing the origin of the nucleon spin and exploring the new QCD frontier of cold nuclear matter [1293,3186].

The EIC will impact our understanding of nucleon spin in many different ways. In the following, we highlight some of these impacts. First, the quark and gluon helicity contributions to the proton spin is the major emphasis of the planned facility. With the unique coverage in both x and Q^2, the EIC would provide the most powerful constraints on $\Delta\Sigma$ and ΔG [1293]. Also shown in Fig. 265 are the projected uncertainty reductions with the proposed EIC machine. Clearly, the EIC will make a huge impact on our knowledge of these quantities, unmatched by any other existing or anticipated facility.

Second, the sea quark polarization will be very precisely determined through SIDIS. With much large Q^2 and x coverage, SIDIS at EIC will provide unprecedented kinematic reach and improve the systematic uncertainties. In Fig. 265, we show the example of sea quark polarization constraints from the EIC pseudo-data simulations.

Third, there will be a comprehensive program on research of GPDs at the EIC. As discussed above, the GPDs provide first hand constraints on the total quark/gluon angular momentum contributions to the proton spin. Moreover, they also provide important information on the nucleon tomography, especially, the 3D imaging of partons inside the proton. With wide kinematic coverage at the EIC, a particular example was shown in Fig. 265 that the transverse imaging of the gluon can be precisely mapped out from the detailed measurement of hard exclusive J/ψ production in DIS processes.

Finally, we would like to emphasize theoretical efforts are as important as the experiments to answer the nucleon spin puzzle. An important question concerns the asymptotic small-x behavior for the spin sum rule. There have been some progresses to understand the proton spin structure at small-x from the associated small-x evolution equations [3231–3239]. More theoretical efforts are needed to resolve the controversial issues raised in these derivations. The final answer to these questions will provide important guidance for the

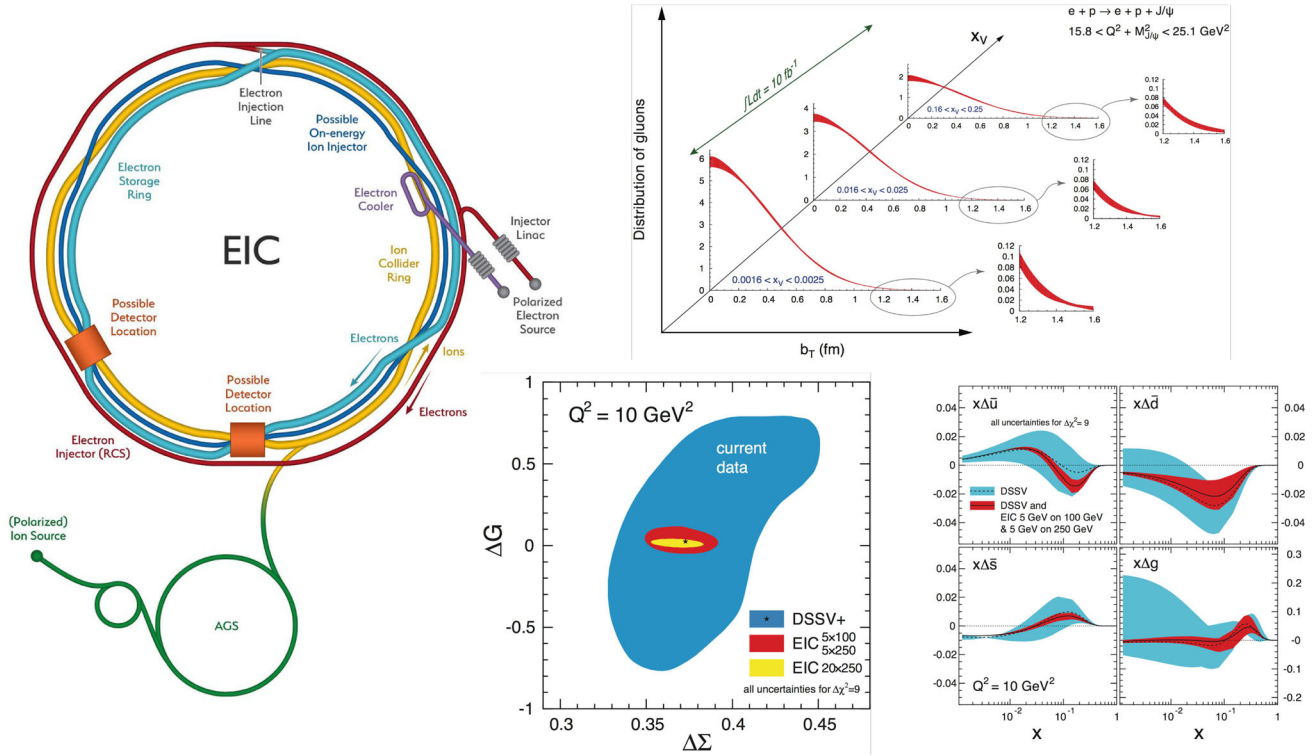


Fig. 265 The planned electron-ion collider (EIC) at BNL, NY, USA. Highlights of the EIC impact on our understanding of nucleon spin: total quark/gluon helicity contributions to the proton spin; sea quark helicity distribution using semi-inclusive deep inelastic scattering; nucleon tomography of the 3D gluon density in the transverse plane for different momentum fractions. (This figure from Ref. [1293])

future EIC, where proton spin rum rule is one of the major focuses.

For additional discussion of these issues, see Sect. 10.2.

10.4 Nucleon tomography: GPDs, TMDs and Wigner distributions

Andreas Sch“afer and Feng Yuan

Exploring the nucleon is of fundamental importance in science, starting from Rutherford's pioneering experiment one hundred years ago where he investigated the internal structure of atomic matter [3240]. Following this effort, the scientific developments in the last century have revealed the most fundamental structure of the matter in our universe: the nucleus is made of nucleons (protons and neutrons) and the nucleon is made of partons: quarks and gluons. In particular, inclusive DIS experiments probe the parton distribution functions which describe the momentum distributions of the partons inside the nucleon, see, Sect. 10.2.

On the other hand, the inclusive measurements of the above processes only probe one dimension of the parton distributions, where the PDF represents the probability distribution of a particular parton (quark or gluon) with a certain fraction x of the nucleon momentum in the infinite momentum frame. In recent years, the hadron physics community is pursuing an extension of this picture to include the transverse direction. The goal is to obtain a three-dimensional tomography of parton densities inside the nucleon. In some sense, these efforts continue the original Rutherford experiment to map out the internal structure of a nucleon in three dimensions.

The nucleon is assumed to move in the $\hat{z}$-direction. Its structure in transverse direction can be either analysed in coordinate space using generalized parton distributions (GPDs) [1081,1286,3241–3246], or in momentum space using transverse momentum dependent parton distributions (TMDs) [1272,1284,3247,3248]. References [964,3249] introduce the impact parameter dependent parton distributions, which are Fourier transforms of GPDs in certain kinematics and which are the desired parton densities in coordinate space.

The information parametrized by GPDs and TMDs is contained in "mother distributions", the so-called Wigner distributions [3250,3251]. Wigner distributions were introduced by Wigner in 1930s as phase space distributions in quantum mechanics,

$$W(r,p) = \int d\eta e^{ip\eta} \psi^* \left(r - \frac{\eta}{2}\right) \psi \left(r + \frac{\eta}{2}\right) , \qquad (10.70)$$

where r and p represent the coordinate and momentum space variables, respectively, and ψ is the wave function. When integrating over r (p), one gets the momentum (probability) density from the wave function, which is positive definite. For arbitrary r and p, the Wigner distribution is not positive definite and does not have a probability interpretation. This reflects the fact that the Wigner distribution contains all quantum mechanical information contained in ψ, which goes beyond probabilities.

Following this concept, we can define the Wigner distribution for a quark in a nucleon with momentum P [3250,3251],

$$W_\Gamma(x,k_\perp,\vec{r}) = \int \frac{d\eta^- d^2\eta_\perp}{(2\pi)^3} e^{ik\cdot\eta} \times \langle P|\overline{\Psi}\left(\vec{r}-\frac{\eta}{2}\right)\Gamma\Psi\left(\vec{r}+\frac{\eta}{2}\right)|P\rangle, \tag{10.71}$$

where x represents the longitudinal momentum fraction carried by the quark, $k_\perp$ is the transverse momentum, $\vec{r}$ the coordinate space variable, and Γ the Dirac matrix to project out a particular quark distribution. The quark field Ψ contains the relevant gauge link to guarantee gauge invariance of the above definition [3250]; see more discussions below. We can also define the Wigner distribution for gluons accordingly.

If we integrate the Wigner distribution over r_z, we obtain the transverse Wigner distribution,

$$\begin{aligned} W^T_\Gamma(x,k_\perp,r_\perp) &= \int \frac{dr_z d\eta^- d^2\eta_\perp}{(2\pi)^3} e^{ik\cdot\eta}\langle P|\overline{\Psi}\left(\vec{r}-\frac{\eta}{2}\right)\Gamma\Psi\left(\vec{r}+\frac{\eta}{2}\right)|P\rangle \\ &= \int \frac{d^2q_\perp d\eta^- d^2\eta_\perp}{(2\pi)^5} e^{iq_\perp\cdot r_\perp} e^{ik\cdot\eta} \times\left\langle P+\frac{q_\perp}{2}\middle|\overline{\Psi}\left(-\frac{\eta}{2}\right)\Gamma\Psi\left(\frac{\eta}{2}\right)\middle|P-\frac{q_\perp}{2}\right\rangle \end{aligned}$$

where we have introduced a wave package for the nucleon state to derive the last equation. The Wigner distribution functions are also referred to as generalized TMDs (GTMDs) [3252,3253]. They can be interpreted as phase space ($r_\perp$,$k_\perp$) distributions of a parton in the transverse plane perpendicular to the nucleon momentum direction.

The Wigner distribution functions reduce to the TMDs and GPDs upon integration over certain kinematic variables. For example, when integrated over $r_\perp$, the above distribution leads to the transverse momentum dependent quark distributions,

$$f(x,k_\perp) = \int \frac{d\eta^- d^2\eta_\perp}{(2\pi)^3} e^{ik\cdot\eta}\langle P|\overline{\Psi}\left(-\frac{\eta}{2}\right)\Gamma\Psi\left(\frac{\eta}{2}\right)|P\rangle. \tag{10.72}$$

On the other hand, if we integrate out $k_\perp$, we obtain the impact parameter dependent quark distribution [964], which is the Fourier transform of the GPDs at $\xi = 0$,

$$\begin{aligned} f(x,b_\perp) &= \int \frac{d^2\Delta_\perp}{(2\pi)^2} e^{i\Delta_\perp\cdot b_\perp}\int \frac{d\eta^-}{2\pi} e^{ik\cdot\eta} \\ &\quad\times\left\langle P+\frac{\Delta_\perp}{2}\middle|\overline{\Psi}\left(-\frac{\eta^-}{2}\right)\Gamma\Psi\left(\frac{\eta^-}{2}\right)\middle|P-\frac{\Delta_\perp}{2}\right\rangle \\ &= \int \frac{d^2\Delta_\perp}{(2\pi)^2} e^{i\Delta_\perp\cdot b_\perp} H(x,\xi,t)|_{\xi=0}\,. \end{aligned} \tag{10.73}$$

Here, $t = -\vec{\Delta}_\perp^2$ and $H(x,\xi,t)$ represents one of the GPDs (definitions will be given below).

The relations between these different functions are often illustrated by the cartoon in Fig. 266 which is, however, somewhat symbolic. Just like the Wigner distribution in quantum mechanics contains the full information of the wave function ψ, a Wigner function in quantum field theory (QFT) contains the full complexity of QFT, including its dependence on the chosen renormalization and factorization scheme. For example TMDs depend on the two scaling variables μ and ζ, while PDFs depend only on μ. Consequently equations like

$$f(x) \overset{?!}{=} \int d^2k_\perp f(x,k_\perp) \tag{10.74}$$

are only valid up to scheme dependent subtraction/renormalization factors or even matching functions. This has significant consequences. For example, usually, the lhs of Eq. (10.74) fulfills a different evolution equation than the rhs. Thus, when comparing the results of different phenomenological TMD fits or lattice calculations one has to convert them into the same scheme.

For other functions there is no such complication. For example, the x integral of GPDs is equal to form factors, e.g., $F_1(Q^2) = \int dx H(x,\xi,t=-Q^2)$. This being said, such complications as well as the μ and ζ dependence are usually suppressed to simplify notation and we do the same in this review.

The status and perspective of both the collinear PDFs and nucleon form factors have been well covered in this review, see, Sects. 10.1 and 10.2.

The tomographical information inherent to Wigner distributions is best illustrated by the resulting intuitive and rigorous method to define the quark/gluon orbital angular momentum (OAM). This follows the concept of the Wigner distribution as a phase-space distribution, i.e., to compute the physical observable, one takes the average over the phase-space as if it were a classical distribution,

$$\langle\hat{O}(r,p)\rangle = \int dr dp W(r,p) O(r,p)\,. \tag{10.75}$$

Since the orbital angular momentum represents the quantity $\vec{r}\times\vec{p}$, we obtain the quark/gluon OAM from the integral of $\vec{r}\times\vec{p}$ multiplied with the Wigner distribution.

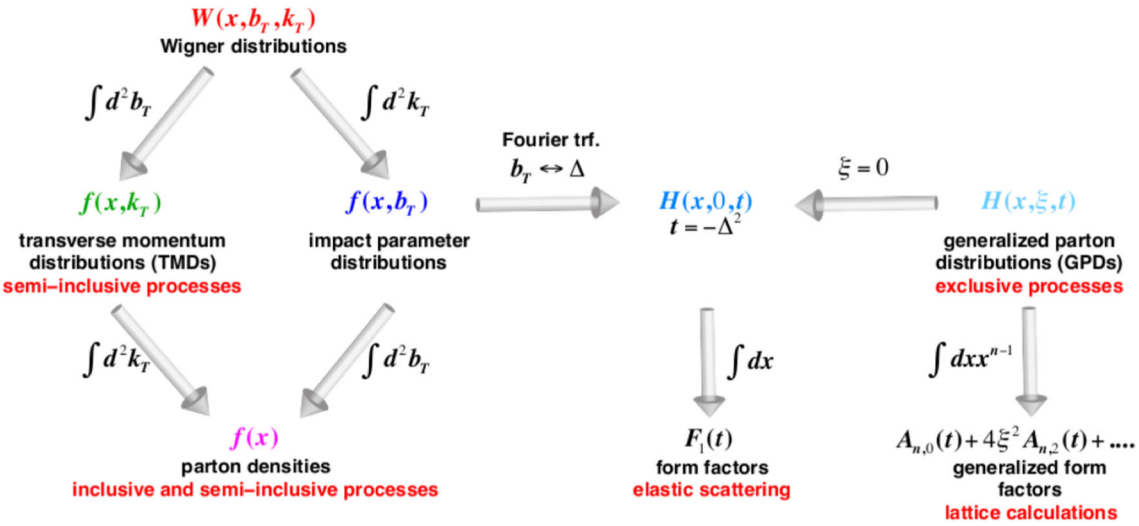


Fig. 266 Transverse momentum dependent parton distributions and the generalized parton distributions are unified in the Wigner distributions. This plot is adopted from Ref. [1293]

For the parton Wigner distribution, one first realizes that a gauge invariant parton distribution must include a gauge link extending from the location of the parton to infinity. An optimal choice for high-energy collisions is a gauge link along the relevant light-cone direction n^μ,

$$\Psi_{LC}(\xi) = P\left[\exp\left(-ig\int_0^{\pm\infty} d\lambda\, n\cdot A(\lambda n+\xi)\right)\right]\psi(\xi)\,, \tag{10.76}$$

where P indicates path ordering. The above defined gauge link can go to $+\infty$ or $-\infty$; see more discussions below. In practical applications, we can also choose a straight-line gauge link along the direction of the spacetime position ξ^μ,

$$\Psi_{FS}(\xi) = P\left[\exp\left(-ig\int_0^{\infty} d\lambda\, \xi\cdot A(\lambda\xi)\right)\right]\psi(\xi)\,. \tag{10.77}$$

This link reduces to unity when $\xi\cdot A(\xi)=0$ (the Fock–Schwinger gauge). With the above definitions, we can write down the quark Wigner distribution as,

$$W_{\mathcal{P}}(k^+ = xP^+, \vec{b}_\perp, \vec{k}_\perp) = \frac{1}{2}\int\frac{d^2\vec{q}_\perp}{(2\pi)^3}\int\frac{dk^-}{(2\pi)^3}e^{-i\vec{q}_\perp\cdot\vec{b}_\perp}\left\langle\frac{\vec{q}_\perp}{2}\left|\hat{\mathcal{W}}_{\mathcal{P}}(0,k)\right|-\frac{\vec{q}_\perp}{2}\right\rangle, \tag{10.78}$$

with the Wigner operator,

$$\hat{\mathcal{W}}_{\mathcal{P}}(\vec{r},k) = \int\overline{\Psi}_{\mathcal{P}}(\vec{r}-\xi/2)\gamma^+\Psi_{\mathcal{P}}(\vec{r}+\xi/2)e^{ik\cdot\xi}d^4\xi, \tag{10.79}$$

where $\mathcal{P}$ denotes the path and is either LC or FS, $\vec{r}$ is the quark phase-space position, and k the phase-space four-momentum.

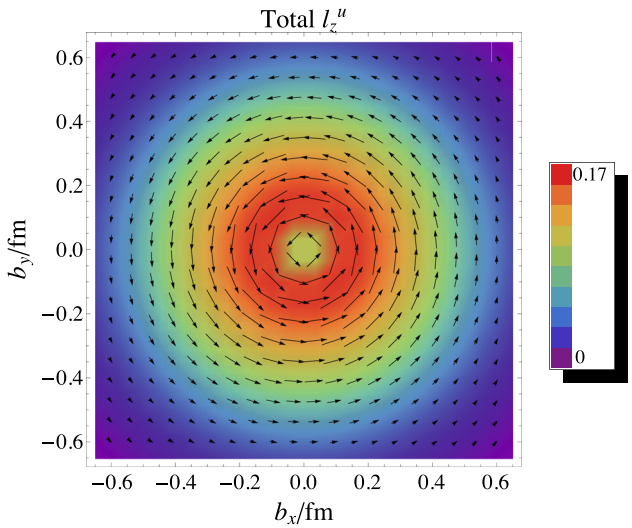


Fig. 267 Distributions in impact parameter space of the mean transverse momentum of an unpolarized u-quark in a longitudinally polarized nucleon, taken from Ref. [3254]. The nucleon is polarized perpendicular to the plane, while the arrows show the size and direction of the mean transverse momentum of the quarks. This gives an intuitive picture of the quark orbital angular motion inside the nucleon

It can be shown that the total OAM is given by the parton's Wigner distribution,

$$L_q = \frac{\langle PS|\int d^3\vec{r}\,\overline{\psi}(\vec{r})\gamma^+(\vec{r}_\perp\times i\vec{D}_\perp)\psi(\vec{r})|PS\rangle}{\langle PS|PS\rangle} = \int(\vec{b}_\perp\times\vec{k}_\perp)W_{FS}(x,\vec{b}_\perp,\vec{k}_\perp)dxd^2\vec{b}_\perp d^2\vec{k}_\perp\,, \tag{10.80}$$

which provides a gauge-invariant expression for the parton's OAM [3196,3255].

Similarly, the canonical OAM in light-cone gauge fulfills the simple but gauge-dependent parton sum rule in the quantum phase space [3253,3254,3256],

$$\ell_q = \frac{\langle PS|\int d^3\vec{r}\ \overline{\psi}(\vec{r})\gamma^+(\vec{r}_\perp \times i\vec{\partial}_\perp)\psi(\vec{r})|PS\rangle}{\langle PS|PS\rangle} = \int (\vec{b}_\perp \times \vec{k}_\perp) W_{LC}(x,\vec{b}_\perp,\vec{k}_\perp) dx d^2\vec{b}_\perp d^2\vec{k}_\perp \,. \quad (10.81)$$

The above two OAMs, L_q and ℓ_q, correspond to the quark OAMs in the Ji and Jaffe-Manohar spin sum rules, respectively, discussed in Sect. 10.3. Similar conclusions hold for the gluon OAMs as well.

Therefore, the Wigner distribution, to some extend, contains the parton OAMs in two different spin sum rules. This further illustrates that the difference between them comes from the gauge link direction. A recent lattice QCD calculation has shown that the quark OAMs can be obtained from the quark Wigner distributions and the difference between L_q and ℓ_q has been demonstrated [3217,3218].

In the last years a number of studies have directly probed the quark/gluon OAM contributions [3257–3261] applying the Wigner distribution for hard exclusive processes. For example, the single longitudinal target-spin asymmetries in hard exclusive dijet production in lepton–nucleon collisions [3257,3258] and the double spin asymmetries in this process [3261] can provide crucial information on the gluon's canonical OAM contribution.

The determination of Wigner distributions is thus an important challenge for future studies; see discussions in the end of this subsection. The crucial point is that there exists a well-defined, standardized way to link nucleon tomography to Wigner distributions constructed from light-cone wave functions [3254]. As an example we show in Fig. 267 the average transverse momentum flow in impact parameter space for u-quarks inside the proton. While this result is model dependent, it has the great advantage of providing an intuitive image of the quark orbital motion distribution inside a hadron.

Generalized parton distributions

The GPDs are one of the projections from the Wigner distributions. They are extensions of the usual collinear parton distributions discussed in Sect. 10.2 and defined as off-forward matrix elements of the hadron. For example, for the quark GPDs, we have [1079,1081,1286,3242–3245]

$$\int \frac{d\lambda}{2\pi} e^{i\lambda x}\langle P'S'|\overline{\Psi}_q\left(-\frac{\lambda}{2}n\right)\not{n}\Psi_q\left(\frac{\lambda}{2}n\right)|PS\rangle = \overline{U}(P')\left[H_q(x,\xi,t)\not{n} + E_q(x,\xi,t)\frac{\sigma^{\alpha\beta}n_\alpha\Delta_\beta}{2M_p}\right]U(P)\,, \quad (10.82)$$

where $\Delta = P' - P$ with $t = \Delta^2$, x is the light-cone momentum fraction of the quark, and the skewness parameter ξ is defined as $\xi = (P - P')\cdot n/(P + P')\cdot n$. In the forward limit, we have $\xi = 0$ and $t = 0$, and the GPDs reduce to the usual collinear PDFs. The x-moments of GPDs lead to not only the electromagnetic form factors but also the gravitational form factors [1081], one of which produces the spin sum rule as discussed in the previous subsection.

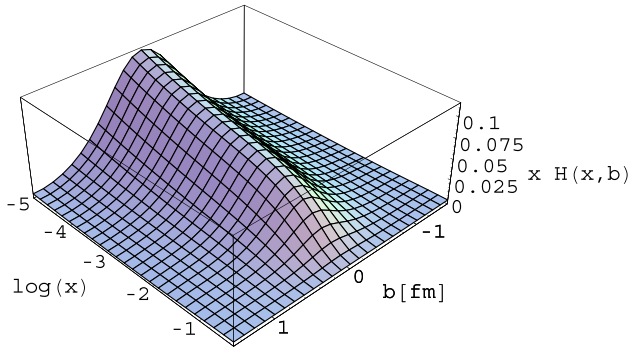


Fig. 268 Transverse profiles for the up quark distribution in transverse coordinate space as a function of x

Depending on the polarization of the quark and the nucleon states, the leading-twist quark GPDs contain eight independent distributions. The GPDs can be measured in many different experiments, for example, DVCS and hard exclusive meson production. Experimental efforts have been made at various facilities, including HERMES at DESY, Jefferson Lab, and COMPASS at CERN. It will be a major focus of the future EIC as well.

Nucleon tomography in terms of the GPDs is best illustrated in the impact parameter dependent parton distribution of Eq. (10.73). From that, we can define the transverse quark density profile [3249]:

$$\rho_q(x,\boldsymbol{b}) = \int \frac{\mathrm{d}^2\boldsymbol{\Delta}}{(2\pi)^2} e^{-i\boldsymbol{\Delta}\cdot\boldsymbol{b}} H_q(x,-\boldsymbol{\Delta}^2)\,. \quad (10.83)$$

An important feature of the above distribution is how it changes with longitudinal momentum fraction x. In Fig. 268, we show the transverse density profile for the up quark from the GPD parameterizations of [3226]. The plot shows that the transverse profile in coordinate space becomes wider at smaller x. At large x, however, it approaches a point-like structure, which means there is no t dependence of the GPD quark distribution, a result consistent with large-x power counting for GPDs [3262]. One of the primary goals of the GPD program at the JLab-12GeV and the EIC is to map out the x-dependence of the GPDs and the tomographic images for both quarks and gluons.

Most interestingly, when the nucleon is transversely polarized, the parton distribution in the transverse plane will be asymmetric due to the contribution from the GPD E [3249],

$$\rho_q^X(x,\boldsymbol{b}) = \int \frac{\mathrm{d}^2\boldsymbol{\Delta}}{(2\pi)^2} e^{-i\boldsymbol{\Delta}\cdot\boldsymbol{b}}\left(H_q(x,-\boldsymbol{\Delta}^2) + \frac{i\Delta_Y}{2M}E_q(x,-\boldsymbol{\Delta}^2)\right)$$

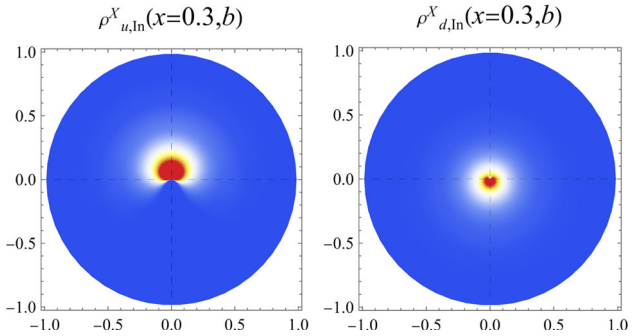


Fig. 269 Plots of the intrinsic quark densities $\rho^X_{q,\mathrm{In}}(x,\boldsymbol{b})$ for both u and d quarks in a transversely polarized proton (in the X direction) at $x = 0.3$. Both the u and d quark densities are shifted in the Y direction and contribute to the angular momentum J^X. While the u contributions are positive ($+Y$ direction) and the d contributions are negative ($-Y$ direction). These plots are adopted from Ref. [3220]

$$= \mathcal{H}_q(x,\boldsymbol{b}) - \frac{1}{2M}\frac{\partial}{\partial b^Y}\mathcal{E}_q(x,\boldsymbol{b}) \,, \tag{10.84}$$

where $\mathcal{H}_q(x,\boldsymbol{b})$ and $\mathcal{E}_q(x,\boldsymbol{b})$ are the 2-dimensional Fourier transformations of $H_q(x,-\boldsymbol{\Delta}^2)$ and $E_q(x,-\boldsymbol{\Delta}^2)$, respectively, and the nucleon is polarized in the X direction. This asymmetric distribution has attracted strong interest in the hadron physics community and it was argued that it might be related to the single spin asymmetry phenomena in hadronic processes [3249]. It has also been found in a lattice simulation [3263].

In order to factor out the transverse displacement from the nucleon's center of momentum and its contribution to the transverse polarization, one can introduce an intrinsic quark density [3198],

$$\begin{aligned}\rho^X_{q,\mathrm{In}}(x,\boldsymbol{b}) &= \int \frac{\mathrm{d}^2\boldsymbol{\Delta}}{(2\pi)^2} e^{-i\boldsymbol{\Delta}\cdot\boldsymbol{b}}\Big[H_q(x,-\boldsymbol{\Delta}^2) \\ &\quad + \frac{i\Delta_Y}{2M}\Big(H_q(x,-\boldsymbol{\Delta}^2)+E_q(x,-\boldsymbol{\Delta}^2)\Big)\Big] \\ &= \mathcal{H}_q(x,\boldsymbol{b}) - \frac{1}{2M}\frac{\partial}{\partial b^Y}\Big(\mathcal{H}_q(x,\boldsymbol{b})+\mathcal{E}_q(x,\boldsymbol{b})\Big) \,,\end{aligned} \tag{10.85}$$

from which one can reproduce the transverse polarization sum rule; see Sect. 10.3. In Fig. 269, we show the intrinsic transverse density for u and d quarks at $x = 0.3$ from the analysis of the GPD quark distribution of [3220]. Clearly, the quarks have non-zero transverse displacement, which contributes to the transverse angular momentum of the nucleon.

The theoretical framework has been well developed for the GPD studies with established QCD factorization for the associated exclusive processes [1288–1290]. Higher order perturbative QCD corrections have been calculated in a number of publications [1290,3264–3272]. The first computation of next-to-next-leading order corrections for DVCS has also been reported recently [3273]. However, since GPDs depend on three variables (x,ξ,t) in addition to the scale variable μ, it is much more difficult to extract them from experiment than PDFs (which only depend on x).

Pioneering phenomenological work has been carried out in Refs. [1286,3226,3230,3269,3274,3275]. In the last years, progress has also been made toward a global analysis of GPDs from a wide range of experiments [3220,3276–3282]. Especially, the twist-2 and twist-3 results were re-derived with an optimal light cone coordinate and full kinematics adopted [3280–3282]. A dedicated program based on earlier developments of Ref. [3274] has been proposed in Ref. [3220]. All these theory advances are crucial for a successful campaign to determine GPDs from DVCS and other hard exclusive processes measured at JLab-12 GeV and the planned Electron-Ion Collider.

Lattice QCD can be used to study these GPDs as well. Employing the LaMET formalism, exciting results on the x-dependence of the GPD quark distributions have already been obtained [671,3283]. We expect many more such simulations to emerge in the future, as well as combined fits to experimental and lattice data.

Transverse Momentum Dependent Parton Distributions

Theoretical studies of TMDs started long ago (see, for example, Ref. [1272]). In recent years great progress was made in the exploration of these distribution functions and the associated single spin asymmetry phenomena. In particular, TMDs provide not only an intuitive illustration of nucleon tomography, as we discussed above, but also the important opportunities to investigate the specific nontrivial QCD dynamics associated with their physics: QCD factorization, universality of the parton distributions and fragmentation functions, and their scale evolutions.

Different from the collinear PDFs discussed in Sect. 10.2, the TMD parton distributions can not be studied in inclusive processes. We have to go beyond that and explore semi-inclusive hard processes, where a hard momentum scale is involved in addition to the transverse momentum of the final-state particle produced. For example, we can study the TMD quark distributions in semi-inclusive DIS (SIDIS), where the virtual photon (with virtuality Q) scatters off the hadron and produces a final state hadron in the current fragmentation region. The hadron's transverse momentum $P_{h\perp}$ has to be much smaller than the hard momentum Q. Because of $P_{h\perp} \ll Q$, this process can be factorized into the TMD quark distribution convoluted with the TMD fragmentation function. Similarly, the Drell–Yan lepton pair production (or W/Z-boson, Higgs boson production) in hadronic collisions can be described by the convolution of two TMD parton distributions with transverse momentum $q_\perp \ll Q$. A related process in e^+e^- annihilation into two back-to-back hadrons can be factorized as a convolution of two TMD fragmentation functions.

	U	L	T
U	f_1		$h_1^\perp$
L		g_{1L}	$h_{1L}^\perp$
T	$f_{1T}^\perp$	g_{1T}	$h_1, h_{1T}^\perp$

Fig. 270 The leading order transverse momentum dependent quark distributions depend on the polarization of the quark (rows) and the nucleon (columns)

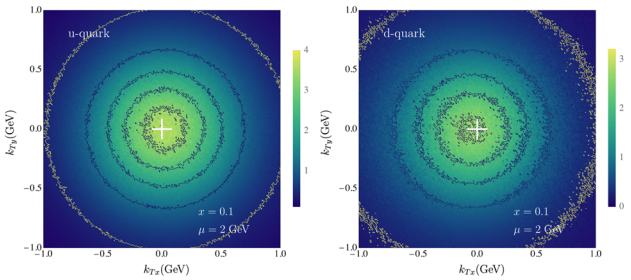


Fig. 271 The quark Sivers function represents the asymmetric distribution of a quark in transverse momentum space when the nucleon is polarized along the $\hat{y}$-direction. This plot is adopted from a phenomenological study [3286]

The TMD quark distributions can be defined by the following matrix [1272,1284,1297,1310,3247,3248],

$$\hat{\mathcal{M}}_{\alpha\beta}(x, k_\perp) = \int \frac{dy^- d^2 y_\perp}{(2\pi)^3} e^{-ixP^+\cdot y^- + i\vec{k}_\perp\cdot\vec{y}_\perp} \times \langle PS|\overline{\Psi}_\beta(y^-, y_\perp)\Psi_\alpha(0)|PS\rangle \,, \qquad (10.86)$$

where x is the longitudinal momentum fraction and $k_\perp$ the transverse momentum carried by the quark. The quark field $\Psi(y)$ contains a gauge link as defined in Eq. (10.76). This definition contains a light-cone singularity from higher order corrections. The regulation and subtraction procedure defines the scheme of the TMD distributions. Obviously, in the theoretical limit in which contributions from all orders and all twists are taken into account, physical quantities have to be scheme independent. (At the most simple level this was actually shown explicitly in Ref. [3284] but it has to be true also non-perturbatively.) Often, however (e.g. in event generators), rather specific models are used for which this is not the case. In these cases the fitted TMDs and thus the result of hadron tomography can be strongly scheme/model dependent (see e.g. Ref. [3285]). Calculating the model-specific matching factors or functions between such a scheme and the usual TMD factorization scheme is often not possible. This should flag a warning that one has to be careful when comparing fitted TMDs from different sources. The leading order expansion of the above matrix contains eight independent quark TMDs, depending on the polarization of the quark (varying horizontally) and the nucleon (varying vertically) in Fig. 270. The gauge link direction plays an essential role in the naive time-reversal-odd TMD quark distributions, including the quark Sivers function $f_{1T}^\perp(x, k_\perp)$ and Boer-Mulders function $h_1^\perp(x, k_\perp)$.

The spin-average quark distributions are symmetric in the transverse plane. However, if the nucleon (or the quark) is transversely polarized, the quark distribution shows an azimuthal asymmetry. In particular, the TMD quark Sivers functions quantify these asymmetries in the transverse momentum space. In Fig. 271, we show one of the resulting distributions for the quark (averaged over x) in transverse momentum space in a transversely polarized nucleon. The TMD distribution comes from the fit to the associated single transverse spin asymmetries in semi-inclusive hard processes [3286].

For the quark Sivers function, because of the initial/final state interaction (represented by the gauge link pointing to $-\infty$ or $+\infty$ in the quark distribution definition) difference, they differ by signs for semi-inclusive hadron production in DIS (SIDIS) and Drell–Yan processes [1296,1297,1309,1310,3287,3288]. This leads to a sign change between the SSAs in SIDIS and Drell–Yan processes,

$$\text{Sivers SSA}|_{\text{DY}} = -\text{Sivers SSA}|_{\text{DIS}} \,. \qquad (10.87)$$

This nontrivial result still holds when gluon radiation contributions are taken into account [1320,3289–3291]. It is very important to test this nontrivial QCD prediction by comparing the SSAs in these two processes. The Sivers single spin asymmetries in SIDIS processes have been observed by the HERMES [3292,3293], COMPASS [3294–3297], and JLab [3298,3299] collaborations. There have been significant efforts to measure the Sivers asymmetries in Drell–Yan process at COMPASS [3300] and that of $W^\pm$ production at RHIC [3301]. The analyses of these data provide an indication for a sign change [3302], but no proof. More precise measurements are needed to confirm this crucial property.

In TMD factorization for semi-inclusive hard processes [1267,1272,1280,1283,1911], collinear and soft gluon radiations are factorized into the TMD parton distributions or fragmentation functions and the associated soft factors. As for the integrated parton distribution functions, these gluon radiation contributions can be resummed to all orders by solving the relevant evolution equations. This resummation is referred to as TMD or Collins–Soper–Sterman resummation [1280]. As a result, the factorization simplifies the differential cross section to a convolution of soft factor-subtracted TMD distributions and/or fragmentation functions [1267], where the hard momentum scale is chosen as factorization scale $\mu_F = Q$. As an example, in Fig. 272, we show the TMD up

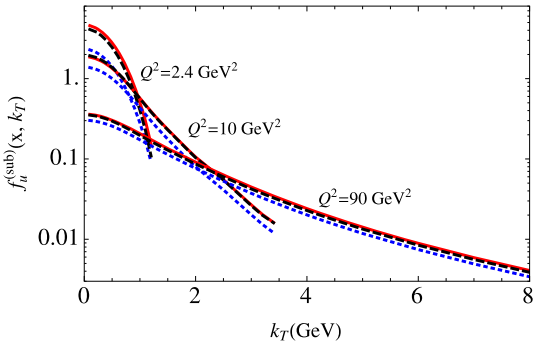


Fig. 272 TMD up-quark distributions $f_u^{(\mathrm{sub})}(x, k_\perp)$ (for $x = 0.1$) as functions of the transverse momentum $k_\perp$ (GeV) at three different scales $Q^2 = 2.4, 10, 90$ (GeV^2). This plot is adopted from Ref. [3284]

quark distribution $f^{(\mathrm{sub})}(x = 0.1, k_T, \mu_F = Q)$ as a function of the transverse momentum at different scales. Clearly, the resummation/scale evolution leads to broadening effects for TMD distributions at higher scales. Based on these developments, recent global analyses have achieved high precision for the unpolarized TMD quark distribution and fragmentation functions fitted to data from various semi-inclusive hard processes [3303–3306]. Of course, further theoretical developments are still needed to answer crucial questions concerning TMD factorization at lower scale SIDIS and the question how non-perturbative effects affect the matching between the TMDs and collinear PDFs [3306–3309]. Upcoming data from the JLab-12 GeV program should significantly improve our understanding of these issues in the near future.

In addition, progress has been made in studying the scale evolution for the quark Sivers function and the associated quark–gluon–quark correlation functions [1317,1318,3310–3314], and the QCD resummation for the SSA observables [3314–3318]. These resummation effects have been taken into account in a recent phenomenological study of all single spin asymmetries associated with the quark Sivers function in a global analysis using (N^{3}LO) evolution for the TMDs [3286,3302].

There have also been significant progresses toward lattice calculations of TMDs [672,675,676,3319–3330] in the last few years, partially again based on the LaMET formalism. The TMD evolution kernel was calculated from lattice QCD [672,674,676,3327,3328] for perturbative and non-perturbative [3305,3306,3331] impact parameters b and the result agreed with that of a fit to experimental data [3285]. This motivates great hopes for future combined TMD fits to experimental and lattice data. We also expect lattice simulation of the single spin asymmetries associated with the quark Sivers function, using the perturbative matching derived in Ref. [3323].

More recently, important developments have taken place addressing the connections between the TMD formalism and small-x saturation physics. Small-x gluon saturation is best described in the color-glass-condensate (CGC)/color-dipole formalism [3332–3336], for which the so-called unintegrated gluon distributions (UGDs) are essential elements. What has been shown in the recent papers [3337–3343] is that these UGDs are the same as the TMD gluon distribution functions at small-x. Meanwhile, considerable progress has also been made in computing Sudakov double logarithms in the small-x formalism [3344–3349]. These computations provide a solid theoretical foundation for further rigorous investigations that probe the dynamics of the saturation regime with hard processes. We anticipate that in the foreseeable future a unified picture of nucleon structure will emerge that covers the whole kinematic domain, including small and large x.

Direct access to the Wigner distributions

It was generally believed that the parton Wigner distributions are not directly measurable in high energy scattering. However, it was realized recently that the Wigner distribution could be measured through hard exclusive processes [3350–3352]. In particular, it was shown in Ref. [3350] that the small-x gluon Wigner distribution is connected to the color dipole S-matrix in the CGC formalism [3332–3336], that diffractive dijet production in ep/eA collisions [3350,3353–3357] may provide a direct probe of this gluon Wigner distribution. Additionally, semi-hard gluon radiation in this process or 'trijet' diffractive production has been shown to probe the color-dipole amplitude in the adjoint representation [3358,3359]. This demonstrates that a new class of diffractive processes, including semi-inclusive diffractive DIS [864] can provide crucial information on the gluon Wigner distributions at small-x. Extension to other processes, in particular, those at moderate and large x will be interesting to follow as well. We expect more research along this direction in the future.

To summarize this subsection: There has been great progress in both experiment and theory for GPD and TMD physics. Of course, challenges are still there in both fields. We would like to emphasize that data from future experiments, including the 12 GeV upgrade of JLab, COMPASS and the planed EIC experiments, together with theory developments, will lead us to a complete 3D tomography of the nucleon.

11 QCD at high energy

Conveners:

Gudrun Heinrich and Eberhard Klempt

The core of high energy collisions consists in a hard scattering of two partons, where the momentum transfer is very large and therefore the process can be calculated perturbatively. The enormous progress in the calculation of QCD corrections beyond the leading order in perturbation theory is described by Gudrun Heinrich. The scattered partons can

emit soft or nearly collinear gluons. In kinematic regions where the phase space for such emissions is restricted, large logarithms arise, which can spoil the perturbative convergence. Due to the universal structure of infrared divergent QCD radiation, such logarithms can be resummed analytically to all orders to restore the predictive power of the perturbative description in these kinematic regions, as described by Simone Marzani.

At the intermediate stage between the hard interaction and hadronization, the radiation of gluons from quarks and the splitting of gluons into secondary quarks and gluons, forming a cascade of emissions, can be described by parton showers. The development of these parton showers and our understanding of these processes are described by Frank Krauss.

Once these showers of partons have evolved to low energies, the process of hadron formation sets in. At these energies, the strong coupling is large, such that bound states are formed, which cannot be described perturbatively anymore. The description of hadronization needs to rely on parameters extracted from data. These parameters are tuned in Monte Carlo simulations. Torbjörn Sjöstrand gives a detailed view of different stages of the collision process and of their simulation.

The reconstruction of jets by reliable jet algorithms and the identification of the primary source, gluons or quarks of a certain flavor, is very important to extract information about the underlying particle dynamics from the data. Jet substructure variables can provide essential information about the decay of heavy particles leading to boosted jets, as described by Bogdan Malaescu, Dag Gillberg, Steven Schramm, and Chris Young.

11.1 Higher-order perturbative calculations

Gudrun Heinrich

11.1.1 Introduction

The property of asymptotic freedom of QCD, together with the fact that short- and long-distance effects in QCD can be factorized up to power corrections, allows us to describe processes with high momentum transfer as a perturbative series in the strong coupling α_s, as illustrated in Eq. (11.1). For example, the cross section for a process such as the production of a Higgs boson through the collision of two protons with momenta p_a and p_b, $p_a + p_b \to H + X$, has the form

$$\sigma_{pp\to H+X} = \sum_{i,j} \int_0^1 dx_1\, f_{i/p_a}(x_1,\alpha_s,\mu_F) \times \int_0^1 dx_2\, f_{j/p_b}(x_2,\alpha_s,\mu_F)\, \hat{\sigma}_{ij\to H+X}(\alpha_s(\mu_R),\mu_R,\mu_F) + \mathcal{O}\left(\frac{\Lambda}{Q}\right)^p , \tag{11.1}$$

where the partonic cross section $\hat{\sigma}_{ij\to H+X}$ can be expanded as

$$\hat{\sigma}_{ij\to H+X} = \alpha_s^2 \hat{\sigma}^{(0)} + \alpha_s^3 \hat{\sigma}^{\mathrm{NLO}} + \alpha_s^4 \hat{\sigma}^{\mathrm{NNLO}} + \cdots \tag{11.2}$$

The renormalization of ultraviolet singularities appearing in loop corrections leads to a dependence of both α_s and the partonic cross section on the renormalization scale μ_R. Similarly, the absorption of collinear singularities into the "bare" parton distribution functions leads to a dependence on the factorization scale μ_F. The functions $f_{i/p_a}(x,\alpha_s,\mu_F)$ are the (physical) parton distribution functions (PDFs), which can be interpreted as probabilities to find a parton of type i with momentum fraction x of the "parent" momentum p_a in a proton (or, more generally, a hadron). This makes an assumption of collinearity of the parton's momentum with p_a, therefore the factorisation described by Eq. (11.1) is also called *collinear factorisation*. For more details about parton distribution functions we refer to Sect. 10.2. Factorization holds up to the so-called power corrections of order $(\Lambda/Q)^p$, where the power p is process-dependent and larger than one for observables that are sufficiently inclusive over the hadronic final state, see, however, Ref. [180].

In Eq. (11.2), the partonic cross section at leading order (LO) in an expansion in α_s is denoted by $\hat{\sigma}^{(0)}$, where for the sake of clarity the powers of the strong coupling have been extracted. The next-to-leading order (NLO) cross section comes with one more power of α_s relative to LO, the next-to-next-to-leading order (NNLO) cross section with two more α_s powers than LO, etc. Of course such an expansion also can be performed for the electroweak corrections, however, as $\alpha/\alpha_s(M_Z) \simeq 0.1$, the QCD corrections are usually larger, except in kinematic regions where logarithms of the form $\alpha \ln(M_W^2/\hat{s})$ grow large. The dependence of the cross section $d\sigma_{pp\to H+X}$ on μ_R and μ_F is an artifact of the truncation of the perturbative series. Therefore, the dependence on these unphysical scales becomes weaker as more perturbative orders are calculated. The variation of the cross section as these scales are varied around a central scale – which should be chosen to be close to the energy at which the hard interaction takes place – therefore can be used as an estimate of the theoretical uncertainty due to missing higher orders.

Higgs boson production in gluon fusion is somewhat special, as the leading order amplitude is already loop-induced, and because the NLO QCD corrections are of the order of 100%, which makes the inclusion of QCD corrections beyond NLO a necessity for a satisfactory description of the data.

The perturbative expansion in powers of α_s is particularly reliable for inclusive observables. If the phase space for QCD radiation is restricted, large logarithms can appear, which spoil the convergence of the perturbative series in α_s. This requires so-called resummation, as described in detail

in Sect. 11.2. Here we will focus on calculations at a fixed order in the strong coupling.

11.1.2 Developments and status

Next-to-leading order QCD corrections

The development of systematic techniques for NLO QCD corrections started in the 1980s with seminal work on e^+e^-–annihilation to jets [3360–3362] and hadron–hadron scattering [197], followed by pioneering developments of techniques for one-loop calculations based on Feynman diagrams and tensor reduction [3363–3366]. In parallel, subtraction methods for soft and collinear real radiation were established [199,3367], leading to the first differential NLO calculations of 2-jet production in hadronic collisions [3368–3370], while the first NLO calculation of 3-jet production in hadronic collisions only appeared 10 years later [3371].

The calculation of NLO cross sections for $(n-2)$-jet production in hadronic collisions (or for $(n-1)$-jet production in e^+e^-–annihilation, as well as the calculation of amplitudes obtained by crossing) involves n-parton one-loop amplitudes and $(n+1)$-parton tree-level amplitudes with up to one unresolved (soft or collinear) parton, see Fig. 273. The efficient calculation of one-loop n-point amplitudes for $n \geq 5$ represented a major challenge in the 1990s and led to the development of more efficient methods to calculate one-loop n-point amplitudes, based on the idea to exploit analytic properties of loop integrals if propagators are put on-shell (so-called "unitarity cuts") [200,3372–3374]. The emergence of methods to perform these cuts numerically [3375–3377], together with the automation of subtraction methods for unresolved real radiation at NLO, led to a new level of efficiency, resulting in the availability of NLO QCD predictions for multi-particle scattering which were considered unfeasible some years before, such as 5-jet production at the LHC [3378], top-quark pair production with up to 3 jets [3379], $Wb\bar{b}$ production with up to 3 light jets [3380], or the NLO QCD and EW corrections to off-shell $t\bar{t}W$ production at the LHC, involving one-loop 10-point integrals [3381]. It also led to the development of automated tools providing one-loop amplitudes for fully differential NLO predictions [3376,3382–3388]. This remarkable jump in efficiency is often called the "NLO revolution".

Beyond NLO

The next step, towards fully differential NNLO predictions, required not only major progress in the calculation of two-loop integrals, but also the development of subtraction schemes for infrared (IR) divergent real radiation where up to two particles can be unresolved.

Multi-loop amplitudes

First, the developments regarding loop integrals with two or more loops will be considered. A very important parameter characterising a Feynman integral besides the number of loops and legs is the number of kinematic/mass scales. As an example, it is instructive to consider the development of the calculation of 2-loop 4-point integrals (2 loops, 4 legs): the analytic calculation of the planar [3389] and non-planar [3390] two-loop 7-propagator diagrams with massless propagators and light-like legs has been performed in 1999, numerical checks of these results in the Euclidean region were performed in Ref. [3391], the calculation of such integrals with one off-shell leg was completed soon after [3392,3393]. The first results for two-loop 4-point amplitudes with *massive* propagators have been achieved by a numerical method [3394,3395], leading to NNLO predictions for top quark pair production in hadronic collisions [3396]. The analytic calculation of two-loop 4-point integrals with two massive legs and massless propagators, entering for example the production of Z-boson pairs or a W^+W^- pair, was completed around the year 2015 for both the on-shell as well as for the off-shell case [3397–3402]. However, the step to include massive propagators leads out of the function class (so-called multiple polylogarithms) describing the above-mentioned objects analytically. Therefore, the calculation of two-loop 4-point integrals with both massive propagators as well as massive final state particles was performed numerically before analytic results appeared, examples are the two-loop QCD corrections to Higgs boson pair production [3403,3404], Higgs+jet production [3405], $gg \to ZH$ [3406–3408] and $gg \to VV$

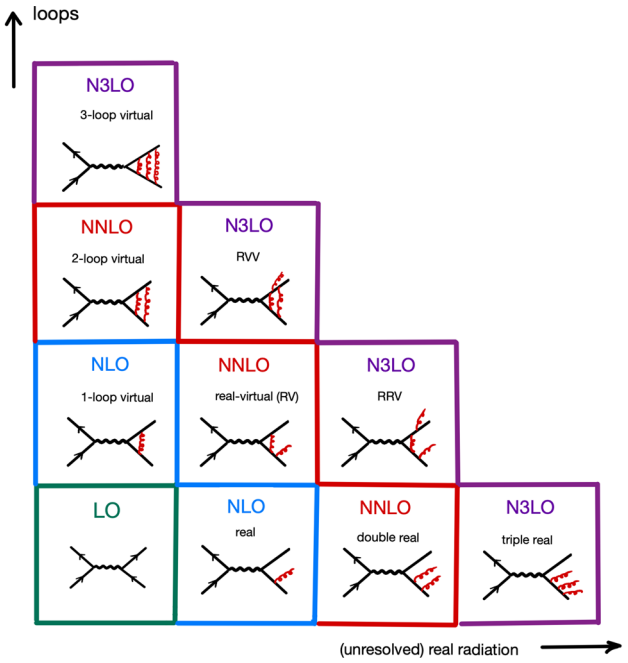


Fig. 273 Building blocks of an N^{x}LO calculation for processes where the leading order is at tree level (in contrast to loop-induced). The higher order diagrams are only representatives of their class, the number of diagrams grows rapidly with the perturbative order

Table 48 Timeline for the availability of full NNLO cross sections at hadron colliders (or NLO cross sections for loop induced processes) based on two-loop four-point or five-point integrals with an increasing number of mass scales. "HTL" denotes the heavy top limit, "QCD-EW DY" denotes mixed QCD-electroweak corrections to the Drell–Yan process

Indep. kinem. scales	Massive/off-shell legs	Internal masses	Process	Full σ
$2 \to 2$				
2	0	0	$\gamma\gamma$	2011
2	0	0	$j\,j$	2017
2	0	0	$\gamma + j$	2017
3	2	1	$t\bar{t}$	2013
3	2	0	VV	2014
4	2	0	VV'	2015
3	1	0	$V + j$	2015
3	1	0	$H + j$ (HTL)	2015
4	2	1	HH	2016
4	1	1	$H + j$	2018
3	0	1	$gg \to \gamma\gamma$	2019
4	2	1	$gg \to ZZ$	2020
4	2	1	$gg \to WW$	2020
5	2	1	$gg \to ZH$	2021
4	2	1	QCD-EW DY	2022
$2 \to 3$				
4	0	0	3γ	2019
4	0	0	$\gamma\gamma\,j$	2021
4	0	0	$3\,j$	2021
5	1	0	$Wb\bar{b}$	2022

with massive loops [3409–3411], where V denotes a massive vector boson. Thus, one can roughly say that it took almost 20 years to increase the number of independent mass scales entering these diagrams from two (s_{12}, s_{23}), to five $(s_{12}, s_{23}, m_t, m_V, m_{V'})$, where $s_{ij} = (p_i + p_j)^2$ and V' denotes a boson with invariant mass different from V. The timeline of available predictions for (differential) cross sections based on these integrals is shown in Table 48, illustrating how an additional mass scale increases the complexity. It is noteworthy that all integrals with massive propagators, i.e. with a non-zero entry in the third column, have been calculated with numerical methods. For processes with jets in the final state, the subtraction of IR divergent real radiation was the bottleneck, not the availability of the two-loop integrals.

More details on the methods employed for these calculations can be found e.g. in Refs. [215,3412,3413].

Real radiation

For many $1 \to 3$ or $2 \to 2$ processes, such as $e^+e^- \to 3$ jets or di-jet production in hadronic collisions, the knowledge of the two-loop amplitudes was not the main bottleneck on the way to fully differential predictions at NNLO. Efficient schemes to treat the infrared (IR) divergent real radiation needed to be developed, and the emergence of several schemes led to an explosion in the availability of NNLO results for LHC processes with up to two particles or jets in the final state after 2015, a development which is sometimes referred to as the "NNLO revolution". The main methods to treat IR divergent real radiation beyond NLO can roughly be classified into two categories: (i) methods based on subtraction, and (ii) methods based on partitions of the phase space into IR-sensitive regions and hard regions, sometimes also called "slicing methods". The latter introduce a dependence on a resolution variable which cancels once the IR-sensitive and hard regions are combined. Subtraction methods aim at a local subtraction of the IR singular structures, i.e. a cancellation of singularities point-wise in the phase space, while for slicing methods the compensations are non-local. This non-locality can lead to large numerical cancellations, however, power corrections in the resolution variable can be included to mitigate their impact. Reviews about recent developments in IR subtraction schemes can be found e.g. in Refs. [214,3413,3414]. The main methods are summarized in Table 49.

The extension of methods to isolate IR divergent real radiation to N^{3}LO, i.e. the case of up to three unresolved partons, in particular in the presence of tagged colored particles or jets, is one of the current challenges in the field of high precision perturbative QCD calculations.

While the complete automation of NNLO calculations is probably not feasible in view of efficiency optimisations that are process specific, libraries with a large collection of codes providing NNLO predictions are available, such as MATRIX [3452], NNLOJET [3419] or MCFM [3459].

Current frontier and recent developments

As shown in Fig. 273, the calculation of N^{x}LO corrections to processes with $(n-2)$ identified particles or jets in the final state in hadronic collisions (where the leading order is a tree amplitude, as contrasted to loop-induced amplitudes such as Higgs boson production in gluon fusion), requires the calculation of amplitudes with $x - j$ loops and $n + j$ legs, where $j = 0, \ldots, x$. The current frontier is, roughly speaking, $x + n \geq 6$, having in mind $2 \to 3$ processes at NNLO, $2 \to 1$ processes at N^{3}LO or 4-loop form factors. However, the type of the involved particles is very important for the complexity of the calculation: all available complete N^{3}LO results to date involve only color singlets in the final state, see e.g. Refs. [3460–3464] for the Drell–Yan process, Refs. [206,3465–3468] for Higgs boson production in gluon fusion in the heavy top limit and Ref. [3469] for VH production. Inclusive N^{3}LO results are also available for Higgs [3442] and Higgs pair [3470] production in vector boson fusion (VBF), Higgs pair production in gluon fusion in the heavy top limit [3471] and Higgs production in bottom quark fusion

Table 49 Methods for the isolation of IR divergent real radiation at NNLO and up to three examples of their application

Method	NNLO examples
Subtraction	
Antenna subtraction [3417,3418]	$e^+e^- \to 3$ jets [3415,3416], $pp \to 2$ jets [3419,3420], $pp \to WHj$ [3421]
Sector-improved residue subtraction [3422–3424]	$pp \to t\bar{t}$ [3396], $pp \to W + c$-jet [3425], $pp \to 3$ jets [3426]
Nested soft-collinear subtraction [3429–3431]	$pp \to VH$ [3427,3428], VBF H [3432], mixed QCD-EW to Drell–Yan [3433–3435]
ColorFul [3436,3437]	$e^+e^- \to 3$ jets [3438], $H \to b\bar{b}$ [3439]
Projection to Born [3440,3442]	VBF H [3440], VBF HH [3441], single top [3443,3444]
Local analytic subtraction [3445–3447]	$e^+e^- \to 2$ jets [3445]
4-dimensional schemes [214,3449]	$\gamma^* \to t\bar{t}$ [3448] (inclusive)
Slicing	
q_T [3450,3451]	VV' [3452], $t\bar{t}$ [3453], mixed QCD-EW to Drell–Yan [3454,3455]
N-jettiness [1816,2020,2021]	$V + j$ [2020,3456], $H + j$ [3457,3458], di-boson [3459]

[3472,3473]. The extension to colored final states requires advances in the treatment of IR divergent real radiation, for example N-jettiness soft and beam functions at this order, see e.g. Refs. [2027,3474–3478] or triple collinear splitting functions [3479,3480], see also Ref. [3414] for more details.

Another ingredient which is needed to be consistent at this order are N^{3}LO parton distribution functions, see Ref. [3101] for recent progress.

For processes such as Higgs boson decays into heavy quarks or the production of heavy quarks at e^+e^- colliders at three loops, massive 3-loop form factors need to be calculated, and the presence of the additional mass scale substantially increases the complexity of the calculation. Analytical and semi-numerical methods have pushed these calculations quite far [3481–3487].

Only very few results for three-loop amplitudes with more than three legs exist. Remarkable recent results are the 3-loop amplitudes for $q\bar{q} \to \gamma\gamma$ [3488], $gg \to \gamma\gamma$ [3489], $q\bar{q} \to q'\bar{q}'$ [3490], $q\bar{q} \to gg$ [3491] and $gg \to gg$ [3492]. For the case of one massive external leg, results for planar master integrals exist [3493].

Another highlight on the 3-loop front is the calculation of the NNLO corrections to Higgs boson production in gluon fusion with full top quark mass dependence [3494], which involves the calculation of 3-loop integrals with two mass scales.

Considering $x = 2, n = 5$, i.e. processes involving 2-loop 5-point integrals, again the number of mass scales is the critical measure of complexity. Results for complete cross sections have been achieved for processes involving only massless particles: $pp \to 3\gamma$ [3495,3496], $pp \to \gamma\gamma j$ [3497–3500] and $pp \to 3$ jets [3426], as well as for the process $pp \to Wb\bar{b}$ [3501,3502].

At four loops, the computation of form factors has seen enormous progress in the past few years [3503], culminating in the calculation of the complete analytic expressions for the photon-quark and the Higgs-gluon form factors at 4-loop order [1959]. These form factors will serve as building blocks for a future complete N^{4}LO calculation of the Drell–Yan process and Higgs boson production in gluon fusion in the heavy top limit. N^{4}LO results for $gg \to H$ in the large-N soft-virtual approximation already exist [3504]; results for

soft corrections to deeply inelastic scattering (DIS) at 4-loop order are also available [3505].

Results at five and more loops mainly involve two-point functions, entering for example the calculation of β-functions, such as the 5-loop β-function in QCD [3506–3509] or in scalar theories [3510]. Five-loop contributions to the anomalous magnetic moment of the electron have been calculated in Refs. [3511–3514]. Results for anomalous dimensions at six [3515,3516] or seven loops and beyond [3517,3518] are available for scalar theories.

11.1.3 Phenomenology

The progress described above concerning precision calculations in QCD has led to a plethora of phenomenological results at unprecedented precision, such as determinations of the strong coupling described in Sect. 3.2, determinations of the W-boson mass, precision measurements in Higgs- and electroweak physics (see Sects. 12.4, 12.3) and top quark physics (see Sect. 12.5). Advances in jet algorithms and jet substructure measurements (see Sects. 12.2 and 11.5)) also play a major role in the LHC precision program. Cross sections for inclusive jet production can be measured at the LHC with an uncertainty of about 5% for central rapidities. This poses challenges on the theory side, in particular it requires a judicious choice of the central scale, as some choices can induce infrared-sensitive contributions [3519]. Furthermore, the transverse momenta of the jets can reach values around 4 TeV, making the combination of NNLO QCD corrections with NLO electroweak corrections indispensable to describe the high-p_T region correctly. In order to make such precision calculations usable efficiently for PDF fits or α_s determinations, it is also important to have them available in a flexible format, for example in the form of fast interpolation grids, see e.g. Ref. [3520] for more details.

Together with ongoing progress in reducing PDF uncertainties, as well as in controlling non-perturbative effects and parton shower uncertainties (see e.g. Sects. 11.4, 11.3), precision phenomenology at hadron colliders has reached a level which was unthinkable 50 years ago when QCD was "born".

11.1.4 Outlook

The calculation of perturbative higher order corrections in QCD at high energies is a success story. Inventive new methods have been developed to deal with the increasing level of complexity at higher perturbative orders. These technical advances were accompanied by a better understanding of important phenomenological concepts, such as infrared-safe observables and jet algorithms, and of the limitations of fixed-order perturbation theory. These developments went hand in hand with increasingly precise measurements of QCD processes at high energy colliders, and they are important pillars of the search for physics beyond the Standard Model.

While the uncertainties due to the truncation of the perturbative series were the dominant theory uncertainties for a long time in the 50-years history of QCD, for processes where the N^{3}LO level of QCD corrections is reached it became clear that other uncertainties, such as PDF uncertainties, parton shower uncertainties, quark mass effects, parametric uncertainties (e.g. in α_s, m_t) or power-suppressed and non-perturbative contributions need to be considered with high priority as well. Being able to control them will play an important role in the next 50 years of QCD and in the search for physics beyond the Standard Model.

11.2 Analytic resummation

Simone Marzani

11.2.1 Large logarithms

QCD processes that involve high-momentum transfer, usually referred to as "hard processes", can be described in perturbation theory. In this framework, theoretical precision is achieved by computing the cross section σ for an observable $\mathcal{V}$, which we assume having the dimension of an energy scale, including higher- and higher-order corrections in the strong coupling α_s, i.e. the so-called fixed-order expansion:

$$\sigma\left(\mathcal{V}\right) = \sigma_0 + \alpha_s\,\sigma_1 + \alpha_s^2\,\sigma_2 + \alpha_s^3\,\sigma_3 + \mathcal{O}(\alpha_s^4), \tag{11.3}$$

where the leading order (LO) contribution σ_0 is the Born-level cross section for the scattering process of interest. Subsequent contributions in the perturbative expansion σ_x constitute the (next-to)^x-leading (N^{x}LO) corrections. In the language of Feynman diagrams, each power of α_s corresponds to the emission of an additional QCD parton, either a quark or a gluon, in the final state or to a virtual correction. Note that, with respect to Eq. (11.2), the explicit dependence on the observable $\mathcal{V}$ has been highlighted.

Calculations of Feynman diagrams are plagued by the appearance of divergences of different nature. Loop-diagrams can exhibit ultra-violet singularities. Because QCD is a renormalizable theory, such infinities can be absorbed into a redefinition of the parameters that enter the Lagrangian. Throughout this discussion, it is understood that such renormalization has already occurred. Real-emission diagrams exhibit singularities in particular corners of the phase-space. More specifically, these singular contributions have to do with collinear, i.e. small-angle, splittings of massless partons and emissions of soft gluons, either off massive or massless particles. Virtual diagrams also exhibit analogous infra-red and collinear (IRC) singularities, and rather general theorems [3521–3523] state that such infinities cancel when real and virtual corrections are added together, thus leading to observable transition probabilities that are free of IRC singularities. Moreover, in order

to be able to use the perturbative expansion of Eq. (11.3), one has to consider observables $\mathcal{V}$ that are "IRC safe", i.e. measurable quantities that do not spoil the above theorems.

The theoretical community has put a huge effort in computing higher-order corrections, as discussed in detail in Sect. 11.1. One of the main challenges in this enterprise is the treatment of the infra-red region and the cancellation of the singular contributions between real and virtual diagrams. Furthermore, the emissions of soft and/or collinear partons are also problematic because they can generate large logarithmic terms in the perturbative coefficients σ_x, thus invalidating the fixed-order approach. The expansion of Eq. (11.3) works well if the measured value of the observable is $\mathcal{V} \simeq Q$, where Q is the scale which characterizes the hard process. However, it loses its predictive power if the measurement of $\mathcal{V} \ll Q$ confines the real radiation into a small corner of phase-space, while clearly leaving virtual corrections unrestricted. For IRC safe observables, soft and collinear singularities cancel, but logarithmic corrections in the ratio $\mathcal{V}/Q$ are left behind, causing the coefficients σ_x to become large, so that $\alpha_s^x \sigma_x \sim 1$. Because these logarithmic corrections are related to soft and/or collinear emissions, one can expect at most two powers of $L = \ln \frac{Q}{\mathcal{V}}$ for each power of the strong coupling:

$$\sigma(\mathcal{V}) = \sigma_0 + \alpha_s \left(\sigma_{12} L^2 + \sigma_{11} L + \dots\right) + \alpha_s^2 \left(\sigma_{24} L^4 + \sigma_{23} L^3 + \dots\right) + \mathcal{O}(\alpha_s^n L^{2n}). \tag{11.4}$$

All-order resummation is then a re-organization of the above perturbative series. For many observables of interest, the resummed expression exponentiates, leading to

$$\sigma(\mathcal{V}) = \sigma_0\, g_0(\alpha_s) \times \exp\left[L g_1(\alpha_s L) + g_2(\alpha_s L) + \alpha_s g_3(\alpha_s L) + \dots\right], \tag{11.5}$$

where g_0 is a constant contribution which admits an expansion in α_s. In analogy to the fixed-order terminology, the inclusion of the contribution g_{x+1}, $i \geq 0$, leads to (next-to)^x-leading logarithmic (N^{x}LL) accuracy.

Fixed-order Eq. (11.3) and resummed Eq. (11.5) expansions are complementary. On the one hand, fixed-order calculations fail in particular limits of phase-space, indicating the need for an all-order approach. On the other hand, all-order calculations are only possible if particular assumptions on the emission kinematics are made. Thus, the most accurate theoretical description for the observable $\mathcal{V}$ is achieved by matching the two approaches

$$\sigma^{\text{matched}}(\mathcal{V}) = \sigma^{\text{f.o.}}(\mathcal{V}) + \sigma^{\text{res}}(\mathcal{V}) - \sigma^{\text{d.c.}}(\mathcal{V}), \tag{11.6}$$

where the third contribution corresponds to the expansion of the resummation to the order we are matching to and it is subtracted in order to avoid double counting. For instance, if we were to match the resummed expression Eq. (11.5), computed to some logarithmic accuracy to a fixed-order calculation, see Eq. (11.4), performed at NNLO, $\sigma^{\text{d.c.}}$ would correspond to the expansion of the resummed result up to second order in the strong coupling, relatively to Born term. Furthermore, we note that, if the resummation is computed at high-enough accuracy, the dangerous logarithmic corrections cancel between $\sigma^{\text{f.o.}}$ and $\sigma^{\text{d.c.}}$ and all the large contributions are resummed in σ^{res}.

All-order resummation is possible because (squared) matrix element and phase-space factorize in certain kinematic limits. Different methods to achieve such factorization have been developed in the literature. For instance, factorization can be obtained by studying directly QCD amplitudes and cross-sections in the soft and collinear limits. Then, resummation can be achieved by iteratively identifying factorization and exponentiation properties of QCD matrix elements and cross-sections [171,3524,3525]. Other approaches instead introduce non-local correlation operators, such as Wilson lines, and exploit their renormalization group evolution [3526]. Finally, one can construct soft-collinear effective field theories (SCET) to describe the soft and collinear degrees of freedom of QCD [1801–1804,1831,1901,1903] (see Sect. 6.4, and, for instance, Ref. [3527] for an extensive review). There is a rich literature describing similarities and differences of the various resummation approaches, see e.g. [3528–3533]. In this presentation we will mostly follow the iterative point of view.

11.2.2 Transverse-momentum resummation

The transverse momentum ($Q_T = p_T^Z, p_T^W, p_T^H$) distribution of electroweak final states at hadron colliders is one of the most extensively investigated observables in QCD. Studies of Q_T spectra and related angular correlations of lepton pairs produced via the Drell–Yan (DY) process provide us with a useful testing ground for an even more interesting Higgs and new physics program. These processes are characterized by the presence of two distinct scales: the measured Q_T and the invariant mass of the final-state Q, which is close to the mass of the electroweak boson for on-shell production. Therefore, if we are interested in phase-space regions where $Q_T \ll Q$, large logarithmic corrections appear. They should be accounted for to all orders, in order to achieve an accurate theoretical description of these observable distributions.

Furthermore, one aspect of physics at hadron colliders that becomes important at small Q_T is the role of non-perturbative effects commonly attributed to the intrinsic transverse motion of partons within the proton. One may therefore view any opportunity to compare precise perturbative predictions with

accurate experimental data for DY lepton pairs as a chance to assess the size of non-perturbative physics; physics which also affects the Higgs Q_T spectrum.

The literature on Q_T resummation is vast and since the seminal papers, which date back to the late 1970s, early 1980s, e.g. [1280,3534], there has been a continuous effort in producing accurate theoretical predictions that can describe the experimental data. For example, high logarithmic accuracy [1947,3535–3542] has been achieved and computer programs that allow one to compute NNLL predictions matched to next-to-leading order (NLO) for the Q_T distribution in case of colorless final states in hadron collision have been available for a long time, e.g. [3536,3537,3543–3548]. Fixed-order predictions have reached NNLO accuracy and the resummation can be now performed to N^{3}LL accuracy [207,1946,3549–3552]. Results with partial N^{4}LL resummation have also been recently obtained [3464].

Moreover, observables such as ϕ^* [3553,3554] that exploit angular correlations to probe similar physics as Q_T, while being measured with even better experimental resolution, have triggered theoretical studies to extend the formalism of Q_T resummation to these new variables [3545,3555–3558].

In this section we review the main ingredients of Q_T resummation for an electroweak final state, i.e. Higgs or DY. For simplicity, we are going to consider distributions which are fully inclusive in the electroweak boson decay products, as well as integrated in the boson's rapidity. The extension to more differential distributions, including fiducial cuts on the final-state particles, is possible. For convenience, we work at NLL and, as further simplification, we explicitly consider only the flavor-diagonal contributions, while restoring full flavor-dependence in the end.

We compute the differential distribution for the transverse momentum of the boson (Higgs or Z/γ^*). At Born level, we have $gg \to h$ or $q\bar{q} \to Z/\gamma^*$, so the boson has no transverse momentum, i.e. the distribution is proportional to $\delta^{(2)}(\mathbf{Q}_T)$, where $\mathbf{Q}_T$ is the two-dimensional transverse-momentum. When computing higher perturbative orders, we must include contributions with additional partons i in the final state. Thus, the boson can acquire a nonzero transverse momentum, such that $\mathbf{Q}_T = -\sum_i \mathbf{k}_{Ti}$. Resummation is relevant when the transverse momentum is much smaller than the mass (or virtuality) of the electroweak boson, $Q_T^2 = |\mathbf{Q}_T|^2 \ll Q^2$. This can happen in two situations: either all recoiling partons have small transverse momenta or their transverse momenta, although not individually small, mostly cancel in their vector sum. Both these mechanisms must be taken into account and, as we shall shortly see, this can be achieved if Q_T resummation is performed in Fourier space. If we denote with $\mathbf{b}$ the conjugate variable to $\mathbf{Q}_T$, then the small-Q_T region corresponds to large $b = |\mathbf{b}|$ and logarithms of Q_T are mapped into logarithms of $1/b$.

Thus, we consider the emission of an arbitrary number of collinear gluons off the incoming hard legs. The partonic cross-section can be written as

$$\begin{aligned}\frac{d^2\sigma}{d\mathbf{Q}_T} = \sigma^{\text{born}}_{c\bar{c}\to F} \sum_{n=0}^{\infty} \frac{1}{n!} \prod_{i=1}^{n} \int [dk_i]\,(2C_c)\frac{\alpha_s(k_{Ti})}{2\pi} \\ \times \left[z_i^{N-1}\,\bar{P}^{\text{real}}(z_i)\,\delta^{(2)}\left(\mathbf{Q}_T + \sum_i \mathbf{k}_{Ti}\right)\right. \\ \left. + \bar{P}^{\text{virtual}}(z_i)\,\delta^{(2)}(\mathbf{Q}_T)\right] \Theta(k_{Ti} - Q_0) \\ \times \Theta\left(1 - z_i + \frac{k_{Ti}}{Q}\right),\end{aligned} \tag{11.7}$$

where we have taken Mellin moments with respect to the longitudinal momentum fractions z_i. The emitted gluon phase space is $[dk_i] = dz_i \frac{dk_{Ti}^2}{k_{Ti}^2}\frac{d\phi_i}{2\pi}$ and $C_c = C_F, C_A$ is the appropriate color factor. The first Θ function expresses the fact that emissions below the cut-off Q_0 belong to the non-perturbative region of the proton wave-function, while the second one correctly accounts for the large-angle soft region of phase-space. In order to achieve NLL accuracy, the strong coupling α_s has to be evaluated at two loops, in the CMW scheme [171]. The emission probability is described by the real and virtual matrix elements (see e.g. App. E of Ref. [3559]):

$$\bar{P}^{\text{real}}(z) = \begin{cases} \frac{1+z^2}{1-z}, & \text{for a quark,} \\ \frac{2z}{1-z} + \frac{2(1-z)}{z} + 2z(1-z), & \text{for a gluon;} \end{cases} \tag{11.8}$$

$$\bar{P}^{\text{virtual}}(z) = (-1) \begin{cases} \frac{1+z^2}{1-z}, & \text{for a quark,} \\ \frac{2z}{1-z} + z(1-z) & \\ +n_f T_R(z^2 + (1-z)^2), & \text{for a gluon.} \end{cases} \tag{11.9}$$

For later convenience, we also introduce the leading order regularized splitting functions

$$\begin{aligned} P_{qq}(z) &= \frac{\alpha_s}{2\pi} C_F \left[\frac{1+z^2}{1-z}\right]_+, \\ P_{gg}(z) &= \frac{\alpha_s}{2\pi} 2C_A \left[\left(\frac{z}{1-z} + \frac{z(1-z)}{2}\right)_+ \right. \\ &\quad \left. + \frac{1-z}{z} + \frac{z(1-z)}{2} - \frac{2}{3} n_f T_R \delta(1-z)\right], \end{aligned} \tag{11.10}$$

and the corresponding anomalous dimensions

$$\gamma_{cc}(N, \alpha_s) = \int_0^1 z^{N-1} P_{cc}(z), \quad c = q, g. \tag{11.11}$$

We note that virtual corrections in Eq. (11.7) do not change the transverse momentum Q_T and trivially exponentiate. The real-emission contribution is also factorized, with the exception of the two-dimensional delta-function constraint. This is where Fourier moments with respect to the two-dimensional vector $\mathbf{Q}_T$ become helpful. We can exploit the relation

$$\delta^{(2)}\left(\mathbf{Q}_T + \sum_i \mathbf{k}_{Ti}\right) = \frac{1}{4\pi^2}\int \mathrm{d}^2\mathbf{b}\, e^{i\mathbf{b}\cdot\mathbf{Q}_T}\prod_{i=1}^{n} e^{i\mathbf{b}\cdot\mathbf{k}_{Ti}}, \tag{11.12}$$

to fully factorize the real-contribution in Eq. (11.7). We obtain

$$\begin{aligned} W^{\text{real}}(b,N) = \sum_{n=0}^{\infty}\frac{1}{n!}\prod_i^n\int [dk_i]\, z_i^{N-1}(2C_c)\frac{\alpha_s(k_{Ti})}{2\pi} \\ \times \bar{P}^{\text{real}}(z_i)e^{i\mathbf{b}\cdot\mathbf{k}_{Ti}}\Theta\left(k_{Ti}-Q_0\right) \\ \times \Theta\left(1-z_i+\frac{k_{Ti}}{Q}\right). \end{aligned} \tag{11.13}$$

The series in Eq. (11.13) sums to an exponential. Thus, the resummed exponent is obtained by putting together real, virtual and PDF ($k_{Ti} < Q_0$) contributions:

$$\begin{aligned} R(b,N) = 2C_c\int [dk]\frac{\alpha_s(k_T)}{2\pi}\Theta\left(k_T-Q_0\right)\Theta\left(1-z+\frac{k_T}{Q}\right) \\ \times\left(-z^{N-1}\bar{P}^{\text{real}}(z)e^{i\mathbf{b}\cdot\mathbf{k}_T}-\bar{P}^{\text{virtual}}(z)\right) \\ +2\int_{Q_0^2}^{Q^2}\frac{dk_T^2}{k_T^2}\gamma_{cc}(N,\alpha_s(k_T)). \end{aligned} \tag{11.14}$$

By rewriting $z^{N-1} = 1+(z^{N-1}-1)$ and using the definitions in Eqs. (11.8), (11.9), and (11.10), we are able to reshuffle the contributions to the resummed exponent as follows

$$\begin{aligned} R(b,N) = -\int_{Q_0^2}^{Q^2}\frac{dk_T^2}{k_T^2}\int_0^{2\pi}\frac{d\phi}{2\pi}\left(1-e^{i\mathbf{b}\cdot\mathbf{k}_T}\right) \\ \times\left[\int_0^{1-\frac{k_T}{Q}} dz\frac{\alpha_s(k_T)\,C_c}{\pi}\bar{P}^{\text{virtual}}(z)\right. \\ \left.-2\gamma_{cc}(N,\alpha_s(k_T))\right]+\mathcal{O}\left(\frac{k_T}{Q}\right). \end{aligned} \tag{11.15}$$

The factor $\left(1-e^{i\mathbf{b}\cdot\mathbf{k}_T}\right)$ essentially acts as a cut-off on the k_T integral. At NLL we have[105]

$$\begin{aligned} R(b,N) = -\int_{b_0^2/b^2}^{Q^2}\frac{dk_T^2}{k_T^2}\left[\int_0^{1-\frac{k_T}{Q}} dz\frac{\alpha_s(k_T)\,C_c}{\pi}\bar{P}^{\text{virtual}}(z)\right. \\ \left.-2\gamma_{cc}(N,\alpha_s(k_t))\right] \\ = -\ln S_c + 2\int_{b_0^2/b^2}^{Q^2}\frac{dk_T^2}{k_t^2}\gamma_{cc}(N,\alpha_s(k_t)), \end{aligned} \tag{11.16}$$

$b_0 = 2e^{-\gamma_E}$, where γ_E is the Euler constant. Thus, we have successfully separated two distinct contributions: the Sudakov form factor (S_c), computed here at NLL accuracy (and systematically improvable) and a DGLAP contribution, which evolves the PDFs from the hard scale Q down to b_0/b. Note that here we have only considered flavor-diagonal splittings. Off-diagonal ones do not alter the Sudakov form factor and they are fully taken into account by the complete DGLAP evolution.

Taking into account all the above effects, the all-order transverse momentum distribution for the production of an electroweak final state F from initial-state partons c and $\bar{c}$ can be written

$$\begin{aligned} \frac{d\sigma}{dQ_T^2} = \sigma^{\text{born}}_{c\bar{c}\to F}\int dx_1\int dx_1\int_0^{\infty} db\frac{b}{2}J_0(bQ_T)S_c(b,Q) \\ \times\int dz_1\int dz_2\,\delta\left(1-z_1z_2\frac{x_1x_2s}{Q^2}\right) \\ \times\left[H^F_{c\bar{c}}\left(\alpha_s(Q)\right)C_{ca_1}\left(z_1,\alpha_s\left(\tfrac{b_0}{b}\right)\right)C_{\bar{c}a_2}\left(z_2,\alpha_s\left(\tfrac{b_0}{b}\right)\right)\right. \\ \left.+\tilde{H}^F_{c\bar{c}}\left(\alpha_s(Q)\right)G_{ca_1}\left(z_1,\alpha_s\left(\tfrac{b_0}{b}\right)\right)G_{\bar{c}a_2}\left(z_2,\alpha_s\left(\tfrac{b_0}{b}\right)\right)\right] \\ \times f_{a_1}\left(x_1,\tfrac{b_0}{b}\right)f_{a_2}\left(x_2,\tfrac{b_0}{b}\right), \end{aligned} \tag{11.17}$$

where we have introduced the Bessel function J_0 and the sum over a_1, a_2 is understood. The functions G_{ab}, C_{ab}, H^F_{ab}, $\tilde{H}^F_{ab}$ can be computed in perturbation theory, while f_a denotes the the parton distribution functions. For Standard Model Higgs production we have $F = h$, $c = \bar{c} = g$, and $H = \tilde{H}$, while for DY production we have $F = Z/\gamma^*$ and $c = q$, and $G_{q,a} = G_{\bar{q},a} = 0$. As already mentioned, different resummation formalisms exist in the literature. They all agree at the perturbative accuracy they claim, but they may numerically differ because of subleading effects. As an example, in Fig. 274 we show a comparison between the resummed and matched calculation of Ref. [3552] and LHC data, collected by the ATLAS collaboration [3561].

11.2.3 Jets and their substructure

All-order techniques not only allow us to probe the dynamics of electroweak bosons that recoil against QCD radiation, as discussed above, but can be employed to study the properties of the radiation itself in great detail. If we look at hadronic final states, we realise that QCD radiation is not uniformly distributed, but rather concentrated in collimated sprays of hadrons that are called jets. Jets really live at the boundary

[105] See Ref. [3560] for a generalization of this approximation to higher-logarithmic accuracy.

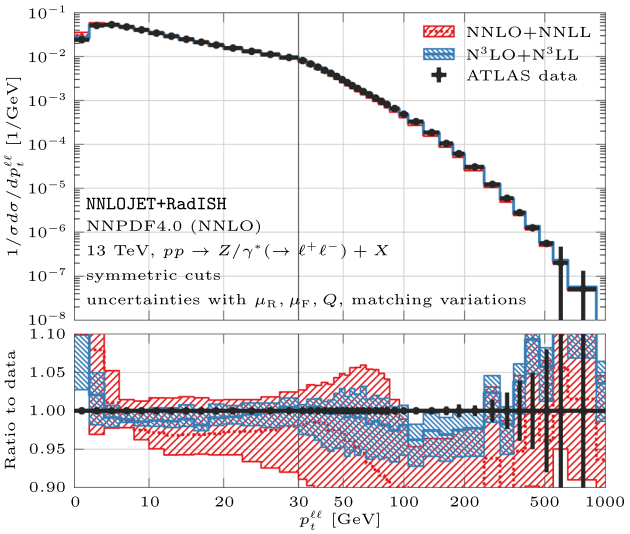


Fig. 274 The lepton pair transverse momentum distribution measured by the ATLAS collaboration at the LHC [3561] is compared to a resummed and matched calculation. Q_T resummation is performed at N^{3}LL logarithmic accuracy and it is matched to the N^{3}LO fixed-order result, with respect to the Born process, which corresponds to NNLO accuracy for the transverse momentum distribution. The plot is taken from Ref. [3552]

between experimental and theoretical particle physics and are abundantly used by both communities. They allow us to describe complex final states in terms of a few objects rather than hundreds of particles. Furthermore, from a theoretical point of view, jets are closely related to quarks and gluons, i.e. the degrees of freedom of perturbative QCD. Thus, the algorithms that are used to define jets must have good theoretical and experimental properties. For instance, jet algorithms should be IRC safe, so that they yield finite cross-sections when evaluated in perturbation theory [185].

Modern jet algorithms are based on the concept of sequential recombination. Pairwise distances (d_{ij}) between particles and so-called distances from the beam (d_{iB}) are evaluated in order to decide whether to recombine two particles. The metric used to evaluate these distances characterizes the jet algorithm. Nowadays, the most popular group of jet algorithm is the generalized k_T family, for which the metric is defined by

$$d_{ij} = \min\left(p_{Ti}^{2p}, p_{Tj}^{2p}\right)\frac{\Delta R_{ij}^2}{R^2}, \quad d_{iB} = p_{Ti}^{2p}, \tag{11.18}$$

where p_{Ti}, p_{Tj} are the particles' transverse momenta and ΔR_{ij}^2 is their distance in the azimuth-rapidity plane. R is an external parameter, which plays the role of the jet radius. Different choices for the parameter p are possible. For instance, $p = 0$ corresponds to the so-called Cambridge–Aachen (C/A) algorithm [189,190], with a purely geometrical distance. For $p = 1$ we have the k_T-algorithm [191,3562], which by clustering particles at low p_T first, is likely to faithfully reconstruct a QCD branching history. Finally, with the choice $p = -1$ we obtain the anti-k_T algorithm [193], which clusters soft particles around a hard core, producing fairly round jets in the azimuth-rapidity plane. It is interesting to note that all algorithms of the generalized k_T family act identically on a configuration with just two particles: they are recombined if $\Delta R_{ij} < R$. More details about jets can be found in Sect. 11.5. Although incredibly useful for phenomenology, jet algorithms introduce resolution parameters, such as the jet radius R, rendering the computation of jet properties a multi-scale problem.

In the past decade, many observables have been devised to study the internal properties of high-p_T jets, see for instance [3563]. The simplest example of such observables is the jet invariant mass, which is defined as

$$m_{\text{jet}}^2 = \left(\sum_{i\in\text{jet}} p_i\right)^2, \quad \rho = \frac{m_{\text{jet}}^2}{R^2 p_T^2}, \tag{11.19}$$

where p_i are the jet constituents' four-momenta and, in order to emphasise the multi-scale nature of the problem, we have also introduced the dimensionless ratio ρ. This ratio is small in the boosted regime $m_{\text{jet}} \ll Rp_t$, which is of particular interest at the LHC. As previously discussed, when scales become widely separated, logarithms (of ρ in this case) become large and in order to obtain reliable predictions for this observable, we need to perform all-order calculations.

We do not report here the details of the resummed calculation for the jet mass distribution, which is closely related to the one of the thrust event shape [3525,3564,3565], but rather we stress similarities and differences with respect to Q_T resummation, described above. Large logarithmic corrections always arise from the emission of soft and/or collinear partons. However, final-state four-momenta are combined differently in the two observables and therefore a different integral transform is needed to diagonalize the invariant mass δ-function. Furthermore, because we are interested in the dynamics of a high-p_T isolated jet, emissions collinear to the incoming legs do not significantly alter the jet properties, leading to a simplified treatment of the PDF contributions. However, there is a major complication that arises when performing calculations with jets. Only emissions that are recombined into the jet contribute to the invariant mass, making it an example of a non-global observable [3566]. As it turns out, the presence of phase-space boundaries noticeably complicates the structure of soft-emissions and essentially invalidates simple exponentiation. Furthermore, the actual shape of the boundary depends on the jet algorithm of choice. For instance, in the presence of many soft emissions together with a hard parton, the anti-k_T algorithm will always cluster all soft gluons to the hard parton, behaving as a rigid cone algorithm, while the choice of different algorithms, such as C/A or k_T, can give rise to more complicated clustering sequences, see e.g. [3567] and references therein. The calcu-

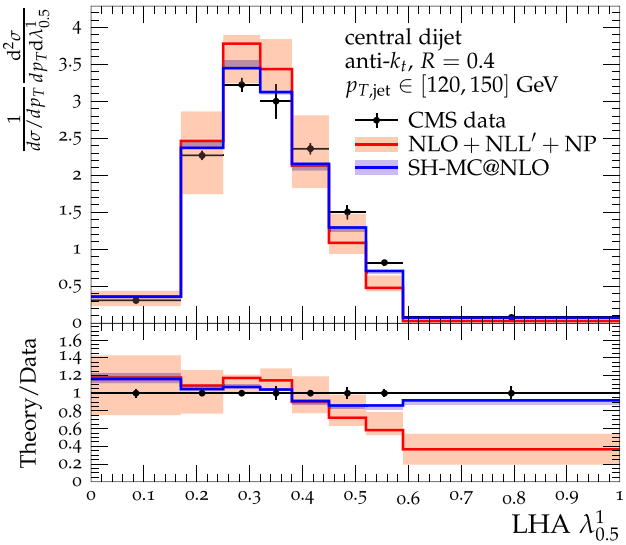


Fig. 275 The Les Houches Angularity (LHA) distribution, which corresponds to $\kappa = 1, \alpha = 0.5$ in Eq. (11.20) measured by the CMS collaboration at the LHC [3576]. The data are compared to the resummed and matched calculation (NLL+NLO), supplemented by non-perturbative corrections [3577,3578] and to the prediction obtain with a state-of-the-art Monte Carlo simulation using Sherpa at NLO QCD accuracy [3579]. The plot is taken from Ref. [3578]

lation of non-global logarithms constitutes the bottle-neck of jet calculations but thanks to an extraordinary effort of different groups, they can now be resummed at high accuracy [1824,2037,3568–3574].

The calculation techniques developed for the jet mass have been extended to other jet substructure observables. An interesting example is the family of jet angularities [3575]. These probe both the angular and the transverse momentum distribution of particles within a given jet. They are defined from the momenta of jet constituents as follows:

$$\lambda_\alpha^\kappa = \sum_{i \in \text{jet}} \left(\frac{p_{T,i}}{\sum_{j \in \text{jet}} p_{T,j}} \right)^\kappa \left(\frac{\Delta_i}{R} \right)^\alpha , \tag{11.20}$$

where

$$\Delta_i = \sqrt{(y_i - y_{\text{jet}})^2 + (\phi_i - \phi_{\text{jet}})^2}, \tag{11.21}$$

is the azimuth-rapidity distance of particle i from the jet axis. Jet angularities are IRC safe for $\kappa = 1$ and $\alpha > 0$. In Fig. 275 we show a comparison between LHC data collected by the CMS collaboration [3576], for the so-called Les Houches Angularity (LHA), which corresponds to setting $\kappa = 1$ and $\alpha = 0.5$, and a resummed calculation performed at NLL accuracy [3577,3578].

Despite the remarkable perturbative accuracy that can be achieved for jet observables, non-perturbative corrections due to the hadronization process or originating from multiple-parton interactions or pile-up, are rather large. Indeed, the resummed curve in Fig. 275 has been corrected for non-perturbative effects, which are important to ensure agreement with the data at small values of the angularity. The situation can be greatly improved if one considers "grooming" and "tagging" algorithms. Broadly speaking, a grooming procedure takes a jet as an input and tries to clean it up by removing constituents which, being at wide angle and relatively soft, are likely to come from contamination, such as the underlying event or pile-up. A tagging procedure instead focuses on some kinematic variable that is able to distinguish signal from background, such as, for instance, the energy sharing between two subjets within the jet, and cuts on it. Many of the algorithms on the market usually perform both grooming and tagging and a clear distinction between the two is difficult. Regardless of their nature, these algorithms try to resolve jets on smaller angular and energy scales, thereby introducing new parameters. This further challenges our ability of computing predictions in perturbative QCD. However, if these algorithms are properly designed, they can effectively reduce contamination from non-perturbative physics, while maintaining calculability. An example of this is SoftDrop [1915]. This procedure steps backward through the C/A clustering tree of a jet and iteratively checks whether the transverse momenta of the two branches satisfy the condition

$$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\text{cut}} \left(\frac{\Delta_{12}}{R} \right)^\beta . \tag{11.22}$$

The difficulty posed by substructure algorithms in general, and SoftDrop in particular, is the presence of new parameters, such as z_{cut} and β, that slice the phase-space for soft gluon emission in a non-trivial way, resulting in potentially complicated all-order behavior of the observable at hand. In the soft limit, the SoftDrop criterion reduces to

$$z > z_{\text{cut}} \left(\frac{\theta}{R} \right)^\beta \quad \Rightarrow \quad \ln \frac{1}{z} < \ln \frac{1}{z_{\text{cut}}} + \beta \ln \frac{R}{\theta}, \tag{11.23}$$

where z is the momentum fraction and θ the opening angle. For $\beta > 0$, collinear splittings always satisfy the SoftDrop condition, so a SoftDrop jet still contains all of its collinear radiation. The amount of soft-collinear radiation that satisfies the SoftDrop condition depends on the relative scaling of the energy fraction z to the angle θ. As $\beta \to 0$, more of the soft-collinear radiation of the jet is removed, and in the $\beta = 0$ limit, all soft-collinear radiation is removed [1916,3580]. Therefore, we expect the coefficient of the double logarithms in observables like the groomed jet mass, the origin of which is soft-collinear radiation, to be proportional to β. In the strict $\beta = 0$ limit, collinear radiation is only maintained if $z > z_{\text{cut}}$. Because soft-collinear radiation is vetoed, the resulting jet mass distributions will only exhibit single logarithms, as emphasized in [1916,3580]. Moreover, non-global logarithms are found to be power-suppressed for $\beta > 0$, and absent for $\beta = 0$. Finally, for $\beta < 0$, there are no logarithmic structures for observables like groomed jet

mass at arbitrarily low values of the observable. For example, $\beta = -1$ roughly corresponds to a cut on the relative transverse momentum of the two subjets under scrutiny.

The above understanding can be formalized into actual calculations and the resummation of a variety of observables measured on SoftDrop jets has been performed to N^{3}LL [1826,1939]. This outstanding theoretical accuracy, together with reduced sensitivity to non-perturbative corrections, make SoftDrop jets a particularly powerful way to probe QCD dynamics and jet formation.

11.2.4 Outlook

In this brief overview we have introduced resummation as a powerful tool that we can use to augment the ability of perturbative calculations to describe the data. We have given two examples of multi-scale processes, namely the transverse momentum of an electroweak boson and the Les Houches (jet) Angularity, for which the inclusion of all-order effects is mandatory in order to be able to describe the data.

Resummation provides us with the right tools to study emergent phenomena in QCD, such as jet formation and it allows us to scrutinise fundamental properties of the theory. The concept of factorization, i.e. the ability of separating physical effects happening at different energy scales, is the foundation of the whole resummation program that we have discussed. Even more generally, we can say any QCD calculation, being it done at fixed-order or at the resummed level, requires some notion of factorization. Of particular importance is the collinear factorization theorem [242] that allows us to separate the perturbative, i.e. calculable, part of a process from the non-perturbative one, which can be described in terms of parton distribution (or fragmentation) functions. Resummation techniques allows us to uncover limitations and possible breakdowns of factorization [1897,3581], which typically happen at perturbative orders that are too high to be reached with fixed-order techniques. Thus, despite resummation being based on the soft/collinear approximation of the perturbative approximation of QCD, it opens up a window to fundamental aspects of the theory:

Resummation just scratches the surface of QCD. But it makes a mark.[106]

11.3 Parton showers

Frank Krauss

11.3.1 Motivation

Producing charged particles in a high-energy collision initiates the emission of secondary bremsstrahlung quanta. Due to the large strong coupling and because of the gluon self-coupling, the radiation of gluons is of particular relevance, and tens or even hundreds of secondary quarks and gluons can be produced in a cascade of emissions.

Apart from the wish to correctly describe particle production at collider experiments in all its facets, and preferably based on first principles, there is another, more practical reason why this process of multiple parton emission is of great phenomenological relevance. The confinement property of QCD prevents quarks and gluons to be directly observed and instead, they manifest themselves through hadrons, which constitute the observable final states. Unfortunately, to date, only phenomenological models for the dynamical transition from quarks and gluons to hadrons in a process aptly dubbed hadronization have been developed, which rely on a large number of parameters which have to be fitted – "tuned" – to experimental data. Clearly, such a programme is sensible only, if the parameters are sufficiently independent from the hard process and rather depend on the properties of the parton ensembles at a common low scale. This is realized by casting the multiple emission of the secondary quanta, the parton cascade, into algorithms that systematically evolve the few original partons in the hard process at a scale of large momenta Q into resulting many–parton ensembles resolved at a lower scale Q_0, at which hadronization sets in. The resulting algorithms are called parton showers, and one might think of them as numerical implementations of a renormalization–group equation that connects these two scales, Q and Q_0. They form an integral part of modern event generators HERWIG [3582], PYTHIA [3583], and SHERPA [3579].

11.3.2 Parton shower realizations

Some first intuition about parton showers can be gained from the (quasi-classical) spectrum of gluons emitted by a fast moving color charge,

$$\mathrm{d}n_g = \frac{\alpha_S}{\pi}\frac{\mathrm{d}\omega}{\omega}\frac{\mathrm{d}^2 p_\perp}{p_\perp^2}, \tag{11.24}$$

exhibiting its characteristic divergent structure in the limit where the emitted gluon becomes soft, with it energy $\omega \to 0$, or collinear with respect to the emitter, with its transverse momentum $p_\perp \to 0$. These well-known soft and collinear divergences, typical for quantum field theories with massless (vector) particles cancel for physically meaningful observables when both real and virtual emissions are taken into account [3584,3585]. In parton showers, which aim to simulate the emission of real quanta, this is implicitly taken into account, by demanding that the emitted partons are resolvable with a minimal energy and transverse momentum; divergences in unresolvable emissions then cancel those from vir-

[106] George Sterman, CTEQ school 2006.

tual corrections. Such a constraint is effectively realized for example by demanding a minimal transverse momentum, $k_\perp > Q_0$ in emissions. The integrated spectrum depends logarithmically on the cut-off, and small values of Q_0 overcoming the smallness of α_s necessitate the resummation of the infrared logarithms.

This physical picture is encoded in probabilistic language, by constructing a Sudakov form factor

$$\Delta_{N\to N+1}(Q,\, Q_0) = \exp\left\{-\int_{Q_0^2}^{Q^2} \mathrm{d}\Phi_{N\to N+1}(t,\, z,\, \phi)\, \mathcal{K}_{N\to N+1}(\Phi_{N\to N+1})\right\}, \tag{11.25}$$

which yields the probability for an N-particle configuration with momenta $\{\tilde{p}\}$ *not* to emit another particle (and therefore *not* to turn it into an $(N+1)$-particle configuration with momenta $\{p\}$). The phase space element for the emission, $\Phi_{N\to N+1}(t,\, z,\, \phi)$, will depend on (1) the ordering parameter t defined below; (2) the splitting parameter z given by the light-cone momentum fraction or energy fraction of the emitted particle; and (3) the azimuth angle ϕ, fixing the orientation of the emitted particle in the transverse plane of the mission. The emission kernel $\mathcal{K}_{N\to N+1}(\Phi_{N\to N+1})$ depends on the phase space of the emission and on the strong coupling $\alpha_S(p_\perp^2)$, with the transverse momentum as preferred scale choice. In the collinear limit, $t \to 0$ with finite z, the kernel for a specific emitter (ij) splits into particles i and j and reduces to the well-known corresponding DGLAP splitting kernels. In the soft limit, $z \to 0$ which also forces $t \to 0$, the kernel should approach the eikonal form,

$$\lim_{z\to 0} \mathcal{K}_{N\to N+1}(\Phi_{N\to N+1}(t,\, z,\, \phi)) \propto \frac{(p_i \cdot p_k)}{(p_i \cdot p_j)(p_j \cdot p_k)}, \tag{11.26}$$

where k denotes the color spectator. Owing to the current standard of using the leading-color approximation in the parton shower construction, k can be uniquely chosen.

Individual simulated events are seeded by the hard process, evaluated at a fixed perturbative order, and dressed afterwards with the parton shower. In marked contrast to the forward evolution of the final-state parton shower, the parton shower in the initial state is described by a backward evolution, back to the initial beam particles and to a fixed, pre-defined state. To enforce that the backward evolution of the parton shower arrives at the correct initial state, while respecting the evolution of its internal structure, emissions are weighted by a ratio of parton distribution functions [3586],

$$\mathcal{K}_{N\to N+1}(\Phi_{N\to N+1}(t,\, z,\, \phi)) \propto \frac{f_i(x_{(ij)}/z,\, \mu_{ij}^2)}{f_{(ij)}(x_{(ij)},\, \mu_{(ij)}^2)}. \tag{11.27}$$

In this way the particle (ij), resolved at scale $\mu_{(ij)}^2$ and with momentum fraction $x_{(ij)}$, is replaced by the new initial-state particle i, resolved at a lower scale $\mu_i^2 < \mu_{(ij)}^2$ and with a larger momentum fraction $x_i = x_{(ij)}/z$.

The choice of a parton-shower realization has an impact on the accuracy with which the radiation pattern is simulated. In first-generation parton-shower implementations, the ordering parameter t is either identified with the virtual mass of the parton before emission, $t = p_{(ij)}^2 = (p_i + p_j)^2$ [3587,3588] or with the (scaled) opening angle of the emission, $t = E_{(ij)}^2(1 - \cos\theta_{ij})$ [3589,3590]. When the regular parts of the (massless) DGLAP splitting kernels at $\mathcal{O}(\alpha_S)$ are used, suitably limiting the allowed range for z accounts for the effect of finite masses. Careful analysis of the radiation pattern indicated that angular ordering is an important ingredient to the correctness of the simulation. The ordering accounts for color coherence effects, and introduces an explicit veto on increasing opening angles of the virtuality-ordered parton showers. In contrast, the dipole shower formulation [3591] in ARIADNE [3592] explicitly fills the Lund plane [3593] in transverse momentum $p_\perp^2$ and rapidity y of emissions. By setting the ordering parameter $t = p_\perp^2$ with the identification of $p_\perp^2$ as the inverse of the eikonal from Eq. (11.26), it fulfils the color coherence requirements that give rise to angular ordering [3594]. A similar approach has been chosen in VINCIA [3595], and extended to include initial state showering and other improvements. The same logic – using a form of transverse momentum as ordering parameter – was usually also chosen in the second-generation parton showers, for example in Refs. [3596–3599]. The explicit inclusion of mass effects in the splitting kernels forces to identify the splitting parameter z with a light-cone momentum fraction. To systematically include universal higher-order effects K from the two-loop cusp anomalous dimension, the customary CMW scheme [171] replaces

$$\alpha_S(p_\perp^2) \longrightarrow \alpha_S(p_\perp^2)\left[1 + K\frac{\alpha_S(p_\perp^2)}{2\pi}\right], \qquad K = \left(\frac{67}{18} - \frac{\pi^2}{6}\right) C_A - \frac{10}{9} T_R n_f, \tag{11.28}$$

where n_f is the number of active flavors and C_A and $T_R = 1/2$ are the usual QCD factors. Once an emission, parameterized by t, z, and ϕ, has been found, the emission kinematics needs to be constructed, including the compensation of the transverse momentum of the emitted particle. Choices range from being local, i.e. contained to the splitter–spectator

pair, to global, i.e. distributed over the full N-particle ensemble. They often reflect a preference for those schemes that lend themselves to a direct matching to infrared subtraction schemes for next-to leading order calculations such as the Catani–Seymour subtraction [199]. While these considerations sound like a minor technical detail, subtle differences in fact have an impact on the overall accuracy, as discussed below.

11.3.3 (N)NLO matching

Despite their success in describing the logarithmically-enhanced soft and collinear regimes of emission phase space, parton showers usually lack accuracy in the hard, wide-angle regions of phase space, the realm of fixed-order perturbative corrections, and they do not capture potentially large higher-order corrections to inclusive cross sections. Therefore the resummation implicit in the parton shower has to be matched to fixed-order calculations. Defining, respectively, $\mathcal{B}_N(\Phi_N)$, $\mathcal{V}_N(\Phi_N)$, and $\mathcal{R}_N(\Phi_{N+1})$ the Born-level, virtual and real corrections to a given process, and suppressing their phase space arguments in the following, a calculation – accurate in next-to leading order (NLO) – can schematically be written as

$$\mathrm{d}\sigma^{(NLO)} = \mathrm{d}\Phi_N \left[\mathcal{B}_N + \tilde{\mathcal{V}}_N\right] + \mathrm{d}\Phi_{N+1}\,\left[\mathcal{R}_N - \mathcal{D}_N\right], \tag{11.29}$$

with the infrared subtracted virtual correction $\tilde{\mathcal{V}}_N(\Phi_N) = \mathcal{V}_N(\Phi_N) + \mathcal{B}_N(\Phi_N) \otimes \mathcal{I}(\Phi_N)$ and the real subtraction term $\mathcal{D}(\Phi_{N+1}) = \mathcal{B}(\Phi_N) \otimes \mathcal{S}_1(\Phi_{N\to N+1})$ both written in factorized form, and where $\mathcal{I}$ emerges from $\mathcal{S}_1$ by analytically integrating over the one-particle emission phase space.

This can be matched to a parton shower along two well-established algorithms. The MC@NLO method [3600] makes use of the fact that the parton shower correctly describes the soft and collinear divergent regions of phase space and the emission kernels $\mathcal{K}$ can thus be matched to the infrared subtraction terms $\mathcal{S}$ required in fixed-order calculations. Events that, at fixed-order, correspond to N-particle final states with Born-level kinematics, are denoted as "soft" events and the parton shower is attached to them in a way exactly like it would be attached to the Born-level leading-order events. Similarly, the $(N+1)$-particle events are dubbed "hard" events, and, again the parton shower starts like it would from any similar tree-level configuration. Simple expansion in α_S reveals that the MC@NLO scheme recovers the fixed-order results, and augments them with the resummation of higher-order terms from the parton shower. Despite its simplicity, the MC@NLO prescription has a practical downside, with the second term in Eq. (11.29) possibly leading to events with a negative weight, a typical feature of practically any higher-order calculation at fixed order.

This is alleviated in the POWHEG method [3601,3602], which defines an NLO-accurate N-particle cross section, and dresses it, for its first emission, with a Sudakov form factor where the parton-shower splitting kernel is replaced with a ratio of real and Born contribution. However, the N-particle phase-space dependent K-factor implicit in the first square bracket is applied to the full $N+1$-particle spectrum, which may overestimate the hard region of emission phase space. To correct for this, in practical applications of the POWHEG method, the real-emission phase space is divided, with a suitable profile function, into a soft and a hard regime, $\mathcal{R}_N = \mathcal{R}_N^{(s)} + \mathcal{R}_N^{(h)}$. Schematically, then

$$\begin{aligned}\mathrm{d}\sigma^{(NLO)} = \mathrm{d}\Phi_N \Bigg[&\mathcal{B}_N(\Phi_N) + \tilde{\mathcal{V}}_N(\Phi_N) \\ &+ \int \mathrm{d}\Phi_1 \left(\mathcal{R}_N^{(s)}(\Phi_N \otimes \Phi_1) - \mathcal{D}_N(\Phi_N \otimes \Phi_1)\right)\Bigg] \\ &\times \exp\left[-\int \mathrm{d}\Phi_1 \frac{\mathcal{R}_N^{(s)}(\Phi_N \otimes \Phi_1)}{\mathcal{B}_N(\Phi_N)}\right] \\ &+ \mathrm{d}\Phi_{N+1}\,\mathcal{R}_N^{(h)}(\Phi_{N+1}).\end{aligned} \tag{11.30}$$

The regular parton shower is then applied to the $(N+1)$-particle configurations. Simple expansion shows, again, the overall cross section and the fixed-order emission spectrum at $\mathcal{O}(\alpha_S)$ are correctly reproduced.

NNLO calculations matched to parton shower so far have been solely available for the production of color singlets, **S**. The first realization was presented in Ref. [3603], based on the POWHEG method above. The underlying idea is to provide a POWHEG matching for $\mathbf{S} + p$ final states, with the additional parton p filling the phase space down to the infrared cut-off of the parton shower and thereby providing NLO accuracy for the overall emission of the hardest particle. This sample is then reweighted to reproduce the inclusive NNLO cross section for the production of the singlet **S** – in the case of a single particle usually achieved by reproducing its rapidity spectrum at NNLO accuracy. Based on multijet merging introduced in the next section, the UNNLOPS method [3604] matches complementary phase spaces of color-singlet production for the emission 0, 1, and 2 additional particles, described by adequately subtracted matrix elements at the two-loop, one-loop, and tree-level respectively. There overall NNLO accuracy is obtained by defining a zero-emission bin and adjusting its cross section accordingly. An alternative approach has been presented in the GENEVA algorithm [1982] which matches the NNLO cross section for **S** production with NNLL resummation of 0-jettiness. Using this observable to define different regions of phase space allows to combine the resulting parton level configurations with a suitably vetoed parton shower.

11.3.4 Multijet merging

Multijet merging provides another way to include exact fixed-order calculations into the parton shower, which is especially useful for the description of samples with large jet multiplicities. The underlying idea is to combine (merge) calculations with 0, 1, 2, etc. additional final state jets into one inclusive sample, by decomposing the parton emission phase space into two regimes, one of hard jet production and one of soft jet evolution. The algorithm achieving this at leading order [3605–3607] proceeds in three steps. Once a parton-level event at fixed order has been produced, the jets are clustered back until a core process corresponding to the 0-additional jet configuration has been found. The differential cross section for this event is reweighted with ratios $\alpha_S(\mu^{(PS)})/\alpha_S(\mu^{(FO)})$ at each emission, with $\mu^{(PS)}$ the scale the parton shower would use and $\mu^{(FO))}$ the fixed-order scale used in the calculation. The cross section is corrected with Sudakov form factors for the internal and external lines, either with analytic expressions or by running the parton shower from the core process and vetoing those events with a emissions leading to additional jets. These steps transform the individual inclusive fixed-order calculations into exclusive calculations for exactly 0, 1, 2 etc. additional jets, and combine them with the resummation in the parton shower. The algorithm outlined above has been extended to a merging of towers of NLO calculations, effectively a merging of multiple MC@NLO simulations with increasing jet multiplicities in [3608,3609].

11.3.5 Current developments

Driven by the ever increasing requirements for improved theoretical accuracy, parton showers have come under increased scrutiny in the past few years, for example in Ref. [3610]. Recent studies revealed that currently used parton showers do not correctly fill the phase space in logarithmically enhanced regions of multiple emissions [3611], limiting their logarithmic accuracy. Criteria to systematically asses the logarithmic accuracy of parton showers and a solution to the problem above was presented in Ref. [202] and led to renewed activity in creating better, parton showers that are accurate at next-to leading logarithmic accuracy for critical observables. Including higher-order terms, i.e. $\mathcal{O}(\alpha_S^2)$ corrections, to the parton showers represents an important step to further increase the accuracy. The treatment of $\mathcal{O}(\alpha_S^2)$ splitting kernels has been discussed in Refs. [3612,3613], and in Ref. [3614] the inclusion of differential two-loop soft corrections has been presented. But higher-order corrections in the strong coupling are not the only ordering parameter – the parton shower implicitly also resides on a leading-color approximation, and the impact of incorporating sub-leading color terms was studied for example in Refs. [3615,3616]. This led to the development of a new paradigm, to describe parton splitting and ultimately to construct a parton shower at the level of amplitudes [3617].

While it is not certain where these activities will lead us in the future, they are testament to the importance and impact of the probabilistic description of the QCD radiation pattern in parton showers, which is nearly as old as QCD itself.

11.4 Monte Carlo event generators

Torbjörn Sjöstrand

A *pp* collision at the LHC may lead to the production of hundreds of particles, via a multitude of processes that can range from the TeV scale down to below the confinement scale. While perturbative calculations can be used at high-momentum scales, currently there is no way to address lower ones directly from the QCD Lagrangian. Instead QCD-inspired models have been developed.

These models typically attempt to break down the full collision process into a combination of relevant mechanisms, that require separate descriptions. Each in its turn often can be formulated as an iterative procedure, where a set of rules are applied repeatedly. These rules are expected to represent quantum mechanical calculations that each gives a range of possible outcomes. The resulting complexity is such that analytical methods are of limited use. Instead the rules are coded up and combined within a bookkeeping framework, where the evolution from a primary perturbative collision to a final multiparticle state is traced. Such computer codes are called Monte Carlo Event Generators (MCEGs), where the "Monte Carlo" part refers to the frequent use of random numbers to pick outcomes according to the assumed quantum mechanical probabilities.

Such generators can be used in phenomenological studies, but the main application is within the experimental community, at all stages of the experiment. When an experiment is designed, it is important to check that the proposed detector has the capability to find key signals. When an experiment is run, triggers have to be optimized to catch the interesting event types. When data is analyzed, the impact of detector imperfections and background processes must be fully understood. In order to address these issues, the output of an MCEG is normally fed into a detector simulation program, that traces the fate of outgoing particles.

11.4.1 Event overview

Events come in many shapes, depending on the collider and the random nature of the collision process. As a starting point, consider a typical LHC *pp* event, and what processes are involved for it. Below these are enumerated, starting from the shortest time/distance scales and progressing towards longer

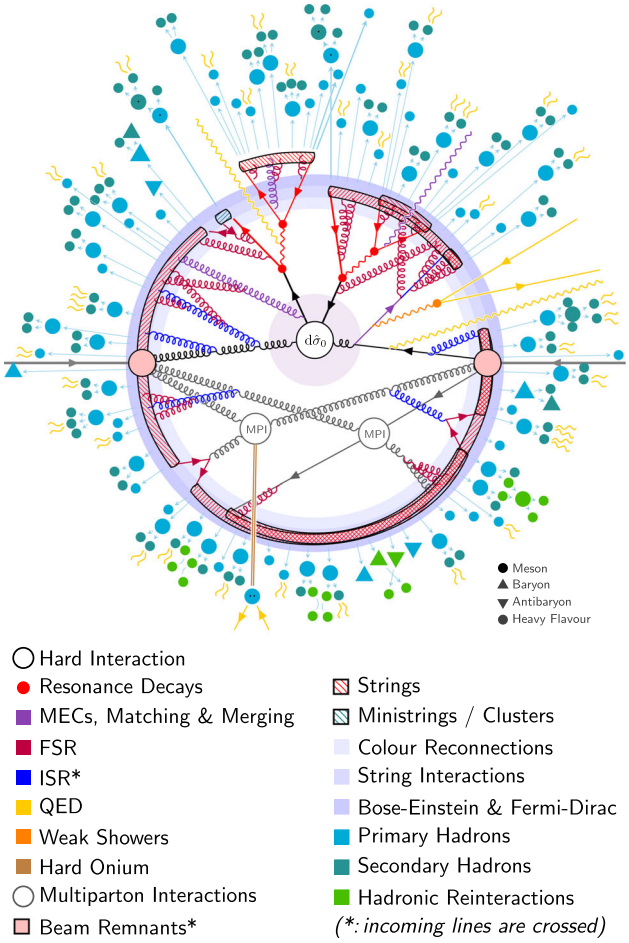


Fig. 276 Schematic illustration of the structure of a $pp \to t\bar{t}$ event. Reproduced from [3583]

ones. This gives rise to a schematic picture with an onion-like structure in some approximate measure of invariant time, Fig. 276.

- At the center of a collision there is sometimes a hard interaction, i.e.one at a high-momentum scale, like in this case the production of a $t\bar{t}$ pair. Its cross section is obtained by a convolution of a matrix-element (ME) expression and parton distribution functions (PDFs). More common are events without any discernible hard interaction.
- The hard interaction may involve the decay of resonances like $t \to bW$, $W \to q\bar{q}'$ as shown in Fig. 276.
- The core hard interaction may be dressed up by higher-order corrections of matrix-elements. This partly overlaps with the subsequent showers, so a consistent transition (matching and merging) is required.
- Perturbative radiation from the scale of the (dressed) hard interaction down to a lower cutoff is usually subdivided into initial-state radiation (ISR) and final-state radiation (FSR). While partonic QCD branchings dominate, QED or even weak branchings may occur. Also some hadron production may be modelled as part of the perturbative stage, e.g. of charmonium and bottomonium.
- Since hadrons are composite objects, several of the incoming partons may undergo (more or less) separate perturbative subcollisions, so-called multiparton interactions (MPIs).
- Parts of the incoming hadrons pass unaffected through the hard-interaction region, and emerge as beam remnants.
- Typically colors are traced through the perturbative stage in the $N_c \to \infty$ limit. Apart from imperfections caused by this approximation, there may also be dynamical processes that lead to color reconnections relative to the naive assignments.
- The color assignments are used to combine partons into separate color singlet subsystems – strings or clusters – that each fragment into a set of primary hadrons.
- To first approximation each subsystem fragments independently, but there may be interactions between them.
- The primary hadrons may be unstable and decay further into secondary particles, in decay chains that span a wide range of time scales.
- Right after the fragmentation the hadrons may also be close-packed and rescatter against each other.

In most of the following subsections these aspects will be described in somewhat more detail. Examples of longer generator overviews are [3618,3619].

11.4.2 A brief history

The first event generator of the QCD era probably is the 1974 one by Artru and Mennessier [3620]. It is based on the concept of linear confinement, originally introduced in pre-QCD string-theory models of hadrons, but later supported by the linear confinement found in quenched lattice QCD, see Sect. 4.3. It was not developed beyond a toy-model stage, however, and was largely forgotten. Instead it was the 1978 article by Field and Feynman [3621] that stimulated an interest in using Monte Carlo methods to simulate jet physics. Their iterative approach for the fragmentation of a single jet was extended to $e^+e^- \to q\bar{q}g$ three-jet events in the Hoyer et al. [3622] and Ali et al. [3623] codes, which played a key role in the discovery of gluon jets, see Sect. 2.2. The Lund string fragmentation model introduced the concept of a color flow in $q\bar{q}g$ events [3624], which was given experimental support by PETRA data [3625]. It helped establish the JETSET implementation as a main generator for subsequent e^+e^- machines.

The first QCD-based generator for $pp/p\bar{p}$ physics was ISAJET [3626], originally intended for the ISABELLE collider, but much used at the S$p\bar{p}$S and Tevatron colliders, and for SSC and LHC preparations. A few other generators were

developed in the early eighties, but left little impact, except for PYTHIA, which was built on top of JETSET, with the same initial objective of modelling the color flow and its consequences. Later on the two programs were merged under the PYTHIA heading.

The earliest generators used leading-order matrix elements to describe the perturbative stage. This was insufficient to describe multijet topologies. The DGLAP equations and their extension to jet calculus [3627] suggested that parton showers could be used to generate multiparton topologies. Several early showers were constructed, but it was only with the Marchesini–Webber angularly-ordered shower [3589] that coherence phenomena were consistently handled. This was the starting point for the HERWIG generator. An alternative was offered by transverse-momentum-ordered dipole showers, proposed by Gustafson [3594] and first implemented in ARIADNE. Today various dipole formulations are the most common shower type.

The combination of hard interactions and parton showers gradually became more sophisticated as various matching and merging techniques were developed. The SHERPA program grew out of such efforts. It was also the first major generator written in C++ from scratch, whereas HERWIG and PYTHIA had to be rewritten from Fortran to C++ to match LHC requirements.

Today HERWIG [3582], PYTHIA [3583], and SHERPA [3579] are the three general-purpose generators used at LHC, or more generally for studies at $e^+e^-/ep/pp/p\overline{p}$ colliders. There also are important dedicated programs, e.g. for matrix element generation, such as MADGRAPH_AMC@NLO [3385] and the POWHEG BOX [3628].

Adjacent physics areas, such as heavy-ion collisions, cosmic ray cascades in the atmosphere, or neutrino interactions, started their generator development somewhat later, and partly under the influence of the general-purpose ones above, e.g. for the high-energy hadronization descriptions. Typically the hard-physics aspects become less relevant, and soft-physics ones more so. These issues will be briefly addressed towards the end.

11.4.3 The perturbative interface

A key task is to generate events of a predetermined type or types. This could be e.g. W + jets, both as a signal and as a background to $t\bar{t}$ production. Typically there is a core hard interaction, that then is complemented by further perturbative QCD activity at varying scales. In such cases the core interaction provides the natural starting point for the description of the rest of the event. As already suggested above, one may discern three main stages:

1. the generation of partonic events purely based on matrix elements and parton distribution functions,
2. the matching and merging stage, where Sudakov form factors generated by parton showers are used to reject some of the events above, so as to avoid double counting, and
3. the subsequent pure parton shower evolution down to a lower cutoff somewhat above the Λ scale.

The first of these is covered in Sect. 11.1, and in Sect. 10.2 for PDFs, while the second two are described in Sect. 11.3.

Of special interest for the continued story are the core $2 \to 2$ pure QCD interactions, $qq' \to qq'$, $q\overline{q} \to q'\overline{q}'$, $q\overline{q} \to gg$, $qg \to qg$, $gg \to gg$ and $gg \to q\overline{q}$. These are by far the dominant perturbative processes at hadron colliders. The main contribution is t-channel gluon exchange, which gives rise to a $\mathrm{d}p_T^2/p_T^4$ divergence in the $p_T \to 0$ limit.

11.4.4 Total cross sections and diffraction

Another key task, at the other extreme, is to generate the inclusive sample of all events at hadron colliders. In practice rare processes are generated separately, so the emphasis comes to lie on common QCD processes.

The total QCD cross section σ_{tot} is finite, related to a finite size of hadrons and a finite range of QCD interactions, owing to confinement. Currently there is no QCD-Lagrangian-based description of σ_{tot}, but instead phenomenological models have been proposed based on Regge theory, with free parameters that have to be tuned to data. At a minimum one Pomeron and one Reggeon term is required to describe the energy dependence [1100], where the former can be associated with a trajectory of exchanged glueball states and the latter with one of mesonic states, see Sect. 8.1. More terms are needed in more realistic models. Notably, recent studies points towards the existence of an Odderon term, see Sect. 12.6.

The total cross section between two hadrons A and B can be subdivided into several partial ones, associated with different event topologies:

$$\begin{aligned}\sigma_{\mathrm{tot}}^{AB}(s) &= \sigma_{\mathrm{el}}^{AB}(s) + \sigma_{\mathrm{inel}}^{AB}(s) \\ &= \sigma_{\mathrm{el}}^{AB}(s) + \sigma_{\mathrm{sd}(XB)}^{AB}(s) + \sigma_{\mathrm{sd}(AX)}^{AB}(s) \\ &\quad + \sigma_{\mathrm{dd}}^{AB}(s) + \sigma_{\mathrm{cd}}^{AB}(s) + \sigma_{\mathrm{nd}}^{AB}(s), \end{aligned} \tag{11.31}$$

where s is the squared collision energy in the rest frame. These topologies are illustrated schematically in Fig. 277. In nondiffractive (nd) events the full rapidity range can be populated by particle production, whereas in single, double or central diffraction (sd, dd, or cd, respectively) only parts of this range are populated, and in elastic (el) events none of it is. The relative composition changes with energy, e.g. such that the elastic fraction is increasing. Roughly speaking,

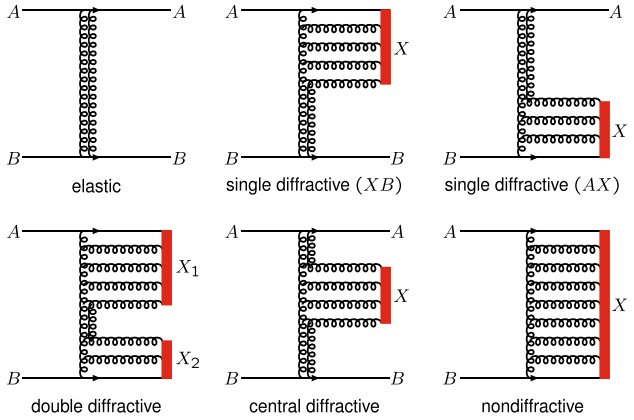


Fig. 277 Main subclasses of the total cross section in AB hadron collisions. The red bars represent the regions in rapidity between A and B where hadrons are produced. Reproduced from [3583]

elastic is 25%, diffractive 20% and nondiffractive 55% at LHC energies.

Many approaches have been proposed to model these partial cross sections, both integrated and differential ones, notably again based on Regge theory. Common is that the mass m_X of a diffractive system obeys an approximate $\mathrm{d}m_X^2/m_X^2 = \mathrm{d}y_X$ behaviour, where y_X is the rapidity range of the X system. An elastically scattered beam particle is also associated with a squared momentum transfer t that obeys an approximate $\exp(Bt)$ shape at low t. The slope B depends on the colliding hadron types, the event topology and the collision energy, but the order of magnitude is 10 GeV^{-2}, i.e.$\langle p_T \rangle \sim 0.3$ GeV.

At low energies also other collision types occur, such as resonant production and baryon annihilation.

The Ingelman–Schlein [3629] ansatz is commonly used for the description of diffractive systems. In it, a Pomeron is viewed as a hadronic state, with its own PDFs. Therefore the Pomeron–hadron subsystem can be described in the same way as we will introduce for nondiffractive events in the following, at least for reasonably large m_X, while a simpler description is called for at small m_X.

11.4.5 Multiparton interactions

All generators assume that a nondiffractive event can contain multiple parton–parton interactions, which can be viewed as the QCD reinterpretation of the cut Pomeron picture of olden days [3630]. MPIs are necessary to explain many aspects of hadronic collisions, such as the wide multiplicity distributions, where most of the multiplicity is related to low-p_T processes. The case of two hard interactions, Double Parton Scattering, is well studied theoretically and experimentally [3631]. Different models have been developed starting from the same basic ideas. This section will begin with the PYTHIA approach, which is also used by SHERPA, and later the differences in HERWIG will be outlined.

It has already been noted that the perturbative QCD $2 \to 2$ cross section is divergent for $p_T \to 0$, on the one hand, and that the total pp cross section is finite, on the other hand. The perturbative picture is based on the assumption of asymptotically free colored partons, however, while the reality is that of partons confined inside color singlet hadrons. Therefore a plausible regularization of the $p_T \to 0$ divergence is provided by color screening, i.e.that partons of opposite color gives destructive interference of scattering amplitudes. A parameter p_{T0} is introduced in PYTHIA as the inverse of the spatial screening distance. This is used to dampen the conventional $2 \to 2$ QCD cross section by a factor

$$\left(\frac{\alpha_s(p_{T0}^2 + p_T^2)}{\alpha_s(p_T^2)} \, \frac{p_T^2}{p_{T0}^2 + p_T^2} \right)^2 , \tag{11.32}$$

which gives

$$\frac{\mathrm{d}\sigma}{\mathrm{d}p_T^2} \sim \frac{\alpha_s^2(p_T^2)}{p_T^4} \to \frac{\alpha_s^2(p_{T0}^2 + p_T^2)}{(p_{T0}^2 + p_T^2)^2} . \tag{11.33}$$

A tune to data gives a p_{T0} of the order of 2 GeV, but slowly increasing with energy, consistent with an increasing screening, as lower-x partons become accessible at higher energies.

The average number of MPIs in nondiffractive events is given by $\langle n_{\mathrm{MPI}} \rangle = \sigma_{\mathrm{pert}}(p_{T0})/\sigma_{\mathrm{nd}}$, neglecting a small correction from the part of σ_{pert} that should be associated with diffraction. Here $\sigma_{\mathrm{pert}}(p_{T0})$ is the integrated damped $2 \to 2$ QCD cross section. At first glance, the n_{MPI} should be distributed according to a Poissonian, with $n_{\mathrm{MPI}} = 0$ removed, since zero MPIs corresponds to the two hadrons passing through without any interactions.

This assumes that all collisions are equivalent, however. More plausibly, the impact parameter b of the collision plays a role, where central collisions generate more activity than peripheral ones. To model this, an ansatz for the matter distribution inside a hadron is required. The simplest choice is a three-dimensional Gaussian, since then the convolution of two hadrons is easily integrated over the collision process to give a two-dimensional Gaussian $\mathcal{O}(b)$. Fits to data prefer a somewhat more uneven matter distribution, e.g. with "hot spots" of enhanced activity around the three valence quarks.

The actual generation of MPIs can conveniently be arranged in a falling sequence of transverse momenta with $\sqrt{s}/2 > p_{T1} > p_{T2} > \cdots > p_{Tn} > 0$. Neglecting the impact-parameter dependence for a moment, the probability for the ith MPI becomes

$$\frac{\mathrm{d}\mathcal{P}}{\mathrm{d}p_{Ti}} = \frac{1}{\sigma_{\mathrm{nd}}} \frac{\mathrm{d}\sigma_{\mathrm{pert}}}{\mathrm{d}p_{Ti}} \exp\left(- \int_{p_{Ti}}^{p_{T(i-1)}} \frac{1}{\sigma_{\mathrm{nd}}} \frac{\mathrm{d}\sigma_{\mathrm{pert}}}{\mathrm{d}p_T'} \mathrm{d}p_T' \right) , \tag{11.34}$$

with a fictitious $p_{T0} = \sqrt{s}/2$. The exponential expresses the probability to have no MPIs between $p_{T(i-1)}$ and p_{Ti}, as comes out of Poissonian statistics and in exact analogy with the Sudakov form factor of parton showers. With impact parameter included, the b is selected proportional to $\mathcal{O}(b)\, d^2b$, and the p_T selection procedure acquires an enhancement/depletion factor of $\mathcal{O}(b)/\langle\mathcal{O}\rangle$. Sequences without any MPIs require a restart with a new b.

So far inclusive nondiffractive events have been considered. Alternatively one specific hard interaction is studied, and an underlying event should be added to it. Then again a b is selected according to $\mathcal{O}(b)$, and an enhancement/depletion factor is defined as before. The upper p_T limit for MPIs now depends on context. If the hard interaction is QCD $2 \to 2$ above some $p_{T\min}$ then its p_T should be equated with a p_{T1} of the MPI sequence, and subsequent ones be below that, or else high p_T scales would be double counted. If the hard interaction is something else, then there is no such double counting, and MPIs can start from the highest possible scale.

The description of n MPIs requires n-parton PDFs, $f(x_1, Q_1^2; x_2, Q_2^2; \ldots; x_n, Q_n^2)$, which are not known from first principles or from data. An approximate approach is to make use of the p_T-ordering of MPIs, such that the first interaction uses conventional PDFs, while subsequent MPIs use more and more modified ones. Thereby standard phenomenology is preserved in the hard region. Subsequently momentum conservation requires a gradually reduced x_i range, within which PDFs are squeezed. Also flavor conservation must be respected. If a valence u quark is taken out of a proton, say, then only one u quark remains, and the valence u distribution must be normalized to 1 rather than 2. If instead a sea u quark is extracted, then the $\overline{u}$ sea must contain one parton more than the u sea, which can be implemented by having one valence-like $\overline{u}$ in addition to the normal u and $\overline{u}$ sea distributions. Finally, when the valence-like distributions have been properly normalized, the gluon and sea distributions are uniformly rescaled so as to obey the momentum sum rule.

With the evolution of ISR and FSR parton showers usually formulated in terms of a decreasing sequence each of p_T values, the MPIs now add a third sequence. In PYTHIA they are fully interleaved into one common sequence. Thus the key evolution equation is

$$\frac{\mathrm{d}\mathcal{P}}{\mathrm{d}p_T} = \left(\frac{\mathrm{d}\mathcal{P}_{\mathrm{MPI}}}{\mathrm{d}p_T} + \frac{\mathrm{d}\mathcal{P}_{\mathrm{ISR}}}{\mathrm{d}p_T} + \frac{\mathrm{d}\mathcal{P}_{\mathrm{FSR}}}{\mathrm{d}p_T}\right) \times S \tag{11.35}$$

where S represents the Sudakov factor, obtained by exponentiation of the real-emission rate, integrated from the previous p_T scale to the current one, cf. Eq. (11.34). In this way the harder part of the event sets the stage for what can occur at softer scales. Notably MPIs and ISR compete for the dwindling amount of momentum in the beams, as represented by the modified PDFs. The p_T evolution should not be viewed as one in physical time; actually all MPIs occur at (almost) the same collision time $t = 0$, while lower p_T scales means earlier times $t < 0$ for ISR and later times $t > 0$ for FSR.

The HERWIG description of MPIs [3632] splits them into a hard and a soft component, separated at a scale $p_T^{\min}(s)$. The perturbative cross section $\mathrm{d}\sigma_{\mathrm{QCD}}/\mathrm{d}p_T$ is recovered above $p_T^{\min}(s)$, while a simple tuneable shape $\mathrm{d}\sigma_{\mathrm{soft}}/\mathrm{d}p_T$ is used for $0 < p_T < p_T^{\min}(s)$, with the constraints that it must vanish at $p_T = 0$ and match $\mathrm{d}\sigma_{\mathrm{QCD}}/\mathrm{d}p_T$ at $p_T^{\min}(s)$. The electromagnetic form factor is used to represent the impact-parameter profile of protons. This gives an overlap function

$$A(b, \mu) = \frac{\mu^2}{96\,\pi}\,(\mu b)^3\, K_3(\mu b), \tag{11.36}$$

where $\int \mathrm{d}^2b\, A(b) = 1$, and μ are used as free parameters, separately set for the hard and soft components, for more flexibility. Combining, an eikonal is defined as

$$\chi_{\mathrm{tot}}(b, s) = \frac{1}{2}\, A_{\mathrm{hard}}(b, \mu_{\mathrm{hard}})\, \sigma_{\mathrm{QCD}}(s, p_T^{\min}) \tag{11.37}$$

$$+ \frac{1}{2}\, A_{\mathrm{soft}}(b, \mu_{\mathrm{soft}}) \sigma_{\mathrm{soft}}(s, p_T^{\min}), \tag{11.38}$$

where σ_{QCD} and σ_{soft} are the respective p_T-integrated cross sections. The number of MPIs at a given b is given by a Poissonian, as in PYTHIA, with $\langle n(b, s)\rangle = 2\chi(b, s)$. The eikonal formalism also predicts other quantities, such as total and elastic cross sections, and the elastic slope, that can be used to constrain the free parameters of the model.

When a hard interaction has been selected in HERWIG, and been associated with an impact parameter b, the number of hard and soft additional MPIs can be selected according to Poissonians. The hard interactions are generated first, and thereafter the soft ones. Unlike PYTHIA they are not ordered in p_T within the hard and soft groups, and there is no rescaling of PDFs. Also the ISR and FSR associated with an interaction are reconstructed before the next is considered. For the hardest interaction the ISR is forced to reconstruct back to a valence quark, while for subsequent ones the ISR evolution is forced back to a gluon. This gluon can then be color-attached to the hardest interaction itself. The MPIs together may take more momentum out of the protons than is available, given the lack of PDF rescaling. When that happens, the latest MPI is regenerated, but if repeated attempts fail the MPI generation may be interrupted with a lower MPI number than intended.

11.4.6 Beam remnants and color reconnection

Since the MPI+ISR machinery in HERWIG reconstructs back the perturbative activity to one single valence quark, having been taken out of an incoming proton, the other two valence quarks together form a diquark remnant, with opposite color

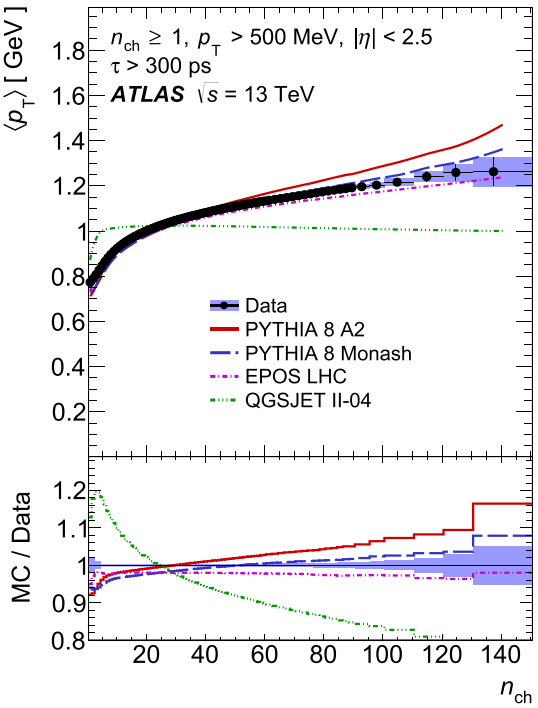


Fig. 278 Average transverse momentum as a function of the charged multiplicity, as measured by ATLAS [3633]. QGSJET exemplifies a model without color reconnection, with a similar flat shape as PYTHIA has with its CR switched off. The two PYTHIA curves are different tunes. EPOS is based on a two-component model, explained later in the section

to the one quark taken out. Four-momentum conservation fixes the remnant momentum.

The situation is more complicated in PYTHIA, since the MPI+ISR can extract a variable number of "initiator" partons out of the incoming proton, leaving behind multiple quarks and antiquarks. Then ad hoc probability distributions are used to share the remnant longitudinal momentum between them. The initiator partons also carry a transverse momentum, a so-called primordial k_T, that is to be compensated by the remnant. When the remnant consists of the several partons, these may also have a relative k_T component. The size of all these transverse kicks should be at or below the hadronic mass scale, though empirically they appear to be at the higher rather than at the lower end of the expected range.

The color lines of the initiator partons naively stretch from the remnants in through the hard interaction at the core of each MPI, i.e.usually fill the whole rapidity range. If so, the average charged multiplicity n_{ch} of an event increases linearly with the number of MPIs, up to corrections from momentum conservation and the details of the remnant handling. Since all MPIs will be equivalent, a constant average transverse momentum per hadron should result, i.e.a flat $\langle p_T \rangle(n_{ch})$ curve. Instead an increasing $\langle p_T \rangle(n_{ch})$ is observed at hadron colliders, Fig. 278.

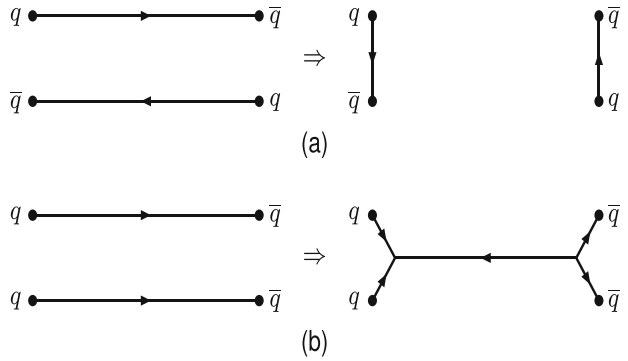


Fig. 279 Schematic illustration of color reconnection. **a** Simple flip. The arrows indicate flow from color to anticolor. **b** Junction reconnection. Note changed direction of the long line, according to $3 \otimes 3 = \overline{3}\,(\oplus 6)$

The natural explanation for this phenomenon is color reconnection (CR). Specifically, it is assumed that the color lines stretched between all final-state partons can be rearranged so as to reduce the overall length. The number of possible rearrangements increases with the number of MPIs, such that the $\langle n_{ch} \rangle$ increase is smaller for each further MPI. The perturbative p_T kick of each MPI remains, however, so when this p_T is shared between fewer particle the result is an increasing $\langle p_T \rangle$.

Many CR models have been implemented over the years, in all three main generators, and it would carry too far to discuss each in detail. A frequent starting point is that standard parton showers operate in the $N_c \to \infty$ limit, and thus miss corrections of order $1/N_c^2$ at each shower branching. One possible approach is to do the evolution in color space more carefully, and thereby be able to formulate CR as a consequence of such subleading corrections. More common is to formulate CR on the nonperturbative level, but then color algebra should be used to restrict the rate at which it can occur.

Also common for many nonperturbative approaches is that a key role is assigned to the string length λ between two color-connected partons i and j

$$\lambda_{ij} \approx \ln\left(1 + \frac{m_{ij}^2}{m_0^2}\right) = \ln\left(1 + \frac{(p_i + p_j)^2}{m_0^2}\right). \qquad (11.39)$$

Here $m_0 \approx m_\rho$ is a typical hadron mass, and 1 has been added to ensure that $\lambda \geq 0$. With this definition λ is a reasonable measure of how many hadrons typically will be produced by the string. A flip of two color-connected pairs (i, j) and (k, l) into (i, l) and (k, j), Fig. 279a, corresponds to a net change $\Delta\lambda = \lambda_{il} + \lambda_{kj} - \lambda_{ij} - \lambda_{kl}$. The assumption is that $\Delta\lambda < 0$ reconnections are favoured.

Further CR variants include ones that change the number of string pieces, say by taking a central gluon connected to both remnants and putting it on an already existing central

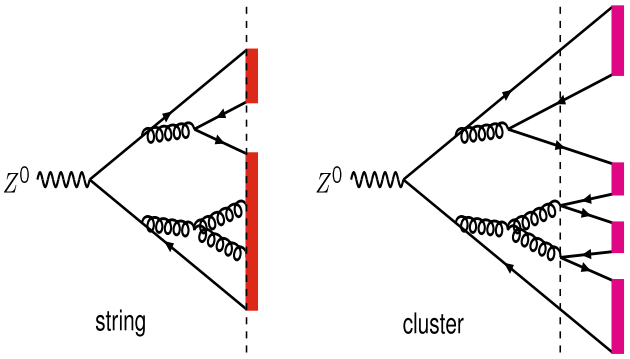


Fig. 280 String versus cluster fragmentation. At the end of the perturbative evolution, the vertical dashed line, strings are directly attached, red regions. Alternatively, a nonperturbative $g \to q\overline{q}$ stage is inserted before clusters are formed, magenta regions

string piece. Of recent special interest is junction reconnection [3634]. A junction is the center of a Y shape where three string pieces come together, and topologically is the carrier of the baryon number. Two strings can collapse to one in a central region with the production of a junction and antijunction near either end, Fig. 279b. This gives an enhanced baryon production. In the cluster model a similar effect can be obtained by letting three aligned $q\overline{q}$ clusters rearrange into one qqq and one $\overline{qqq}$ cluster.

CR ought to be possible only when the strings concerned overlap in space–time. For normal pp collisions this is almost automatic, since most strings run more-or-less parallel with the collision axis within a small transverse region. Space–time should be taken more seriously e.g. in $e^+e^- \to W^+W^- \to q_1\overline{q}_2q_3\overline{q}_4$, where the $W^\pm$ decay angles will influence the amount of overlap. Models have been developed to this end, and predictions agree well with the combined LEP data [3635]. The best description is obtained with an $\sim 50\%$ CR rate, but unfortunately statistics is limited and a no-CR scenario is only disfavoured at the 2.2σ level.

11.4.7 Hadronization

There are two main fragmentation models in common use: strings and clusters. Both start out from the color flow topologies set up according to the previous sections, in the $N_c \to \infty$ limit. Specifically, each $q \to qg$ and $g \to gg$ leads to a new uniquely defined color line between the two daughter partons. The string model retains all the gluons produced in the perturbative stage. A string can therefore be stretched e.g. like $q - g_1 - g_2 - g_3 - \overline{q}$, where each color line between two adjacent partons is unique. In the cluster model the perturbative shower is followed by a semi-perturbative step where all gluons branch by $g \to q\overline{q}$. The system therefore subdivides into smaller $q\overline{q}$ clusters.This key difference is illustrated in Fig. 280. The string is central in PYTHIA, while HERWIG and SHERPA implements clusters. In the latter program there is an interface to PYTHIA strings to allow comparisons. In the following key features will be presented in some more detail.

The string approach is based on the assumption of a linear confinement potential, as supported by quenched lattice QCD phenomenology. In a simple $q\overline{q}$ system studied in the rest frame, e.g. from a Z^0 decay, the potential can then be written as $V(r) = \kappa r$, where r is the separation and κ is the string tension. Empirically $\kappa \approx 1$ GeV/fm, determined mainly from hadron spectroscopy. The mathematical one-dimensional string stretched straight between the q and the $\overline{q}$ can be viewed as defining the center of a physical chromoelectric flux tube, with transverse dimensions comparable to hadron sizes, i.e.with a radius of around 0.7 fm. It is not settled whether this tube should be viewed in analogy with a vortex line in a type II superconductor, or with an elongated bag in a type I one, or with an intermediate behaviour, but for the basic considerations this is also not important.

If the string does not break, it will undergo a yo-yo-like oscillatory motion, where initially the quarks carry the full energy of the system, but gradually lose it to the string being stretched out between them. Massless quarks will reach a maximal separation where all the energy is stored in the string, and then the string tension will pull them back, so that they again cross, carrying the full energy. The key relation is that the massless endpoint quarks, moving out along the $\pm z$ axis, obey

$$\left|\frac{\mathrm{d}E}{\mathrm{d}t}\right| = \left|\frac{\mathrm{d}E}{\mathrm{d}z}\right| = \left|\frac{\mathrm{d}p_z}{\mathrm{d}t}\right| = \left|\frac{\mathrm{d}p_z}{\mathrm{d}z}\right| = \kappa \tag{11.40}$$

(with $c = 1$). If such a system is boosted along the z axis the q and $\overline{q}$ start out with different energies, which means that their turning points occur at different times, which gives the expected net motion of the system as a whole. The string tension remains unchanged by the boost, and a string piece in the new frame still carries no three-momentum. This may seem counterintuitive, but note that the boost will take an equal-times string piece to one where the endpoints are at different times, and if viewed this way the boosted string piece will pick up the expected momentum.

Now introduce the possibility for a string to break by the production of a new $q_i\overline{q}_i$ pair somewhere along the string. In lattice QCD this corresponds to going from the quenched to the unquenched situation. Each break splits the original color singlet system into two separate smaller ones. A sequence of breaks thus gives an ordered singlet chain $q\overline{q}_1 - q_1\overline{q}_2 - q_2\overline{q}_3 - \ldots - q_{n-1}\overline{q}$, and these singlets can be associated with the primary (i.e.before any decays) hadrons. Such a sequence of breaks is illustrated in Fig. 281. Notice that, in this picture, each produced hadron undergoes a yo-yo motion of its own.

If the q_i have $m = p_T = 0$ then a $q_i\overline{q}_i$ pair can be produced on-shell in a single vertex, and afterwards be pulled apart. The partons are virtual initially when this is not the

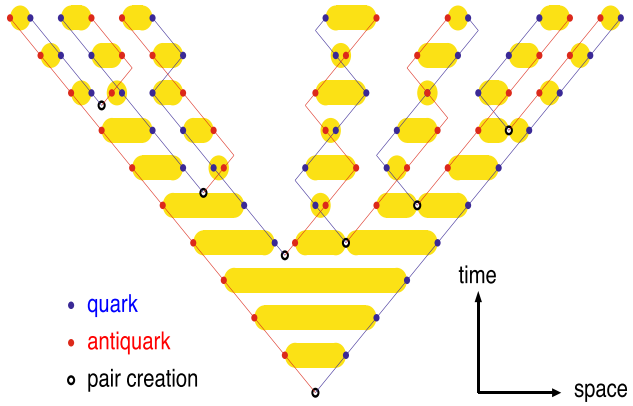


Fig. 281 String fragmentation of a $q\overline{q}$ system, where yellow regions represents snapshots in time of the string pieces being stretched out. Reproduced from [3636]

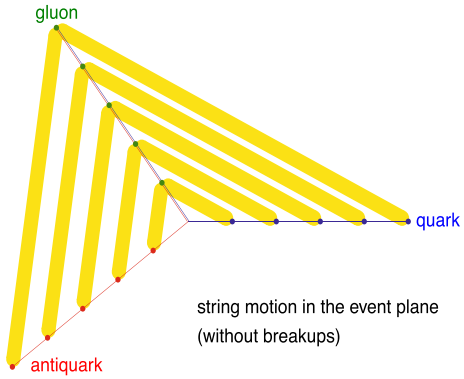


Fig. 282 String motion in a $q\overline{q}g$ system. Yellow regions represent snapshots in time of the string pieces. The fragmentation of the strings is suppressed for clarity. Reproduced from [3636]

case, and each has to tunnel a distance m_T/κ until it has absorbed enough energy from the string to come on-shell. This leads to a suppression factor

$$\exp\left(-\frac{\pi\, m_T^2}{\kappa}\right) = \exp\left(-\frac{\pi\, m^2}{\kappa}\right) \exp\left(-\frac{\pi\, p_T^2}{\kappa}\right). \quad (11.41)$$

The transverse momentum kick can be compensated locally between the q_i and $\overline{q}_i$, which defines a vector sum for each $q_i\overline{q}_{i+1}$ hadron. Empirically the observed $\langle p_T \rangle$ is somewhat higher than predicted this way, which could be related to the cutoff of soft gluons in the parton shower, so for tuning purposes the p_T width of primary hadrons is considered a free parameter.

The tunneling also implies that nonperturbative production of heavier quarks is suppressed, for charm and bottom to a negligible level. For the strangeness suppression it is not clear what quark masses to use – the observed $s/u \approx 0.25$ production ratio is in between results for current algebra and constituent masses – so again it is considered a free parameter.

Neglecting orbitally and radially excited states, a produced meson belongs either to the pseudoscalar or to the vector multiplet. Naive spin counting would imply a 1 : 3 production rate, but vectors are suppressed by the heavier mass, to an extent that is not easily calculated from first principles, so further parameters are introduced. The many flavor-related parameters is the biggest weakness of the string model.

For baryon production antidiquark–diquark string breaks are introduced, in analogy with quark–antiquark ones, with the diquark in a color antitriplet. Again tunneling, spin and mass factors are combined in production-rate parameters. The overall diquark break fraction needed to describe the observed baryon production rate is around 10%. A modified scenario is the popcorn one. In it, a $q\overline{q}$ pair can be produced with a color that does not screen the endpoint ones, such that it does not break the string. Inside that pair one or two further breaks may occur, where the latter would allow a meson to be produced between the baryon and the antibaryon.

String breaks on the average ought to be produced along a hyperbola of fixed invariant time, which translates into a flat rapidity plateau of produced hadrons. Then particle production would start in the middle of the event and spread outwards, Fig. 281. But actually all string breaks have a spacelike separation to each other, so there is no Lorentz-frame-independent definition of what comes first. It is then more convenient to begin at an endpoint quark and work inwards. The final result should be independent of the order used, which is satisfied for an almost unique fragmentation function shape

$$f(z) = \frac{1}{z}\,(1-z)^a\, \exp\left(-\frac{b\, m_T^2}{z}\right). \quad (11.42)$$

Here a and b are two free parameters and m_T the transverse mass of the hadron considered. The z variable parametrizes the fraction of remaining lightcone momentum that the hadron takes. That is, if the quark is moving in the $+z$ direction, then z is the fraction of $E + p_z$ taken, with $1 - z$ the fraction remaining to be used in subsequent steps. Note that heavier hadrons on the average take a larger fraction z than lighter ones. The $f(z)$ shape may be generalized slightly to take into account the effect of different quark masses, notably for massive (c or b) endpoint quarks.

The extension to more complicated string topologies involves no new principles or parameters. In a $q\overline{q}g$ event, Fig. 282, the $N_c \to \infty$ color algebra implies that one color is shared between q and g, and another between g and $\overline{q}$, with no direct connection between q and $\overline{q}$. The strings pulled out now have a transverse motion and thus a higher energy per unit length, but less length is pulled out per unit time, and these two effect exactly compensate to give $|\mathrm{d}E/\mathrm{d}t| = \kappa$ for the quarks and twice that for the gluon. This is to be compared with the QCD color-charge ratio $N_c/C_F = 9/4$.

Again the two string pieces will fragment along their length, with one hadron formed around the gluon kink, obtaining four-momentum contributions from both pieces. The generalization to multigluon systems is obvious, and also closed gluon loops can be addressed in this way.

The string motion becomes more complicated when a gluon is soft or collinear, with new string regions arising. The punch line, however, is that the total string motion is only mildly affected by such gluon emissions, although technical complications arise.

The occasional low-mass string also needs some special care, and in the extreme case it may become necessary to force the endpoint flavor content to form a single hadron, with four-momentum conservation being ensured by exchange with another parton or hadron.

LHC pp data has revealed several unexpected features, notably that fragmentation properties change when the multiplicity is increased, towards more strangeness and baryon production, and with signs of collective flow, both in the direction of the heavy-ion behaviour. Possibly a quark–gluon plasma is being formed, but in a string context it is also worthwhile to consider how the environment, i.e.the close-packing of strings at high multiplicities, could perturb the standard fragmentation picture. One such potential effect is color rope formation, i.e.that several parallel strings combine into an object in a higher color representation [3637]. Then baryon and strangeness production is enhanced. Baryons can in addition be enhanced by the aforementioned junction CR mechanism. There can also be a repulsive force between strings, so-called shove, that can give rise to collective flow [3638]. It remains to be seen whether these ideas can be combined into a new coherent framework in agreement with LHC observations.

The cluster model is based on the concept of preconfinement [3639] during the parton-shower evolution. That is, each color line (for $N_c \to \infty$) tends to correspond to a low-mass system, with only a small tail towards larger masses. The model becomes even more suggestive if it is assumed that all gluons branch into quarks, $g \to q\overline{q}$, at the end of the cascade, such that each color line is associated with a separate color singlet cluster. This would occur naturally if constituent masses obey $m_g \geq 2m_u = 2m_d$, as is supported by lattice QCD. Several cluster studies have been presented over the years. Here the generic features are outlined.

A gluon decays into any kinematically allowed $q\overline{q}$ pair according to its phase-space weight, which implies a dependence on the choice of gluon and quark constituent masses, notably whether $s\overline{s}$ can occur at this stage. Thereafter each $q_1\overline{q}_2$ cluster decays isotropically into a two-body state, hadrons $q_1\overline{q}_3$ and $q_3\overline{q}_2$, where $\overline{q}_3$ may also represent a diquark, resulting in baryon production. The hadrons are picked at random among all possibilities consistent with the flavors, according to relative weights. These weights are the product of the spin factor $2s+1$ for each final hadron and the phase-space factor $2p^*/m$, where p^* is the common magnitude of the three-momentum of the hadrons in the rest frame of the cluster with mass m. In some cases, such as $\pi^0-\eta-\eta'$, also the mixing of identical-flavor states needs to be included in the weight. It is also possible to allow an overall extra factor for a multiplet, notably to enhance baryon production.

A number of improvements have been introduced to this basic picture, as follows.

When the four-momenta of the cluster constituent q_1 and $\overline{q}_2$ are combined into the four-momentum of the cluster, the tail to large cluster masses is suppressed, but it is not completely absent. It is therefore assumed that such clusters can fission into two smaller ones, preferentially aligned along the $q_1\overline{q}_2$ axis. Flavor-dependent parameters are introduced to provide the mass above which a cluster must break, and others to describe the mass spectrum of the daughter clusters. The fission procedure can be repeated on the daughters, if necessary. In e^+e^- events ∼15% of the clusters need to be split, but these account for ∼50% of the final hadrons.

If baryons only are produced as baryon–antibaryon pairs inside isotropically decaying clusters then that does not agree with observed anisotropies in e^+e^- events. One solution is to allow $g \to qq + \overline{q}\overline{q}$ branches in the final stages of the shower. This has been implemented both for HERWIG and SHERPA, but has now been replaced by the next approach in HERWIG. (The possibility to rearrange three mesonic clusters into two baryonic ones has already been mentioned, but is not relevant for e^+e^-.)

Isotropic cluster decays also give too soft charm and bottom hadron spectra in e^+e^-. Therefore such cluster decays are treated anisotropically, such that the heavy hadron is preferentially near the heavy-quark direction, when viewed in the cluster rest frame. Some further improvements can be obtained if also other cluster decays preferentially favor hadrons closer to the cluster end with the matching flavor.

There may be a small fraction where the cluster mass is not large enough to produce two hadrons with the required flavor content. In such cases the cluster can be allowed to collapse into a single hadron, with excess four-momentum shuffled to another nearby cluster. For heavy quarks one may also allow some such collapses above the two-body threshold, to further harden the heavy-hadron spectrum.

Further procedures exist both in HERWIG and SHERPA to handle other special cases.

11.4.8 Decays, rescattering and Bose–Einstein

Many of the primary produced hadrons are unstable and decay further. Often the decay channels and their branching ratios are well-known, but for charm and especially bottom hadrons the picture is incomplete. Higher resonances are poorly known also in the light-quark sector. Furthermore,

inclusive measurements of a given final state may need to be translated into a potential sequence of intermediate states, e.g. $K\pi\pi$ may receive contributions from ρ and K^* resonances. Once the decay sequence has been settled, angular correlations in the decays should be considered, where feasible. Especially for bottom the EVTGEN package provides a large selection of relevant matrix elements, as does TAUOLA for τ lepton decay. The standard event generators also handle such nonisotropic decays to a varying degree.

The main pp generators assume that particles are free-streaming once formed. This is not the case in heavy-ion collisions, where the particle density remains high a while after the hadronization stage, and hadrons therefore can rescatter against each other. Studies show that also pp collisions are affected by rescattering, but not to a dramatic degree.

Another issue is Bose–Einstein (or Fermi–Dirac) correlations, present in the production of identical bosons (or fermions). Empirically this results in an enhancement (or depletion) of nearby pairs. Typical deduced emission source sizes range from somewhat below 1 fm in e^+e^- to somewhat above that for pp. Such scales obviously overlap with the hadronization ones, but also with the decays of short-lived resonances such as ρ, and with hadronic rescattering. The modelling therefore is far from trivial, and no traditional generator includes Bose–Einstein effects by default.

11.4.9 Other collision types

While the emphasis of the description above has been on the three main pp generators at LHC, a few words on adjacent fields and other generators are in place [3619]. Many of these other programs do not address hard physics, but are intended to describe inclusive events dominated by low-p_T QCD processes. Via an MPI machinery they may or may not contain a tail of harder QCD events.

Fields that can be covered by the e^+e^-/pp generators include Deeply Inelastic Scattering and photoproduction in ep or μp. The latter is largely based on the concept of Vector Meson Dominance (VMD), i.e.that a real photon can fluctuate into a vector meson state. The transition between the two regions of photon virtualities remains less easily modelled. The VMD picture can also be used e.g. for ultraperipheral $\gamma\gamma$ collisions in heavy-ion beams. Work remains to extend the ep collision framework to eA, as required for the simulation EIC physics.

Generators for heavy-ion physics span a broad range. In one extreme models introduce nuclear geometry and multiple $pp/pn/nn$ collisions, but with each collision similar to a regular pp one, up to energy–momentum conservation effects and the like. The earliest such example is FRITIOF, the ANGANTYR descendant of which is now included in PYTHIA. Others are SIBYLL, QGSJET and DPMJET. Such models can be run reasonably fast, and the latter three therefore are commonly used for the hadronic part of cosmic-ray cascades in the atmosphere.

In the other extreme the formation and evolution of a quark–gluon plasma (QGP) is the key feature. This requires the combination of models for several stages of the evolution, notably the hydrodynamical evolution of the plasma, which can be quite time-consuming. JETSCAPE offers a common framework where models for the different stages can be combined at will.

A successful intermediate program is the core–corona EPOS one [3640]. In it, peripheral $pp/pn/nn$ collisions (corona) occur more-or-less separated from each other, while the central higher-density core region may form a QGP, which then decays to hadrons according to a statistical model. The core QGP component gains in relative importance when going from pp to pA to AA, and from peripheral to central collisions. This gives a behaviour largely consistent with data.

Finally, generators for neutrino physics, like GENIE, are largely separate from the ones above, in that an emphasis lies on interactions with nuclei at low energies. The separation into a primary physics process followed by a simulation of detector effects thereby is blurred.

11.4.10 Standardization

The main generators discussed here largely are separate codes. This allows for cross-checks where results should agree, and a spread of predictions where the physics is not well-specified. Comparisons are greatly simplified by common standards.

The oldest standard is the PDG particle numbering scheme, whereby observed and postulated particles are assigned unique integer numbers.

The transfer of information from matrix-element generators to general-purpose generators is defined in the Les Houches Accord (LHA), and the associated Les Houches Event File (LHEF) [3641]. It specifies in particular a listing of all incoming and outgoing partons of a hard interaction, with their four-momenta. Extensions include multiple event weights to represent scale and PDF variations.

The transfer of the much bigger complete events from generators to detector simulation, or straight to users, is handled by the HEPMC standard [3642]. Again PDG particle codes and four-momenta provide the basic information. Also the step-by-step event history is documented, but cannot be made completely generator-independent since different physics is involved, e.g. strings versus clusters.

Parton distributions are widely used in generators, for hard interactions, MPIs and ISR. Today each new PDF set typically consists in the order of a hundred members, to provide a representation of the correlated uncertainties. Each member is stored as a file with the PDF value of all relevant partons

in a grid in (x, Q) space. LHAPDF [3643] specifies the file format, such that common interpolation routines can be used for the PDF evaluation for arbitrary x and Q values.

A major issue in the interpretation of data, not least for generator development and tuning, is the difficulty to reproduce all the methods and cuts used in the analysis, even after the data has been corrected for detector inefficiencies and smearing. Here the RIVET framework [3644] allows a standardized way for experiments to submit a code that takes generated (HEPMC) events and analyzes them in such a way that the output can be directly compared with published data.

11.4.11 The future

Before the start of the LHC, we believed to have a fair understanding of the physics at high-energy $e^+e^-/pp/p\overline{p}$ colliders. The hadronization description developed in the light of PETRA worked surprisingly well also at LEP. By jet universality – the assumption that the same hadronization mechanisms are at play in different collision types – the same should hold also for hadron colliders, when extended by aspects such as multiparton interactions and color reconnection.

The shock of LHC then was that high-multiplicity pp events were shown to behave surprisingly like heavy-ion collisions, with strangeness and baryon enhancement, both in the light-quark and in the charm/bottom sectors, and signs of collective flow such as ridge effects. From what we have been able to learn so far, it seems that high-p_T physics remains unaffected, such that there perturbation theory still can be reliably combined with LEP-tuned hadronization models. This makes sense, in that partons in that region mainly evolve in a vacuum. But, at low p_T, we already knew that the multiparton interactions lead to a close-packing of fragmenting systems, whether strings or clusters. We just had not fully appreciated its consequences, in part lulled by the common belief in the heavy-ion community that time scales in pp collisions are too short for a quark–gluon plasma to form. Now we are in the process of rethinking hadronization. One approach is the core–corona one, where a core part of the pp event indeed behaves like a plasma, while the corona part does not. The alternative is to avoid the plasma and introduce other possible mechanisms, such as junctions, ropes and shove. While some progress has been made, still no such coherent alternative exists. Anyway, the bottom line is that LHC has reinvigorated the study of the soft-physics aspects of event generators, in addition to obviously driving the hard-physics evolution, see Sect. 11.3.

While there is still much more to be learned from the LHC, attention is also turning to other future colliders. The one that may require most new generator development is the EIC, since it involves new physics scenarios not addressed before.

11.5 Jet reconstruction

Bogdan Malaescu, Dag Gillberg, Steven Schramm, and Chris Young

A QCD interaction at a very high energy, such as the hard process of an LHC collision, produces quarks and gluons that are asymptotically free at very short distances, but often result in a final state of hundreds of particles at the distance scales of detectors (> 1 mm). It is highly desirable to reduce the complexity of the hadronic final state and map it onto a representation that mimics the kinematics of the short-distance hard process. This is the goal of *jet algorithms*. Jet algorithms are a set of rules used to group directionally nearby particles to form jets. A jet can hence be thought of as a collimated group of particles that might correspond to a high energy parton of the hard process. The particles used as input to form jets can be of several types: a set of partons, a consistent set of hadrons, or a set of detector objects such as reconstructed charged-particle tracks or localized calorimeter energy measurements.

11.5.1 Jet algorithms

There are a number of desirable features for a jet algorithm. It should be computationally robust and well specified, ideally with few parameters. It should be theoretically well behaved, and exhibit both *infrared* and *collinear safety*. The former refers to adding one or several particles with infinitesimal energy, and the latter to split any input particle into two. For both these kinds of alterations of the input particles, the resulting jet four-momenta will be identical if the jet algorithm is safe against said effects. The jet algorithm should further behave equivalently at different orders of the QCD evolution: at the parton, hadron and detector levels. Furthermore, it should not be tailored to a specific detector, but be useful and used both by theorists and by experimental collaborations.

One of the early jet algorithms was the Snowmass Cone Algorithm of 1990 [3645]. This algorithm, which used E_T and operated in (η, ϕ)-space,[107] wrestled with several of the issues mentioned above. Complication arose due to choice of seeds and overlapping cones, which were dealt with by a merging and splitting stage of the jet algorithm, and which tried to find 'stable cones'. Similar cone algorithms with various improvements were employed by the CDF and DØ collaborations at Fermilab [3646,3647]. The k_t algorithm [191] was developed in 1993, inspired by QCD splittings scales (see Sect. 11.2). The advantages of the k_t algorithm are that it has no split/merge stage, and jets are uniquely defined; dis-

[107] $E_T \equiv E \sin(\theta)$ and the pseudorapidity $\eta = -\ln(\tan(\theta/2))$, where θ is the angle to the beam pipe.

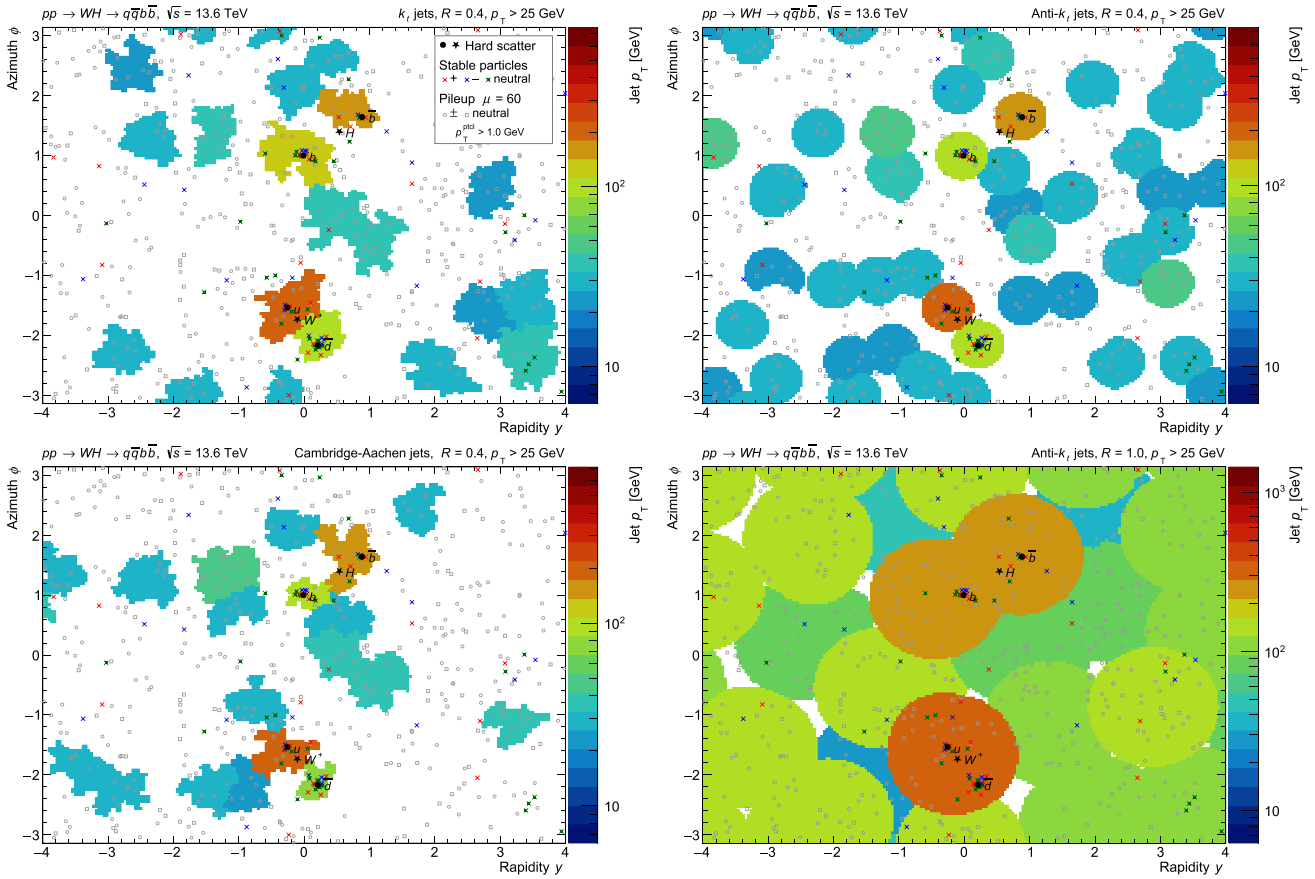


Fig. 283 The same simulated $pp \to W^+H \to u\bar{d}\,b\bar{b}$ event, reconstructed with four different jet algorithms: k_t (top left), anti-k_t (top right) and Cambridge–Aachen (bottom left), all with radius parameter $R = 0.4$, and anti-k_t with $R = 1.0$ (bottom right). The hard process particles are shown as black markers, while the final set of stable particles are displayed as crosses. Particles from pileup interactions, generated using a mean of $\mu = 60$ inelastic pp collisions, are shown as grey open markers. Particles with $p_T < 1\,\text{GeV}$ are not displayed. The solid colored areas show the extension (catchment area) of each jet with $p_T > 25\,\text{GeV}$, and their colors indicate the jet p_T. The code needed to produce this plot is available as the example program `main95` in recent PYTHIA distributions

advantages include the irregular jet shapes, and the difficulty to experimentally reconstruct and calibrate the jets.

Today, the most common method to build jets is the anti-k_t algorithm [193], defined very similarly to the k_t algorithm. Both algorithms start from a set of particles, each with associated four-momenta, and the following distance measures are calculated

$$d_{ij} = \min(p_{T,i}^{2p}, p_{T,j}^{2p})\frac{\Delta R_{ij}^2}{R^2}, \qquad d_{iB} = p_{T,i}^{2p}, \tag{11.43}$$

where R is a radius parameter, $\Delta R_{ij}^2 = \Delta y_{ij}^2 + \Delta\phi_{ij}^2$ is the distance squared in (y, ϕ)-space between particles i and j, and the parameter p is 1 for the k_t algorithm, 0 for the Cambridge–Aachen [189] algorithm and -1 for the anti-k_t algorithm. The distance d_{ij} is calculated for all combinations of pairs of particles, and d_{iB} once per particle. The smallest distance is found; if this is a d_{iB} value, then particle i will define a jet. If it is a d_{ij} value, then particles i and j are merged, normally by four-momentum addition ($p_k = p_i + p_j$). In both cases, the list of particles and the associated distances are updated, and the algorithm proceeds with one less particle per iteration until all particles have been used. When finished, each input particle is uniquely part of a jet. An illustration of the produced jets for these three k_t-style jet algorithms is presented in Fig. 283, where the jets are built for stable particles produced by a simulated $pp \to W^+H \to u\bar{d}\,b\bar{b}$ event at the LHC with a pileup contribution corresponding to a mean number of inelastic pp interactions of $\mu = 60$.

As is clear from Fig. 283, jets do not provide a unique interpretation of any given event, rather they are a tool that can be optimized to best address the needs of a given task. Even if jet algorithms are intended to represent the underlying hard process of a given collision, the variety of possible hard processes necessitates the consideration of different jet algorithm configurations. In other words, a jet algorithm

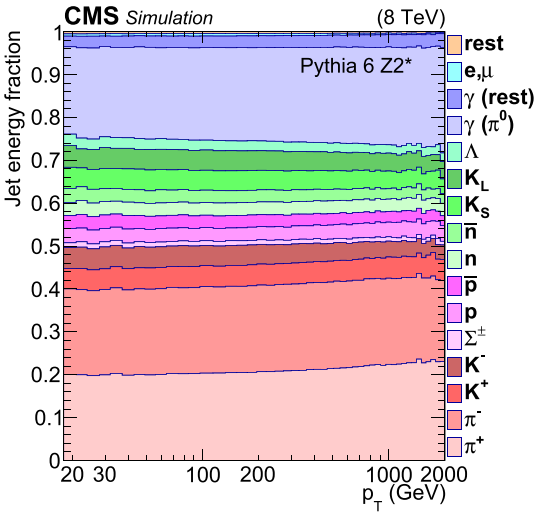


Fig. 284 The fraction of the total jet energy carried by different types of particles of particle-level jets produced in simulated LHC dijet events. Particle-level jets are built from particles ($c\tau > 10$ mm). The ratio of charged-to-neutral pions is 2:1 due to isospin symmetry, while for baryons it is 1:1; the overall charged-to-neutral fraction of particles in a jet roughly averages between these two expectations [3648]

defines an event organization concept and it can be adapted for different physics processes.

The most common usage of jets in the collider context is to represent the collimated group of final state particles originating from individual quarks or gluons of the hard scatter. For this task, the preferred jet radius parameter has slightly changed during the last decades. Values of $R = 0.6$ or 0.7 have often been used for studies of events with well separated jets (e.g. dijet production), while smaller radii (0.4 or 0.5) are more appropriate to resolve more complex final states, such as $t\bar{t}$ or the example shown in Fig. 283. As is further discussed in Sect. 11.5.2, smaller radii makes jets less susceptible to pileup, which has become an important consideration at the LHC. Since the start of LHC Run 2, the anti-k_t algorithm with a radius parameter of $R = 0.4$ has been the standard choice largely due to these reasons. The resulting jets are then interpreted as a set of quark- and gluon-initiated showers. Such jets are primarily composed of charged and neutral pions, but baryons and other types of mesons contribute a moderate fraction of the total jet energy, as shown in Fig. 284. Small energy fractions of electrons and muons can also be seen that originate from semi-leptonic heavy hadron decays.

A natural second-level question relating to such jets is to determine their underlying production mechanism. Is a given jet produced by a light-flavor quark ($u/d/s$), a gluon, a heavy-flavor quark (c or b), or by some other process? Heavy-flavor jets are typically easier to define at all levels: they can be identified by whether or not the list of constituents the jet is composed of contains b or c quarks at parton-level; B or D hadrons (or their decay products) at particle-level; or, have associated charge-particle tracks originating from collision-point-displaced vertices at the experimental level. The difference between light-flavor-quark- and gluon-initiated showers is more subtle, and is not rigorously defined for particle- or experiment-level jets. Instead, the expected properties of quarks and gluons can be used to differentiate between such jets on average, noting that quarks have a single color charge and are thus expected to radiate less, resulting in more narrow showers containing fewer constituents than showers produced by gluons.

Another important concept, and which is of great relevance at the LHC, is to use jets to represent complex energy flow processes rather than individual showers. The high energy collisions at the LHC can result in the production of massive particles, such as W, Z and H bosons and top quarks, with high transverse momentum. Therefore they have a sizable Lorentz boost in the rest frame of the detector, and their decay products will be collimated. In the case of hadronic decay products, each daughter particle further produces showers of hadrons, which can overlap. Rather than reconstructing this complex structure of overlapping hadronic showers as separate jets, the entire decay of the massive particle can be treated as a single jet, and properties of that jet can be used to infer the nature of the originating particle [3649]. In such a scenario, it is useful to increase the distance parameter used to build jets to contain the entire hadronic decay, as shown for the anti-k_t algorithm with $R = 1.0$ in the bottom right plot in Fig. 283 where the W boson decay is within a single jet, while the H boson decay is split between two jets. The collimation of the decay particles is related to the mass and momentum of the parent particle; for a two-body decay, this becomes:

$$\Delta R \gtrsim \frac{2\, m_{\text{parent}}}{p_{\text{T}}^{\text{parent}}}\,, \tag{11.44}$$

where ΔR is the angular separation between the decay products in (y, ϕ)-space. From this equation, it is clear that increased collision energies producing higher-momentum massive particles will result in increasingly collimated decays, and thus the importance of using a larger value of R to represent a complex energy flow is related to the energy scale of the process under study. Jets built with this context in mind are typically referred to as large-radius or large-R jets, where typical modern values are $R = 0.8$ for CMS or $R = 1.0$ for ATLAS; this is in contrast to the previously discussed $R = 0.4$ jets, which are referred to as small-radius or small-R jets.

Using a larger distance parameter comes with several complications, both experimental and theoretical. From the purely algorithmic perspective, one challenge is that the catchment area [3650] of an individual jet grows dramatically, as clearly visible when comparing the top right and bottom right plots in Fig. 283. Among other effects, this

increases the amount of energy from the underlying event included in the jet, which can hide the features of interest: for example, the mass of the jet should peak at the mass of the parent particle, but this is not the case due to the presence of the underlying event. This can be mitigated through the use of a variety of different grooming algorithms [1915,1916,3649,3651,3652]. These algorithms are typically applied after building the initial jets. The objects clustered into the jet are then subject to a further selection, and those which appear to be inconsistent with originating from a hard-scattering process are removed, thus suppressing the underlying event and other undesired contributions while retaining the physics features of interest.

11.5.2 Jet reconstruction

Inputs to jet reconstruction:
Particle-level jets, often referred to as *truth jets*, are used as a theoretical reference for experimental measurements. These are jets built from *stable particles*, defined as those with lifetime τ such that $c\tau > 10\,\text{mm}$ ($\tau > 33\,\text{ps}$), which can be thought of as "what a perfect detector would see". It should be noted that neutral pions are not considered stable and hence their decay products (photons) will be used as input to truth jets (see Fig. 284). Only particles produced in the proton–proton interaction of interest are considered. These jets also form the reference for the calibration of reconstructed jets.

Experimental reconstruction of jets requires the definition of a given set of inputs, which will ideally represent the true particles of the jet or the energy flow. As jets consist of both charged and neutral hadrons, the simplest reconstruction makes use of the energy flow captured in a calorimeter, which measures the energy of both charged and neutral incident particles. However, as we will see in this section, the accuracy of jet reconstruction can be improved through the use of additional information from tracks reconstructed from charged particles.

At a hadron collider such as the LHC, a wide range of energies of jets need to be accurately reconstructed: from 20 GeV to above 4 TeV in p_T. This represents a significant challenge for the design of the detectors. Both ATLAS and CMS surround the interaction point with a tracking detector immersed in a magnetic field, such that the momentum of charged particles can be measured. Around this are the calorimeters. The innermost calorimeters are designed to reconstruct electromagnetically showering particles, such as electrons and photons, and will also capture some energy from charged and neutral hadrons. Radially outward of these detectors are hadronic calorimeters that measure the energy of showers from remaining charged and neutral hadrons.

An additional complication at the LHC is *pileup*. Each time two bunches of protons cross, multiple pairs of protons can collide. This is referred to as in-time pileup. The beam-spot, the region of interactions, is typically 30–50 mm in length along the beam direction. This means that such collisions are typically separated in this dimension and tracks originating from different collisions can be identified. A second effect is out-of-time pileup. The bunches of protons cross every 25 ns in the LHC, therefore there are still residual effects in many of the detectors from the previous (and subsequent for some systems with large integration times) bunch crossings. These residual signals are referred to as out-of-time pileup.

Throughout Run 1 of the LHC (2010–2012), ATLAS used solely calorimeter inputs to build their jets. The ATLAS calorimeters consist of over 100,000 cells. This fine cell granularity is used to suppress noise by constructing clusters of cells, which represent the energy flow. Cells with energy significantly greater than the expected background noise are used to seed such clusters, and adjacent cells are added iteratively, forming topologically connected clusters representing a shower [3653]. This process means that most cells in the calorimeter are not included in the event reconstruction, and hence their noise does not contribute to the jet resolution. As the calorimeters are non-compensating, showers caused by electromagnetically and hadronically interacting particles of identical initial energies have different energy responses. The jet resolution can therefore be improved by identifying which type of shower each cluster contains and calibrating it appropriately. In ATLAS the energy density of the cluster and its position in the calorimeter are used for classification and subsequent calibration [3653]. These calibrated clusters were the input signals to jet reconstruction for ATLAS in Run 1.

CMS has employed a particle-flow approach both in Run 1 and Run 2 [3655], and ATLAS also developed such an approach for Run 2 [3654]. The principle of particle flow is to supplement the information from the calorimeter with tracking information. Both collaborations match tracks reconstructed in the inner detector to calorimeter energy deposits from the same particle. The ability to do this depends on the granularity of the detector, the small transverse size of the showers in the calorimeter, and the separation of the particles. Figure 285 shows how this can be achieved by extrapolating tracks through the magnetic field to the calorimeter and matching them to calorimeter energy deposits. The CMS algorithm combines the measurements of tracks and matched calorimeter-energy deposits to create combined reconstructed charged hadrons with improved resolution. Calorimeter deposits without tracks are then identified as neutral hadrons. Situations where the showers of a charged hadron and a neutral hadron are overlapping are identified by the excess of energy in the calorimeter above what would be expected from the charged hadron. In ATLAS a choice is made between the calorimeter and track reconstruction. For low p_T tracks, where the track resolution is significantly bet-

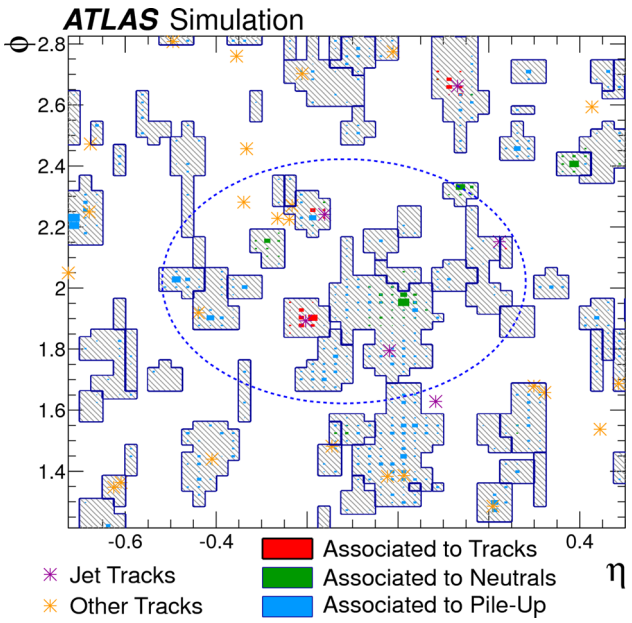


Fig. 285 The simulated signals from a $p_T = 30\,\text{GeV}$ jet in the (η, ϕ) plane of the second layer of the ATLAS electromagnetic calorimeter. The shaded cells are those included in calorimeter topoclusters. Green deposits are from neutral hadrons within the jet, red deposits are from charged hadrons within the jet, and blue deposits are from pileup particles. The purple $*$ represents the tracks of charged hadrons within the jet after being extrapolated to the calorimeter, and the yellow $*$ represents tracks from pileup [3654]

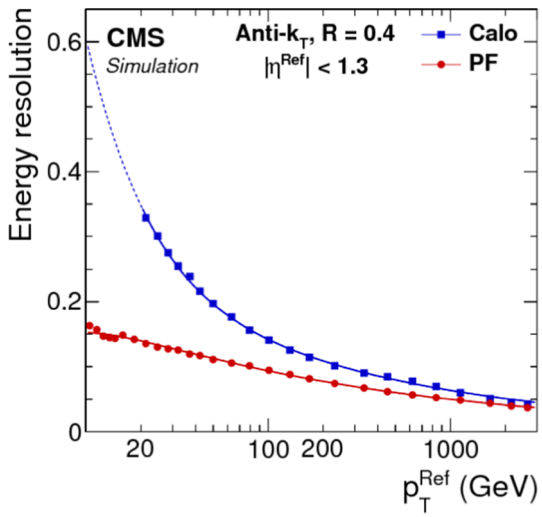


Fig. 286 The jet resolution in the central region of the CMS detector when jets are reconstructed using calorimeter signals (Calo) or particle flow objects (PF). The simulated QCD events have $\sqrt{s} = 13\,\text{TeV}$ and there are no pileup effects present [3655]

ter than that of the calorimeter, the momentum measurement is taken from the reconstructed track and the corresponding shower created by that particle is removed from the calorimeter. The remaining calorimeter energy deposits then represent the energy flow from particles without tracks and those where the track is not selected.

Both collaborations see significant improvements in the p_T and angular resolutions of jets reconstructed using particle flow. Figure 286 shows the dramatic improvement in the energy resolution in CMS. In ATLAS the improvement is smaller, and primarily at lower p_T, due to the superior calorimeter resolution. However, the gains from the use of particle flow increase at higher pileup motivating its use in Run 2 and beyond.

Jet algorithms in the experimental context:
Having reconstructed either clusters or a set of particle flow objects, the jet algorithms featured in Sect. 11.5.1 can be used to build jets. A key advantage of using particle flow objects is that prior to building the jets, charged particles that are from in-time pileup interactions can be excluded. This is known as Charged Hadron Subtraction, and is performed by both experiments' particle flow algorithms [3654,3655]. This removes the majority of the effects of charged pileup particles but the effects due to neutral pileup particles and out-of-time pileup remain. This explains why ATLAS observes increasing benefits of the particle flow approach at higher pileup. Additionally CMS employs PUPPI [3656] which uses the local information to try to identify neutral pileup energy deposits and weight these to lower significance prior to jet finding [3656,3657].

Some small-R jets reconstructed from either calorimeter or particle flow inputs will consist of only signals from pileup particles. These are referred to as pileup jets and can be the result of QCD jets from other in-time collisions, multiple particles from different in-time collisions, out-of-time pileup signals, or a combination of several of these effects. These jets will not have tracks pointing at them from the interaction vertex of interest, while they will in some cases have tracks from other vertices. These features are used by both ATLAS and CMS to reject such jets such that they are not used in analyses [3657,3658].

Large-radius jets are much more susceptible to pileup, due to their larger catchment area. Most large-radius jets at the LHC will therefore contain a mixture of energy originating from multiple collisions (either in-time or out-of-time), and thus it is impractical to reject entire jets. Moreover, large-radius jet are typically used in situations where the internal structure of the large radius jet is of interest, and thus any constituents originating from other processes than the hard-scatter interaction must be suppressed to observe the jet's internal structure. Charged hadron subtraction, from particle flow algorithms, can help to remove charged contributions for separate collisions, but alternative strategies are needed to remove overlapping neutral contributions. Grooming algorithms, previously motivated in the context of suppressing underlying event contributions, are also useful in this context: the same criteria of suppressing soft and wide-angle

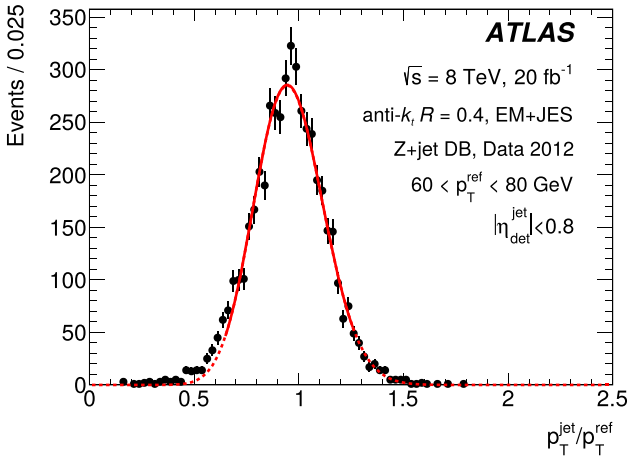


Fig. 287 Distributions of $p_T^{\text{jet}}/p_T^{\text{ref}}$ in Z+jet events, where p_T^{ref} is defined by the reconstructed Z boson p_T and is required to be in the range (60, 80) GeV. The dashed line shows the fitted distribution, from which the means are taken as the response measurements. The solid line indicates the fitting ranges. The markers are the data counts with error bars corresponding to the statistical uncertainties. Figure from Ref. [3662]

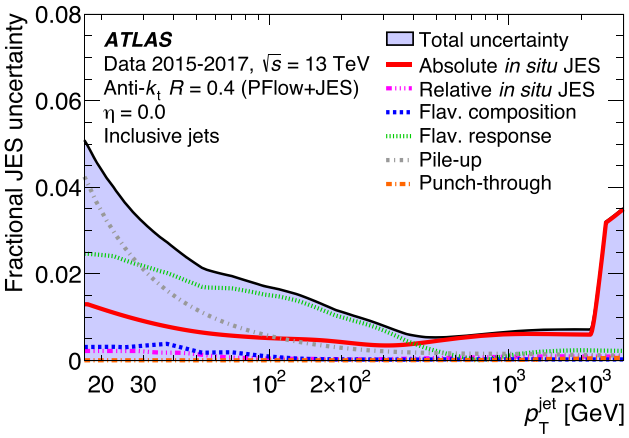


Fig. 288 Fractional jet energy scale systematic uncertainty as a function of p_T for jets reconstructed from particle-flow objects. The total uncertainty, determined as the quadrature sum of all components, is shown as a filled region topped by a solid black line. flavor-dependent uncertainty components shown here assume a dijet flavor composition. Figure from Ref. [3663]

radiation is also useful for mitigating pileup contributions. These grooming algorithms are applied after the jet is built, but the inputs to jet algorithms can also be corrected; various criteria can be used to suppress neutral jet inputs from vertices other than the one of interest, such as Constituent Subtraction (CS) [3659], Soft Killer (SK) [3660], PUPPI, or combinations thereof such as CS+SK. Currently, ATLAS makes use of CS+SK to modify the inputs to large-radius jet reconstruction [3661], while CMS makes use of PUPPI [3657].

11.5.3 Jet calibration

Energy scale and resolution:
Once jets are reconstructed, they need to be calibrated such that on average the reconstructed jet four-momenta match those at the particle level within the assigned uncertainties. At hadron colliders, the jet energy-scale (JES) calibration-correction is typically applied in a sequence of steps. Those account for (the mitigation of) contributions from additional proton–proton collisions, energy losses in the dead material of the detector, calorimeter non-compensation (where applicable), angular biases, etc. Several of these calibration steps rely on a detailed Monte Carlo simulation (MC) of detector effects. Modern techniques use jet and event properties (e.g. jet area, jet width, fraction of energy in the various layers of the calorimeters, average p_T density) to improve resolution and to mitigate the dependence of the JES response on the jet flavor. The latter are sizable mainly at low jet transverse momentum (p_T) and yield one of the main modeling uncertainties impacting the JES calibration.

The calibration chain is completed by in situ corrections that are most commonly derived by exploiting momentum balance between jets and well-measured reference objects. Selection criteria are applied to suppress extra radiation and obtain a sample of events where a probe jet is back-to-back with the reference object. A correction is then derived by comparing the measured balance in data relative to the expectations of MC simulation, and correcting for the residual difference:

$$\left(p_T^{\text{jet}}/p_T^{\text{ref}}\right)^{\text{data}} \Big/ \left(p_T^{\text{jet}}/p_T^{\text{ref}}\right)^{\text{MC}}. \tag{11.45}$$

This principle was developed for the calibration of small-radius ($R \in [0.4, 0.7]$) jets [3648,3662,3663] and has now also been used for large-radius jets [3664].

These in situ methods employ, as reference objects, photons, Z bosons decaying to charged leptons, and one or several pre-calibrated jets. Fig. 287 presents an example of $p_T^{\text{jet}}/p_T^{\text{ref}}$ distribution in data, the mean of which is used to derive the jet calibration. They also provide the main uncertainties impacting the JES calibration(see Fig. 288), reaching nowadays sub-percent precision across a broad phase-space, while being larger for relatively low- and large-p_T jets, as well as in the forward region of the detectors. While for large-p_T jets these approaches are limited by the available statistics, for low-p_T and forward jets they are limited by modeling effects, related to e.g. the emission of extra radiation impacting the p_T balance. The use of in situ techniques have allowed for significant improvements in precision compared to jet calibrations based on test-beam studies. The latter are still used in phase-space regions with little/no statistics coverage for the in situ approaches.

Statistical combinations, with a full propagation of uncertainties and correlations, are generally employed and yield the necessary inputs for physics analyses. In these studies, uncertainties on the uncertainties and on the correlations have also been evaluated (see e.g. Ref. [3665]). This is an example where QCD studies trigger developments that set new standards on a topic of interest in other scientific areas too.

The width of the $p_T^{\text{jet}}/p_T^{\text{ref}}$ distributions, such as the one exemplified in Fig. 287, provides information about the jet energy resolution (JER). Indeed, the JER is determined in various p_T^{jet} ranges and detector regions, after subtracting statistically the smearing effect induced by the presence of extra radiation in the events. Afterwards, a statistical combination of several in situ methods through a fit allows for the extraction of a parameterization of the JER in data, together with its uncertainties, readily usable in physics analyses accounting for detector smearing effects.

Mass scale and resolution:
To first order, calibrations derived to correct the energy of a jet are also important to use when correcting the mass of a jet, as these two quantities are related. However, the mass calculation includes both energy and angular components, and thus the jet mass must be further corrected after the energy has been addressed. Similarly to the energy, calibrating the mass of a jet begins with corrections based on simulated samples, to correct the average simulated jet mass to the particle-level scale. In the context of large-radius jets, it is very important to apply the same grooming algorithms to the truth jets and the reconstructed jets, as the grooming algorithm has a substantial impact on the mass of the jet built from particles, primarily due to the suppression of the underlying event contributions.

Following these simulation-based corrections, the resulting mass must be compared between data and simulation, but the strategies to evaluate differences between data and simulation necessarily differ. The jet energy corrections exploited the conservation of momentum in the transverse frame through the balance between probe and reference objects to obtain a precise calibration. There is no equivalent conservation law for jet mass, so a different approach is needed. Instead, the mass has a well-defined expectation value in specific cases, notably if a pure sample of W, Z or H bosons or top quarks can be obtained. W bosons and top quarks are the easiest particles to identify in this context: semi-leptonic $t\bar{t}$ events provide an ideal means of identifying a high-purity selection of hadronically decaying top quarks, and the distinction between a full top decay and a W boson decay can be made by requiring the b-quark from the top decay to be either inside or outside of the large-radius jet of interest. The resulting high-purity selection of W bosons or top quarks can be compared between data and simulated events, where differences in the mass peak's central value

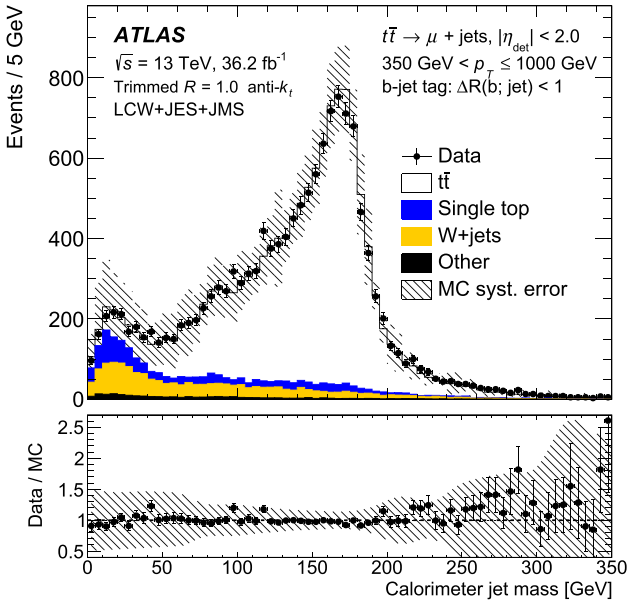


Fig. 289 The mass of large-radius jets in a final state targeting semi-leptonic decays of $t\bar{t}$ events, where a b-tagged jet overlaps with the large-radius jet. This selection primarily identifies large-radius jets containing the decays of boosted top quarks, as is clear from the dominant peak structure consistent with the $t\bar{t}$ simulation expectation. Differences between data and simulation in both the jet mass scale and resolution can be extracted from such a plot. [3664]

(mass scale) and width (mass resolution) can be evaluated and corrected for; an example of the top quark selection is shown in Fig. 289, where it is clear that the selected events are very pure in the signal of interest.

This approach works well, but is limited to only a few possible jet mass values where we have a well-defined Standard Model expectation. Correcting the scale and resolution for other mass values is a much more complex task, and a robust, high-precision method to provide a general mass correction remains an open challenge.

11.5.4 Classifying hadronic decays of massive particles

The use of large-radius jets is overwhelmingly linked to the desire to represent the entire hadronic decay of a massive particle, such as (but not limited to) a $W/Z/H$ boson or a top quark. If the jet does contain all of the daughter particles and their corresponding showers, then the mass of the jet now has a well-defined prior, namely the mass of the parent particle. This prior holds so long as the large-radius jet represents only the process of interest: underlying event and pile-up contributions falling within the jet's catchment area can both obscure the internal structure of the jet, and must thus be mitigated, as previously discussed. The mass then becomes an excellent means of classifying jets based on the parent particle that they originate from.

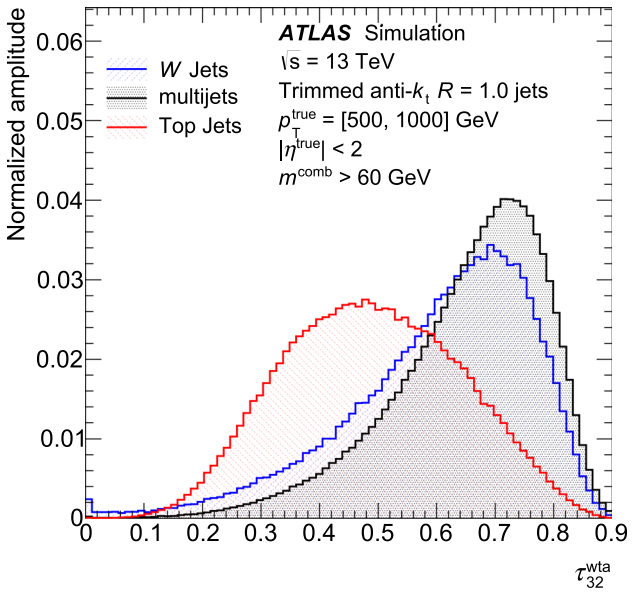


Fig. 290 The N-subjettiness ratio τ_{32}, with the *winner takes all* (wta) axis definition [3666], showing the separation of jets containing three-body decays (top jets) against jets containing either two-body decays (W jets) or individual quarks/gluons (multijets). This is after a selection criterion is applied on the jet mass, and thus the separation shown between the three jet types provides additional classification power. [3667]

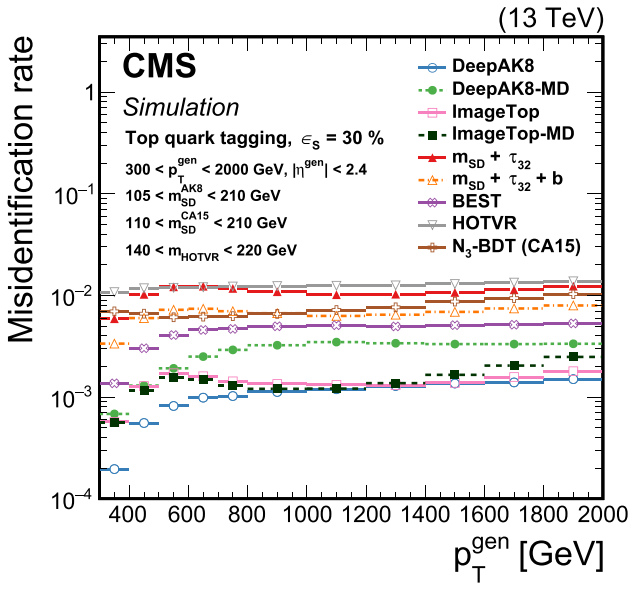


Fig. 291 A comparison of many different algorithms designed to identify jets originating from hadronic decays of top quarks. A simple tagger based on the jet mass combined with the N-subjettiness ratio τ_{32} ($m_{SD} + \tau_{32}$) is shown alongside many alternative classifiers providing significantly better performance. This can be seen as they have much lower quark/gluon (background) misidentification rates for a fixed top quark (signal) efficiency. The large majority of the alternative classifiers make use of machine learning techniques [3668]

While the jet mass provides a robust means of differentiating between different possible sources of large-radius jets, in many cases it is not sufficient, as light quarks and gluons from QCD multijet processes are produced in extreme abundance compared to the massive particle decays of interest. The mass distribution of light quarks and gluons is peaked at low values, well below the $W/Z/H$ boson or top quark masses, but the tail of the mass distribution extends to high masses, and these tails are still more probable than the production of the target massive particles.

Additional jet properties can be used to further classify the origin of a given large-radius jet. These properties are referred to as *jet substructure variables* and are designed to quantify the internal angular energy structure of a jet. Substructure variables are almost always correlated with the jet mass, and thus it is important to identify variables that are sufficiently distinct to provide further separation power. One commonly used example is the N-subjettiness ratios, $\tau_{xy} = \tau_x/\tau_y$, where τ_n is a projection of the constituents of a jet along n axes, thereby evaluating the consistency of the jet containing n or fewer decay particles. As an example, τ_{32} is commonly used to identify jets containing top quarks, as it differentiates 3-body decays from 2-or-fewer-body decays, as shown in Fig. 290. This is only one example out of the many different types of jet substructure variables that have been used to complement the jet mass in classifying the origin of large-radius jets.

While the jet mass and substructure variables provide a solid baseline, modern large-radius jet classifiers make use of machine learning techniques to maximally discriminate between different possible jet origin interpretations. There is a wide variety of machine-learning-based classifiers in use by both the ATLAS and CMS Collaborations, and they continue to become more powerful; a comparison of several such algorithms as used by CMS is provided in Fig. 291.

Similar to the jet energy and mass calibrations, the difference between data and simulation must also be quantified when classifying the origins of large-radius jets. The algorithms used are usually optimized using simulated events, and there is no guarantee that the simulation properly describes the data, especially for the complex angular energy structure within a jet, which is what such classifiers rely upon to differentiate between different jet categories. Similar to the jet mass scale calibration, semi-leptonic $t\bar{t}$ events provide a useful signal-enriched region to evaluate the performance of both W boson and top quark classifiers in simulation and data; other signal categories remain more challenging, as we do not yet have sufficiently signal-pure regions to perform similar comparisons. In contrast, comparing the differences between data and simulation for the misidentified background events is straightforward, as the QCD multijet and γ+jet processes have such large cross sections that they are highly background enriched by default. Any differences between data and simulation in the fraction of background

events passing a given large-radius jet classifier can thus be evaluated using these two processes.

11.5.5 Summary

Jets are crucial tools for numerous physics analyses performed at hadron colliders. During the last four decades, there has been significant development in this field and jet definitions that are robust for both experimental measurements and theoretical predictions have been identified. In addition, the improved particle detectors, with highly granular calorimeters and high resolution reconstruction of charged tracks are enabling reconstruction of the full jet four momentum, investigation of the jet internal structure and classification of jets via tagging. These developments are allowing us to expand the knowledge about QCD and to look for signatures of BSM physics, yielding greatly improved searches and measurements.

12 Measurements at colliders

Conveners:
Karl Jakobs and Eberhard Klempt

The development of QCD and of colliders are intimately linked. The formation of jets, of streams of collimated hadrons, was first observed at SPEAR; soon after, gluons were investigated in three-jet events at TASSO (see Sect. 3.2). The Large-Electron–Positron (LEP) collider, operating from 1989 to 2000, allowed the collaborations to determine the energy-dependence of the strong-interaction coupling constant α_s, to confirm the gauge structure of QCD and to test QCD systematically (see Sect. 12.1). The identification of jets with quarks and gluons is of prime importance to understand the dynamics of the first hard reactions (see Sect. 11).

The study of the structure of protons in deep inelastic electron–proton scattering at SLAC and neutrino-proton scattering with Gargamelle at CERN led to experimental evidence for quarks with electric charge assignments as predicted by the quark model. Later, these studies were continued at HERA, the only electron–proton collider, and at other places. These studies are presented in Sect. 10. Three quarks (c, b, t) were discovered in collider experiments (the b-quark at least co-discovered). The physics of the Brookhaven Relativistic Heavy Ion Collider (RHIC) is discussed in Sect. 7.

Since the 1980s the high-energy frontier of particle physics was defined by the $Sp\bar{p}S$ collider at CERN and the Tevatron at Fermilab. As outlined by Daniel Britzger, Klaus Rabbertz and Markus Wobisch, the production of jets developed to a QCD testing ground to searches for new phenomena up to the largest accelerator-based energies at the Large Hadron Collider (LHC). Jets initiated by gluons, quarks – including the heavy quarks c and b – can be produced jointly with the vector bosons $W^{\pm}$ and Z^0. The measured cross sections of all these processes are precisely reproduced by QCD calculations (Monica Dunford). The discovery of the Higgs boson in 2012 was a milestone for particle physics. Chiara Mariotti describes with which surprising precision the properties of the Higgs boson follow the predictions of the SM. The top quark, discovered in 1995 at the Tevatron and discussed here by Marcel Vos, is identified in a large variety of production processes, from top–anti-top production to $t\bar{t}$ production associated with a vector boson or the production of two $t\bar{t}$ pairs. The cross sections for these processes span a wide range from nearly 10^3 pb down to a few 10^{-2} pb.

12.1 The legacy of LEP

Stefan Kluth

The large electron positron collider LEP was conceived and designed at CERN in the 1980s to study the then just discovered massive vector bosons of the Standard Model (SM), the neutral Z and the charged $\mathrm{W}^{\pm}$ bosons [3669]. The four LEP experiments ALEPH, DELPHI, L3 and OPAL collected more than four million Z decays and about 10,000 W pairs each. The LEP 2 runs at centre-of-mass (cms) energies above the Z resonance up to 209 GeV provided samples of about 1000 hadronic final states from off-shell $(Z/\gamma)^*$ decays at each cms energy. These data, together with the extremely accurate LEP beam energy determination, established "electro-weak precision observables" (EWPO) and the confirmation of the SM at very high precision [3670]. Previous studies of QCD in e^+e^- annihilation at the PEP and PETRA colliders and in other experiments are summarized e.g. in [3671–3674].

It was thus clear that hadronic final states at LEP are a great laboratory to study a large spectrum of QCD predictions. The missing initial- and final-state interference and the comparatively high energy lead to clearly interpretable hadronic final states and usually small corrections from non-perturbative effects. All LEP experiments have among their first few publications papers on properties of hadronic Z decays.

The detectors of the LEP experiments were significant improvements on their predecessors and offered an almost complete coverage of the solid angle with efficient and precise tracking and finely grained calorimeters with layers for electromagnetic and hadronic showers. All LEP experiments had silicon micro-vertex detectors and full coverage with muon detection systems outside of the calorimeters.

The e^+e^- initial state with well known beam energies provides a strong constraint to improve energy measurements. For example the scaled jet energies in Z decays to 3-jets can be determined from jet angles only [3675]. Even without using the beam energy directly in a constraint the use of quantities scaled to the cms energy reduces dependence

on the absolute energy scale of the detector. As explained below, jet definitions, event shape observables and particle spectra are normalized to the cms energy $Q = \sqrt{s}$. Note that in the measurements the normalization to Q is replaced by the measured total visible energy E_{vis} which also partially removes the influence of statistical fluctuations.

Compared to previous experiments, the LEP data have much larger event samples on the Z peak, low experimental systematic uncertainties and higher cms energies leading to smaller and well controlled hadronization corrections.

The data taken on the Z peak (LEP 1) have favorable experimental conditions. The trigger efficiency for hadronic final states is essentially 100% and can be measured using redundant triggers. Backgrounds from hadronic decays of τ-lepton pairs are suppressed by demanding more than four charged particles. Requirements on balance of observed momentum along the beam direction and total visible energy remove backgrounds from $e^+e^- \to 2\gamma \to$ hadrons interactions. There are corrections for initial state photon radiation effects, but on the Z peak these are small.

The data taken at $\sqrt{s} > m_Z$ but below the threshold for W^+W^- pair production (LEP 1.5) at $\sqrt{s} = 130$ and 136 GeV already contain a substantial fraction of so-called "radiative return" interactions $e^+e^- \to \gamma_{\rm ISR} + Z \to$ hadrons.[108] Simply speaking, instead of a high-energy interaction near the nominal $\sqrt{s}$, a Z decay to hadrons recoiling against the ISR photon γ_{ISR} is produced. The LEP collaborations developed algorithms to reconstruct the effective cms energy $\sqrt{s'}$ of the observed hadronic system by assuming a 2-body decay together with one or more high-energy ISR photons.

The data taken at $\sqrt{s} \geq 2m_W$ (LEP 2) include an increasing fraction of so-called "4-fermion" final states including quarks. These 4-fermion final states are dominated by W^+W^- pair production in the all-hadronic or lepton+jets channel depending on the decays of the W-bosons. After a hadronic preselection the di-lepton channel is a rather small background. The LEP collaborations developed sophisticated selections for the W^+W^- pairs for the precise measurements of W-boson properties designed to reject "2-fermion" final states with quarks $e^+e^- \to (Z\gamma)^* \to q\bar{q} \to$ hadrons [3676]. These results are then the basis for selections of hadronic final states produced via a $(Z/\gamma)^*$ at high energy. The remaining 4-fermion background in the data increases with $\sqrt{s}$ to about 10% at the highest LEP 2 energies but contributes mostly in regions dominated by multi-jet topologies, see e.g. [3677]. Figure 292 shows the distribution of $\sqrt{s'_{rec}}$ observed for hadronic final states at $\sqrt{s} = 200$ GeV by DELPHI [3677]. The peak at $m_Z \simeq 91.2$ GeV is due to hadronic Z decays recoiling against photon ISR. The analysis imposes a cut on $\sqrt{s'_{rec}}$ to select the peak near the nominal $\sqrt{s} = 200$ GeV. The yellow shaded area shows the simulated background contribution of W^+W^- and ZZ final states with hadrons.

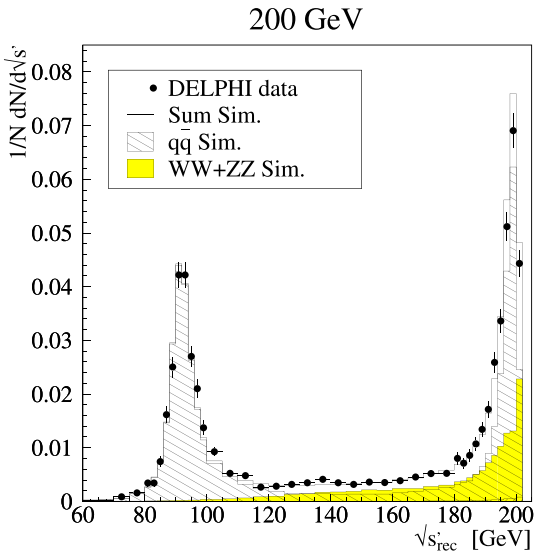


Fig. 292 The figure shows the distribution of reconstructed effective cms energies $\sqrt{s'_{rec}}$ in hadronic final states in e^+e^- collisions at $\sqrt{s} =$ 200 GeV. The data are compared with simulations of hadronic final states mediated by a single $(Z/\gamma)^*$ ($q\bar{q}$ Sim.) and W^+W^- or ZZ pair production (WW+ZZ Sim.) [3677]

12.1.1 Gluon and quark properties

The gluon was established as one of the elementary particles of the SM by the PETRA experiments, see Sect. 2.2. QCD requires for its gauge bosons that they have spin-1, and that they carry color charge themselves manifesting in three- and four-gluon vertices of the QCD Lagrangian.

The phenomenological analysis of the jet axes in Υ decays to three gluons provided evidence for the spin-1 assignment [3678]. The QCD predictions for spin-0 and spin-1 gluons were the basis of an analysis by OPAL [3678] using the energy distribution of the 2nd jet after energy ordering in hadronic Z decays to three jets [3675]. The 2nd-jet-energy distribution after correction for experimental and hadronization effects was in good agreement with a NLO QCD prediction while a MC based LO calculation with scalar gluons showed an estimated $\chi^2/dof = 44/14$. This is clearly well above the requirements for a discovery. A similar study is discussed in [3679].

The search strategy for observable effects of the three-gluon vertex was discussed in [3680], but convincing results could only be obtained after NLO calculations became available for the angular correlations between four jets in hadronic Z decays [3681]. The QCD predictions at NLO decompose into contributions proportional to (products of) the color factors C_F, C_FC_A, C_FC_F and $C_FN_FT_F$, and two of them can be determined together with the strong coupling $\alpha_S(M_Z^2)$.

[108] ISR stands for initial state radiation.

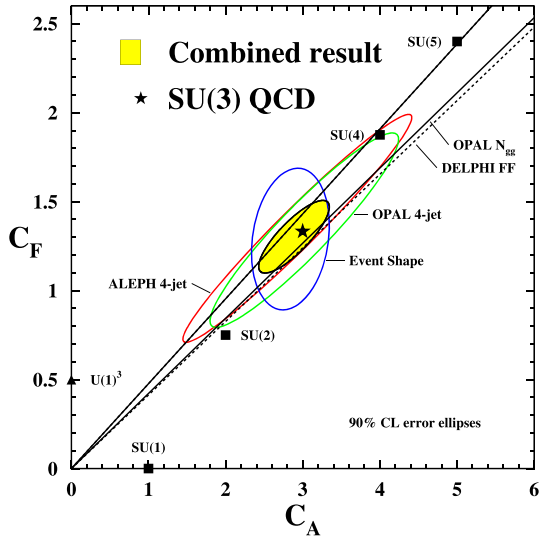


Fig. 293 The figure shows results for the color factors C_A and C_F from various analyses as indicated [3685]

The analyses by OPAL and ALEPH [3682,3683] determine C_A and C_F corresponding to the contributions of three-gluon or quark–gluon vertices to the NLO predictions. The contribution of the three-gluon vertex proportional to C_A is clearly observed. Since the result for the second color factor product can be recast as C_F, the color charge of quarks at the strength required by QCD is observed as well.

The analysis of the event shape observables Thrust and the C-parameter (see below for details) at several cms energies from re-analysed JADE (at PETRA) data and from OPAL data is based on the same decomposition of the NLO QCD prediction and also results in a clear observation of the three-gluon vertex contribution [3684]. A combination of these and other results for determinations of the color factors is discussed in [3685]. Figure 293 shows a summary of the results for C_A and C_F from 4-jet angular correlations, event shapes, and other analyses [3685].

The properties of quarks in the SM such as their spin-1/2 assignment and their electric charges have been studied at LEP and earlier collider experiments [122,3686,3687]. Another quark property is their mass, which will be discussed below in Sect. 12.1.4.

12.1.2 Jets and event shapes

Jet and event shape observables have been designed to study properties of hadronic final states at colliders. The aim generally is to classify hadronic final states according to their topology by introducing an additional energy scale. E.g. for clustering hadronic final states in e^+e^- annihilation with the JADE algorithm [3688] $m_{ij}^2 = 2E_iE_j(1-\cos\theta_{ij})$ is the distance between two objects i and j with energies E_i and E_j. At each iteration the pair ij with the smallest distance m_{ij} is merged by adding the pair's 4-vectors.[109] One can introduce the scaled quantity $y_{cut} = m_{cut}^2/s$ and count how many events have three jets when the clustering is stopped at y_{cut}. Alternatively, the value of $y_{23} = m_{23}^2/s$ can be used to classify events where in each event the clustering goes from three to two jets [3689]. In the first case, jet rates are studied and in the second an event shape observable is used. The Thrust observable $T = \max_{\vec{n}} \sum_i |\vec{p}_i \cdot \vec{n}| / \sum_i |\vec{p}_i|$ quantifies the coherence or "jettiness" of an event. Here, i runs over all particles in the hadronic final state, $\vec{p}_i$ are the particle 3-momenta, and the thrust axis $\vec{n}$ is a unit vector that maximizes T. The nominator $\sum_i |\vec{p}_i|$ defines an energy scale.

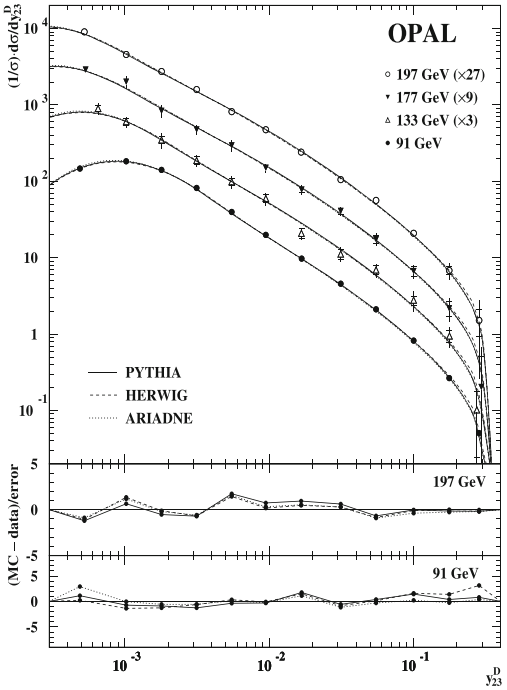
The value of an event shape observable is the classifier which can distinguish between e.g. collimated 2-jet like events and broader 3-jet (or multi-jet) like events. Their distributions reflect the proportion of 2-jet like vs. 3-jet or multi-jet like events in the data in a similar way as the fraction of 3-jet events at a fixed value of y_{cut}.

As discussed by Dokshitzer in Sect. 2.3, it is the property of infrared-collinear safety which allows for stable prediction by perturbative QCD (pQCD) and thus for a meaningful comparison between experimental observations and pQCD predictions. However, before a successful quantitative comparison of experiment and theory can be made, the transition from the partons of pQCD calculations to the observed hadrons (hadronization) must be accounted for. If there was a major redistribution of 4-momenta between partons and hadrons in a given final state due to hadronization, a comparison of pQCD predictions with data would be highly problematic. Turning this argument around we must have a hadronization process which is local in phase space. This is discussed as "local parton hadron duality" (LPHD) by Dokshitzer in Sect. 2.3. Experimental evidence for the LPHD collected by the LEP experiments and previous studies is discussed below.

Figure 294 (left) shows as an example the measurements by OPAL of the event shape observable y_{23}^D at cms energies $\sqrt{s} = 91.2$, 133, 177 and 197 GeV. The cms energies are weighted averages of combined LEP runs with similar cms energies. The observable y_{23}^D is the value of the jet distance in the Durham algorithm [187] $y = 2\min(E_i, E_j)^2(1-\cos\theta_{ij})/s$ where the number of jets changes from three to two. Figure 294 (right) shows measurements by ALEPH [3690] of n jets, $n = 1,\ldots,6$, or more production fractions using the Durham algorithm. These data show that at LEP hadronic final states with complex jet topologies can be measured well.

The reasonably successful comparisons of the data with simulations by the Monte Carlo event generators PYTHIA, HERWIG and ARIADNE validate the experimental corrections derived using these simulations after passing them through the simulations of the detectors. Furthermore, they

[109] This is the E-scheme, other merging schemes exist.



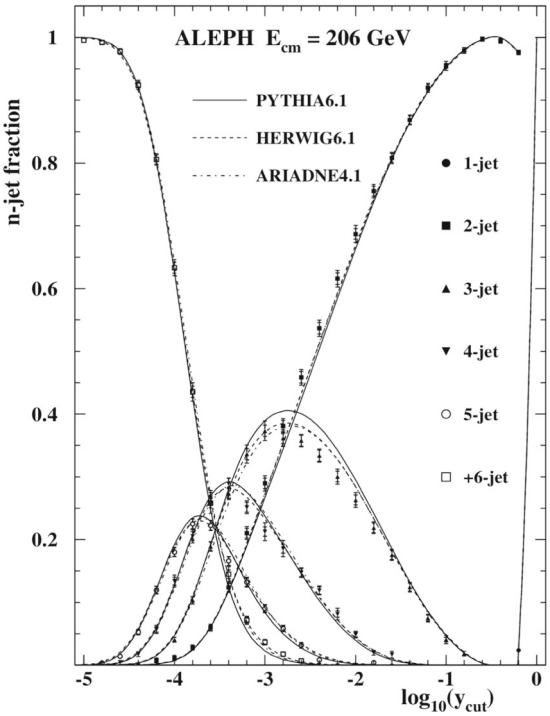


Fig. 294 (left) The figure shows measurements of the event shape observable y_{23}^D by OPAL at average cms energies as indicated. The measurements are corrected for experimental effects and are compared with simulations as indicated [3691]. (right) The figure displays measurements of n-jet production fractions as a function of y_{cut} using the Durham algorithm by ALEPH at $\sqrt{s} = 206$ GeV. The measurements are compared with simulations [3690]

pave the way for using these simulations to derive the hadronization corrections needed to compare pQCD predictions with the data. The final LEP measurements and their comparison to the then relevant NLO+NLLA QCD predictions and determinations of $\alpha_s(m_Z)$ are discussed in [3685]. Improved determinations of $\alpha_s(m_Z)$ using NNLO QCD predictions combined with resummed NLLA calculation appeared soon after the NNLO predictions became available [3692–3695]

The QCD analyses of some jet rates and event shape distributions (starting at 3-jet final states) from LEP and previous e^+e^- experiments today has reached percent level precision using pQCD predictions at NNLO combined with resummation up to N3LL. For example in [302] distributions of Thrust at $\sqrt{s} = 35$ to 200 GeV are analysed in a global fit based on NNLO+N3LL QCD predictions.[110] The hadronization corrections are applied using an analytic model integrated into the prediction. The final result is $\alpha_s(m_Z) = 0.1135 \pm 0.0011$ and has a relative uncertainty of 1%. A similar measurement using the C-parameter is [303], the energy-energy correlation EEC was analysed in NNLO+NNLL accuracy and the 2-jet rate with the Durham algorithm was studied with N3LO+NNLL predictions [3696].

Limitations for the ultimate accuracy of these studies are currently the uncertainties connected with hadronization corrections, see e.g. [3697] for a recent study. An early study [3690] based on event shapes at all LEP energies and NLO+NLL pQCD found differences in $\alpha_s(m_Z)$ of about 10% between results using MC simulations or an analytic model to derive hadronization corrections. These differences became smaller with more complete QCD predictions such as NNLO+NNLL or NNLO+ N3LL. They also tend to reduce when MC simulations with NLO calculations matched to the parton shower are used. In both cases a larger fraction of the prediction is contributed by pQCD and thus only a smaller difference w.r.t. the data is left to be covered by hadronization corrections. New studies show that the hadronization corrections in an improved analytic model depend on the event shape value [3698], in contrast with the analytic models used so far.

The analyses of final states with four or more jets are based on the accurate measurements of multi-jet rates and corresponding event shape distributions at LEP. Similar to NLO QCD predictions for angular correlations in 4-jet final states also NLO predictions for 4-jet rates became possible [3699]. It is important to realize that for 4-jet final states the NLO QCD prediction is $O(\alpha_s^2) + O(\alpha_s^3)$ which implies a sensitivity to α_s larger by about a factor of 2 compared with a prediction

[110] The exact power counting is explained in [302].

for 3-jet final states. The higher sensitivity can compensate for the larger experimental uncertainties of the 4-jet measurements w.r.t. 3-jet measurements. The corresponding analyses with LEP data are [3677,3700] while [3701] is a study based on re-analysed data from JADE at PETRA.

Automated NLO QCD calculations allowed predictions for 5-jet observables [3702] and the corresponding determination of $\alpha_s(m_Z)$. By the same argument as above the sensitivity to $\alpha_s(m_Z)$ is enhanced w.r.t. 3-jet observables which compensates for larger measurement uncertainties.

The review of measurements of $\alpha_s(Q)$ in Sect. 3 shows clearly that the strong coupling strength decreases with increasing energy scale of the process, i.e. asymptotic freedom. Here we discuss direct experimental evidence without performing measurements of α_s. Figure 294 (left) shows distributions of y_{23}^D measured at cms energies from 91 to 197 GeV and a change in the distribution is clearly visible. A more direct way to observe a change of the strong coupling strength with the energy scale of the process $Q = \sqrt{s}$ is to use inclusive observables such as jet production rates at a fixed value of y_{cut} or moments of event shape observables.

In QCD in LO the prediction for the mean value of e.g. the Thrust $1 - T$ distribution is $\langle 1 - T\rangle(Q) = \alpha_s(Q^2)A_{1-T}$ while the running coupling follows $\alpha_s(Q^2) = \alpha_s(\mu^2)/(1 + \alpha_s(\mu^2)\beta_0 \ln(x_\mu^2))$, $\beta_0 = (11C_A - 4T_F N_F)/(12\pi)$, $x_\mu = Q/\mu$, μ is the renormalization scale. This implies $1/\langle 1 - T\rangle \sim \ln Q$ at LO with $O(\alpha_s^2)$ corrections. Figure 295 displays data from DELPHI and lower energy experiments for $1/\langle 1 - T\rangle$ as a function of Q on a logarithmic scale confirming the QCD prediction for the running coupling as measured by $\langle 1 - T\rangle$. Hadronization corrections to $\langle 1 - T\rangle$ are predicted using simulations to only change the logarithmic slope, see e.g. [3703]. Earlier studies using JADE (at PETRA) data for 3-jet rates using the JADE algorithm as a function of cms energy had already proven the running strong coupling at the 4-σ level [3704].

12.1.3 Fragmentation

Here, we use the term fragmentation to refer to measuring and predicting properties of the hadrons produced in hadronic final states. In studies of fragmentation of hadrons the energies or momentum components w.r.t. an event orientation or jet axis, or their multiplicity, are studied.

The scaled momentum fraction of a hadron with momentum p is defined as $x = 2p/Q$. One expects in the quark–parton model, i.e. in the absence of strong interactions of the partons, that the x-spectra of hadrons are independent of $\sqrt{s}$. This is analogous to the prediction of scaling for x_{Bj} in lepton–hadron DIS, i.e. that distributions of x_{Bj} are independent of the 4-momentum transfer Q^2 of the DIS process. Scaling violations are then due to scale-dependent strong interactions of the partons. Figure 296 shows as an example

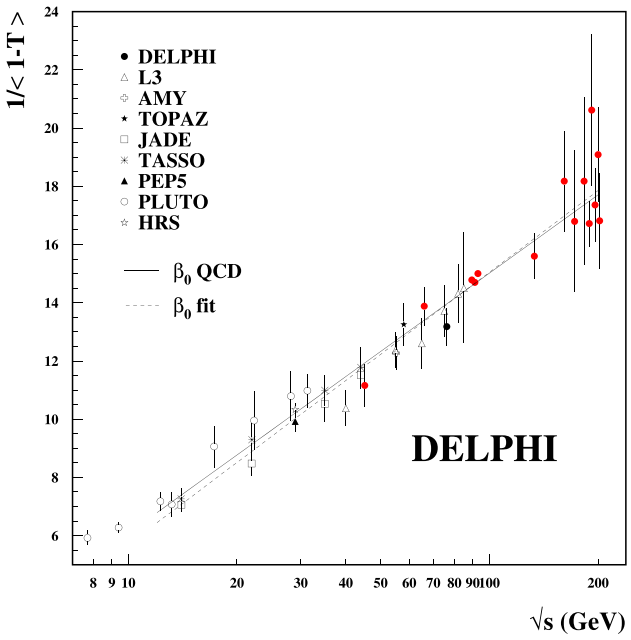


Fig. 295 The figure shows measurements of $1/\langle 1 - T\rangle$ as a function of $\sqrt{s} = Q$ on a logarithmic scale by DELPHI and lower energy experiments. The lines show a NLO QCD prediction and fit by DELPHI [3703]

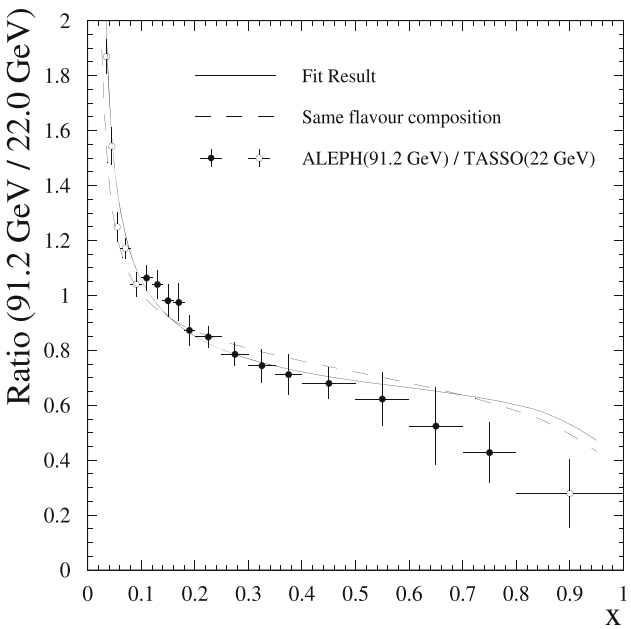


Fig. 296 The figure shows the ratio of the scaled momentum spectra $1/\sigma_h d\sigma_h/dx$ of charged particles measured by ALEPH at $\sqrt{s} \simeq 91.2$ GeV to data from TASSO measured at $\sqrt{s} = 22$ GeV [3705]

the ratio of measurements of x-spectra measured by ALEPH on the Z peak to corresponding measurements by TASSO (at PETRA) measured at $\sqrt{s} = 22$ GeV [3705]. The scaling violations are clearly visible.

The QCD analysis of scaling violations of scaled momentum distributions measured at different cms energies is the e^+e^- analog of the analysis of structure functions $F_2(Q^2, x_{Bj})$ in lepton–hadron DIS. The scaled momentum distribution is described by

$$\frac{1}{\sigma_h}\frac{d\sigma_h}{dx} = \int_0^1 \sum_f C_f(z, \alpha_s(\mu), x_\mu) D_f\left(\frac{x}{z}, \mu\right)\frac{dz}{z} \quad (12.1)$$

with the flavor index $f = u, d, s, c, b, g$. The C_f are coefficient functions known in NNLO QCD, and the D_f are nonperturbative fragmentation functions. The D_f correspond to the probability to obtain a hadron with momentum fraction x from a parton f analogous to the parton density functions (PDF) of DIS. The rate of change with changing momentum scale μ of the D_f is described by the DGLAP equations, see 2.3. A first NNLO framework for the analysis of scaled momentum distributions in e^+e^- annihilation to hadrons is [3706].

It is interesting to focus on low momentum hadrons. To this end the variable $\xi = \ln(1/x)$ is introduced. The majority of hadrons is produced at low values of x, and by transforming to ξ their properties can be studied in more detail. As an example Fig. 297 shows measurements of ξ for charged hadrons at LEP by OPAL and also from previous experiments at lower energies [3707]. The distributions show a maximum and drop quickly towards small ξ corresponding to large hadron momenta. At large ξ, i.e. for small momenta, the distributions fall off faster than expected from the kinematic limits from hadron masses.

This can be explained by destructive interference of multiple soft gluon radiation in the parton shower, often named soft gluon coherence. Under the assumption of LPHD the production of soft hadrons is driven by the production of soft gluons from the parton shower. The "QCD Chudakov effect" means that soft gluons cannot resolve the individual parton color charges and instead the smaller color charge before branchings is relevant. Based on these ideas detailed pQCD predictions for multiple soft gluon radiation are calculated. For Fig. 297 such predictions [3708] are shown by the solid and dashed lines, where the solid lines are fitted to the data and the dashed lines are extrapolations. The extrapolated QCD predictions at small ξ (large x) are not expected to be a good approximation while at large ξ (small x) the data are well described. The evolution of the peak position with cms energy extracted from the fits also follows the pQCD prediction, see e.g. [3707]. These measurements provide convincing experimental confirmation of the LPHD and the corresponding pQCD calculations. A recent analysis of the ξ spectra measured in e^+e^- annihilation and other processes including higher order corrections is presented in [3709].

The interpretation of [3710], based on simulations with and without soft gluon interference (coherence) effects, that

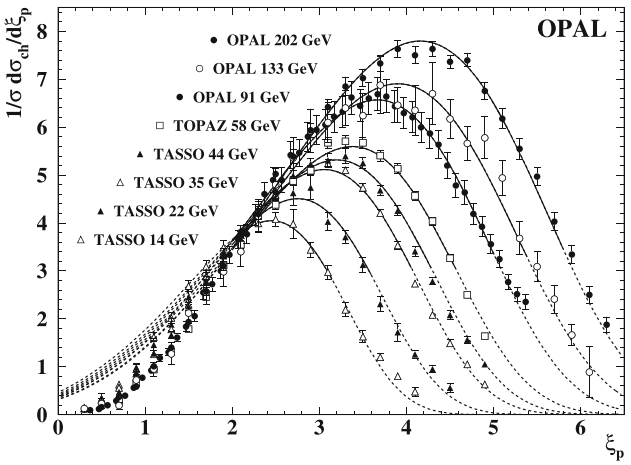


Fig. 297 The figure shows the spectra of $\xi = \ln(1/x)$ measured by OPAL, TOPAZ and TASSO [3707]. The data are compared with fitted QCD predictions, see text for details

the data of scaled momentum spectra do not provide evidence for coherence has been discussed in [3711]. There it was pointed out that the hadronization models of the simulation programs will compensate for the lack of coherence effects to still give a reasonable description of the data. The confirmation of the LPHD lies in the successful comparison of the corresponding QCD calculations with the data involving only two free normalization and scale parameters.

The QCD parton shower picture, i.e. the idea of high-momentum partons radiating many times a gluon, and also gluons producing a $q\bar{q}$ pair, is the basis of the simulation programs, because it allows implementations as iterative probabilistic branchings. The implementations of the parton shower picture are approximations correctly summing leading logarithmic terms (LLA). In the LLA the soft gluon interference effects correspond to the angular ordering phenomenon: a subsequent parton branching must occur at branching angles smaller than the previous one. There are limitations to the angular ordering approximation for less inclusive observables [3712], in particular some which are used for tuning (optimization of agreement with data) of the simulation programs.

The legacy of LEP in this area is the wealth of precise data on event shape observables, jet production, spectra of inclusive and identified hadrons, and multiplicities which can in many cases be interpreted with little ambiguities. These data are to a large part the basis for parameter settings of the popular simulation programs used in our field and in particular at the LHC [3618,3636,3713].

The topic of color reconnection (CR) concerns possible changes to hadronization effects if several color singlet sources are produced in a collision. The question is: do the final partons in parton showers of different color singlet sources merge to form hadrons together or not. At LEP 2

the production of $e^+e^- \to W^+W^- \to hadrons$ final states was an important contribution to the LEP 2 measurements of the mass and other properties of the W-boson [3676]. The modeling uncertainties of CR effects gave rise to significant systematic errors on the W boson mass and width. Later, measurements of particle flow between the four jets of the two hadronic W decays were used to constrain different CR models. New models for CR were discussed in [3634] and compared with the LEP 2 measurements. CR also affects measurements of the top quark mass due to the intermediate color singlet W-boson in the top quark decay [3714] and due to interactions of proton remnants (multi parton interactions MPI) in pp collisions. Recent measurements from LHC take this into account [513,3715]. The CR model with the biggest impact on the results of [3715] is also the only one in tension with LEP 2 data in [3634]. This shows that the LEP data can still help to constrain CR models.

12.1.4 Heavy quarks

In QCD with massless quarks the coupling constant is the only free parameter. Asymptotic freedom of the running strong coupling is one of the defining features of QCD and is well confirmed by experiments [3716] since LEP results contributed. Quark masses are also free parameters of the theory and subject to similar phenomena as asymptotic freedom for the strong coupling. The quark masses are predicted to depend on the energy scale of the process through so-called mass anomalous dimensions, the quark mass analogous of the beta-function.

The two main phenomenological predictions are first, that effective quark mass values decrease with the energy scale of the interaction implying asymptotic freedom for quark masses, see e.g. [3717] for a review. The second prediction is the suppression of gluon radiation from massive quarks with angle $\Theta < \Theta_0 = m/E$, where m is the heavy quark mass and E the heavy quark energy. This is referred to as the "dead-cone" effect of QCD [3718].

The first prediction of running quark masses was studied at LEP using large samples of O(10^5) hadronic Z decays with b-tags. Mass effects can be enhanced for observables like the 3-jet rate $R_3(y_{cut})$ due to their additional energy scale y_{cut} [3719]. In order to reduce common experimental uncertainties a double ratio $B_3 = R_3^b/R_3^l$ is defined, with $R_3^{b(l)}$ the 3-jet rate in Z decays to b (light) quarks. Figure 298 shows data for B_3 from ALEPH compared with NLO QCD predictions for values of the running b-quark mass in the $\overline{\text{MS}}$ scheme $m_b(M_Z) = 3$ or 5 GeV [3720]. The data are consistent with the lower value of $m_b(M_Z)$.

The analyses by ALEPH, DELPHI, OPAL and SLD are summarized in [3685,3721] with $m_b(M_Z) = 2.82 \pm 0.28$ GeV [3721]. With $m_b(m_b) = 4.18^{+0.03}_{-0.02}$ [513], the s-

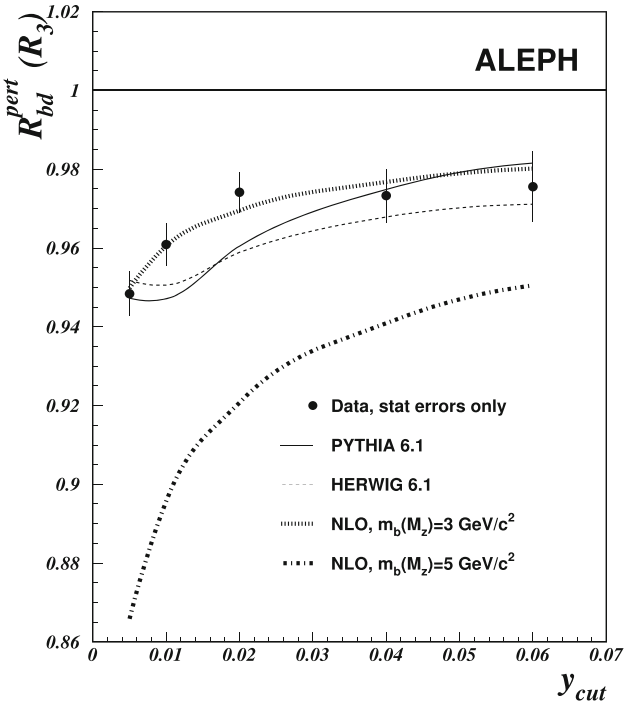


Fig. 298 The diagram shows data for $B_3(y_{cut})$ by ALEPH corrected for experimental and hadronization effects using the Durham algorithm. The lines show NLO QCD predictions for $m_b(M_Z)$ values as indicated, as well as predictions from simulations [3720]

dependence of the b-quark mass is observed with a significance of more than four standard deviations. The analysis [3721] adds a determination of $m_b(M_H)$ from measurements of the branching ratio of the Higgs boson to b quarks by the LHC experiments ATLAS and CMS assuming the Yukawa coupling of b quarks at its SM value. Figure 299 presents results for $m_b(m_b)$, $m_b(M_Z)$ and for $m_b(M_H)$ together with the QCD prediction for the running $m_b(Q)$ [3721]. There is good agreement between the measurements and the QCD prediction.

The dead-cone effect is not straightforward to study at LEP or other colliders. For example for b-jets from on-peak Z decays at LEP the dead-cone angle is expected to be $\Theta_0 \simeq 2m_b/m_Z \simeq 0.1$ which is well inside typical jet energy profiles in hadronic Z decays [3722]. A recent analysis by ALICE has found evidence for reduced particle production inside angular regions consistent with the dead cone for charm-tagged jets produced in pp collisions at the LHC [195]. The key to this observation was reversing a sequential jet clustering history using an angular distance definition[111] which enforces angular ordering by construction.

Predictions for phenomenology of the dead cone effect at LEP concentrate on multiple soft gluon production and thus on particle spectra or multiplicities [3718,3723]. The

[111] The Cambridge/Aachen (C/A) algorithm.

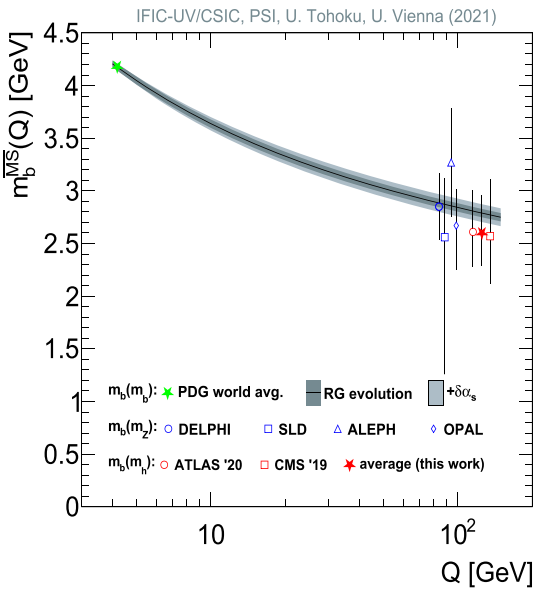


Fig. 299 The figure shows determinations of $m_b(m_b)$, $m_b(M_Z)$ and $m_b(M_H)$ together with the QCD prediction for the running $m_b(Q)$ [3721]

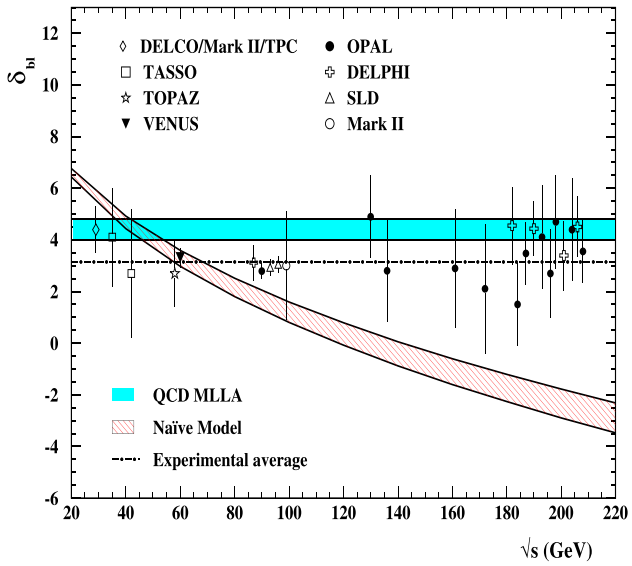


Fig. 300 The figure presents measurements of δ_{bl} compared with QCD predictions and an alternative model as indicated [3724]

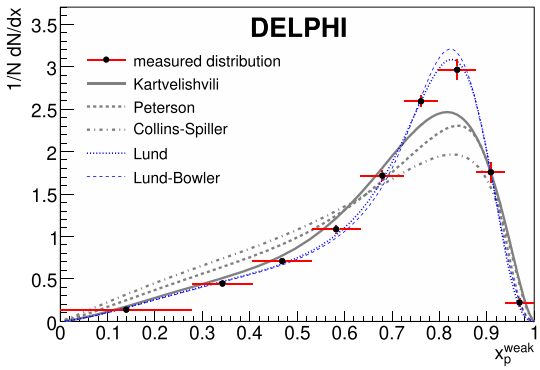


Fig. 301 The figure shows the b quark to B hadron fragmentation function for weak b decays. The lines display predictions by simulations with a fixed perturbative component and different models for the fragmentation functions [3725]

so-called "leading particle effect" refers to large and mass dependent average scaled momenta of heavy hadrons (c or b). The leading particle effect is derived from pQCD as a direct consequence of the dead cone and shown to be consistent with data from LEP and previous e^+e^- colliders [172].

The particle multiplicity in Z decays to b or light (u, d, s) quarks is sensitive to the dead cone effect due to its impact on soft gluon radiation, which is directly related to particle production via the LPHD. The pQCD prediction in the MLLA for the charged particle multiplicity difference in hadronic Z decays to b or light quarks is $\delta_{bl} = 4.4 \pm 0.4$ [3724]. A different model for δ_{bl} without dead cone contributions predicts a fast decrease with cms energy $\sqrt{s}$. The predictions for δ_{bl} and measurements by LEP experiments and previous experiments at different $\sqrt{s}$ are shown in Fig. 300 [3724]. The blue band corresponding to the QCD dead cone prediction is in agreement with the data within theoretical and experimental uncertainties. The alternative model is excluded by the high energy LEP 2 measurements at $\sqrt{s} \geq 183$ GeV with an estimated $\chi^2/dof \simeq 100/11$. The hypothesis that $\delta_{bl} \to 0$ for large $\sqrt{s}$ leads to an estimated $\chi^2/dof \simeq 43/11$ and is thus also clearly excluded.

Another example of precision measurements in the heavy flavor sector is the b quark to hadron fragmentation function. The measurement by DELPHI is shown in Fig. 301 [3725]. The quantity x_p^{weak} refers to the scaled momentum of the B hadron reconstructed from its weak decay. In this way possible preceding strong decays of excited B hadrons are accounted for. In the figure the data are compared with several models for the fragmentation functions folded with a fixed perturbative component. The data can clearly separate the different models. Recent parameter optimizations of e.g. the PYTHIA simulation take these results into account [3713].

12.1.5 Zedometry and hadronic τ decays

The EWPOs measured by the LEP experiments and by SLD at Stanford are the main legacy of the e^+e^- collider program. The EWPOs are also a valuable legacy for the understanding and experimental verification of QCD. All EWPOs connected with quarks will have SM predictions with QCD corrections reflecting gluon radiation. Corrections to pure electroweak processes involving quarks scale typically like $1 + C\alpha_s(m_Z)/\pi$, where C is a process specific constant, and are thus expected to modify electroweak EWPO predictions by a few %.

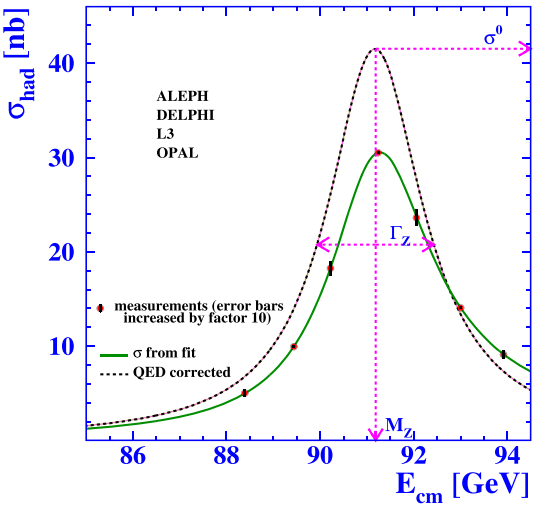


Fig. 302 The figure displays measurements of the hadronic cross section in e^+e^- annihilation at cms energies E_{cm} around m_Z measured by the LEP experiments. The lines show the model-independent fit to extract EWPOs before and after QED corrections [3670]

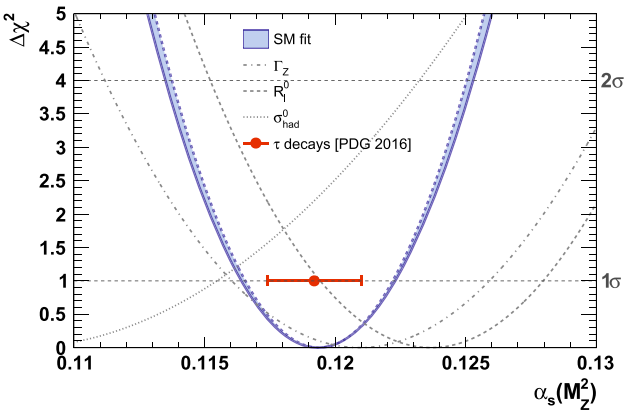


Fig. 303 The figure shows with the blue band the χ^2 profile of a global SM fit as a function of the value of the strong coupling $\alpha_s(m_Z)$. The grey lines are similar profiles for individual EWPOs as indicated. The red data point shows the value of $\alpha_s(m_Z)$ determined from hadronic τ lepton decays [3726]

Figure 302 shows measurements of cross sections for the process $e^+e^- \rightarrow hadrons$ at cms energies around $E_{cm} = m_Z$ by the LEP experiments [3670]. The measurements map out the Z boson resonance in e^+e^- annihilation in the hadronic channel. The lines show the result of a model-independent fit before and after QED corrections to these and other measurements to extract the Z boson resonance parameters such as the mass m_Z, the total width Γ_Z, the R-ratio $R_l^0 = \Gamma_{Z,had}/\Gamma_{Z,ll}$ and the hadronic pole cross section σ^0_{had}.

The extracted parameters are part of the set of EWPOs which can be compared with predictions by the SM including the QCD corrections. The QCD corrections for the EWPOs connected with the Z lineshape are known to N3LO, the corrections due to mixed and non-factorising electroweak and strong interaction diagrams are known up to $\alpha\alpha_s$ terms, and the QCD corrections for massive quarks are known up to $(m_q/Q)^4\alpha_s(Q)^3$, see [3726] for details.

Figure 303 shows the χ^2 profile of a recent SM global fit as a function of the strong coupling $\alpha_s(m_Z)$ using the LEP data and other data for the masses of the top quark, the W-boson and the Higgs boson [3726]. The blue band shows the χ^2 of the global fit around the best value of $\alpha_s(m_Z)$. The grey lines show the contributions to this result of the most sensitive EWPOs. The width of the band reflects the theoretical uncertainties of the global SM fit. A comparison of the grey bands shows the consistency between the QCD corrections to the different EWPOs. The red data point is a direct measurement of $\alpha_s(m_Z)$ from the hadronic branching ratio of τ lepton decays measured mostly using LEP data.

In the SM description, weak decays of τ-leptons to hadrons proceed via a virtual W-boson decaying to quarks. Similar to hadronic Z-boson decays, QCD corrections to the final state modify the predictions. At the scale of the τ lepton mass $m_\tau \simeq 1.78$ GeV the strong coupling $\alpha_s(m_\tau) \simeq 0.3$ such that large corrections are expected. The QCD corrections are also known to N3LO due to the similarity of the calculations. In addition, non-perturbative effects are significant, while they are strongly suppressed for hadronic Z decays.

A recent analysis of the important theoretical issues for the extraction of α_s from hadronic τ lepton decays is [3727]. The data point shows the average of $\alpha_s(m_Z)$ determinations by the PDG from 2016 which has since been updated with only small changes [513]. The good consistency between these related determinations of $\alpha_s(m_Z)$ is a strong test of the consistent application of QCD corrections in the SM, as well as of the understanding of the evolution equations for the running of the strong coupling including the treatment of quark mass thresholds.

The large collection of measurements from the LEP experiments, SLD, and also the previous and partially re-analysed experiments at e^+e^- colliders are a cornerstone of the experimental validation of the theory of strong interactions, QCD. Possible future e^+e^- colliders are designed to deliver at least 1000 times the integrated luminosity w.r.t. LEP and with more advanced detectors. In addition large samples of Higgs and W^+W^- bosons, and possibly of top–antitop quark pairs will open the door to many more tests of the SM including its QCD sector, and its proposed extensions.

12.2 High-p_{T} jets

Daniel Britzger, Klaus Rabbertz, and Markus Wobisch

12.2.1 Introduction

One of the most fundamental testing grounds for the predictions of perturbative QCD (pQCD) are studies of the production rates of collimated sprays of hadrons, so-called *hadronic jets*. Even though such jets are neither fundamental entities of the theory nor single particles measured in experiment, the notion of jets proved to be an extremely useful concept, because it allows to make the connection between the objects of pQCD, the quarks and gluons or, generically, *partons*, and the tracks and energy depositions in a detector. In a measured collision event, high-energetic jets can roughly be identified by eye for example when looking at an event display in the radial or the transverse plane. However, for an unambiguous attribution of each track and energy deposit to a jet, a mathematical prescription is required: a *jet algorithm*. Equally, to relate experimental measurements of such jets to production rates predictable in perturbative QCD, a precise definition of *partonic jets* is needed. To close the gap, for good jet algorithms it must also be demonstrated that corrections are under control that on the one hand unfold for detector effects to the level of stable hadrons as in Monte Carlo event generators, and on the other hand account for the non-perturbative transition from partons to the same stable-hadron level. History has shown that jet algorithms can be found that are suitable simultaneously for all three levels, measured tracks and energy clusters, the partons of perturbative calculations, and the hadrons of Monte Carlo event generators used in detector simulations. Alas, it took time approximately halfway through "the first 50 years of QCD" to evolve from first ideas to mature jet definitions used in today's precision phenomenology. In the following sections, the authors describe the essential steps of this evolution from their perspective of working at the LEP, HERA, Tevatron, and LHC colliders.

12.2.2 A hint of color: quark- and gluon-initiated jets

Establishing QCD as the theory of the strong interaction requires us not only to investigate the pattern of *colorless* hadronic particles and their properties, but to go beyond confinement and search for signs of the underlying dynamics of this asymptotically free quantum field theory. In other words, we need to find hints of color even though the confining property of QCD does not allow us to directly measure colored quarks – let alone gluons. Indirect evidence came in 1968 from the observation of Bjorken scaling in Deep-Inelastic Scattering (DIS) at SLAC [110,167], where inelastic scattering of electrons on nucleons at large momentum-transfer squared, Q^2, is well described by the assumption of a virtual photon interacting with point-like constituents inside a nucleon. These constituents, named *partons* by Feynman, were later identified with the (valence) quarks of Gell-Mann and Zweig [18,3171].

It is conjectured that the struck parton should manifest itself in the form of a collimated stream of hadrons moving along the direction of the primary parton with only a few hundred MeV of transverse momentum, like defined as *jet* in the introduction. This brings us to the second question implicit in this section's title "high-p_{T} jets": How high is "high"? The center-of-mass energies of a few GeV available at the time were insufficient to clearly observe well separated jets simply because the opening angles of the hadron streams were far too large and the "jets" interleaved with each other even though the back-to-back orientation of the primary $q\bar{q}$ pair should guarantee their maximal separation. A way out was found by focusing on the main interest to differentiate between a two-jet like structure favored by QCD and the expectations from other models. Instead of reconstructing jets or jet quantities explicitly, the strategy rather consists in searching for a principal event axis along which most of the momentum of each produced hadron is aligned. In 1975, the SLAC-LBL Mark I experiment at the e^+e^- storage ring SPEAR used *sphericity* [3728,3729], which defines such an event axis by minimizing the sum of squares of all momenta with respect to this axis. The *event shape* sphericity, S, is then defined as

$$S = 3 \sum_i \left(p_{\mathrm{T},i}^2 \right) \, / \, 2 \sum_i |\vec{p}_i|^2 \,, \tag{12.2}$$

where the sum is over all particles i in the event with 3-momenta $\vec{p}_i$ and transverse momenta $p_{\mathrm{T},i}$ with respect to the sphericity axis. Each event is characterized by one number S ranging from zero, when all particles are fully aligned along the axis, up to unity for isotropic events. By means of defining such an event axis for their measurements at 3.0, 3.8, 4.8, 6.2, and 7.4 GeV center-of-mass energy, the Mark I experiment found first evidence for quark-initiated jet production emerging when going to the higher center-of-mass energies [122]. Moreover, profiting from transversely polarized beams at 7.4 GeV center-of-mass energy, by comparing the angular distribution of the sphericity axis of $q\bar{q}$ production to the one of $e^+e^- \to \mu^+\mu^-$ they concluded that the potential partons must have spin 1/2 rather than spin 0.

How about gluons then, the exchange quanta of QCD? Do they exist and, if yes, how do they manifest themselves? In 1976 Ellis, Gaillard, and Ross [121] argued gluon bremsstrahlung $e^+e^- \to q\bar{q}g$ to be the leading correction to $q\bar{q}$ dijet production. As a consequence, with increasing center-of-mass energy one of the two quark-initiated jets should exhibit signs of widening up with higher multiplicity until finally a third gluon-initiated jet emerges leading to

planar 3-jet events. The center-of-mass energies available at SPEAR and also DORIS at DESY, however, were not sufficient to provide evidence for 3-jet production, although valuable results could be achieved by investigating the conjectured dominant decay of the upsilon resonance into three gluons, $\Upsilon \to ggg$, confirming predictions by QCD including the vector character of the gluons [3730]. Only the much higher center-of-mass energy of 27 GeV reached by the PETRA collider at DESY in spring 1979 could provide sufficiently high-energetic e^+e^- collisions such that clearly identifiable 3-jet events could be produced. The first event display of the TASSO Collaboration was presented by Wiik at the "Neutrino 79" conference in Bergen [108] and, of course, is also reproduced in this commemorative work, see the section by S.L. Wu for a more personal recollection of events. Subsequently, all four experiments at PETRA published clear evidence for planar 3-jet events affirming the discovery of the gluon and gluon-induced jets [109,126–128].

The increasing e^+e^- center-of-mass energies at PETRA, TRISTAN, SLC, and LEP up to $\sqrt{s} = 209$ GeV allowed a plethora of (multi-)jet measurements to be performed, all confirming the conjectures of QCD as theory of the strong interaction. Notably, the rate of events with three jets as compared to dijet production is to first order proportional to the strong coupling, which then can be extracted at each energy point to demonstrate its energy dependence or *running* as predicted by QCD.

Finally, angular correlations in 4-jet events are sensitive already at leading order (LO) to the color factors $C_A = 3$ and $C_F = 4/3$ of the non-Abelian special unitary group SU(3) of QCD and thus are probing its non-Abelian nature as described in the previous section. A compilation of constraints on these color factors is presented in Ref. [3685], where world average values are quoted that are in perfect agreement with the expectations from QCD.

12.2.3 Jets at hadron–hadron colliders

Despite great new insights obtained thanks to high-precision measurements at e^+e^- colliders, the term of *discovery machines* generally is reserved for hadron–hadron colliders. Because of the much larger mass of protons as compared to electrons, the huge loss of energy per turn in circular storage rings due to synchrotron radiation can be avoided enabling much higher collision energies of e.g. $p\bar{p}$ accelerators than possible with circular e^+e^- beams. The benchmark observable of jet physics at hadron–hadron colliders is the inclusive jet production cross section and in the early days the phase space was divided up into intervals of the jet transverse energy E_T and the jet pseudorapidity η defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$. Measured jet yields are transformed into a double-differential cross section via

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}E_\mathrm{T}\,\mathrm{d}\eta} = \frac{1}{\epsilon \cdot \mathcal{L}_\mathrm{int}} \cdot \frac{N_\mathrm{jets}}{\Delta E_\mathrm{T} \Delta\eta}, \tag{12.3}$$

where N_jets is the number of jets counted within a bin, corrected for detector distortions, ϵ is the experimental efficiency, and ΔE_T and $\Delta\eta$ are the respective bin widths.

The first such measurement of inclusive jet production was published in 1982 by the UA2 Collaboration with data recorded in the so-called *jet run* at the Sp$\bar{\mathrm{p}}$S collider operating at 540 GeV center-of-mass energy [3731]. The observed steep decrease of the jet E_T spectrum proportional to E_T^{-n} with $n \approx 9$ was correctly predicted by QCD at LO [3732]. Firm conclusions on the absolute normalization, however, were not possible because of large experimental and theoretical uncertainties, and lack of a well-defined jet algorithm. The UA2 Collaboration employed a cell-based clustering of energy deposits in the calorimeters, where neighboring cells could be merged into one cluster. A "final" cluster could be split up again, if it contained multiple, well separated maxima. Instead of referring directly to cell geometry, the UA1 experiment used an algorithm based on cones of radius R equal to unity in (η, ϕ) space in order to decide whether cells are merged or not [3733]. Here, ϕ is the azimuthal angle. To initiate a jet, cells exceeding a minimal transverse energy are taken in decreasing order of E_T as "seeds", around which cells within the defined cone are combined with this seed to form the jet. This algorithm corresponds already to a cone jet algorithm; alas, it suffers from a number of shortcomings like unclustered energy or sensitivity to collinear splittings further described in the next section. Nevertheless, at the level of the limited experimental precision and with only order-of-magnitude predictions at LO, jet measurements conducted at the Sp$\bar{\mathrm{p}}$S and at the Intersecting Storage Rings ISR [3734] were in agreement with expectations from QCD.

12.2.4 The evolution of jet algorithms

Until the end of the 1980s, a vast amount of jet data from hadron colliders were collected, reaching a level of precision of 10 %. Predictions at LO in pQCD, however, were very limited in precision by the uncompensated dependence on the *renormalization scale*, μ_r, through the running strong coupling. The calculation of next-to-leading-order (NLO) corrections to jet production advanced the accuracy of perturbative predictions to a comparable level. This progress required a careful re-evaluation of the concept of jets and resulted into new classes of jet algorithms, since several shortcomings of previous jet definitions were identified, which limited their usability in higher-order pQCD predictions or in hadron-induced processes. Let us have a closer look into the evolution of jet algorithms over time.

The first jet algorithm was described in 1977 by Sterman and Weinberg for e^+e^- collisions [185]. In their algorithm,

particles with momenta pointing towards the same direction within some opening angle were clustered together. Most importantly, their jet definition made the result insensitive to the emission of either soft or collinear particles. This is called *infrared and collinear safety*, which is crucial to produce finite results at all orders in perturbation theory. Otherwise the cancellation of soft and collinear singularities associated with such partonic emissions in calculations of pQCD is spoiled leading to infinite results. To be useful in comparisons to pQCD, the outcome of a jet algorithm therefore must neither depend on the addition of arbitrarily soft clustering objects to the set of inputs, nor on the merging of two collinear input objects or the splitting of an input object into two collinear ones.

The following decade saw the proposal by Sterman and Weinberg to be generalized in order to analyze hadron–hadron collisions in terms of a number of cone-shaped jets of a chosen jet radius, R, pointing into the directions of the highest energy or momentum densities in an event. In the same period the JADE Collaboration at the PETRA collider introduced another type of jet algorithm based on iterative pairwise clusterings for the analysis of e^+e^- events [3688]. Hence, two classes of jet algorithms emerged:

1. cone algorithms that assign objects to the leading energy-flow objects in an event based on geometrical criteria;
2. sequential-recombination algorithms that iteratively combine the closest pairs of objects.

A summary of jet algorithms discussed at the time is presented in the proceedings of the *Snowmass* "Summer Study on High Energy Physics" [3735].

Although introduced only in 2008 in its general form, one can determine the so-called *catchment area* of a jet, often just named *jet area*, for both classes provided the algorithm is infrared- and collinear-safe [3650]. For cone algorithms defined in (η, ϕ) space as used already by the UA1 Collaboration, this jet area formerly was identified with the circular area with jet radius R, which simplified considerably the task of jet energy calibration at hadron–hadron colliders.

In e^+e^- collisions all final-state particles emerge from the hard subprocess. Therefore, in e^+e^- measurements exclusive jet algorithms were applied, which assign each final-state particle to one of the high-p_{T} jets. Hence, a collision event is classified as an exclusive jet final state, e.g. $e^+e^- \to n$ jets and nothing else.

Although being more costly in terms of computing time, it was affordable to use successive recombination algorithms because of the low multiplicity in e^+e^- annihilations. Initially, the JADE algorithm was favored, where pairs of particles are clustered in the order of increasing invariant di-particle masses, assuming this would result in jets with small invariant masses. In the phenomenology of e^+e^- physics, it

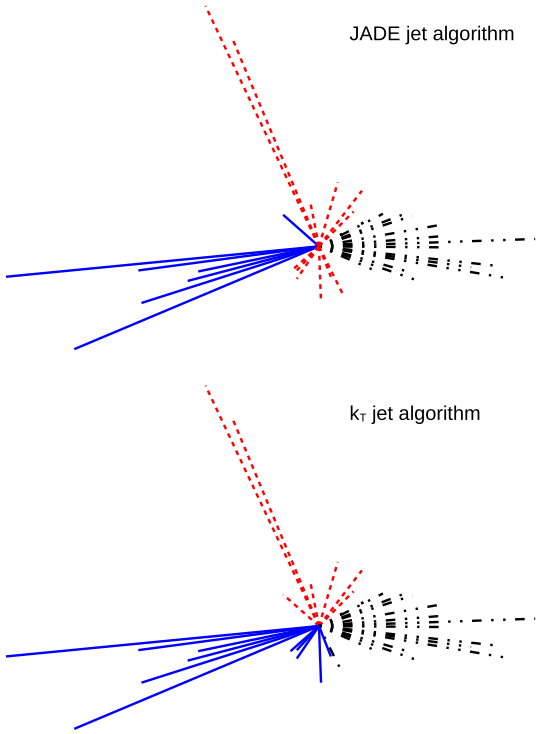


Fig. 304 A 3-jet final state in e^+e^- collisions as defined by the JADE (upper) and k_{t} (lower) jet algorithms. The particle assignments to the three jets according to the algorithms are indicated by blue full, black dash-dotted, and red dashed lines. Figure redrawn from Ref. [3737]

was, however, discovered that the JADE algorithm frequently clusters soft particles at large angles, cf. also Fig. 304, which is very disadvantageous for precision calculations [3736]. This problem was addressed in the k_{t} or "Durham algorithm" [187], [112] in which the distance measure was changed from the invariant di-particle mass to the relative transverse momentum, k_{t}, of the particle pair. This version, also called the (exclusive) k_{t} algorithm, was confirmed to have superior properties than the JADE algorithm in e^+e^- annihilation.

When HERA, the first and only electron–proton (ep) collider, started in 1992, "standard" jet algorithms had been defined already for e^+e^- annihilation as discussed. In hadron–hadron collisions cone-type algorithms were favored over sequential-recombination algorithms to avoid time-consuming repeated iterations over many final-state particles. Nothing yet had been developed for physics at an ep collider such that many physicists coming from LEP experiments tried to adopt methods as they were used in e^+e^- physics. So in the early HERA jet analyses, a modified version of the JADE algorithm was used (the "mJADE algorithm" [3738]), in which the proton beam remnant is treated by introducing a pseudo-particle (carrying the missing longitudinal momen-

[112] Originally, $k_\perp$ was used as label instead of k_{t}. For simplicity we use k_{t} throughout.

tum in the event), to which particles can be clustered. At the end, all particles are either assigned to the high-p_T jets, or to the jet including the pseudo-particle. The former are considered as the n high-p_T jets, while the latter is considered to be the (one) beam remnant. The final states are therefore classified as exclusive $(n+1)$-jet final states.

In reactions with initial state hadrons, i.e.*ep* and hadron–hadron collisions, collinear singularities in the matrix elements of the hard subprocess are factorized into process-independent parton distribution functions (PDFs), which depend on the *factorization scale*, μ_f, that defines the limit between attribution to the perturbative hard process or the non-perturbative hadron structure in form of the PDFs. This factorization, however, only works, if it is not spoiled by the definition of the measured quantity that must not depend on the beam-remnant(s). For the mJADE algorithm, it was the inclusion of the kinematics of the beam-remnant that made the algorithm non-factorizable. This issue was fixed in the exclusive k_t algorithm for *ep* and hadron–hadron collisions by treating the beam remnant(s) as particles of infinite momentum and thus independent of their actual kinematics. This exclusive k_t algorithm was in use for some time within the HERA experiments and later was replaced by its inclusive counterpart.

Hadron–hadron and *ep* collisions share the common feature of having activity in their final states related to the remnant(s) of the beam hadrons. Therefore, the jet definitions used in hadron–hadron physics were based on the cone-type proposal by Sterman and Weinberg to define a jet by the transverse energies through a cone, which is moved so as to maximize the transverse energy flow through it. In this approach, only selected final-state particles are included in jets. Those, which are not assigned to jets are effectively interpreted to stem from the so-called *Underlying Event* that is related to soft processes involving interactions with the beam remnants. The jet final-states are thus classified as *inclusive* with respect to additional unclustered particles, e.g. $pp \to n$ jets plus additional activity, which could consist of additional jets and/or unclustered particles.

Another difference between e^+e^- and hadron–hadron physics consists in the choice of variables. In hadron–hadron collisions, the center-of-mass frame of the hard subprocess is boosted longitudinally, i.e.along the beam direction with respect to the detector rest frame. Hence, instead of energies and angles as used in e^+e^- collisions, transverse momenta and/or transverse energies are used, together with azimuthal angles and either the pseudorapidity η as defined before, or the rapidity $y = 1/2 \cdot \ln\left[(E+p_z)/(E-p_z)\right]$, which coincides with η for massless objects. As a consequence, cone-jet algorithms in hadron–hadron collisions are used with cone radii R defined in the plane of azimuthal angle and (pseudo)rapidity.

Cone algorithms are, however, not as easy to implement as one would naïvely think. The basic idea of a cone-jet algorithm sounds rather simple: Decide on a cone radius, R, place it in the plane of azimuthal angle and (pseudo)rapidity, compute the transverse energy/momentum flow through the cone, and move the cone over the plane so as to maximize this flow. Before the end of the 1990s, experimental jet measurements used a large number of different implementations. These early cone algorithms suffered from a number of problems. Many were not infrared or collinear safe, while others had undesired features. Some of the problems arise from the fact that a true, continuous maximization procedure of the energy flow through the cone required too much computing resources, and short-cuts were applied. Some versions simply defined the final jets by building cones around the particles/detector clusters of highest energy. Other versions used these clusters as starting points, or "seeds" for an iterative procedure. All of these algorithms were either not infrared- or not collinear-safe, or even both. Other undesired features emerged through the treatment of overlapping cones. Sometimes, it happens that two resulting jet cones share a number of particles. To have a unique assignment of particles to jets, an *overlap treatment* is added to the algorithm, which assigns the particles in the overlap regions uniquely to one of the two jets. This overlap treatment depends on additional parameters (adding to the complexity of the algorithm) and in most cases it also introduced additional violations of infrared or collinear safety. These problems were ultimately addressed and solved with the Seedless Infrared-Safe Cone (SISCone) jet algorithm [3739]. By eliminating seeds, and using a refined overlap treatment, SISCone became the first and so far only cone jet algorithm that is infrared- and collinear safe.

The SISCone algorithm was, however, never widely used since the rather late time it was introduced. Jet measurements had moved on to different jet algorithms. Soon after the introduction of the exclusive k_t algorithm for e^+e^- physics and the above-mentioned modifications for processes with initial-state hadrons, a similar inclusive algorithm was introduced: the "Cambridge algorithm" [189]. This algorithm transferred the basic concepts of the exclusive k_t algorithm consistently to hadron–hadron collider physics. In the same way that the Cambridge algorithm was a modification of the exclusive k_t algorithm, a corresponding modification of the inclusive k_t algorithm was introduced, called the "Aachen algorithm" or, later, the "Cambridge–Aachen algorithm" [190]. This algorithm recombines pairs of particles simply in the order of increasing distances in (y,ϕ) space. Both algorithms can be specified in a unified way by defining the pairwise distance d_{ij} between any two objects i and j, and the *beam distance* d_{iB} of each object i as:

$$d_{ij} = \min\left(p_{\mathrm{T},i}^{2p}, p_{\mathrm{T},j}^{2p}\right)\frac{\Delta R_{ij}^2}{R^2}, \tag{12.4}$$

$$d_{iB} = p_{\mathrm{T},i}^{2p}. \tag{12.5}$$

Here, the power p is the algorithm-defining parameter, and ΔR_{ij} is the purely "angular" distance in (y, ϕ) space between i and j:

$$\left(\Delta R_{ij}\right)^2 = \left(y_i - y_j\right)^2 + \left(\phi_i - \phi_j\right)^2. \tag{12.6}$$

Then, each time the minimal distance of all pairwise and beam distances is a d_{iB}, object i is declared a final jet and removed from the list of clustering objects. If the minimal distance is a d_{ij} instead, the two objects are merged using four-vector addition into a new object that is added to the clustering list. This is repeated until no more input objects are left.

Setting p equal to unity gives the k_t algorithm, while $p = 0$ corresponds to the Cambridge–Aachen one that only considers ΔR_{ij} in the clustering and is frequently used for studies of jet substructure. Interestingly, as discovered in Ref. [193], the choice of $p = -1$ is also a valid option, where in contrast to the k_t algorithm the clustering starts with the highest-p_T objects and hence this third "family member" was dubbed the "anti-k_t algorithm", which leads at least for the leading p_T jets to round-shaped jet areas as if from a cone jet algorithm! In comparison to the more fractal-like jet areas of the k_t algorithm the cone-like anti-k_t jets were much easier to calibrate reusing recipes for previous cone jet algorithms. In particular the subtraction of additional energy within a jet cone from further proton–proton collisions in the same or neighboring bunch crossings (pile-up) was much facilitated. As a consequence the anti-k_t algorithm was quickly adopted as the main jet algorithm for jet physics at the LHC.

12.2.5 New physics with jets: excesses in jet cross sections

The next stage of establishing QCD as the theory of the strong interaction was triggered by two developments: the arrival of predictions at NLO in pQCD also for hadron–hadron collisions, and the start of the Tevatron collider at Fermilab with a $p\bar{p}$ center-of-mass energy ranging from 540 GeV up to 1.96 TeV. The by far dominating theoretical uncertainty caused by the large μ_r scale dependence of LO predictions was reduced from factors of roughly two to 10–30% [3740,3741]. Additional uncertainties from non-perturbative effects and from the proton structure were estimated to lie between 5 and 20%, respectively. The latter uncertainty was derived from calculations using different extractions of the proton PDFs from data of deep-inelastic scattering of leptons on fixed targets [3742–3745]. First comparisons of these NLO predictions to $p\bar{p}$ collider data from UA2 and from the new CDF experiment at Tevatron exhibited a very nice agreement.

This picture changed suddenly in 1996 when the CDF Collaboration reported an excess in inclusive jet data at high E_T beyond 200 GeV as shown in Fig. 305 [3746]. A possible explanation could be new phenomena at an energy scale Λ far beyond reach to allow e.g. resonant production of new particles. Similarly to Fermi's low-energy four-fermion coupling to approximate weak interactions at scales well below the W boson mass, such an excess can be described in terms of *contact interactions (CI)* [3742,3747]. Speculations about such contact interactions as a possible explanation were, however, quickly dismissed and the results were scrutinized for effects not properly covered by uncertainties. With respect to the proton structure there was no other means than taking the spread in predictions using different proton PDFs, also shown in Fig. 305, as a proxy for the uncertainty, which now had become very relevant. As all the PDFs known at the time were potentially prone to the same biases, the association of the spread in the corresponding predictions with a PDF uncertainty could only be considered an educated guess or, in the words of Soper [3748]: "This is similar to estimating the size of a French mountain valley by taking the r.m.s. dispersion in the locations of individuals in a flock of sheep grazing in the valley."

The way forward was described in the seminal paper Ref. [3749], where a systematic approach was presented to derive parton distributions with reliable uncertainty estimations. Using the preliminary PDFs including experimental uncertainties derived in Ref. [3750] from DIS data, the authors demonstrated that the excess reported by the CDF Collaboration can be absorbed in updated parameter values for the strong coupling constant and the gluon distribution. While the quark parton distributions are directly determined in DIS, in particular with data from the new HERA collider as used in Ref. [3750], the DIS data are insufficient to also fix $\alpha_S(M_Z)$ and the gluon content in the proton. For both, jet cross sections measured at the Tevatron and at HERA are valuable input to the PDF fits.

12.2.6 The running coupling and the gluon content of the proton

HERA, which was approved in 1984, became operational in 1992, coinciding with the 20th anniversary celebration of QCD in Aachen [3751].[113] At that time, QCD was in a "transition from the stage of early exploratory studies to high precision analyses in QCD" as noted by Zerwas and Kastrup in the introduction to this workshop [78]. A milestone for testing QCD was achieved by demonstrating experimentally the *running* of the strong coupling from the τ mass of around 2 GeV up to the Z boson mass at 91 GeV using

[113] This was the very first conference participation of KR triggering his profound interest in jets and QCD.

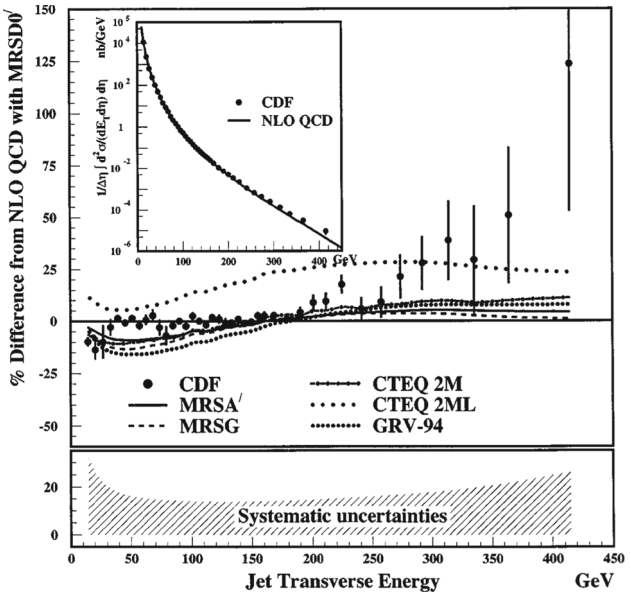


Fig. 305 Percentual difference between the CDF inclusive jet cross section (points) and a prediction at NLO QCD using MRSD0' PDFs. Additional lines show predictions for a selection of alternative PDFs available at the time. The error bars represent uncorrelated uncertainties, while the quadratic sum of the correlated systematic uncertainties are shown in the bottom panel. The inset compares the absolute cross sections. Figure taken from Ref. [3746]

various observables and data from different experiments as reported in Ref. [3752]. A first summary of determinations of $\alpha_S(M_Z)$ was presented by Altarelli [3753] who concluded on $\alpha_S(M_Z) = 0.118 \pm 0.007$.

HERA was constructed as the paramount extension to the series of previous, very successful fixed-target lepton–nucleon scattering experiments, which have led the way to conceiving QCD as the theory of strong interactions. The increase by a factor of ten in the lepton–proton center-of-mass energy promised rich new data for testing many aspects of QCD. In particular, with $\sqrt{s} = 300$ GeV, HERA allowed to elucidate the structure of the proton and the running of $\alpha_S(\mu_r)$ by means of unique and detailed measurements of the hadronic final state in addition to the scattered lepton.

The HERA collider at DESY consisted of two independent accelerators designed to collide 30 GeV electron and 820 GeV proton beams. Two multi-purpose detectors, H1 and ZEUS, were conceived to precisely measure the hadronic final state with almost hermetic coverage. The main difference between the two experiments with respect to jets is given by their calorimeters. The H1 collaboration opted for a liquid argon calorimeter with electromagnetic and hadronic sections, both inside the solenoïd providing the magnetic field [3754,3755]. The ZEUS collaboration optimized their calorimetric system for hadronic measurements and employed a compensating uranium plastic-scintillator sampling calorimeter [3756]. The overconstrained kinematics of neutral-current DIS events enabled precise in situ calibrations for the electromagnetic and hadronic energy scales such that both collaborations could report a jet energy scale uncertainty of only 1 % for jets with transverse momenta exceeding 10 GeV in the laboratory rest frame [3757,3758].

Already the first HERA data brought striking QCD results, like the confirmation of the logarithmic violation of Bjorken scaling shown by the F_2 structure function in dependence of the parton fractional momentum x as predicted by QCD [3759,3760], or support for the presence of a hadronic structure of quasi-real photons as a result of dijet events observed in photoproduction [3761,3762]. Hence, jets were an integral part of the HERA physics program from the very beginning. The term *jet physics* quickly extended well beyond the simple picture of one "DIS jet", which is initiated by the struck quark in the Quark–Parton-Model (QPM) picture, or of dijet topologies in photoproduction. Studies of further properties like jet charge, substructure, fragmentation, or the heavy flavor content of jets led to many more interesting results, which, however, cannot be covered here. In the following we will limit ourselves to high-p_T jets in neutral-current DIS and will refer the interested reader to other sections in this book or to review articles [3763–3767].

At HERA, for the first time, it became possible to study large numbers of dijet events in neutral-current DIS, so-called (2+1) jet events. In pQCD the cross section for hard processes in DIS is given up to order n in the perturbative expansion in α_S through the factorization theorem

$$\sigma = \sum_{n}^{\text{order}} \sum_{i}^{q,\bar{q},g} \left(\frac{\alpha_S(\mu_r)}{2\pi}\right)^{k+n} \int \mathrm{d}x f_{a/i}(x,\mu_f) \mathrm{d}\hat{\sigma}_i^{[n]}(x,\mu_r), \tag{12.7}$$

where i denotes the parton flavors in the proton PDF f_a, and k corresponds to the power in α_S at leading order. The universal proton PDFs are convoluted in x with the hard coefficients at a selected factorization scale μ_f. At LO, pQCD predicts the (2+1) jet events to be produced proportional to α_S ($k = 1$). At HERA, this process is mainly initiated by a gluon inside the proton and thus dijet data provide direct access to the gluon content of the proton down to $x \sim 10^{-3}$. A second LO contribution arises from gluon radiation off one of the quark-lines in the QPM diagram and becomes dominant at large x.

The first measurement of (2+1) jet rates by the H1 Collaboration [3768] employed the JADE jet algorithm [3688], while the ZEUS Collaboration [3769] opted for a cone jet algorithm following the *Snowmass convention* [3645]. The hadronization corrections were found to be reasonably small and the measured jet profiles could be directly related to the underlying hard process and the gluonic content of the proton. These early data strongly supported the QCD picture of jet-production in DIS and the data were found to be well

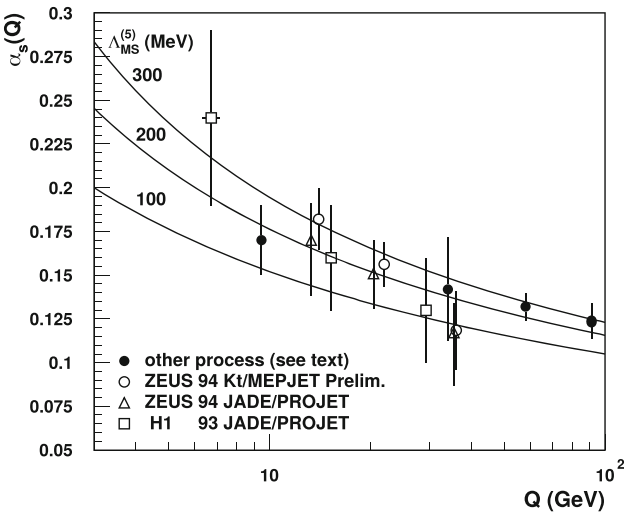


Fig. 306 Dependence on the energy scale Q of the strong coupling, $\alpha_S(\mu_r = Q)$, from early HERA data in comparison to other processes, see text for details. The predictions of QCD for three values of the α_S-equivalent $\Lambda_{\overline{\text{MS}}}$ parameter are superimposed as lines. Figure taken from Ref. [3770]

described by first order QCD calculations supplemented with leading-logarithmic parton showers as an approximation of higher-order QCD corrections. Already at this stage a running coupling was significantly favored over a constant value of α_S.

The inclusion of NLO QCD corrections in dijet calculations [3771] and an improved understanding of hadronization corrections of jet data together with refined and enlarged data sets, allowed for the first time the study of the running of the strong coupling constant in a single process using (2+1) jet rates based on the JADE algorithm [3772,3773]. A summary of these results from H1 and ZEUS in comparison to measurements in e^+e^- collisions is displayed in Fig. 306. The additional points are determined from Υ decays (Γ_Υ), the ratio R of hadronic over total cross section ($\sigma_{\text{had}}/\sigma_{\text{tot}}$), event shapes, and the ratio of hadronic over leptonic decay width of the Z boson ($\Gamma_{\text{hadron}}/\Gamma_{\text{lepton}}$) as described in Ref. [3774]. An insight gained from these data and from subsequent studies with improved NLO calculations [3775] was that cone or k_t jet algorithms seem to be preferred over the JADE algorithm for precision QCD analyses due to their improved perturbative stability in hadron-induced processes [191,3776,3777], as already outlined in the previous section. In addition, it became apparent that the choice of suitable renormalization and factorization scales is crucial to achieve reliable results for multi-scale processes such as jet production in DIS.

Despite these first successes it became rapidly clear that for jet measurements in the laboratory rest frame theoretical shortcomings prevent optimal comparisons to theory. Firstly, it is highly desirable for the jet observables to respect factorization, and secondly it is highly non-trivial to separate the hadronic final state from the beam remnant. A way forward is found by boosting every event to the Breit frame of reference [3777] using the reconstructed DIS kinematics. In the Breit frame the incoming parton collides head-on with the exchanged electroweak boson along the z axis of this reference frame. Any significant transverse momentum is generated from QCD effects. High-p_T jets primarily occur in dijet topologies, for which the LO QCD diagram is of $\mathcal{O}(\alpha_S^1)$, whereas LO DIS or the beam remnant do not contribute. First measurements of jet cross sections in the Breit frame using variants of the longitudinal invariant k_t jet algorithm have been conducted by the H1 and ZEUS collaborations with a distance parameter of $R = 1.0$ [3778–3780]. This choice promises high accuracy of pQCD predictions and small non-perturbative corrections for hadronization effects. From data at high $Q^2 \gtrsim 150\,\text{GeV}^2$, where scale choice ambiguities are reduced, since jet transverse momenta are of a similar size as the virtuality of the exchanged boson $\sqrt{Q^2}$, both collaborations determined $\alpha_S(M_Z)$ with NLO pQCD predictions at a precision of around 4 %. The uncertainty in $\alpha_S(M_Z)$ was comparable to the level of the LEP experiments [3781] and considerably outperformed the ongoing experiments CDF and D0 at the Tevatron. Moreover, the running of α_S could be successfully tested in the scale range from about 7 to 50 GeV. Together with inclusive neutral- and charged-current DIS data, even the first combined determination of the proton PDFs together with $\alpha_S(M_Z)$ was performed from data of a single experiment [3782].

In 1998, the beam energy of the HERA protons was raised to 920 GeV, corresponding to $\sqrt{s} \simeq 320\,\text{GeV}$. The large amount of data recorded from 1998 to 2000, and during the HERA-II running period from 2003 to 2007, led to a multitude of measurements i.a. investigating the dependence of jet cross sections on the type of jet algorithm and the jet size R, or the benefits of normalizing to the DIS cross section. With respect to the strong coupling constant the development culminated in a determination of $\alpha_S(M_Z)$ with only 0.4 % of experimental uncertainty [3783]. Yet, in all these QCD analyses, the NLO scale uncertainties of roughly 5 % in the jet predictions remained the dominant uncertainty and, hence, the limiting factor preventing a higher precision for $\alpha_S(M_Z)$. The next decisive progress, then, should come from theory. After more than 15 years, the next-to-next-to-leading order (NNLO) corrections to jet production in DIS were finally calculated in Refs. [3784,3785], which allowed for a reduction in the scale dependence of the predictions for the interpretation of the HERA jet data. The latest improved HERA-II measurements were then the first to be confronted with the new NNLO cross section predictions, which proved the corrections to be sizeable reaching up to 40 % at low scales. Yet, the NNLO predictions provided a very good description of the data over the entire accessible kinematic range [3783] and a significant improvement as compared to the long-standing

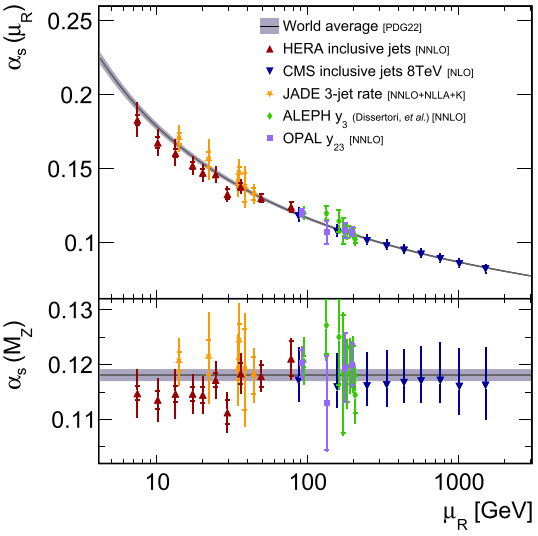


Fig. 307 Tests of the running of the strong coupling from HERA and CMS inclusive jet cross section data using NNLO or NLO pQCD predictions, respectively. The data are compared to the expectation from QCD and measurements of jet-rates in e^+e^-. The lower panel displays the respective value of $\alpha_S(M_Z)$ for the representative value μ_r of the data

NLO predictions. This NNLO revolution for single-jet inclusive predictions was the ultimate step to reduce the theoretical uncertainties to a level comparable to the experimental uncertainties. A full analysis of all inclusive jet data from H1 [3786], and an analysis of data from H1 and ZEUS [3787] demonstrated an excellent agreement between the data and the NNLO pQCD predictions. A comparison of selected inclusive jet cross section data with NNLO predictions is displayed in Fig. 308.

From inclusive jet data the value of $\alpha_S(M_Z)$ was finally determined at NNLO to be

$$\alpha_S(M_Z) = 0.1178 \pm 0.0015\,(\text{exp}) \pm 0.0021\,(\text{theo}) \quad (12.8)$$

with percent level experimental and theoretical uncertainties of similar size. Surprisingly, although jet data were believed to have a significant sensitivity to the gluon PDF, a complete analysis of jet data together with HERA inclusive DIS data at NLO [3059] or NNLO [3100,3786,3788] showed only little impact on the gluon density.

Finally, the inclusive jet data from HERA were able to unfold their full potential to test the running of the strong coupling from a single process using NNLO pQCD predictions [3786,3787]. The results are found to be in excellent agreement with expectations from pQCD and are shown in Fig. 307, where the extracted values of $\alpha_S(\mu_r)$ from these data are compared additionally with the $\alpha_S(\mu_r)$ determinations from inclusive jet data of the CMS experiment [3789] and with analyses using jet-rate measurements in e^+e^- collisions [3692,3790,3791].

Thus, the HERA inclusive jet data improve significantly over measurements from the JADE experiment in a similar region of μ_r, and bridge the gap between low-scale determinations of α_S from τ-decays and the precision measurements at the Z-pole in e^+e^- collisions.

12.2.7 Highest-p_T jets at the LHC

From early exploratory up to the latest results, jet measurements have accumulated numerous successes: the gluon discovery at PETRA, the confirmation of the gauge structure of QCD at LEP, or the running of the strong coupling constant at HERA. So what is in store with the next-to-next hadron–hadron collider, the LHC? After 25 years from first concepts discussed in 1984, cf. Ref. [3792], up to first collisions at the LHC in 2009, and a similar timespan between the availability of NLO calculations for jet production in hadron–hadron collisions in 1989/1990 [3741,3793] and the arrival of NNLO predictions in 2017 [3419] we are now in a much better position for precision comparisons. The dependence of the NNLO predictions on the choice of the renormalization scale is significantly reduced as compared to NLO. The required proton PDFs have much smaller uncertainties and were determined from a lot more and more accurate data in a more systematic way that considers and provides systematic uncertainties. The modern experiments at the LHC deliver more precise data than at any other hadron–hadron collider before and include correlations as well as the full decomposition of systematic uncertainties. Figure 308 provides an overview of data-theory comparisons for the inclusive jet cross section versus jet p_T as measured at the LHC and previous hadron colliders. Overall, the description of the data at various center-of-mass energies and covering many magnitudes in inclusive jet cross section and jet p_T is excellent. Figure 309 summarises such measurements at the LHC in the form of a total inclusive jet cross section within a suitably defined fiducial phase space as a function of $\sqrt{s}$.

Despite the great success of pQCD for the description of jet data, a few concerns in particular on the theory side still persist. The scale dependence is just a proxy to estimate the effect of missing higher orders (MHO) and can be misleading if not combined with other insights into the process of interest like the relative sizes of the higher-order corrections or the absence of new process types at a given perturbative order. A newer approach [3817] makes use of Bayesian models assuming a specific behaviour of the coefficients of the perturbative series to estimate MHO uncertainties with the advantage that a proper description in statistical terms like credibility intervals becomes possible. Newer work in this direction can be found in Refs. [3818–3820], while Ref. [3821] follows a different technique to approximately complete the perturbative series. With respect to PDFs this uncertainty of purely theoretical nature only starts

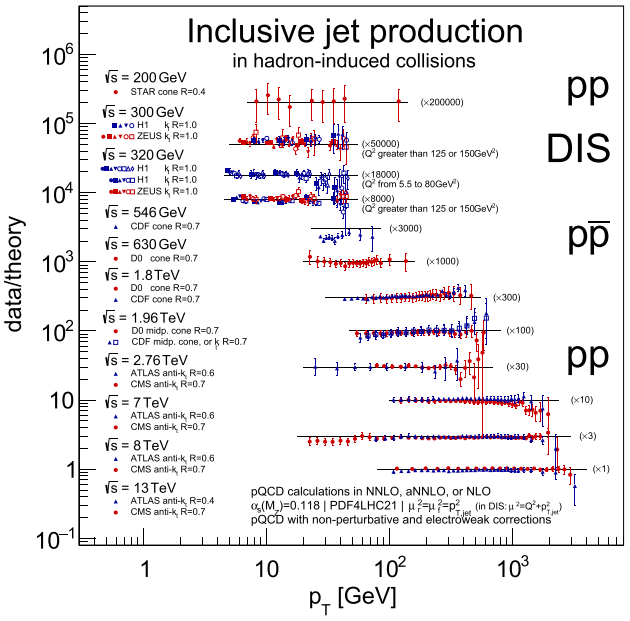


Fig. 308 Ratios of cross-section measurements to predictions in perturbative QCD for inclusive jet production at central (pseudo-)rapidity as a function of the jet p_{T} or E_{T}. The data were taken in pp, $p\bar{p}$, or ep collisions by the ATLAS, CDF, CMS, D0, H1, STAR, and ZEUS experiments, at the RHIC, HERA, Tevatron, and LHC colliders [3758,3778,3780,3783,3789,3794–3809]. From data available for multiple jet algorithms and/or distance parameters one particular choice has been made as indicated. The vertical error bars indicate the total experimental uncertainty of the data. The pQCD predictions are derived using the PDF4LHC21 PDF set [3810] for a value of $\alpha_S(M_Z) = 0.118$ at NNLO in QCD [3419,3520,3784,3785,3787,3811–3813] unless indicated otherwise. The renormalization and factorization scales μ_r and μ_f are identified with p_{T} at hadron–hadron colliders, and $\sqrt{Q^2 + p_{\mathrm{T}}^2}$ in DIS. The predictions for $p\bar{p}$ are only in NLO QCD supplemented with 2-loop threshold corrections (*aNNLO*) [3371,3814–3816], since most of the jet algorithms are IRC-unsafe. For STAR, the predictions are at NLO QCD only. The pQCD predictions are complemented with correction factors for non-perturbative and electroweak effects where applicable

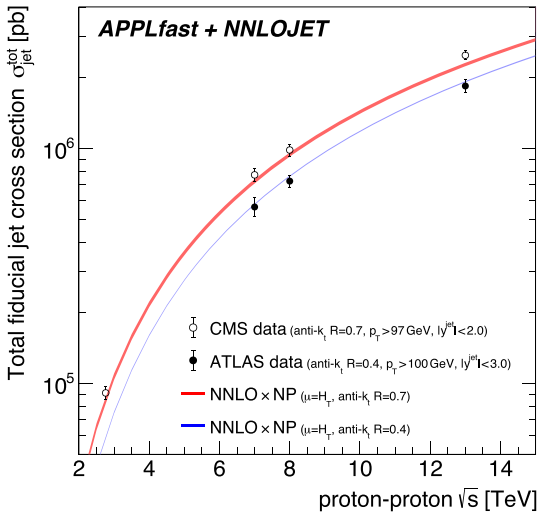


Fig. 309 The total jet cross section as a function of the pp center-of-mass energy for anti-k_t jets with $R = 0.4$ and 0.7. The predictions are compared to data from ATLAS ($R = 0.4$) and CMS ($R = 0.7$). The size of the shaded area indicates the scale uncertainty. Figure taken from Ref. [3520]

being considered in fits and the corresponding uncertainties [3101,3822,3823]. Another point of concern, which limits the precision of phenomenological analyses, is related to the uncertainties of non-perturbative effects, which are important specifically for small transverse momenta. Currently, they are "guesstimated" in a similar manner as PDF uncertainties 25 years ago, i.e. essentially the predictions by a number of MC event generators and their model parameter tunes are compared without systematic account of potential biases or correlations.

With the data from the LHC, it became possible for the first time to probe QCD and the running of the strong coupling from 100 GeV up to the TeV scale as shown in Fig. 307 using CMS inclusive jet data at $\sqrt{s} = 8$ TeV from Ref. [3789]. Notably, beyond 1 TeV of jet p_{T}, electroweak effects become significant and must be considered. Also, in a search for new phenomena with the so-called dijet angular distribution $\chi = \exp(|y_1 - y_2|)$ it was found that small deviations at low χ for dijet masses beyond 2 TeV could be accommodated by electroweak corrections [3824]. Otherwise such deviations from a mostly flat behaviour that is expected from Rutherford-like parton–parton scattering could again be an indication for contact interactions as an expression of new phenomena at a scale Λ. Similarly, excesses at large jet p_{T} like the one by CDF discussed in Sect. 12.2.5 have to be considered carefully to avoid premature conclusions on new phenomena, or, much worse, fitting away first signs of new physics by absorbing them into PDFs! Again Ref. [3749] provides advice: "Note that once data is used in the PDF fit, it cannot be used for other purposes: specifically, setting limits on possible physics beyond the standard model. In that case, one should fit the PDFs and the new physics simultaneously." In the latest publication on inclusive jet production at $\sqrt{s} = 13$ TeV [3809] the CMS Collaboration performed such a fit in the framework of the effective field theory-improved standard model (SMEFT), where a perturbative coefficient c_1 representing potential contact interactions was used as a free fit parameter. It was found that the data are well described by the standard model alone and the c_1 coefficient was compatible with zero. A modification of the gluon PDF as before was not required as shown in Fig. 310. Once, it has been assured that new LHC jet data are consistent with the standard model, they can be used in combination with HERA data to simultaneously extract PDFs and the strong coupling constant at NNLO to

$$\alpha_S(M_Z) = 0.1166 \pm 0.0014\,(\mathrm{exp}) \pm 0.0009\,(\mathrm{theo}). \quad (12.9)$$

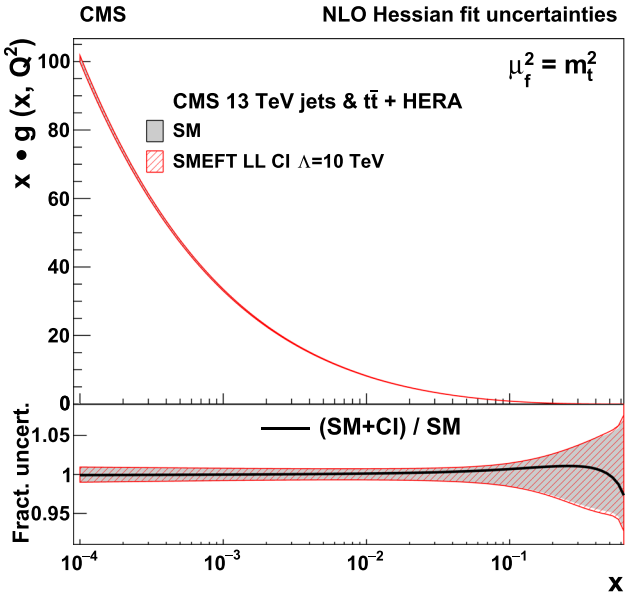


Fig. 310 The gluon distribution as a function of the proton fractional momentum x resulting from fits with and without contact interaction terms to HERA DIS data combined with data from $t\bar{t}$ and inclusive jet production as measured by CMS. The SMEFT fit here is performed with the left-handed CI model with $\Lambda = 10$ TeV. The gluon distribution is shown at a factorization scale μ_f chosen to be the top quark mass ($\mu_f^2 = m_t{}^2$). Figure taken from Ref. [3809]

Also data from multiple reactions can be combined in PDF determinations as recently demonstrated by the ATLAS experiment [3825]. Yet, the best results of the LHC run 2 are still to come, since the data recorded from 2015–2018 are still in preparation by the collaborations for final calibration and publication.

12.2.8 Final words

The presented article tries to recount the story of jet measurements and their relevance for QCD. Specifically, we addressed what has been learned in the course of time from the interplay between theory and measurement at the highest jet p_T available at each moment in time. We have selected a few key measurements for this purpose from a plethora of results achieved at the various colliders. For a more complete overview other sources may be consulted [3826,3827,3827].

For the future, of course, we expect to see more precise jet measurements at even higher jet p_T with corresponding studies of their impact on searches for new phenomena, the running of the strong coupling, or the proton structure. Before concluding, we would like to point out explicitly three developments that might change how future analyses will be performed.

First, not only gluons can be radiated in large numbers by a (color) charge, but also photons by electric charges. So whenever comparing electrons in the final state to predictions including radiative corrections, one has to account for the effect that calorimeter cells add up the energies of e.g. an electron and all surrounding photons hitting the same cell. To avoid a potential mismatch between what experimentally is considered an electron and what is calculated in theory, one can define a cone around the electron and include all photon-like objects into the definition of the electron. This is then called a *dressed* electron or, more generally, a *dressed* lepton, since the same concept can be applied to muons, although the latter radiate less and are measured predominantly in tracking detectors. Essentially, this is again a kind of jet algorithm, but applied to leptons as primary particles [3828], raising the question "What is not a jet?".

Secondly, enormous technical progress not only allows us to produce jets at unprecedented transverse momenta of several TeV instead of GeV, we can also measure with much better precision such high-p_T jets of order hundred or more tracks and clusters. This is especially important, since high-p_T jets may not only be categorized into quark- or gluon-initiated jets, but also into *boosted jets* meaning that such jets may additionally contain the whole decay chain of massive boosted objects from either standard model W and Z bosons, and top quarks up to new hypothetical particles. A whole new field of QCD-focused analyses has been opened up here looking in detail into the substructure of jets asking the question "What is in a jet?".

Finally, progress in computing technology enabled large-scale application of neural network techniques and machine learning methods to jet physics and jet substructure. For order hundred and more jet components with kinematic properties and other characteristics, deep learning techniques allow us to study all available information in its high dimensionality. This development has considerably increased the discrimination power among different jet types, and has the potential to genuinely improve our understanding of perturbative QCD, cf. for example the review in Ref. [1829].

In summary, even after 50 years of QCD, we still have exciting new developments in front of us.

12.3 Vector boson + jet production

Monica Dunford

Measurements of single vector boson production in association with jets (V+jets production) play a central role in particle physics as they are sensitive probes to several different aspects of the Standard Model. With these measurements the predictions of perturbative QCD can be tested in new regions of phase space and with small statistical and systematic uncertainties. In many places, the experimental accuracy is better or comparable to that of the theoretical predictions. The studies of W and Z boson production with additional jets are sensitive tests of the dynamics of higher-order diagrams in the QCD predictions, of models of heavy-flavor produc-

tion and of parton density functions (PDFs). These measurements are used to test the accuracy of the wide range of theoretical models available. This is especially important since W and Z boson productions are dominant backgrounds to measurements and to search for New Physics. Accurate simulations are necessary for everything from the calibration of the detector to modeling of the signal process of interest. Measurements of jets in V+jets production is one of the main processes used for simulation, defining event parameters (tuning), and the validation of the theoretical model. Excellent knowledge of QCD-related variables is also critical for precision measurements at hadron collider, such as measurements of the W boson mass, which rely upon accurate modeling of the W boson p_T spectra.

12.3.1 Results from S$\bar{p}$pS and the Tevatron

The W and Z vector bosons were both discovered in 1983 by the UA1 and UA2 experiments at the Super Proton Synchrotron (S$\bar{p}$pS) at CERN. By today's standards, the number of vector boson events collected was miniscule; the UA1 detector for example collected 240 $W \to e\nu$ events and 57 $W \to \mu\nu$ events at a center-of-mass energy of 0.630 TeV [3829]. The data from these detectors permitted first tests of QCD in vector boson production. One of the immediate conclusions drawn from the data was that higher-order QCD corrections such as gluon radiation from an initial-state quark or anti-quark are needed to explain events where the vector boson has large a momentum in the transverse plane (p_T).

Since the dominant production of V+jets at the S$\bar{p}$pS collider is due to gluon radiation from an initial-state quark or anti-quark, these events are an ideal sample of gluon-initiated jets. Using W+1-jet events, measurements of the angular distribution of the jet are consistent with the expected bremsstrahlung-like radiation [3829]. In addition, the spin of the gluon was measured via the polarization of the W boson. When a scalar gluon is radiated by an incoming quark or anti-quark, the helicity of the quark will be changed since the axial coupling is not conserved. In contrast, in the case of a vector gluon which conserves helicity, the quark's helicity will be preserved. The two cases lead to different polarizations of the W boson. Although the gluon spin was determined at PETRA [3830] and using two-jet events at UA1 [3831], this test was an important confirmation that the gluon has a spin of one. Finally, the value of the strong coupling (α_s) was determined by measuring the ratio of the number of W+1-jet events to W+0-jet events [3832]. Although the precision of these measurements could not compete with contemporary results from electron–positron colliders [3833], they verified that the value of α_s for events where a gluon is radiated in the initial state is consistent with other measurements.

The Tevatron collider, which ran at center-of-mass energies of 1.8 TeV and 1.96 TeV ushered in the era of large data

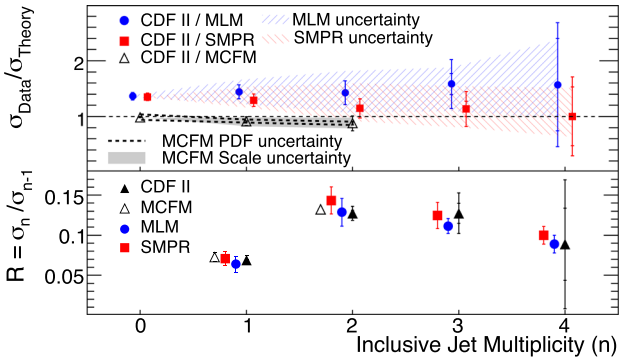


Fig. 311 CDF [3835]: The top panel shows the ratio of data to the predictions for the cross section of W+jets production for different inclusive jet multiplicities. The bottom panel shows the ratio of the cross section for n jets to $(n-1)$ jets. The NLO predictions (MCFM) are shown by the open triangles and the LO predictions (MLM, SMPR) are shown by the blue circles and red squares. The uncertainties on the data are indicted by the error bars, where the inner bars are the statistical uncertainty and the outer bars are the total uncertainties. The uncertainties on the predictions in the top panel are indicated by the hashed lines

samples of W and Z boson events and of increasing sophistication of the theoretical predictions used to describe that data. Since V+jets production is a dominant background to $t\bar{t}$ measurements and searches for the Higgs boson, the focus of the measurements shifted away from tests of the properties of QCD, such as α_s measurements, to tests of the dynamics of V+jets events. The large data samples collected by the CDF and D0 experiments allowed for measurements of W and Z boson production with up to four associated jets [3834]. Studies from the CDF and D0 experiments were expanded to include, for example, measurements of the differential cross sections as a function of the transverse momenta and rapidities of the jets, the angular separation of the two highest energy jets and the transverse momentum of the Z boson.

To describe these data, increasing sophisticated theoretical predictions were developed. The experimental and theoretical status at the time is nicely summarized in Fig. 311, which compares a next-to-leading-order (NLO) calculation and two leading-order (LO) calculations to the data. The LO calculations, which included multiple partons in the matrix-element calculations, are able to describe the shape of many kinematic distributions up to an overall normalization factor for high numbers of associated jets but are plagued by large uncertainties. In contrast, the theoretical uncertainties for the NLO calculation are much improved but the predictions do not extend over the full kinematic range of the data. For many years this figure represented the state-of-the-art in theoretical predictions for V+jets production.

The large W and Z boson data samples produced at the Tevatron also allow detailed studies of vector boson production in association with heavy-flavor jets, where heavy-flavor jets refers to c- or b-quark initiated jets. These measurements

are extremely important as these events provided the largest background contribution to measurements of $t\bar{t}$ production and searches of the Higgs boson via $WH(H \to bb)$ production. From the CDF and D0 collaborations, measurements of W production in association with a charm quark and W and Z production in association with b quarks were performed [3834]. One most notable result is the first measurement of $W + b$-jets production, which was done by the CDF collaboration, the measured cross section is 2.5–3.5 times larger than the various predictions with significance of 2.8 standard deviations. While the theoretical predictions used in this comparison did not fully account for b quarks in the initial state, this is not expected to explain the difference. The data sample itself was too small to allow measurements of kinematic distributions to resolve the source of the discrepancy.

In summary, the experiments at the S$\bar{p}p$S and the Tevatron colliders provided important tests of QCD theory in V+jets production. However, the scope of these measurements, with the exception of Ref. [3836] focused largely on measurements of the cross section for different jet multiplicities and a handful of differential cross section measurements. These measurements are important in validating QCD theory for topologies with multiple low energy jets, where the highest jet energies are not much greater than the mass of the vector boson itself. Rare processes such as W+b-jets production were measured for the first time but the statistical precision of the data samples is not sufficient to probe the kinematic distributions of these events.

12.3.2 V+jets at the LHC

In V+jets production at the LHC, measurements of jets with a transverse momentum greater than 1 TeV, which is much beyond the mass of the vector boson, are now accessible. At these high energies, the calculations from perturbative QCD suffer from large logarithmic corrections and are themselves potentially unreliable [3837]. With the large data samples available from the LHC, we have entered an era of high precision differential measurements with which we can explore QCD at higher-orders and in extreme corners of the phase space. For the first time, we also have sufficient data samples to measure in detail heavy-flavor production in multiple differential distributions. These measurements also provide for better understanding of the PDFs.

In pace with the increase in data samples, a plethora of new, more precise theoretical predictions, all with slightly different focuses, exist today for V+jets production. A more detailed summary of the available predictions can be found in Refs. [3412,3413]. In addition to LO matrix-element calculations, NLO calculations matched to parton shower models are now available; most notable for V+jets production are SHERPA, MADGRAPH5- AMC@NLO, MC@NLO and MEPS@NLO. NNLO calculations with next-to-next-to-leading logarithmic resummation and with parton showering are available using GENEVA. For fixed-order calculations, NLO predictions to five jets or more are available, such as BLACKHAT- SHERPA calculations, approximate NNLO predictions for jets with up to one jet, such as LOOPSIM calculations and NNLO predictions, such as N_{jetti}. Another calculation, HEJ, focuses on large rapidity separation and uses a resummation method to give an approximation to the hard-scattering matrix element for jet multiplicities of two or greater; in the limit of large rapidity separation between partons, this approximation becomes exact.

12.3.3 Tests of higher-orders

For our theoretical understanding of particle physics to keep pace with the improved accuracy of the measurements, theoretical predictions which include higher-order corrections are indispensable. Most of the measurements and searches performed today involve very high momenta jets, leptons or large amounts of missing transverse energy. In these regions, the high-order corrections play large and vital roles.

One important variable to test contributions from higher-order corrections is the observable of H_T, which is defined as the scalar sum of transverse momenta of the leptons, the missing transverse energy (for W+jets events) and the transverse momenta of all jets passing the selection criteria. At large values of H_T the average number of associated jets in the event increases. LO matrix-element calculations which do not provide higher-order terms drastically underestimate the average jet multiplicity. Here NLO predictions are needed to fully model these distributions. These distributions, among others, have been measured for both W+jets and Z+jets production [3838–3841]. Compared to the previous colliders the increase in kinematic reach at the LHC is dramatic; Tevatron results reach up to H_T values of 500 GeV, while the LHC results extend to 2 TeV.

The necessity of high-order corrections can readily be seen in measurements of the balance between the Z boson and the jet transverse momenta. The so-called jet-Z balance (JZB) is defined as the difference between the sum of the jet p_Ts (with $p_T > 30$ GeV and rapidity within 2.4) and the Z boson p_T. When all hadronic activity is contained within the selected jets, the JZB variable is zero. Figure 312 shows the measured data for events with $p_T(Z) < 50$ GeV compared to a LO and NLO MADGRAPH predictions and the GENEVA predictions [3842]. As seen in the figure, the distribution is not symmetric, with hadronic activity more dominantly pointing in the direction of the Z boson (i.e. positive values in this definition). The low $p_T(Z)$ region is interesting: While larger Z boson momenta can be described by fixed-order calculations, small values require resummation of multiple soft-gluon emissions to all orders in perturbation theory [1280,3843] (see Sects. 11.1 and 11.2). Different

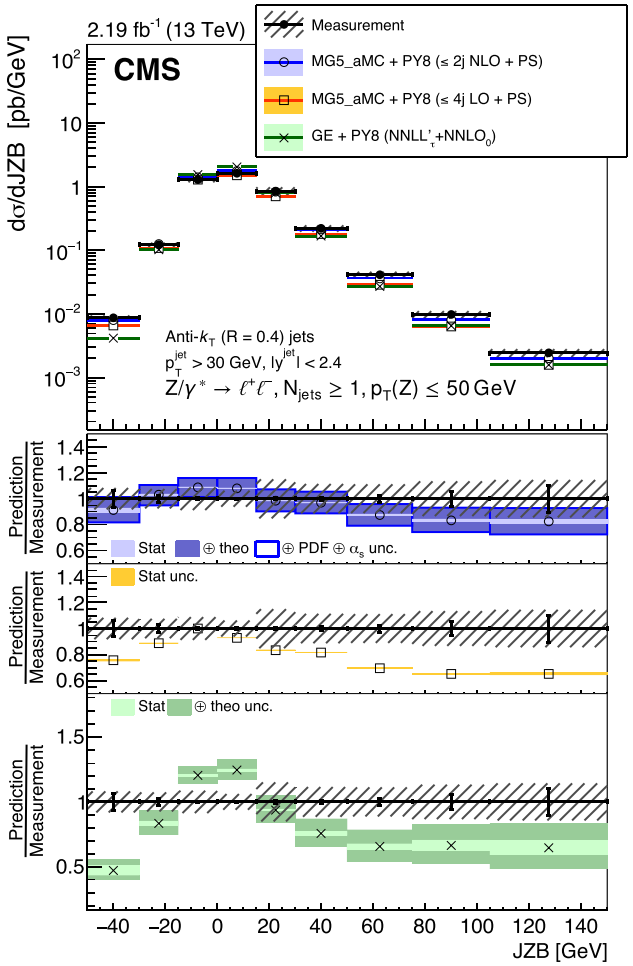


Fig. 312 CMS [3842]: Measured Z+jets cross section as a function of the JZB variable for $P_T(Z) < 50$ GeV. The data are compared to LO, NLO and NNLO predictions. The lower panels show the ratio of the three predictions to the data. The error bars on the measurement represent the statistical uncertainty and the grey hatched bands represent the total uncertainty. The uncertainties included on the theory predictions are listed in the labels

$p_T(Z)$ regions therefore test different theoretical treatments. The NLO predictions best describe the data and indicate that the NLO corrections are important to describe all hadronic activity in the event.

12.3.4 Modeling in extreme phase spaces

Extreme phase space regions, including events with high-p_T jets or high boson momenta or events with small angular separation between objects in the final state, tend to be governed by a complex mixture of the number of jets contributing to the final state and contributions from QCD as well as EW processes. This makes measurements of this nature an ideal test bed for studying multiple theory approaches.

The study of V+jet production with small angles between the boson and jets is one such critical test [3844–3846]. At LO, V+1-jet production is simply described by a V boson recoiling, in a back-to-back configuration with a jet. However, at NLO both real and virtual contributions to V+1-jet production appear via QCD and EW corrections. For real emissions of a V boson, either from an initial- or final-state quark, these contributions lead to an enhancement in production that is proportional to $\alpha_s ln^2(p_{T,j}/m_V)$, where α_s is the strong coupling, $p_{T,j}$ is the transverse momentum of the jet, and m_V is the mass of the V boson. This effect becomes largest with high transverse-momentum jets and can be isolated by selecting events with small angular separation between a jet and the V boson. In this region, the cancellation between real and virtual correction is incomplete, making the region ideal to probe. However, other processes such as V+2-jet production also play a role in this region and must be considered.

To study these effects, the ratio (so-called $r_{Z,j}$) of the Z boson p_T to the closet-jet p_T is defined. For collinear radiation of a Z boson, where the Z boson is expected to have a lower transverse momentum, this ratio results in smaller values. Figure 313 shows the $r_{Z,j}$ distribution for events where the angular separation, ΔR, between the jet and the Z boson is less than 1.4, corresponding to the region where the Z boson has a small angular separation from the jet (the collinear region) [3844]. While regions with back-to-back topologies (not shown here) are better modeled by predictions, the figure shows that higher order predictions model the collinear region best.

There are many other examples of tests in extreme regions including specific tests to isolate matrix element and parton shower effects [3847], measurements of probability of producing an additional jet in a rapidity gap of two jets [3848–3853], measurements in the forward region [3854,3855] and tests of QCD with sensitivity to physics beyond the Standard Model [3856,3857]. All of these measurements are critical for understanding QCD in these difficult-to-model regions.

12.3.5 Tests of QCD emissions

As demonstrated by the results from the UA1 and UA2 experiments, radiation of additional quarks and gluons is necessary in order to describe the events where the vector boson has a large transverse momentum. These higher-order QCD corrections consist of two classes; terms with a virtual loop which do not directly affect the boson p_T and terms with a real emission which do so but result in a jet which could be recorded by the detector. Measurements of the V+jet cross section for each jet multiplicity is therefore a direct test of QCD radiation. Measurements of the jet multiplicity ratios at the Tevatron and then at the LHC showed a striking feature: the ratio of the n-jet cross section to the $(n+1)$-jet cross section is a constant. This implies that the V+jet cross section

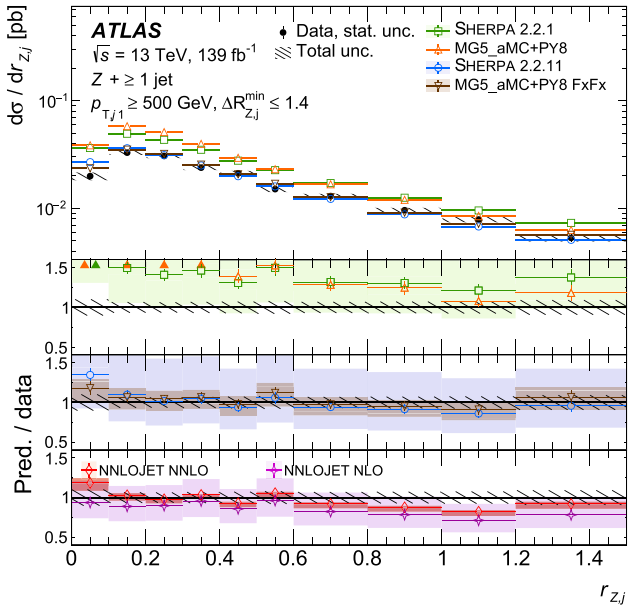


Fig. 313 ATLAS [3844]: Z+jets cross section in the collinear region as a function of the $r_{Z,j}$ variable. The data are compared to LO and NLO predictions. The lower panels show the ratio of the predictions to the data. The uncertainties shown on the data include the statistical uncertainty, as indicated by the error bar and the total uncertainty, as indicated by the hatched region. The uncertainties on the predictions are indicated by the colored bands

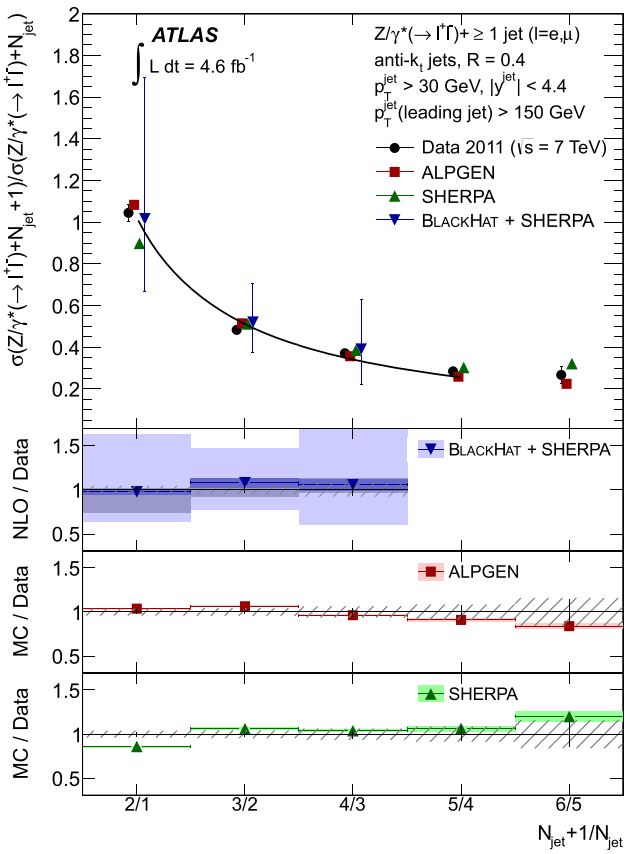


Fig. 314 ATLAS [3859]: Measurement of the ratio of the exclusive n-jet and $(n+1)$-jet cross sections for events where the highest p_T jet must have an energy above 150 GeV. The data are compared to the predictions from BLACKHAT- SHERPA, ALPGEN and SHERPA. The lower panels show the ratio of the three predictions to the data. The error bars indicate the statistical uncertainty on the data, and the hatched bands the statistical and systematic uncertainties on data. Uncertainties on the theory predictions are statistical only except for those of BLACKHAT- - SHERPA

follows the form

$$\sigma^{LO}_{V+n-jet} \sim \sigma_0 e^{-an} \tag{12.10}$$

with a is an experimentally-determined constant values and depends on the exact definition of the jets and σ_0 is representing the zero-jet exclusive cross section.

Although the constant scaling is a well established experimental observation, this behavior is not a priori expected [3858]. Assume, for example, that the probability of radiating a gluon from a quark follows the theory of Quantum Electrodynamics (QED), such that the gluon cannot radiate another gluon. In this case, the probability of radiating a gluon is dictated by a Poisson distribution, which implies that the cross section for an n-jet exclusive final state is

$$\sigma^{LO}_{V+n-jet} \sim \frac{\bar{n}e^{-\bar{n}}}{n!}\sigma_{tot}, \tag{12.11}$$

where σ_{tot} is the total cross section and $\bar{n}$ is the expectation value of the Poisson distribution, which also depends on the exact definition and selection of the jets. However, the gluon follows the non-abelian QCD theory and can radiate an additional gluon from itself. Therefore, at higher jet multiplicities the scaling would become constant.

The observation of a constant scaling for all jet multiplicities is instead a subtle cancellation of two different and opposite-sign effects. At low jet multiplicities, the Poisson scaling is present but cancelled by effects from the PDFs. To understand this effect, consider the case of high jet multiplicities with a cross section ratio of n-jet events to $(n+1)$-jet events, where n is a large number of jets. Here, the parton momentum fraction, x, for the involved partons is similar between the two jet multiplicities and therefore any effects on the cross section due to the PDFs essentially cancel in the ratio. In contrast, at low jet multiplicities, the relative difference in x for the involved partons between 2-jet events and 1-jet events is larger. Due to the steeply falling x-distribution of the gluon PDF, this implies that the production of 2-jet events compared to 1-jet events is suppressed by the PDFs. Depending on the exact selection criteria, this suppression cancels the increase in the production cross sections, which arises from the Poisson scaling.

Based on the work of Ref. [3858], the Poisson nature can be seen directly by selecting events with one very energetic jet. In these events, the effect on the cross sections from the PDFs is reduced; as seen in Fig. 314, the jet multiplicity cross

section follows the expected Poisson distribution. This measurement is a nice validation of the nature of QCD emissions from first principles.

12.3.6 Differential heavy flavor results

The associated production of vector bosons with heavy flavor jets is an important precision test of perturbative QCD in the presence of two mass scales – the vector boson mass and the c- or b-quark masses. Measurements of this nature also provide critical input to charm and strange distributions inside the proton, as discussed more in the next section. At LO, heavy-flavor production stems from either a gluon in the final state splitting into a heavy-flavor quark–antiquark pair or a heavy-flavor quark produced in the initial state. At the Tevatron, corrections to the cross section from the latter contributions are small, but at the LHC, these processes can lead to corrections of up to 50% [3860–3863]. Theory predictions for heavy-flavor production consist of 5-flavor-scheme models, where the b-quark is included in the PDF itself or 4-flavor-scheme models, where it is not. The two schemes, however, are equivalent if the calculations included all orders of α_s.

With the large data samples available from the LHC, these processes can be studied for a variety of differential observables [3834,3864–3869] and also in the forward region and in phase spaces with very energetic, boosted jets [3870–3873]. In general, 5-flavor-scheme predictions are better at describing the data compared to 4-flavor ones. However, there are sizable differences even between predictions of a similar nature. Figure 315 shows the separation between the two b-quarks, which is a sensitive variable to gluon splitting [3864]. The NLO SHERPA simulation estimates this observable well but fails to get the overall cross section correct (not seen in this figure). In contrast, the LO and NLO MADGRAPH predictions are less able to model the shapes of the kinematic observables but estimate well the overall cross section. In regimes where the vector boson has a large p_T, the predictions generally perform worse; for example they underestimate events with high m(bb) by about half [3866]. Work is on-going to combine massless NNLO with a massive NLO computation, with promising comparisons to data [3874].

12.3.7 Probes of parton density functions

A major source of uncertainty in all hadron collider measurements stems from knowledge of the PDFs. As our knowledge of QCD deepens, better knowledge of the PDFs are needed to continue to be sensitive to deviations from Standard Model predictions [3057,3058]. Deep inelastic scattering data from the HERA experiments provided some of the best data for PDF determination over a wide range of Q^2 and x. In addition to these data, data from various experiments, such as

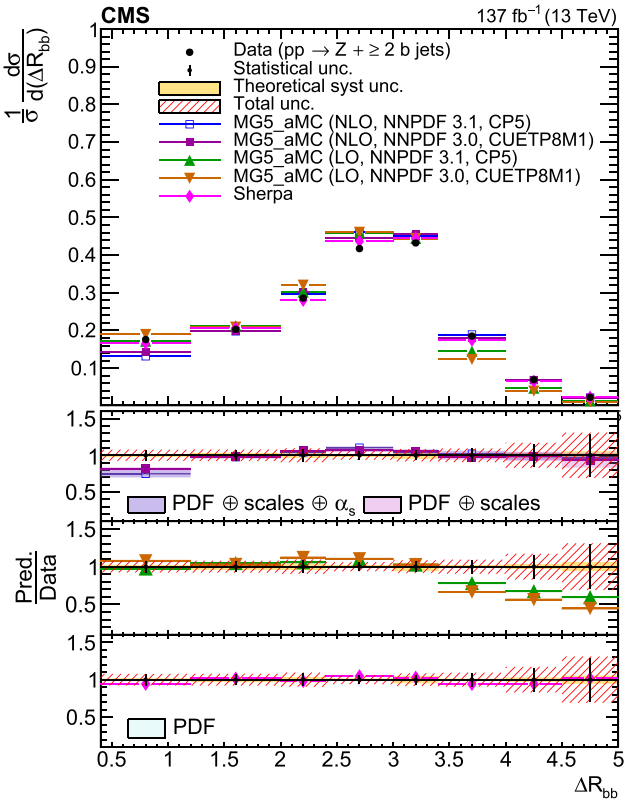


Fig. 315 CMS [3864]: Normalized differential cross section as function of the angular separation between two b jets, ΔR_{bb} for $Z+ \geq 2$ b-jets events. The uncertainties in the predictions are shown as colored bands in the bottom panel. The statistical, theoretical, and total uncertainties in data are indicated by the vertical bars and the hatched bands

those from neutrino and hadron collider experiments. The LHC offers a unique opportunity in that it provides a diverse set of processes, such as jet, photon, vector boson or top production, which can be used to constrain different regions with the PDFs. Today, PDFs can be determined at up to NNLO precision in perturbative QCD. The input data span the range of $10^{-5} \lesssim x \lesssim 1$ and $1 \lesssim Q^2 \lesssim 10^6$ GeV^2.

Measurements of V+jet production are particularly important since these processes can probe u and d quarks and also contributions from s, c and b quarks. By considering vector boson processes with additional jets, the measurements are sensitive to higher values of x, accessing $x \approx 0.1-0.3$ [3875], compared to inclusive W and Z measurements. Measurements of this nature constrain the light-quark sea at higher x as well as the strangeness contributions and help to better understand the gluon distribution at high x [3825]. The LHCb experiment, with its precision tracking coverage in the forward region, offers new possibilities here in that its V+jets measurements are sensitive to PDFs at different x ranges compared to the ATLAS and CMS experiments [3854]. These measurements probe PDFs at x as low as 10^{-4} and at high $x > 0.5$ [3876].

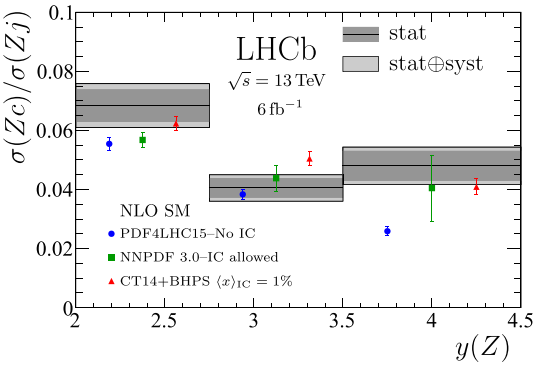


Fig. 316 LHCb [3871]: Measured cross section ratio of $Z + c/Z$+jet production for three intervals of forward Z rapidity, compared to NLO predictions with and without IC, and with IC as predicted by BHPS with a mean momentum fraction of 1%

The contributions to the proton from strange quarks can be probed through measurements of W boson production in association with c-quark, as was done at the Tevatron and the LHC [3132,3834,3865,3872]. In past years, whether or not strangeness contributions are suppressed in the proton is a topic of debate, with mainly the ATLAS data preferring less suppression as compared to neutrino scattering and CMS data. However, with more data, in particular measurements of W+jets and $W + c$ production have provided powerful input on the strangeness contribution. Today, there is general agreement by modern PDFs that strangeness is not strongly suppressed at low x but has substantial suppression at high x.

It has been a decade long debate if the proton may contain an 'intrinsic' charm component in addition to that from gluon splitting, which decreases sharply at large values of x [3877]. Such models, like the BHPS model, predict that protons would have a valence-like charm content. Global PDF analyses are generally inconclusive and therefore more direct probes are needed [3141,3878]. Since intrinsic charm contributions are enhanced in Z+jet production where the Z boson has large rapidity, the LHCb experiment is perfectly suited for these measurements [3871]. As seen in Fig. 316, the data at forward Z rapidities from a recent $Z + c$ measurement, are consistent with an intrinsic charm contribution as predicted by BHPS models. Future analysis of the effect of these data within PDF fits themselves, however, is still needed.

Since their discovery in the early 1980s, the W and Z bosons are important probes to understanding QCD. The early measurements at the S$\bar{p}p$S and the Tevatron were critical in establishing the dynamics of these processes, while at the LHC, V + jets production is now explored at the highest available energies. To step up with experimental precision, a suite of versatile and precise theory predictions have been developed to compare to the data. Future measurements of V+jets production are needed to better understand QCD theory in very energetic regions of phase space, to measure electroweak corrections, to improve PDFs and for a better understanding of heavy flavor production.

12.4 Higgs production

Chiara Mariotti

In July 2012, the ATLAS and CMS Collaborations at the CERN Large Hadron Collider (LHC) announced the discovery of the last missing piece of the Standard Model (SM) of elementary particles: the Higgs boson [139,140,3880]. The discovery arrived about 50 years after theorists had postulated its existence to explain the mechanism by which the elementary particles acquire mass.

The Higgs boson is the excitation of a field, called Brout–Englert–Higgs (BEH) field. The field name comes from the theoreticians who first introduced the mechanism [43,44,3881]. The BEH field filled the entire universe less than a picoseconds after the Big-Bang. The elementary particles interacting with it acquire mass. Without this field the world would not be the same, as an example the electron would be massless and atoms could not be formed.

The Higgs boson has unique quantum numbers: $J^{PC} = 0^{++}$, since the field must be the same everywhere in the space and should not depend on the reference frame.

Since the time of the discovery, the ATLAS and CMS experiments have accumulated data during the Run 1 (2009–2012) at 7 and 8 TeV proton–proton center-of-mass energy and Run 2 (2015–2018) at 13 TeV. The two collaborations observed the Higgs boson in numerous bosonic (ZZ, WW, $\gamma\gamma$), and fermionic decay channels ($\tau^+\tau^-$, $b\bar{b}$ quark), measured its mass and width, determined its spin-parity quantum numbers, and measured its production cross sections in various modes (gluon–gluon fusion, vector boson fusion, associated production with a W or a Z, associated production with 2 top quarks). Within the uncertainties, all these observations are compatible with the predictions of the SM.

Finding the Higgs boson has been very demanding. Its production cross section is 12 orders of magnitude smaller than the proton–proton inelastic cross section at LHC energies. Few hundreds of particles are produced at each collision, and there can be several simultaneous proton–proton collisions at each proton bunch crossing (pileup). It is thus fundamental to have a very good understanding of the resonant and non-resonant hadronic background: production of background processes via QCD interactions has to be well understood and modeled.

Because of its large mass, the Higgs boson could not be discovered at LEP [3882] at CERN, and because of its very low production cross section it was very challenging to observe it at the Tevatron [3883] at Fermilab. Only at LHC, thanks to the energy available in the center-of-mass, and to the exceptionally high luminosity, it was possible to produce it with a rate sufficient to discover it.

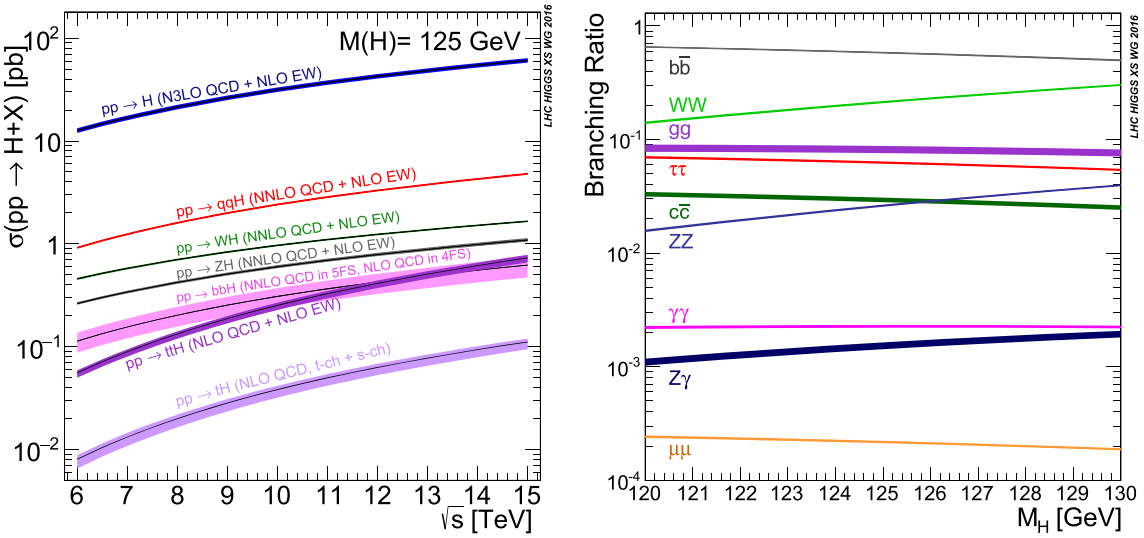


Fig. 317 (Left) Cross section for a Higgs boson of 125 GeV of mass as a function of the center-of-mass energy at the LHC [3879]. (Right) Branching ratios as a function of the Higgs boson mass [3879]

Precise theoretical calculations for the Higgs boson production modes and decay channels have been performed; the results are shown in Fig. 317. The dominant production mode at the proton–proton LHC collider is the gluon–gluon fusion (ggF, or $pp \to H$) as shown in Fig. 317(left), followed by the vector boson fusion (VBF, or $pp \to qqH$), the associated H production with vector bosons ($pp \to ZH, WH$), and the associated production with two b quarks or two top quarks or just one top quark. Many groups contributed to the computation of these production cross sections over many years [3879,3884–3886]. The perturbative order of the calculations in QCD and EW is indicated in the figure. The thickness of the line represents the uncertainty of the calculation.

The cross section of the ggF process is known at N3LO with very good precision (5% in total, of which 3% are due to missing higher order effects). The calculation of the higher perturbative orders in QCD, as well as the resummation (see Sects. 11.1 and 11.2), contribute substantially to the precision as shown in Fig. 318 [3879]. The parton distribution functions (PDFs) have been determined with very good accuracy by several groups at NNLO in QCD and reached a precision of $\sim 2\%$ for the gluon–gluon luminosity over a wide range of energy [3879].

The strength of the Standard Model Higgs boson coupling is proportional to the mass of the fermions, and to the mass squared of the vector bosons. Thus it will decay predominantly to the available elementary particle with larger mass: for a Higgs boson of $m_H = 125$ GeV, the largest branching ratio (BR) is to $b\bar{b}$, followed by W^*W. The various BRs have been computed at least at NLO precision for both QCD and EW corrections, and are shown in Fig. 317(right) [3879].

Calculation of the background processes for the various Higgs boson decay channels have been and are being computed with increasing precision at higher order in perturbation theory. In parallel, experiments have developed methods to estimate the various sources of background in a data-driven way, not to depend on the availability of Monte Carlo (MC) simulations, or on precise theoretical calculations and modeling.

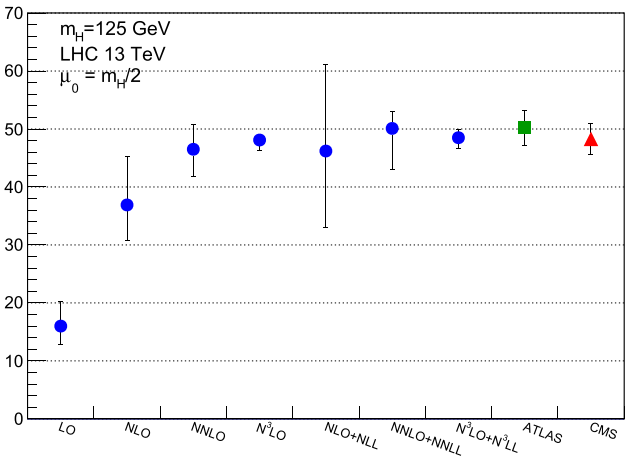


Fig. 318 Calculated theoretical ggF cross-section values (blue circles) at various perturbation orders [3879]. The latest ATLAS (green square) [3887] and CMS (red triangle) [3888] results from Run2 are also shown

12.4.1 Higgs boson properties

The ATLAS and CMS experiments, with the data collected during the Run 1 and Run 2, measured with very good precision the properties of the Higgs boson: the mass is measured with a precision better than 0.2% in the $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$ final states:

ATLAS ($H \to ZZ \to 4\ell$ final state only) [3889]:

$m_H = 124.94 \pm 0.17$ GeV

CMS [3890]:

$m_H = 125.38 \pm 0.14$ GeV.

As an example, Fig. 319 shows the diphoton invariant mass distribution targeting the study of the decay channel $H \to \gamma\gamma$ in ATLAS [3891], and the invariant mass distribution of four charged leptons targeting the study of the decay channel $H \to ZZ^* \to 4\ell$ in CMS [3890].

The spin and parity have been measured and found to be compatible with the SM prediction, $J^P = 0^+$, at $> 99.9\%$ confidence level (CL) [3892,3893]. The width of the Higgs boson has been measured to be $\Gamma_H = 3.2^{+2.4}_{-1.7}$ MeV by using off-mass-shell and on-mass-shell Higgs boson production, in final states with four charged leptons or $2\ell 2\nu$ [3894], with the assumption that on-shell and off-shell effective couplings are the same.

All the production modes (except tH and bbH) have been observed with a significance larger than 5σ, as well as several decay channels: WW, ZZ, $\gamma\gamma$, $\tau\tau$, $b\bar{b}$. A 3σ evidence for the $\mu\mu$ final state was found by the CMS experiment [3888]. ATLAS and CMS have recently presented results on the search for the $\ell\bar{\ell}\gamma$ final state, reaching about 3σ significance [3895,3896].

The experiments test the compatibility of their measurements with the SM, and the results are generally presented in two ways: by means of *signal-strength modifiers* μ (defined as $\mu = \sigma \times BR/(\sigma \times BR)_{SM}$, or *coupling-strength modifiers* κ (defined as $\kappa^2 = \sigma/\sigma_{SM}$, or $\kappa^2 = \Gamma/\Gamma_{SM}$) [3886]. Fitting the data from all production modes and decay channels with a common signal strength μ, the experiment found the following results:
ATLAS [3887]:

$\mu = 1.05 \pm 0.04(th) \pm 0.03(exp) \pm 0.03(stat)$,

CMS [3888]:

$\mu = 1.002 \pm 0.036(th) \pm 0.033(exp) \pm 0.029(stat)$,

showing a very good agreement with the SM, within the uncertainty. The theoretical (th) uncertainty has decreased by about a factor of 2 with respect to Run1, thanks to the huge effort of the theoretical community; the huge increase in statistics (i.e. 30 times more Higgs boson events), a better understanding of the detector, and more sophisticated methods (like Boosted Decision Trees, Deep Neural Network and Advanced Machine Learning) have helped to decrease the experimental (exp) and statistical (stat) uncertainty by a factor of more than two.

For a given production and decay, $i \to H \to f$, two parameters μ_i and μ_f are defined as $\mu_i = \sigma_i/(\sigma_i)_{SM}$ and $\mu^f = BR^f/(BR^f)_{SM}$. Many initial states i and final states

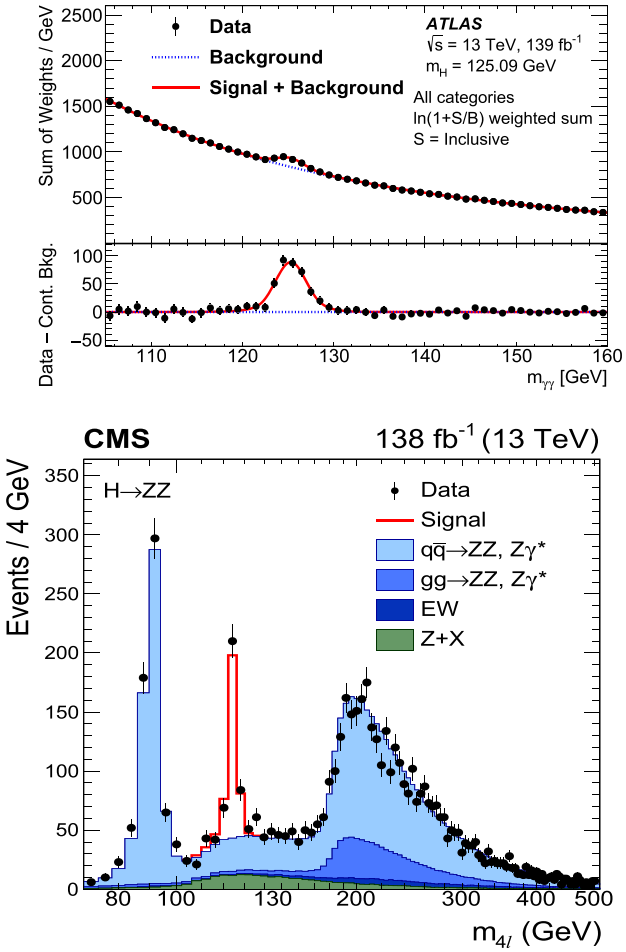
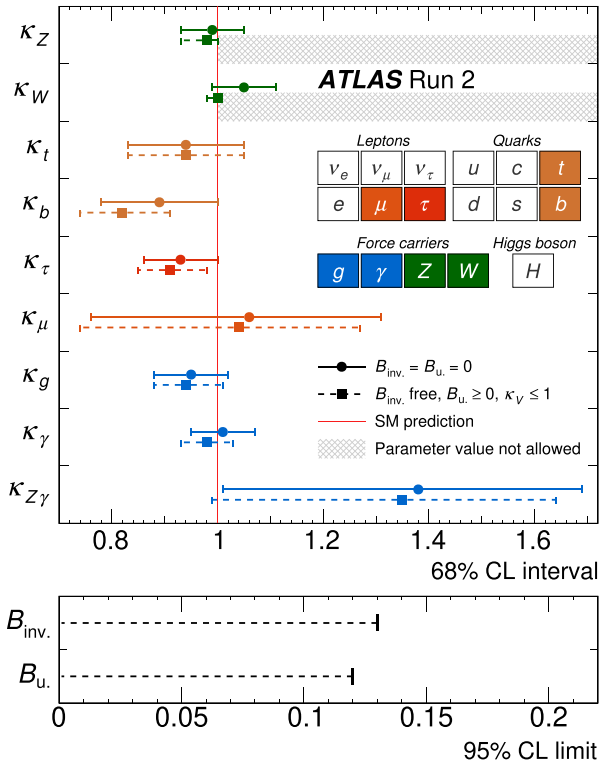


Fig. 319 (Upper) The diphoton invariant mass distribution in ATLAS [3891]. The data events (dots) are weighted by $\ln(1 + |S|/|B|)$, where S and B are the expected signal and background. (Lower) The invariant mass distribution of four charged leptons targeting the study of the decay channel $H \to ZZ^* \to 4\ell$ in CMS [3890]

f share the same coupling, e.g. VBF H production and $H \to VV$ decay both involve the HVV coupling ($V = W, Z$). Another example is the $H \to \gamma\gamma$, that proceeds via a loop of W bosons or top quarks, thus involving the HWW and Htt couplings. Each iiH and Hff coupling is multiplied by a scaling factor κ, thus defined as $\kappa_j^2 = \sigma^j/\sigma_{SM}^j$, or $\kappa_j^2 = \Gamma^j/\Gamma_{SM}^J$. The experiments have presented results on the κ_j with the full Run2 statistic [3887,3888].

In the presence of new physics, new particles could contribute to the loops, affecting the various couplings and modifying the SM relations. Thus an alternative fit could be performed assuming non resolved loop for the coupling of the Higgs boson with photons or gluons, and thus assuming effective couplings κ_γ and κ_g. The results are shown in Fig. 320 [3887,3888]. Moreover, in the fit the possibility of the Higgs boson decaying to invisible particles (i.e. neutrinos or dark matter candidates), B_{inv}, or to undetected particles,



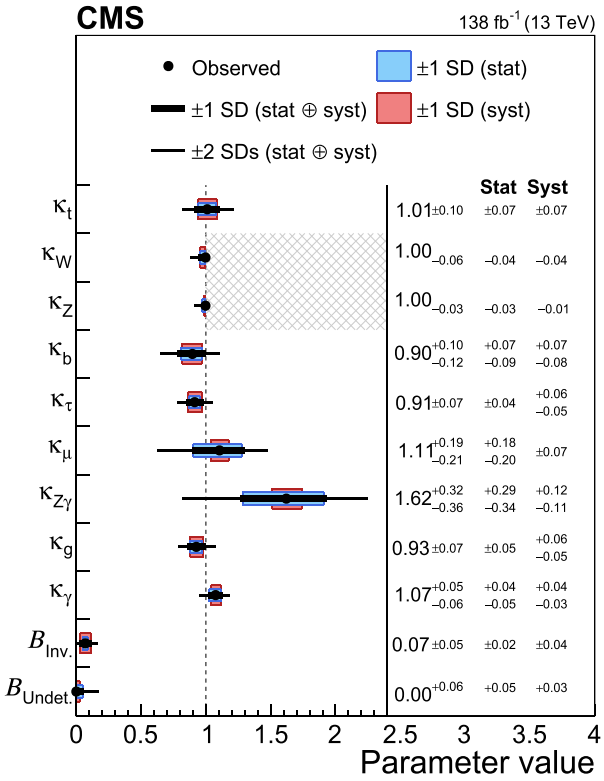


Fig. 320 (Left) Coupling-strength modifiers and their uncertainties per particle type with effective photon, $Z\gamma$ and gluon couplings in the ATLAS experiment [3887]. The horizontal bars on each point denote the 68% confidence interval. The scenario where $B_{inv.} = B_{u.} = 0$ is assumed is shown as solid lines with circle markers. The p-value for compatibility with the SM prediction is 61% in this case. The scenario where $B_{inv.}$ and $B_{u.}$ are allowed to contribute to the total Higgs boson decay width while assuming that $\kappa_V \leq 1$ and $B_{u.} \geq 0$ is shown as dashed lines with square markers. The lower panel shows the 95% CL upper limits on $B_{inv.}$ and $B_{u.}$. (Right) Results of a fit to the coupling-strength modifiers κ allowing both invisible and the undetected decay modes, with the SM value used as an upper bound on both κ_W and κ_Z in the CMS experiment [3888]. The thick (thin) black lines indicate the 1 (2) standard deviation confidence intervals, with the systematic and statistical components of the 1 standard deviation interval indicated by the red and blue bands, respectively. The p-value with respect to the SM prediction is 33%

B_u or $B_{Undet.}$ (i.e. particles that may or may not leave a trace in the detector, and the experiments do not have dedicated searches looking for these) is allowed. The presence of invisible or undetected decays can be inferred indirectly from a reduction in the branching fraction for SM decays or by an increase in the total Higgs boson width. In this interpretation, the total width becomes $\Gamma_H = \sum \Gamma_f(\kappa)/(1 - B_{inv.} - B_u)$.

Figure 321(left) shows that indeed the Higgs boson couples with the fermion and boson masses as predicted by the SM. The very good agreement spans over many orders of magnitude. The results are shown for CMS [3888], and ATLAS has presented similar results [3887]. Figure 321(right) shows the observed and projected values resulting from the fit in the κ-framework in different data sets: at the time of the Higgs boson discovery, using the full data from LHC Run 1, in the Run2 data set (*"This paper"*), and the expected 1 standard deviation uncertainty at the high-luminosity run (HL-LHC) for an integrated luminosity of 3000 fb^{-1} [3888].

12.4.2 Cross section measurements

With the data collected during Run 1 and Run 2, the ATLAS and CMS experiments measured the Higgs boson ggF production cross section with about 6% precision. The total cross section measurement from ATLAS [3887] at $\sqrt{s} = 13$ TeV is 50.2 ± 3.0 pb, and CMS measures 48.3 ± 2.7 pb [3888], both in agreement with the SM prediction of $48.5^{+1.5}_{-1.9}$ pb, as shown in Fig. 318.

Figure 322 shows the cross sections for different production processes and the branching fractions for different decay modes, as measured by the ATLAS experiment [3887].

12.4.3 The simplified template cross section

The simplified template cross section (STXS) method has been developed at the Les Houches 2015 workshop, and within the LHC Higgs Cross Section Working Group [3879] with the aims to separate more cleanly measurement and

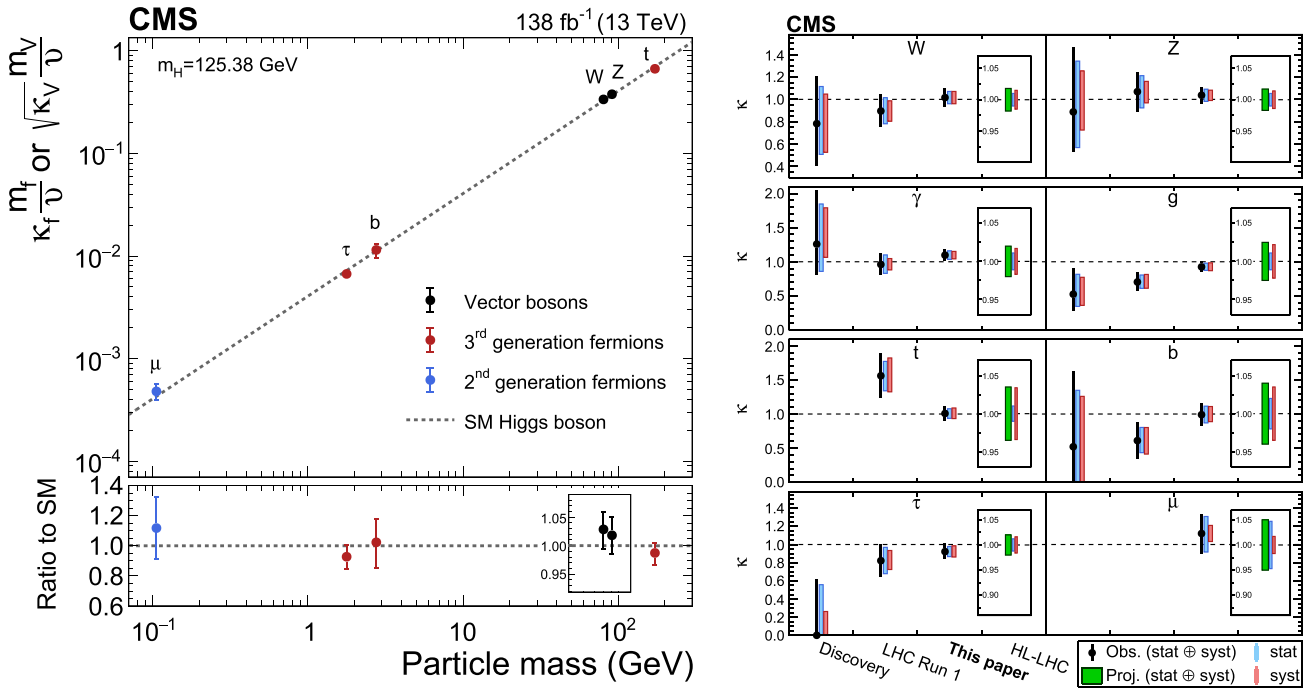


Fig. 321 (Left) Measured Higgs boson couplings to fermions and gauge bosons as a function of the fermion or gauge boson mass, where υ is the vacuum expectation value of the BEH field, and κ_i are the coupling modifiers as described in the text [3888]. (Right) Observed and projected values resulting from the fit in the κ-framework in different data sets: at the time of the Higgs boson discovery, using the full data from LHC Run 1, in the Run 2 data set (*this paper*), and the expected 1 standard deviation uncertainty at the HL-LHC for an integrated luminosity of 3000 fb^{-1} [3888]. These results assume that no contributions from BSM is present in loops

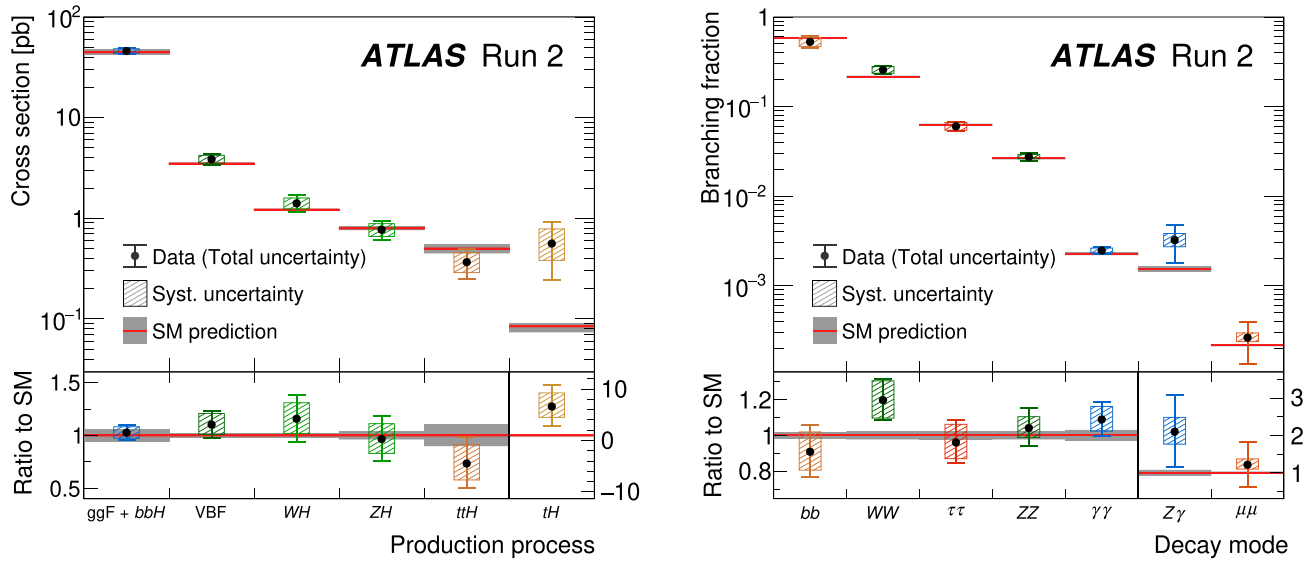


Fig. 322 (Left) Observed and predicted cross section for different Higgs boson production modes, measured assuming SM values for the decay branching fractions in ATLAS [3887]. (Right) Observed and predicted branching fractions for different Higgs boson decay channels. The lower panels show the ratio of the measured values to their SM predictions [3887]

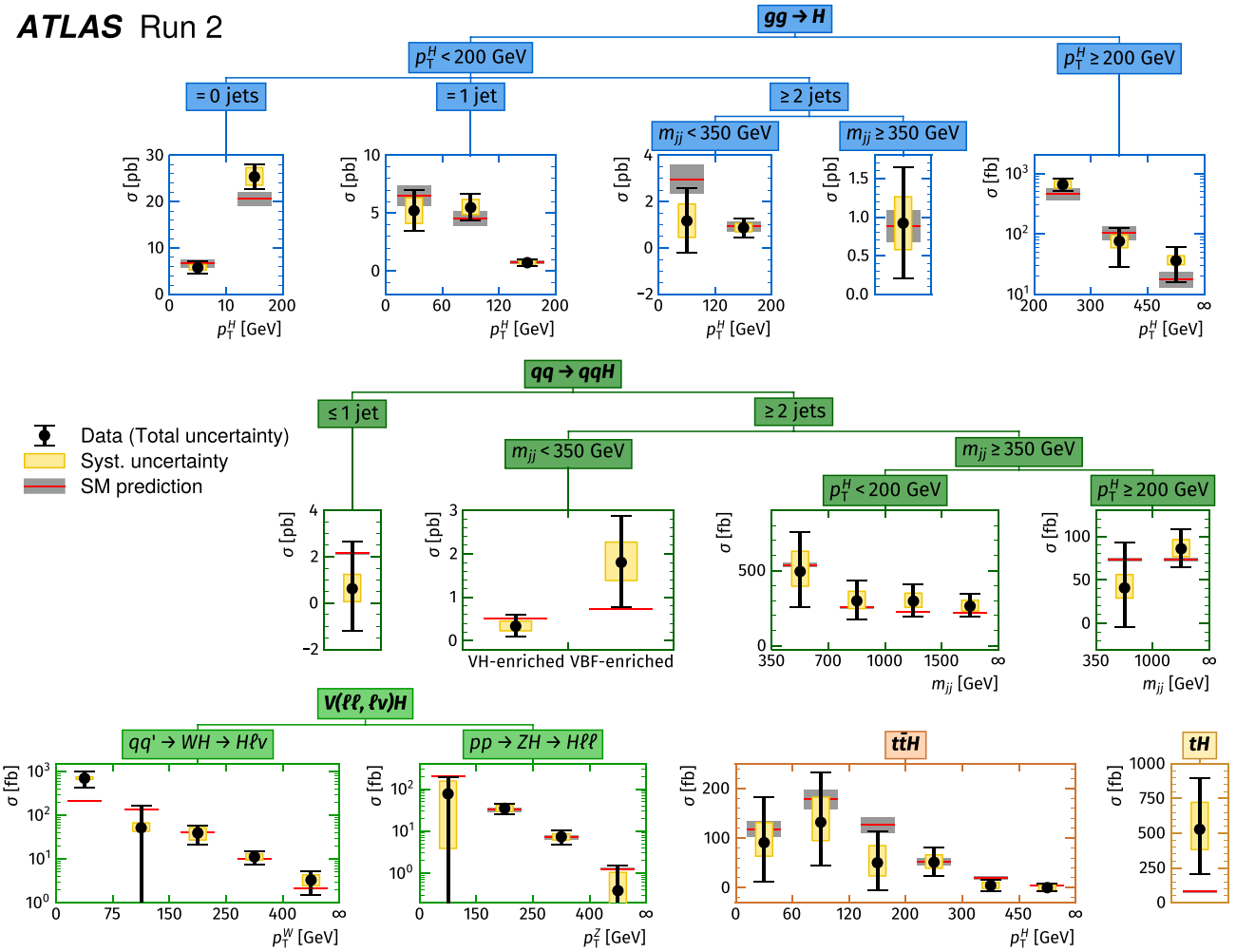


Fig. 323 Observed and predicted Higgs boson production cross sections in different kinematic regions [3887]. The vertical bar on each point denotes the 68% confidence interval. The p-value for compatibility of the combined measurement and the SM prediction is 94%. Kinematic regions are defined separately for each production process, based on the jet multiplicity, the transverse momentum of the Higgs boson $p_T(H)$ and vector bosons $p_T(W)$ and $p_T(Z)$ and the two-jet invariant mass (m_{jj}). The VH-enriched and VBF-enriched regions with the respective requirements of $60 < m_{jj} < 120$ GeV and $m_{jj} < 60$, $m_{jj} > 120$ GeV are enhanced in signal events from VH and VBF productions, respectively

interpretation steps in order to reduce the theory dependencies that are folded into the measurements (including the dependence on theoretical uncertainties and on the underlying physics model). Its primary goals are to maximize the sensitivity of the measurements and to minimize their theory dependence. The method is designed to measure cross sections separated into production modes (instead of signal strengths), in mutually exclusive regions of phase space, and to be inclusive over Higgs boson decays, allowing to perform a global combination of all decay channels and to ease interpretation and search for BSM phenomena. Figure 323 shows the results of ATLAS for the LHC Run2 data [3887].

12.4.4 Differential distributions

The large data set accumulated during the LHC Run 2 allowed the experiments to do the first studies of differential distributions. A convenient set of kinematic variables to describe the Higgs boson production in hadronic collisions, and to test QCD consists of the transverse momentum p_T, the rapidity y, and the azimuthal angle ϕ. The first two variables allow to understand many important QCD effects. The p_T distribution is sensitive to perturbative QCD, and at low value it is strictly connected with the resummation of the leading logarithms, while at large values new physics could manifest. The y distribution is sensitive to the parton distribution functions. At LHC the processes should not depend

on ϕ. Two important additional variables, that probe the theoretical modeling of high-p_T QCD radiations in Higgs boson production, are the number of jets in the event, N_{jet}, and the transverse momentum of the leading jet, $p_T^{lead.jet}$.

Differential distributions are usually measured unfolding the detector resolution and efficiency effects and calculating "fiducial" cross sections. Cross sections are measured in a fiducial phase space, which is defined to closely match the experimental acceptance in terms of the physics object kinematics and topological event selection. This approach is chosen in order to reduce the systematic uncertainty associated with the underlying model and with the extrapolation to non-measured regions. As an example, the fiducial phase space for $H \to 4\ell$ constitutes approximately 50% of the total phase space. The fiducial differential cross sections are then compared with the various MC simulations and analytical calculations.

Figure 324 shows the differential cross section for the processes $pp \to H \to 4\ell$, $pp \to H \to \gamma\gamma$, and their combination as a function of the Higgs boson transverse momentum, its rapidity, the number of jets in the event, and the leading jet p_T as measured by the ATLAS experiment [3897]. The data are compared with various theoretical predictions, all normalized to the total cross section, where the dominant ggF contribution is calculated at fixed order N3LO.

Figure 325 shows the double differential fiducial cross section measured in bins of $p_T^{\gamma\gamma}$ and n_{jets} for $H \to \gamma\gamma$ events in the CMS experiment [3898]. The data are compared to the predictions from different setups of the event generator MadGraph5_aMC@NLO (version 2.6.5) [3385].

12.4.5 The Higgs boson and heavy quarks

The dominant decay of the SM Higgs boson is into pairs of b quarks, with an expected branching fraction of approximately 58% for a mass of 125 GeV, but the large background from multi-jet (QCD) production makes the search in ggF very challenging. The decay of the Higgs boson to $b\bar{b}$ was observed during Run 2 by ATLAS and CMS, in events where the H is produced in association with a vector boson, i.e. in the WH and ZH production modes [3899,3901]. In these events, the leptonic decay of the vector boson allows for efficient triggering and a significant reduction of the multi-jet background. In addition, two identified jets coming from the hadronization of b quarks from the Higgs boson decay are required. The dominant background processes after the event selection are V+jets, $t\bar{t}$, single-top, diboson process and multi-jets.

Benefiting from multivariate techniques (MVA) and new machine learning algorithms, the experiments are now developing analyses to search for $H \to b\bar{b}$ inclusively in the production mode. Highly Lorentz-boosted Higgs bosons decaying to $b\bar{b}$, recoiling against a hadronic system, are reconstructed as single large-radius jets, which are identified using jet substructure algorithms and a dedicated b tagging technique based on a deep neural network (see Sect. 11.5). The jet mass is required to be consistent with that of the observed Higgs boson, and the jet transverse momentum is required to be $p_T > 400{-}450$ GeV. The analyses are validated using $Z \to b\bar{b}$ events. The measured cross section is compatible with the SM one, but for the moment the uncertainty is still very large, i.e. around 10% [3902,3903]. Figure 326(left) shows the reconstructed $b\bar{b}$ invariant mass for the selected VH events in the ATLAS experiment [3899].

The decay branching fraction of the SM Higgs boson into a pair of c quarks is slightly less than 3%. The difficulties to measure this channel are even larger than for the b quark final state, because the main background to c jet identification is indeed from b jets. Higgs boson candidates, produced in association with a W or a Z boson, are constructed from the two jets with the highest p_T, with at least one jet identified as originating from a c quark. [3900,3904]. In CMS the search is extended to events in which the H boson decays to a single large-radius jet. Additionally, a $b-$jet identification algorithm is used to veto b jets. Novel charm jet identification and analysis methods using machine learning techniques are employed. In Fig. 326(center) the $c\bar{c}$ tagging efficiency is shown versus the efficiency of misidentifying quarks and gluons from V+jet and $H \to b\bar{b}$ in CMS. The analysis is validated by searching for $Z \to c\bar{c}$ decays in the VZ process, leading to the first observation of this process at a hadron collider with a significance of 5.7 standard deviations, as shown in Fig. 326(right) [3900]. The observed upper limit on $\sigma(VH)BR(H \to c\bar{c})$ is ranging from 14 to 26 times the SM prediction, for an expected limit that ranges from 7 to 31 for CMS and ATLAS, respectively.

The $t\bar{t}H$ and tH production channels probe the coupling of the Higgs boson to the top quarks. The large mass of the top quark may indicate that it plays a special role in the mechanism of electroweak symmetry breaking. Deviations from the SM prediction would indicate the presence of physics beyond the SM. The measurement of the Higgs boson production rate in association with a top quark pair ($t\bar{t}H$) provides a model-independent determination of the magnitude of the top quark Yukawa coupling y_t. The sign of y_t is determined from the associated production of a Higgs boson with a single top quark (tH). The $t\bar{t}H$ and tH production channels are studied in the case where the Higgs boson and the top quarks subsequently decay into final states with several leptons (including taus, also when they decay hadronically), complementing dedicated studies of the $H \to \gamma\gamma$, $H \to ZZ \to 4\ell$, and $H \to b\bar{b}$ decay modes. Several MVA techniques are employed to better separate the $t\bar{t}H$ and tH production modes. The ttH production modes has been observed in Run 2 [3905,3906]. The precision on the

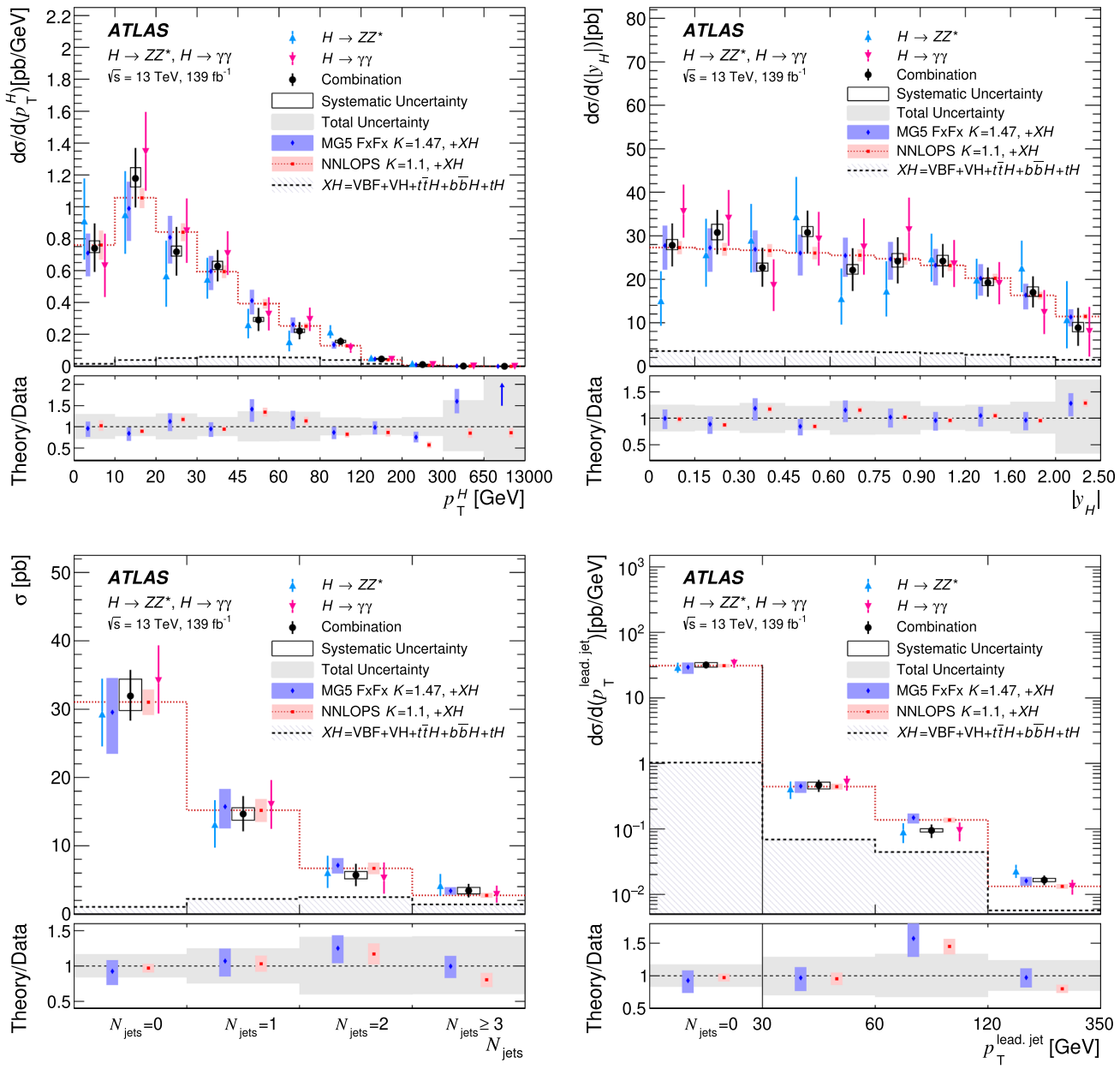


Fig. 324 Differential $pp \to H + X$ cross-sections, in the full phase space, as a function of variables characterising the Higgs boson kinematics in ATLAS [3897]: **a** Higgs boson transverse momentum p_T^H, **b** Higgs boson rapidity y^H, **c** number of jets and **d** p_T of the leading jet, compared with the Standard Model prediction. The $H \to ZZ^* \to 4\ell$ (blue triangles), $H \to \gamma\gamma$ (magenta inverted triangles), and combined (black squares) measurements are shown. The error bars on the data points show the total uncertainties, while the systematic uncertainties are indicated by the boxes. The measurements are compared with two predictions, obtained by summing the ggF predictions of NNLOPS or MG5 FxFx, normalized to the fixed order N3LO total cross-section, and MC predictions for the other production processes XH. The shaded bands indicate the relative impact of the PDF and scale systematic uncertainties in the prediction. These include the uncertainties related to the XH production modes. The dotted red histogram corresponds to the central value of the prediction that uses NNLOPS for the modelling of the ggF component. The bottom panels show the ratios between the predictions and the combined measurement. The grey area represents the total uncertainty of the measurement. For better visibility, all bins are shown as having the same size, independent of their numerical width

top Yukawa coupling and on $t\bar{t}H$ cross section measurement is presented in Figs. 320 and 322.

12.4.6 Precision Higgs boson physics

The Higgs boson was discovered by the ATLAS and CMS experiments in 2012 at the LHC. With the data taken dur-

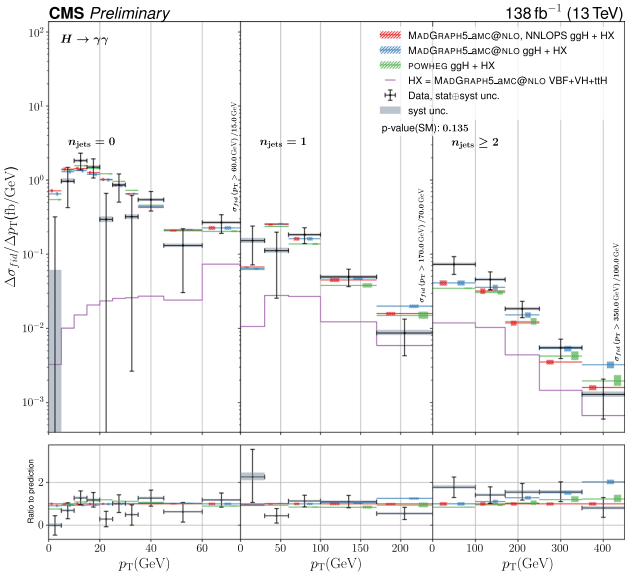


Fig. 325 Double differential fiducial cross section measured in bins of $p_T^{\gamma\gamma}$ and n_{jets} [3898]. The observed differential fiducial cross section values are shown as black points with the vertical error bars showing the full uncertainty, the horizontal error bars show the width of the respective bin. The grey shaded areas visualize the systematic component of the uncertainty. The colored lines denote the predictions from different setups of the event generator. All of them have the $HX = VBF + VH + ttH$ component from MadGraph5_aMC@NLO in common. The green lines show the sum of HX and the ggF component from MadGraph5_aMC@NLO reweighted to match the nnlops prediction. For the orange lines no nnlops reweighting is done and the purple lines take the prediction for the ggF production mode from POWHEG. The hatched areas show the uncertainties on theoretical predictions. Only effects coming from varying the set of PDF replicas, the α_S value and the QCD renormalization and factorization scales that impact the shape are taken into account here, the total cross section is kept constant

ing the Run1 and Run2 the two experiments successfully tested the SM Higgs boson. The precision on the couplings is ranging from 3% for the coupling to the Z, to 10% for the coupling to $b\bar{b}$ and $t\bar{t}$, to 20–30% for the couplings to muons and $Z\gamma$. Run3 and the high-luminosity LHC (HL-LHC) will deliver approximately 3000 fb^{-1} of luminosity to each experiment. By the end of HL-LHC, rare decays channels such as $H \to \mu\mu$ and $H \to Z\gamma$ will be observed and studied, the SM Higgs boson pair production is estimated to be observed with a significance of 4 to 5 standard deviations, when combining the results of the two experiments, as well as the Higgs boson coupling to charm quarks. As of today, the experiments have analysed only 3% of the Higgs boson events that they will have at the end of LHC. By then, most of the couplings measurements will reach the 2 to 3% precision, sufficient to start exploring contributions from physics beyond the SM in the Higgs boson area. A detailed discussion on the physics reach at HL-LHC is given in Sect. 14.9.

12.5 Top quark physics

Marcel Vos

12.5.1 A brief history of the top quark

The late 1960s and early 1970s established the quark model, as described in Sect. 1. After the discovery of the charm quark [91,92] in 1974, and the bottom quark [3907] in 1977, the hunt for the sixth quark was open. An intensive search at PETRA for a $t\bar{t}$-resonance or a jump in cross section was carried out with the result that the top quark, if it exists, must be bigger than 23 GeV. Higher order predictions of electroweak quantities in the Standard Model, being a gauge theory, depend on all model parameters, in particular on the masses of the yet unknown top quark and the Higgs boson [3908]. At the end of the 1980s sufficiently precise measurements existed to predict the top quark mass. It came as a big surprise that the predictions indicated a value in the range between 100 and 200 GeV [3909,3910]. The new measurements of the Z-shape parameters at LEP and SLC resulted in precise predictions of the top quark mass : 125 ± 40 GeV [3911], 164^{+16+18}_{-17-21} GeV [3912] and 177^{+11+18}_{-11-19} GeV [3913]. One year later, 1995, the two collaborations at the Tevatron, CDF and D0, discovered the top quark at the predicted value. Their measured values are CDF : 177^{+11+18}_{-11-19} GeV [85] and D0 $199^{+19}_{-21} \pm 22$ GeV [3914]. The present best value is 172.69 ± 0.30 GeV [616].

The top quark pair production process – leading-order diagrams are shown in the top row of Fig. 327 – is the dominant process at hadron collider. The two Tevatron experiments could also demonstrate the existence of the electroweak single-top-quark production processes in the t-channel [3915,3916] and s-channel [3917]. Feynman diagrams are shown in the bottom row of Fig. 327. Precise measurements confirmed key SM predictions, such as the forward-backward asymmetry in $t\bar{t}$ production [3918] and the W-boson helicity fractions in top quark decay [3919]. Last but not least, the Tevatron legacy includes a top quark mass combination with a sub-GeV precision [3920].

The Large Hadron Collider [3921] at CERN entered operation in 2010 with pp runs at 7 TeV and 8 TeV. Data taking resumed at 13 TeV in 2015, and the ATLAS and CMS experiments had harvested 140 fb^{-1} by 2018. At the time of writing, in summer 2022, run 3 has just started with pp collisions at 13.6 TeV. The large center-of-mass energy strongly enhances the top quark production cross section. Top quark pair production is primarily from the gluon-initiated diagrams of Fig. 327 in pp collisions at LHC energies, while in the $p\bar{p}$ collisions at the Tevatron, quark–anti-quark production dominated. In combination with the large instantaneous luminosity, the LHC is a genuine "top factory". More than

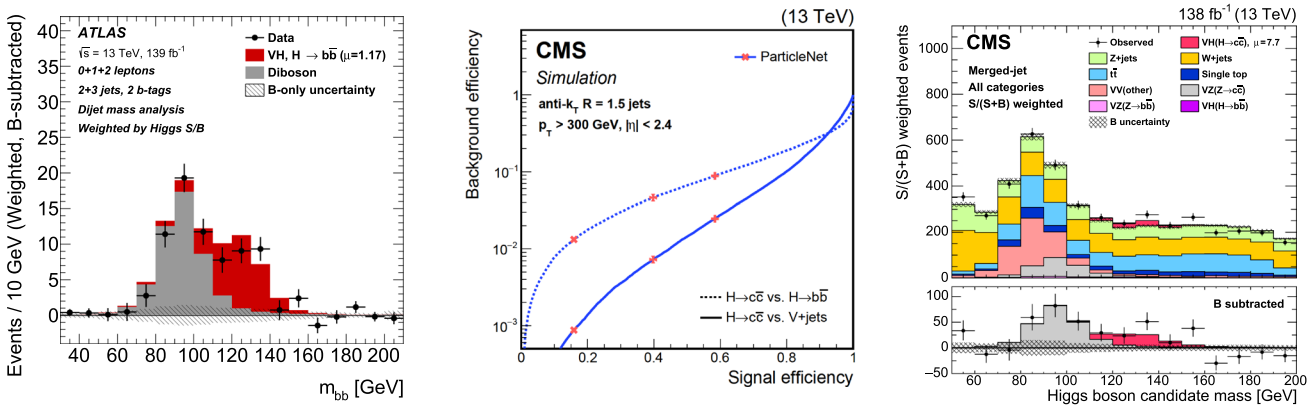


Fig. 326 (left) The distribution of m_{bb} in data after subtraction of all backgrounds except for the WZ and ZZ diboson processes. All the contributions are summed and weighted by their respective S/B ratios, with S being the total fitted signal and B the total fitted background. The expected contribution of the associated WH and ZH production of a SM Higgs boson with $m_H = 125$ GeV is shown scaled by the measured signal strength ($\mu = 1.17$). The size of the combined statistical and systematic uncertainty for the fitted background is indicated by the hatched band [3899]. (center) Performance of CMS algorithm ParticleNet for identifying a $c\bar{c}$ pair for large-radius jets with $p_T > 300$ GeV. The solid (dashed) line shows the efficiency to correctly identify $H \to c\bar{c}$ vs the efficiency of misidentifying quarks or gluons from the V+jets process ($H \to b\bar{b}$). The red crosses represent the three working points used in the large-radius jet analysis [3900]. (right) Invariant mass distribution of the selected $c\bar{c}$ events [3900]. The lower panel shows the data (points) and the fitted $VH(H \to c\bar{c})$ (red) and $VZ(Z \to c\bar{c})$ (grey) distributions after subtracting all other processes. Error bars represent pre-subtraction statistical uncertainties in data, while the gray hatching indicates the total uncertainty in the signal and all background processes

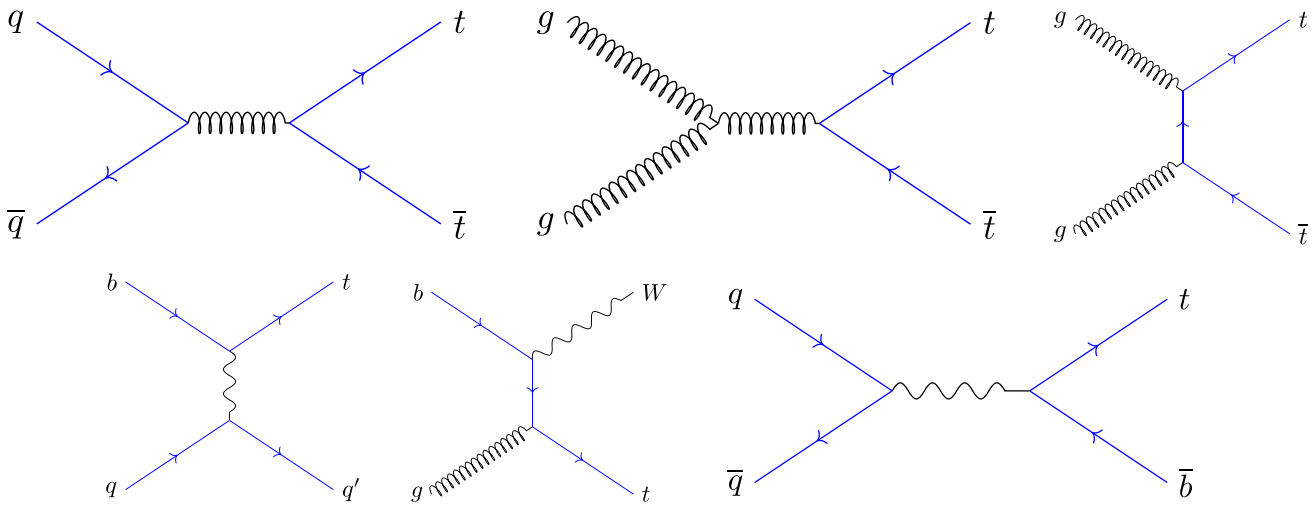


Fig. 327 Leading-order Feynman diagrams for top quark pair production and (top row) and electro-weak single top production (bottom row)

100 million $t\bar{t}$ pairs have been produced in run 1 and 2 and more than 1 billion are expected in future runs. The LHC therefore marks a new era in experimental top-quark physics and dominates the summary in this chapter.

Its properties make the top quark an ideal laboratory for studies of the electro-weak and strong interactions. As the top quark mass of approximately 172 GeV exceeds that of the W-boson, the decay $t \to Wb$ is kinematically allowed and makes up nearly 100% of the branching ratio (with sub-% fractions of top quarks decaying to Wd and Ws in the Standard Model). The subsequent $W^+ \to l^+\nu_l$ and $W^- \to l^-\bar{\nu}_l$ decays of the W-boson yield an isolated charged lepton $l^\pm = e^\pm, \mu^\pm, \tau^\pm$. These are a key signature to trigger and select events with top quarks at hadron colliders. The charge of the lepton furthermore reveals whether the decay corresponds to that of a top-quark or anti-quark, providing an efficient "tag" for asymmetry measurements [3922]. Finally, the charged lepton is an efficient polarimeter that enables studies of top quark polarization [3923], spin correlations [3924,3925] and quantum entanglement [3926]. All these features lead to a rich and varied experimental top quark physics programme.

12.5.2 Precise predictions for top quark physics

The calculability of top quark production is one of the keys to the top quark physics programme at hadron colliders. The large top quark mass regulates perturbative calculations, enabling precise predictions of QCD processes with colored objects in the final state.

The fully differential top quark pair production cross section at hadron colliders is known to NNLO accuracy in the strong coupling [3396,3453,3927]. Electro-weak corrections are available at NLO [3928] and NNLL resummations are available. Predictions of the inclusive $pp \to t\bar{t}$ production rate reach an uncertainty of 4–5%. The uncertainty is dominated by the scale uncertainties, that estimate the impact of higher-order QCD corrections, followed closely by the PDF uncertainties.

While the NNLO calculation of top quark pair production is a major milestone, it remains a considerable challenge to meet the experimental precision that can be achieved at the LHC. The most precise measurements reach an uncertainty a bit over 2%, half that of the predictions. The NNLO QCD corrections have a sizable impact on the shape of differential measurements, in particular on the top PT and related distributions [3453,3927–3929]. Fully differential NNLO fixed-order calculations and Monte Carlo generators are required to provide an adequate description of the data collected by ATLAS [3930–3932] and CMS [3933–3936].

Associated top quark production processes with electro-weak bosons become accessible at the LHC and provide a direct probe of the top quark interactions with the Higgs boson and the neutral electro-weak gauge bosons (see for instance Ref. [3937] and references therein). The $t\bar{t}X$ processes at the LHC are known to NLO accuracy, and uncertainties on the inclusive production rates are well below 10%. The experimental results for these rare processes are improving rapidly and already challenge the precision of the best SM predictions. Resummation to NNLL and NLO electroweak corrections are available [1993,3938] and elements of the NNLO calculations for $t\bar{t}H$ production are known [3939]. A complete NNLO description is required for all $t\bar{t}X$ processes to take full advantage of the HL-LHC programme [3940].

NLO calculations are available for $2 \to N$ processes that include top quark decays and off-shell effects [3941]. These provide sizable corrections for the top quark pair production process and associated production processes.

Predictions at the particle- and detector-level play an important role in measurements of top quark cross sections and properties and in searches for rare processes. State-of-the-art Monte Carlo generators match NLO matrix elements to the parton shower and hadronization models implemented in Pythia8 [3942] or Herwig [3943]. The work horse implementations for the LHC programme during run 1 and run 2 are provided by the Powheg-box [3601,3602,3628], where resonance-aware matching is an important recent addition for top physics [3944], and the MG5_aMCNLO [3385] package, that can include also NLO electroweak corrections [3387]. SHERPA [3579] offers multi-leg generation for top quark pair production and other high-jet-multiplicity processes involving top quarks. The MINNLOps package [3945,3946] provides a Monte Carlo event generator at NNLO accuracy for top quark pair production that can be interfaced to Parton Shower and hadronization programmes.

12.5.3 Precision measurements at hadron colliders

The measurements of top quark production cross sections in the ATLAS experiment are summarized in Fig. 328. The measurements cover four different center-of-mass energies (5, 7, 8 and 13 TeV) and span over five decades in production rate: from $\mathcal{O}$(1 nb) for top quark pair production to $\mathcal{O}$(10 fb) for $t\bar{t}t\bar{t}$ production. The experimental results indicated by the colored markers are compared to the best available Standard Model predictions in grey.

The measurements of the production cross section for the classical top quark production processes have become precision measurements, with the measurement of the inclusive cross section reaching 2.4% precision [3931]. The result is limited by the knowledge of the integrated luminosity delivered by the LHC. Progress in the understanding of the luminosity calibration is expected to reduce this uncertainty to about 1%, but this is likely to remain the limiting factor for the most precise inclusive measurements.

Also electro-weak single top production is characterized precisely in the t-channel and tW associated production channel. With a precision of less than 7% for the t-channel [3947], the Cabibbo–Kobayashi–Maskawa matrix element V_{tb} is determined as: $|f_{LV}V_{tb}| = 1.02 \pm 0.04$, where the uncertainty includes contributions from experiment and predictions and f_{LV} is a form factor, identical to 1 in the SM, that parameterizes the possible presence of anomalous left-handed vector couplings. This result is in good agreement with the determinations from b-physics [300].

The LHC programme has eclipsed the Tevatron measurements in nearly all processes and measurements. However, the Tevatron legacy remains very relevant, as the different initial states ($p\bar{p}$ instead of pp) and center-of-mass energy lead to important complementarities and Tevatron data continue to provide important inputs for global analyses. Several highlights of the Tevatron programme remain unrivalled to this day, as the dominance of $q\bar{q}$-initiated production provides an ideal laboratory for certain measurements. Good examples are the study of $s-$channel single top quark production [3917] and the measurement of the forward-backward asymmetry in $t\bar{t}$ production, that reached a high significance for the SM effect at the Tevatron [3918].

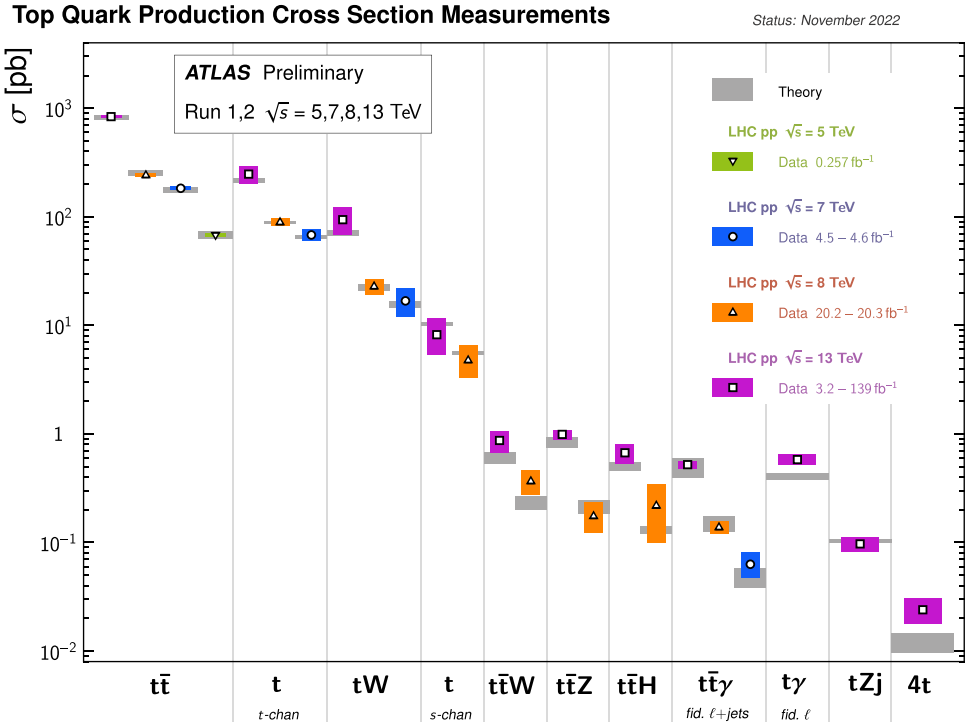


Fig. 328 Measurements of production cross sections for processes with top quarks in the final state by the ATLAS and CMS experiments. The experimental results indicated by the colored markers are compared to the best available Standard Model predictions in grey. Figure courtesy of the ATLAS experiment

12.5.4 Boosting sensitivity

The enormous sample of top quark pairs collected at the LHC enables precise differential cross section measurements. Many measurements extend well into the *boosted regime*, where the top quark transverse momentum significantly exceeds the top quark mass and the collimated hadronic top quark decays are reconstructed as a single large-radius jet.

From the first observation of boosted top quark candidates at the start of the LHC, the study of their production has come a long way. An avalanche of new techniques has been developed [3948], from pile-up mitigation to top tagging algorithms, and the experiments have carefully characterized jet substructure [3949,3950] and the experimental response [3664]. With the large samples of boosted top quarks, these developments have enabled precise measurements of top quark interactions in the most energetic collisions at the LHC. The most recent measurements of top quark pair production in the boosted regime [3951,3952] yield precise bounds on the Wilson coefficients of the $q\bar{q}t\bar{t}$ operators in the Standard Model Effective Field Theory (see Refs. [3953,3954] and references therein), as the energy-growth of their impact boosts the sensitivity of these measurements.

12.5.5 New rare top quark production processes

The right half of Fig. 328 is devoted to the new, rare associated top quark production processes that were observed by the LHC experiments in run 2. Many of these measurements scrutinize aspects of the Standard Model description of the top quark interactions that were not, or not directly, tested at previous facilities. The associated production processes of a top quark pair with a photon [3955,3956] or Z-boson [3957,3958] offer a new, direct probe to the neutral-current interactions of the top quark [3937]. These processes are well-established and differential measurements are available. More recently, the single top production process in association with a Z-boson [3959,3960] and a photon [3961] were observed as well. The observation of the $pp \to t\bar{t}H$ production process [3905,3962] confirms unambiguously that the heaviest SM particle indeed couples to the Higgs boson. The combination of rate measurements in different production and decay channels yields a robust estimate of the top-quark Yukawa coupling [3887,3888]. In LHC run 2 the first evidence was found for four-top-quark-production. Refined analyses of the run 2 data by ATLAS [3963] and CMS [3964] could establish the existence of this very rare and spectacular process with a significance well over 5 σ per experiment by the time of the Moriond conference in 2023. With more data from run 3 of the LHC and improved experimental techniques precise measurements of the production cross section will provide a probe for the four-heavy-quark vertex and an alternative determination of the top quark Yukawa coupling.

Rare top quark production processes provide qualitatively new information on previously unprobed interactions and form a valuable input to fits of the Standard-Model-Effective-Field-Theory parameters to collider data [3937,3953,3954].

12.5.6 New physics searches with top quarks

Beyond Standard Model (BSM) searches in final states with top quarks have pioneered the development of tools for boosted object tagging. Thus prepared, the experiments have been able to take advantage of the LHC centre-of-mass energy to push bounds on new massive states beyond 1 TeV and in many cases into the multi-TeV range. The $t\bar{t}$ resonance searches by ATLAS and CMS indicate that new narrow massive states that decay to top quark pairs or a top and bottom quark cannot have a production cross section times branching ratio greater than 0.1 pb in the mass region from 1.5 to 4–5 TeV. Concrete scenarios such as the bulk RS KK gluon [3965] and W' bosons are excluded for resonance masses up to 4 TeV[3966]. Searches for vector-like quarks decaying to a top quark and Higgs or gauge boson yield lower limits greater than 1 TeV for the mass of the vector-like quarks.

The integrated luminosity is a key for the search for flavor-changing-neutral-current (FCNC) interactions of the top quark. The branching fractions $t \to qX$ decays (with $X = \gamma, Z, g, H$) are suppressed well beyond the experimental sensitivity in the SM, but can be enhanced to $\mathcal{O}(10^{-5})$ in several extensions [3967]. An even larger branching ratio $BR(t \to cH) \sim 10^{-3}$ can be present in certain two-Higgs-doublet models [3968,3969]. The observation of these FCNC interactions would be an unambiguous sign of physics beyond the Standard Model.

Searches have advanced rapidly in sensitivity in run 2 and the exclusion bounds in Fig. 329 are reaching $\mathcal{O}(10^{-4}-10^{-5})$, scratching the surface of the branching ratios predicted in viable models. The inclusion of single top production in association with a Higgs or gauge boson have been important to improve the bounds, in particular for the FCNC vertex with an up-quark.

12.5.7 The top quark mass

The top quark mass is a fundamental parameter of the SM Lagrangian that must be determined experimentally. As any other quark mass, its definition generally depends on the renormalization scheme and the value of the renormalization scale at which it is evaluated. The pole mass scheme is used in Monte Carlo generators and many fixed-order calculations. The $\overline{\text{MS}}$ mass is extracted from the top quark pair production cross section [3970].

Three main classes of measurements of the top quark mass at hadron colliders are discussed below. A selection of results obtained with each approach is presented in Fig. 330.

The first class of measurements extracts the top quark mass from the comparison of top quark pair cross section measurements (corrected to the parton level) to SM predictions at NNLO+NNLL accuracy. The uncertainty of the mass determined from the total cross section a 13 TeV is around 2 GeV.

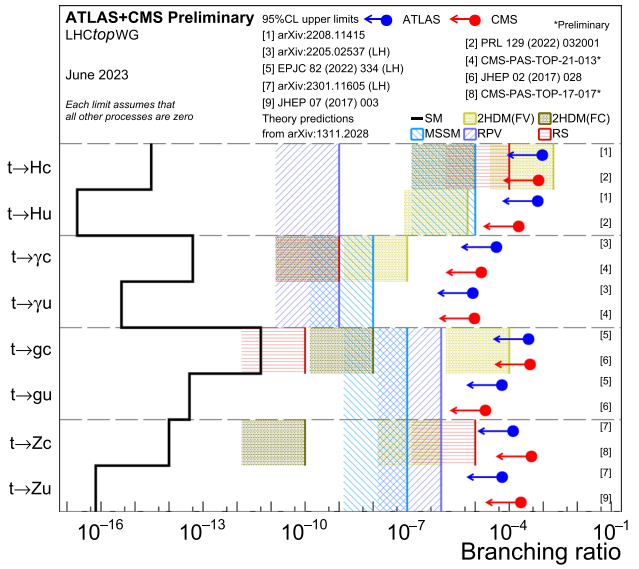


Fig. 329 Leftmost panel: summary of the searches for FCNC interactions of the top quark with the Higgs boson, photon, gluon and Z-boson. 95% confidence limits are derived on the equivalent branching ratio $t \to Xu$ and $t \to Xc$, and in some cases for left-handed and right-handed couplings (left-handed couplings are assumed for the limits collected in the summary plot in those cases). Figure courtesy of the LHC top Working Group. Rightmost panel: selection of top quark mass measurements at the Tevatron and LHC, by category. ATLAS measurements are indicated with blue markers, CMS measurements in red and the Tevatron or combined Tevatron-LHC results in black. The 2014 world average is given by the pink bar, and the indirect determination of the top quark mass from the electroweak fit with the cyan band. Figure prepared by the author based on data collected by the LHC top Working Group

This includes a theoretical uncertainty, estimated by varying the renormalization and factorization scales and propagating uncertainties from the parton distribution functions of the proton. Importantly, recent cross section measurements have a much reduced dependence on the mass assumption in the correction of detector acceptance and efficiency, such that in practice the dependence on the MC mass parameter can be ignored to good precision. There is a broad consensus that this method yields a solid measurement with a rigorous interpretation. Future progress is expected from improving fixed-order calculations and PDFs, and from a reduction of the luminosity uncertainty on the experimental side.

A more precise determination is possible based on measurements of differential cross sections [3971]. These enhance the mass sensitivity in e.g. the threshold region. In the shape analysis of the differential cross section important uncertainties in the absolute cross section and integrated luminosity cancel, leading to a precision of about 1 GeV for the most precise measurements [3972]. The theory uncertainty is accounted for in the same way as in the inclusive measurements and the method retains some flexibility in the choice of the mass scheme. More work is required, however,

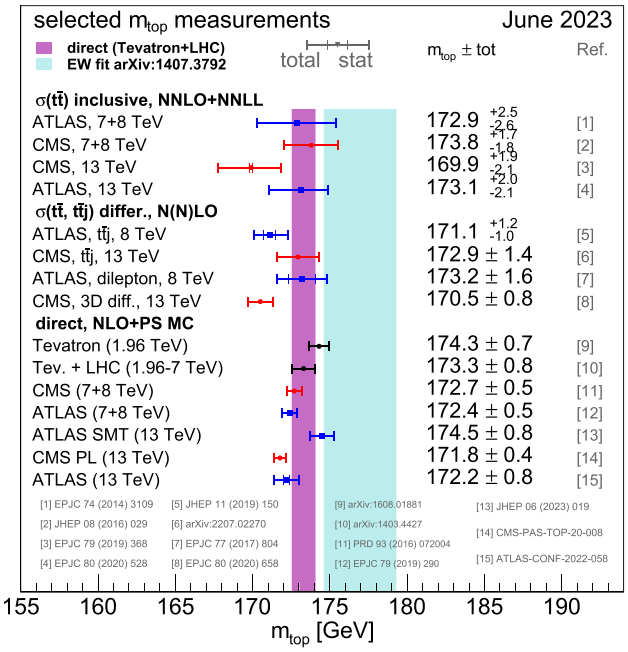


Fig. 330 Selection of top quark mass measurements at the Tevatron and LHC, by category. ATLAS measurements are indicated with blue markers, CMS measurements in red and the Tevatron or combined Tevatron-LHC results in black. The 2014 world average is given by the pink bar, and the indirect determination of the top quark mass from the electroweak fit with the cyan band. Figure prepared by the author based on data collected by the LHC top Working Group

to account for bound-state effects in the threshold region [3973].

The third, and experimentally most precise, approach determines the mass parameter of the Monte Carlo generator that yields the best fit to the observed distribution of top quark decay products. The 2014 combination of Tevatron and LHC run 1 results yields $m_t = 173.3 \pm 0.8$ [3974] and is used as a reference in Fig. 330, indicated by the magenta area. The most precise single measurements by CDF, D0, ATLAS and CMS have since reached an uncertainty of approximately 400–600 MeV.

The results of direct mass measurements are interpreted as the top quark pole mass. An additional uncertainty is assigned to cover the ambiguity in this interpretation. Analytical investigations into the relation between top mass parameter of the Herwig Monte Carlo and the mass in field-theoretical mass scheme observe that the parton shower cut-off in the Monte Carlo generator (typically set to 1 GeV) alters the mass definition in a non-negligible way [3975]. A quantitative relation of the top mass parameter with the pole mass is obtained by comparing particle-level observables in the Monte Carlo generator with first-principle calculations[3976,3977]. Based on these studies, Ref. [3978] proposes a 500 MeV uncertainty in the interpretation of direct mass measurements. Important theory work is ongoing to improve the Monte Carlo templates [3979,3980], including top quark decay at NLO and full off-shell effects.

The analysis of run 2 results is in full swing. The last two points in the rightmost panel of Fig. 330 correspond to two innovative analyses on partial run 2 data. ATLAS published an analysis based on a purely leptonic observable, the invariant mass of the system formed by the prompt lepton from W-decay and the soft muon found in the b-jet, shifting the systematic uncertainties from jet response to fragmentation and B-decay modelling. A recent preliminary result by CMS [3981] based on a profile-likelihood fit reaches an uncertainty below 400 MeV. This result demonstrates the power of the profile-likelihood-fit approach in top quark mass measurements, but also emphasizes the importance of a robust uncertainty model for MC-related uncertainties. A combination of all measurements collected can reach an experimental precision of 300 MeV.

Projections of future improvements are notoriously hard in this area, where a detailed understanding of the limitations of Monte Carlo generators is key. Direct measurements can potentially be improved to an experimental precision below 200 MeV in the remainder of the LHC programme [3940], while cross-section-based mass extractions can reach a total uncertainty below 500 MeV [3982].

12.5.8 Top quark data in global analyses

The top quark and the results in the top quark sector presented in previous sections are inevitably part of the "global" interpretations of collider data. In this section, three examples are briefly discussed.

Most recent analyses of the parton density functions [3810] consider also data on top quark production,[114] that provide an important constraint on the high-x gluon content of the proton. The ATLAS PDF fit [3825] includes differential measurements of $t\bar{t}$ production, using NNLO predictions with a fixed top quark mass. The CMS experiment has performed a global analysis [3983] with partial run 2 data, where the top quark pole mass is floated, as well as the PDFs and the strong coupling. The analysis is based on NLO predictions for the top quark pair production process and threshold corrections remain to be included.

Radiative effects connect electroweak precision observables at the Z-pole to precise measurements of α_s and the W-boson, Higgs boson and top quark masses. The electroweak fit tests the relations among these parameters predicted by the SM and forms a stringent check of the internal consistency of the theory. In global electro-weak fits [3726,3984]

[114] To avoid absorbing potential BSM contributions to top quark production in the PDFs, care is taken to select differential measurements that are less likely affected. PDF results without top data are available in at least one PDF set and allow for important cross-checks.

before 2022 there is a very mild tension between direct top quark mass measurements and the mass inferred from precision electroweak data (e.g. the magenta and cyan bands in Fig. 330). Inclusion of the 2022 CDF W-mass measurement [3985] leads to considerable tension in the fit [3986,3987]. The main avenue towards a tighter test should focus on an understanding and improvement of the W-boson mass measurements, but eventually also the precision of measurements of other parameters, among which the top quark mass, must be improved.

The legacy data from collider experiments are collected in the framework of the Standard Model Effective Field Theory. Global fits of the top quark sector have been performed by several groups [3988,3989]. State-of-the-art fits include a combined analysis of Higgs, electroweak and top data [3953, 3954], showing an interesting interplay between the top and Higgs sectors through the effect of operators involving top quarks in loops, for instance in $gg \to H$ and $H \to \gamma\gamma$, through the box diagram contribution to di-Higgs production and through the associated production of top quarks and a Higgs boson.

12.5.9 Outlook

A vibrant top quark physics programme was started at the Tevatron and has culminated in a broad and rich programme at the LHC. Direct searches in final states with top quarks explore the multi-TeV regime looking for signs of new resonances and exotic phenomena. Precise measurements of the classical top quark production processes and many new rare processes involving top quarks and the Higgs and gauge bosons form a powerful set of constraints on top quark couplings. The top quark mass is known to a precision of less than 0.5%.

The upcoming runs of the LHC and its high-luminosity upgrade are expected to improve considerably on current run 2 results [3940], increasing the precision, pushing differential measurements further into the high-energy regime, and probing ever more rare processes. Projections are particularly encouraging for rare associated production processes, where statistical limitations remain important and theory predictions are currently only available at NLO accuracy.

Top quark physics is an important consideration also for a new facility in high energy physics beyond the HL-LHC. A new electron–positron collider is identified as the highest-priority installation in the European, American and Asian road maps for particle physics. All projects for such a Higgs/EW/top factory envisage operation at and above the $t\bar{t}$ production threshold. This enables scrutiny of the top quark in e^+e^- collisions and provides precision measurements of the top quark mass, with $\mathcal{O}(50\,\text{MeV})$ precision [3990], and electroweak couplings, that improve by an order of magnitude compared to the HL-LHC projections [3991].

A new pp collider at the energy frontier can potentially push the discovery reach for massive particles by a further order of magnitude. It can also unlock ultra-rare SM processes, such as six-top-production and $t\bar{t}HH$. Quantitative projections remain to be made, as well as more detailed studies of top quark reconstruction in this challenging environment. Also the top quark physics of a multi-TeV lepton collider, be it the CLIC high-energy stage, a muon collider or a novel installation based on plasma-wakefield acceleration, remains to be explored in detail. High-energy lepton collisions, well above 1 TeV, offer the possibility to constrain four-fermion operators with two light particles and two top quark to unprecedented precision [3992], and provide an exquisite precision probe for new physics [3993].

12.6 Soft QCD and elastic scattering

Per Grafstrom

12.6.1 Introduction

Soft QCD has become a term covering many different topics. Elastic scattering and diffraction are central topics associated with soft QCD but in addition there is a long list of different areas associated with the term Soft QCD e.g. particle correlations, multiple parton interactions, particle densities, the underlying event. The list is not inclusive and could be made longer. It covers an enormous amount of different processes and concepts and just the elastic and diffractive part represents by itself more than 30–40% of the total cross section (σ_{tot}) at high energy hadron colliders. What basically unifies all those different processes is a large distance scale or equivalently a relative small momentum transferred in the reaction.

Another way of expressing the same criteria is to say that "Soft QCD" deals with processes for which the perturbative approach of QCD is not applicable due to the size of the strong coupling at small momentum transfer. This is a direct consequence of the running of the strong coupling α_s. In this low momentum transfer regime more phenomenological approaches have to be applied. However, while using phenomenological methods, the aim is always to try to provide a smooth transition to harder and thus perturbative QCD processes.

Soft QCD processes have an interest in their own right representing a particular challenging part of QCD. However, there are a number of other reasons that motivate trying to achieve a better understanding of Soft QCD processes. The Soft QCD processes represent often the most significant correction in searches for new physics. The so called underlying event stands for everything which is produced in a pp collision except for the hard scatterer. The better one understands the underlying event the easier it is to extract signals for new

physics. There is also the phenomena of pile-up at modern colliders. In order to push the instantaneous luminosity to such high values that very rare processes can be detected, the colliders have to be operated with such high bunch population that about fifty separate interactions occur during one and the same bunch crossing. Most of those interactions are soft and produce what is called "pile-up" background in the different detectors and this background has to be separated from the signal.

Understanding of Soft QCD processes are also important in the context of cosmic rays. Monte-Carlo event generators used for simulation of the forward cascades in air showers have to be tuned in order to extract the essential physics parameters in cosmic ray studies.

Here we will start with a discussion of elastic scattering and the total cross section in proton–proton collisions, and in a second part some other typical Soft QCD topics will be addressed. It will be impossible to cover all the topics nowadays associated with Soft QCD in this short review and we have to make a biased selection. A very good and more extensive summary of Soft QCD is given in the article "High Energy Soft QCD and Diffraction" written by V.A. Khoze, M. Ryskin and M. Taševský published in PDG [616].

12.6.2 *Elastic proton–proton scattering and the total cross section*

First principles

Elastic scattering is the simplest process possible at a hadron–hadron collider. Two incoming protons scattering at the Interaction Point (IP) giving two outgoing protons and nothing more. It is the most simple process possible involving strongly interacting particles but still it can not be described directly by QCD. However there are first principles or fundamental concepts which are relevant for elastic scattering and the total cross section. Those principles have to be fulfilled by any theory of strong interactions and must obviously also be fulfilled by QCD. Principles like unitarity, crossing symmetry and analyticity of the elastic scattering amplitude are of importance. Those principles connect elastic scattering with the total cross section in different manners.

The most straight forward is the optical theorem that connects the total cross section with the imaginary part of the scattering amplitude in the forward direction. The high energy form of the optical theorem can be written:

$$\sigma_{tot} = \frac{ImF_{el}(t=0)}{s}, \tag{12.12}$$

where t is the four momentum squared which at high energies can be written as $-t = (p\theta)^2$ with p being the momentum and θ the scattering angle. The Mandelstam variable s represents the centre of mass energy squared. The optical theorem is based upon probability conservation in the scattering process and is easily derived using Quantum Mechanics.

The optical theorem has been used to experimentally determine the total cross section via measurement of the differential elastic cross section from ISR times to LHC today. From the optical theorem one derives the formula

$$\sigma_{tot}^2 = \frac{16\pi}{1+\rho^2}\frac{d\sigma_{el}}{dt}(t=0), \tag{12.13}$$

where $\frac{d\sigma_{el}}{dt}(t=0)$ is the elastic differential cross section extrapolated to $t=0$ and ρ is the ratio of the real to imaginary part of the elastic scattering amplitude in the forward direction i.e.

$$\rho = \frac{ReF_{el}(t=0)}{ImF_{el}(t=0)}. \tag{12.14}$$

However the optical theorem is not the only connection between σ_{tot} and elastic scattering. Using the concepts of analyticity and crossing symmetry, dispersion relations for elastic scattering can be derived. Dispersion relations connect the ρ-parameter at a certain energy to the energy evolution of σ_{tot} both below and above this energy and are a very powerful tool which play a crucial role in the understanding of elastic scattering. Dispersion relations thus imply that the ρ-value at a certain energy is sensitive to the energy evolution of σ_{tot} beyond the energy at which ρ is measured. The ρ-value is accessible experimentally and can be measured by measuring elastic scattering as such small angles where the Coulomb amplitude starts to be significant. The Coulomb amplitude is proportional to $1/t$ and dominates in the very forward direction. The interference between the Coulomb amplitude and the strong amplitude permits a measurement of ρ. Using the measurement of ρ and dispersion relations one can make predictions of σ_{tot} to an energy of the order 10 times higher than the energy at which ρ has been measured. This has been done several times in the past [3994,3995].

The Froissart–Martin bound [3996,3997] is another example of an important consequence derived from first principles. Based upon axiomatic quantum field theory it was shown that σ_{tot} can not grew faster than

$$\sigma_{tot} < \frac{\pi}{m_\pi^2} ln^2 s. \tag{12.15}$$

As will be discussed in the paragraph "The total cross section" below, this bound is not very constraining given the energy scales available today.

The Regge approach and QCD

The principles discussed above generate bounds and relations between important entities but do not lead to a concrete proposal for the scattering amplitude. For this, one still has to rely upon phenomenological approaches. The phenomenology of Regge theory dominated the description of high energy scattering process in pre-QCD times (see

for instance [3998,3999] and references therein). With the advent of QCD as the theory for strong interaction in the 70th, Regge theory started to loose its role. The obvious wish was of course to try to derive Regge theory from QCD. Due to the non-perturbative character of low p_T reactions this turned out to be extremely difficult and still today Regge concepts are the basis of the phenomenology used to describe soft processes. However, whenever possible one tries to connect to QCD in a smooth way.

The key concept in Regge theory is singularities of the amplitude in the complex angular moment plane, the so called j-plane. The most straightforward singularity is a simple pole. Using this concept for a given scattering process has as consequence that the scattering amplitude in the t-channel can be be calculated using an exchange of so called Regge-trajectories which replaces a single particle exchange. A Regge trajectory composes of many particles with the same quantum numbers except for the spin. The particles are organized in increasing spin with increasing mass on the trajectories. A trajectory is represented by the function $\alpha(t)$ where $\alpha(t)$ is the pole position in the j-plane and is usually parameterized as a linear function of t :

$$\alpha(t) = \alpha(0) + \alpha' t. \tag{12.16}$$

The exchange of a Regge trajectory or a Reggeon leads to a power-like growth of the amplitude with s and an exponent $\alpha(t)$ i.e.

$$A(s,t) \propto s^{\alpha(t)}. \tag{12.17}$$

Using the optical theorem one then gets for the corresponding cross section

$$\sigma \propto s^{\alpha(0)-1}. \tag{12.18}$$

The contribution of a given Regge trajectory factorizes in general, i.e. the amplitude is a product of two factors depending only on the coupling of the exchanged object to the scattered particles at each vertex.

At energies around 20–50 GeV, corresponding to the ISR and at energies below, several different leading Regge trajectories contribute to the amplitude. Experimentally it turns out that the leading trajectories in pp scattering have an intercept $\alpha(0) \approx 0.5$ (see section 51 in [4000]). Thus the corresponding contributions all vanish with increasing energy in an inverse power law according to Eq. 12.18. At higher energies only the so called Pomeron trajectory survives. The Pomeron carries the quantum numbers of vacuum with $CP = ++$ and was proposed in the 1960s to explain the asymptotic behaviour of the total cross section as will be discussed in the following paragraph. The Pomeron is a good example how Regge theory connects to QCD. The Pomeron has now been identified as a two gluon state in QCD (see e.g. references in [4001]) and some of the properties of the Pomeron has been derived in QCD.This will be discussed in Sect. 12.6.4. QCD also predicts the possible existence of a three gluon state with $CP = --$. Such a state corresponds to a trajectory proposed in Regge theory in the 1970s [4002], the so called Odderon. The Pomeron and the Odderon will be discussed more in detail later.

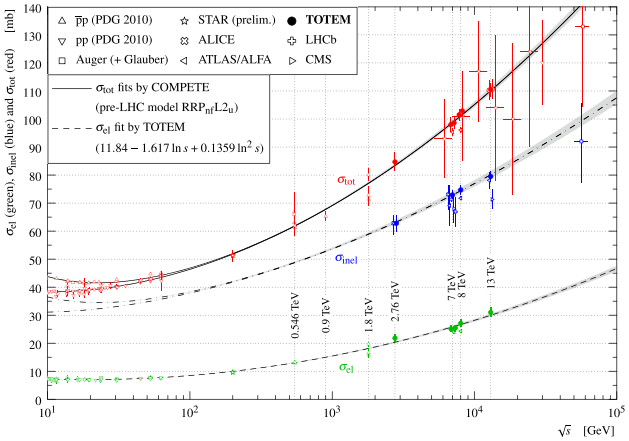


Fig. 331 The total cross section for pp and $\bar{p}p$ as a function of the center-of-mass energy. In the figure is also shown the energy dependence of the elastic and inelastic cross sections. Figure is taken from Ref. [4003]. More details about the figure can be found in [4003]

The total cross section

In Fig. 331 all data are shown of the total pp and $\bar{p}p$-cross section from the ISR to LHC. Low energy proton data from fixed target experiments are also shown in the figure. The total cross section starts to rise at ISR. The rise of σ_{tot} was not at all expected. Still today this rise can not be derived from first principles and is not satisfactory solved within QCD.

At the time of the discovery of the surprising rise of the total cross section with increasing center-of-mass energy, Regge theory was the relevant theory for strong interaction. The Pomeron was thought to dominate the high energy behaviour of σ_{tot} and the s-dependence was thus given by (see Eq. 12.18)

$$\sigma_{tot} \propto s^{\alpha_P(0)-1}, \tag{12.19}$$

where $\alpha_P(0)$ is the intercept of the Pomeron trajectory. The intercept was believed to be exactly 1 thus giving a constant total cross section asymptotically. Later when it was discovered that σ_{tot} was starting to rise an intercept $\alpha_P(0)$ just above 1 was introduced. Taking into account the data from the SPS collider and the Tevatron, σ_{tot} was parameterized in terms of an "effective" Pomeron trajectory $\alpha(t) = 1 + \Delta + \alpha' t \approx 1 + 0.08 + 0.25t$ where t is given in GeV^2 [1100]. However it was always clear that at some very high energy such a power growth in s will violate unitarity and the Froissart bound (see Eq. 12.15). This problem is addressed by also considering cuts in the j-plane in addition to the simple poles. The cuts describe multi-pomeron exchanges and it turns out that

those multi-pomeron exchanges tame the growth of the total cross section and thus restore unitarity.

Actually the possibility of a rising total cross section had been outlined already by Heisenberg in 1952 [4004]. He used a very simple argument based upon the range of the strong interaction and the pion mass to indicate a possibility of a $ln^2(s)$ rise of σ_{tot}. This argument had fallen into oblivion in the mid's of the seventies. Now it has turned out at each new collider energy that σ_{tot} essentially rises as $ln^2(s)$. The full line drawn in Fig. 331 represents one of many $ln^2(s)$ fit to the data. In this case it is one of the COMPETE parametrisations [4005]. In Regge theory such a $ln^2(s)$behaviour can only appear if the the singularity in the j-plane is a pole of order 3 i.e. a triple pole.

Does the $ln^2(s)$ rise mean that the Froissart–Martin bound mentioned in the previous paragraph is saturated? Actually we are far away from a saturation today. The coefficient in front of $ln^2(s)$ term in the Froissart–Martin bound is 60 mb and typically $ln^2(s)$ fits to the data give coefficients O(0.1 mb). Thus the bound is far away from being saturated at LHC energy.

The rise of σ_{tot} as $ln^2(s)$ cannot be derived from QCD today. However it is interesting to note that there have been some attempts in lattice QCD with indications of possible asymptotic $ln^2(s)$ behaviour [4006]. There has also been attempts to generate a $ln^2(s)$ behaviour using gluon saturation in color Glass Saturation models [4007]. This is a good example of how perturbative and non-perturbative physics meet giving an interesting result.

Elastic scattering

Figure 332 shows a couple of examples of the differential elastic cross section and its t-dependence at different energies at the LHC. The measurements have been done by the TOTEM collaboration [4008]. As mentioned in the previous paragraph the Pomeron trajectory dominates at energies of the LHC. In terms of QCD this means a dominance of two gluon exchange. The gross features of the t-dependence of differential elastic cross section at high energies can be described in terms of the Pomeron or a two gluon exchange. The cross section falls close to exponential in the forward direction. This means that the Pomeron–proton coupling has an exponential fall off. There are small deviations from the exponential that are not completely understood but might at least partly be due to multi-pomeron exchanges.

The exponential fall-off parameter, often called the t-slope B, also has an energy dependence. The energy dependence of B is plotted in Fig. 333. The straight line corresponds to a linear dependence in $ln(s)$

$$B = B_0 + 2\alpha'_P \ln(s). \tag{12.20}$$

This linear dependence in $ln(s)$ is a direct consequence of the exchange of a Pomeron in the t-channel. The α'_P parameter represents the slope of the Pomeron trajectory $\alpha_P(t) = \alpha_P(0) + \alpha'_P t$.

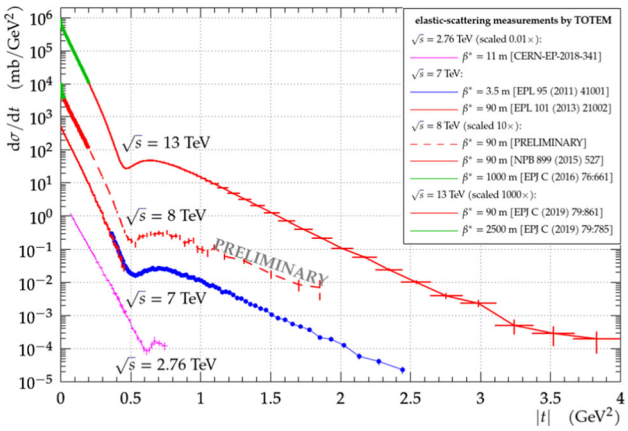


Fig. 332 The differential elastic cross section as a function of the four momentum transfer t for different energies at the LHC as measured by the TOTEM collaboration. Figure is taken from Ref. [4009]

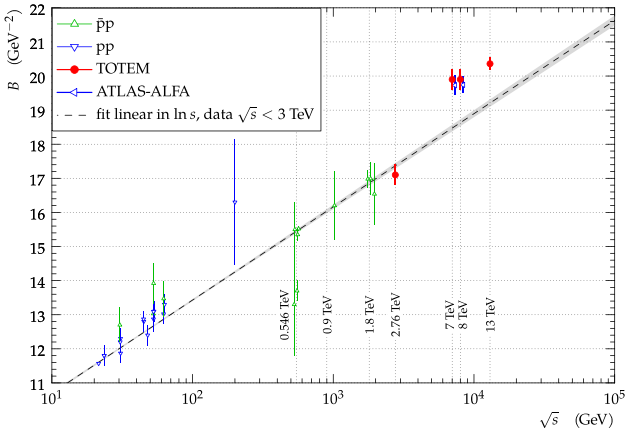


Fig. 333 Measurements of the slope-parameter B as a function of the center-of-mass energy $\sqrt{s}$ for pp and $\bar{p}p$ scattering . The straight line represents a linear fit in $\ln(s)$ of of the data below $\sqrt{s} = 3$ TeV. Figure taken from Ref. [4010]

As can be seen in the figure this linear relation works well for energies below LHC. However at the LHC the increase with s starts to accelerate. Also this can be explained in terms of multi-pomeron exchanges mentioned above [4011].

In Fig. 332 one can also see that after the exponential decrease, the differential cross section exhibits a dip that moves towards smaller t-values when the energy increases. In the Pomeron language this is interpreted as an interference between one pomeron exchange amplitude and multi-pomeron exchange amplitudes making essentially the imaginary part of the total amplitude disappear at the dip. This mechanism generates a dip which correctly moves towards smaller t-values with energy.

At high values of t, beyond the dip, the cross section decreases further in a smooth way. Here one moves away

from the non-perturbative regime and instead one might see signs of perturbative QCD. The triple gluon exchange proposed in Ref. [4012] could be a manifestation of this.

The Odderon

As seen in the previous paragraphs the Pomeron plays an essential role in the description of elastic scattering and the total cross section. The situation is very different concerning the Odderon. The Odderon is the $CP = --$ counter-partner of the Pomeron and contributes with a different sign to the amplitude for pp-scattering relative to $\bar{p}p$-scattering. The Odderon is both controversial and non-controversial. It is non-controversial in the sense that no one really doubts its existence. It is a firm prediction of QCD and represents a three gluon state in contrast to the two gluon state of the Pomeron. What is somewhat controversial is the size of its coupling and its importance in the elastic amplitude. To what extent the Odderon really has manifested itself in the available experimental data is debatable (see e.g. Ref. [4013]) though the authors of Ref. [4014] claim a discovery.

Experimentally there are two different signals that have been evoked as a sign of an Odderon. The most convincing is probably the difference between $\bar{p}p$-scattering and pp-scattering observed in the dip region of elastic scattering. The $\bar{p}p$ data from the D0 experiment at the Tevatron at 1.96 TeV have been compared to the pp data at 2.76 TeV from the TOTEM experiment at the LHC [4014]. The dip is supposed to be filled partly by the real part of the Odderon amplitude having a different sign for pp and $\bar{p}p$-scattering. The two distributions are shown in Fig. 334. Ideally the comparison pp and $\bar{p}p$ should be done at the same energy. However, the authors have taken great care to compare the D0 measurement with TOTEM data extrapolated to the 1.96 TeV of the Tevatron. They find a 3.4σ difference between the two distributions in Fig. 334.

The second possible experimental manifestation of the Odderon is a measurement of the TOTEM experiment which has measured the ρ parameter at 13 TeV to be $\rho = 0.09 \pm 0.01$ [4003]. This result is in contradiction to dispersion relation calculations assuming that the standard $ln^2(s)$ behaviour of σ_{tot} continues beyond LHC and assuming that the elastic amplitude only contains the Pomeron contribution. Those calculations give $\rho = 0.13-0.14$ (see Fig. 335 and Ref. [4005]) thus significantly higher than the TOTEM result.

The TOTEM result could therefore be an indication that σ_{tot} starts to grow somewhat slower beyond the LHC energies. However an alternative explanation might well be that the low ρ value is produced by an Odderon effect. An Odderon contribution to the amplitude can modify the dispersion relation calculation in a way to give a better agreement with the data. The effect depends on the size of the Odderon contribution at a certain energy. The so called maximal Odderon

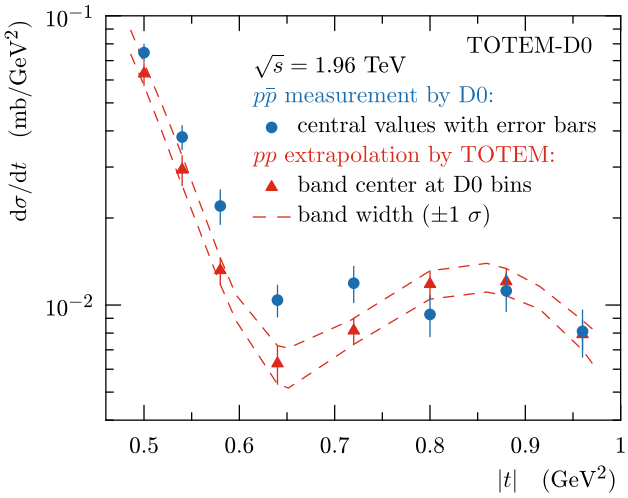


Fig. 334 Comparison between the D0 $\bar{p}p$ measurement at $\sqrt{s} = 1.96$ TeV and the extrapolated TOTEM pp cross section rescaled to match the optical point of the D0 measurement. The dashed lines show the 1σ uncertainty band. Figure is taken from Ref. [4014] where more details are given

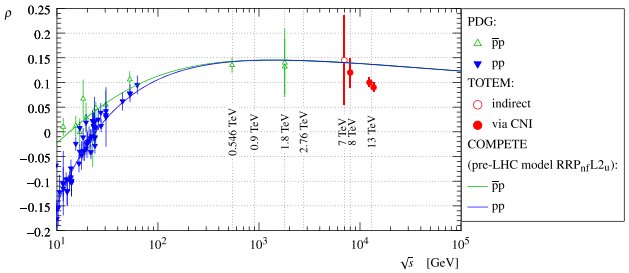


Fig. 335 Dependence of the ρ-parameter on center-of-mass energy. The pp(blue) and $\bar{p}p$(green) curves are taken from Ref. [4005]. Figure is taken from Ref. [4003]

[4015] is one example that actually produces an effect agreeing with the TOTEM data.

To summarise: the measurement of ρ at 13 TeV may be an indication of the Odderon but the fact that an alternative explanation exists means that this signal can not be taken as a hard proof of the Odderon.

12.6.3 Diffraction

In this article we have separated the discussion of elastic scattering and diffraction but actually elastic scattering is the dominant diffractive process. There is no unique definition of diffraction, neither theoretically nor experimentally. A key concept when talking about diffraction is rapidity gaps. For elastic scattering the size of the rapidity gap (a rapidity[115] region void of particles) between the two outgoing protons is

[115] When dealing with a particle whose mass is negligible compared with its energy, the pseudorapidity = $-\ln(\tan(\theta/2))$ is a good approximation to the rapidity. Here θ is the polar angle of the particle. In this

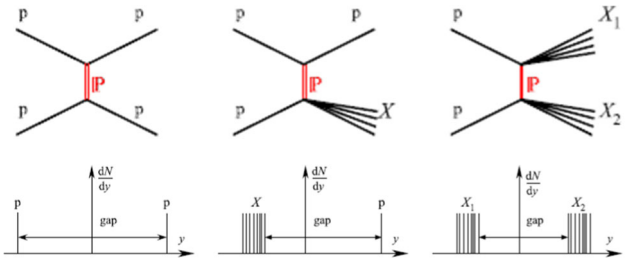


Fig. 336 Feynman diagrams for different diffractive topologies. *IP* stands for Pomeron and *p* for proton while *X* represents the diffractive systems. Below each diagram is also shown the corresponding rapidity distribution of the outgoing particles. Figure taken from Ref. [4016]

at its kinematical limit. In general a diffractive event is characterized by a rapidity gap which is significant larger than possible fluctuations in the hadronization process.Typical this means rapidity gaps bigger than 4–5 units of rapidity at LHC energies. Depending on the topology of the rapidity gaps one talks about different types of diffractive events. This is illustrated in Fig. 336 where the topologies, elastic, single dissociation and double dissociation are shown. Below the Feynman diagrams are also illustrated schematically the corresponding rapidity distributions of the outgoing particles.

All these topologies are characterized by the exchange of the Pomeron in the t-channel or in other words an exchange of a a color singlet state of two gluons.

It is not always possible to map the experimental data directly to the different topologies seen in Fig. 336. In general it is difficult to measure diffraction at high energy colliders, especially diffractive system with a low mass. For low mass systems a large fraction of the diffractively produced particles are emitted in the very forward direction and lost in the beam pipe.

Experimentally there are two ways to select diffractive events. Either to look for rapidity gaps or use so called proton tagging. Proton tagging implies that one or both of the two intact protons actually are detected and measured. This requires small and sophisticated detectors situated as close as possible to the beam line. In practice, the detectors have to be placed at distances of a mm or smaller from the beam and thus the vessels containing the detectors have to be integrated in the beampipe. This technique, using so called Roman Pots, was introduced by the CERN–Rome group at the ISR half a century ago and is still used as the main technique to approach the beam [3994].

Here we will limit ourselves to discuss the simplest topology in Fig. 336, i.e. single diffraction. As an example we show in Fig. 337 the distribution of the experimental gap size $\Delta\eta_f$ measured by ATLAS at 7 TeV for particles with $p_T > 200$ MeV[4017]. The true gap size is $\Delta\eta = \Delta\eta_f + 4$ in

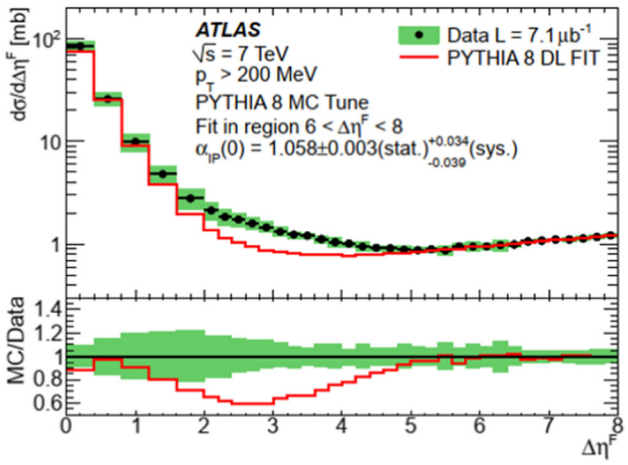


Fig. 337 Inelastic cross section differential in the experimental gap size $\Delta\eta_f$ measured by ATLAS at 7 TeV for particles with $p_T > 200$ MeV. Figure taken from Ref. [4017]. See text in Ref. [4017] for further details

this example. The size of the rapidity gap is directly related to the mass M_x of the diffracted system.

$$\Delta\eta \backsimeq -ln(\xi_x) \tag{12.21}$$

with

$$\xi_x = \frac{M_x^2}{s} \tag{12.22}$$

The larger the rapidity gap is the smaller is the produced mass.

In Fig. 337 one can clearly see the difference between non diffractive and diffractive events. At small gap sizes $\Delta\eta_f < 2$, non diffractive events dominate and the expected exponential decrease of the cross section with increasing gap size which characterize the fluctuations of the hadronisation is the dominant feature. On the other hand for gap sizes $\Delta\eta_f > 3$ there is a rather flat plateau, which corresponds mainly to single diffractive processes. The largest rapidity gap size bin at the end of the plateau in Fig. 337 corresponds to diffractive masses larger than about 15 GeV and thus the plateau corresponds to masses above and equal 15 GeV. The cross section on the plateau is roughly 1 mb per unit of rapidity gap size. In the Regge theory, such type of high mass diffraction is characterized by a triple Pomeron coupling which actually predicts such a plateau. In perturbative QCD a triple Pomeron coupling of the same order of magnitude is found [4018].

Measurements at masses much lower than this are very difficult at high center of mass energies. There exist measurements down to masses of 3–4 GeV but they are scarce and often contradictory. Moreover also theoretically the estimates of low mass diffraction are notorious difficult and uncertain. Actually the uncertainties related to low mass diffraction constitute the largest uncertainty of the total cross section mea-

article we do not make the distinction between rapidity and pseudorapidity.

surements by the TOTEM experiment using the so called luminosity independent method which requires an estimate of the total inelastic cross section including low mass diffraction (see e.g. Ref. [4010]).

12.6.4 "Soft" and "hard" diffraction

This article deals with "soft" diffraction but this is of course a somewhat arbitrary classification and in this section the concept of "hard" diffraction will also briefly be touched upon. A part of diffractive events has a hard scale present.The hard scale is often given by a diffractively produced heavy system such as for example dijets, W or Z bosons, or heavy quarks. With such a hard scale present, perturbative QCD is applicable. There is actually no sharp distinction between what is called "soft" diffraction and "hard" diffraction but rather a smooth transition between the two. Often the perturbative approach is extended into the soft domain in a gradual manner using a unified framework. This has led to the concept of a "soft" Pomeron and a separate and different "hard" QCD Pomeron. This distinction between different Pomerons is very likely an oversimplification of a more complex situation.

The data seem to indicate a "hard" Pomeron with the intercept $\alpha(0) \backsim 1.3-1.5$ with a small slope α' in contrast to the "effective" Pomeron which is relevant for elastic scattering and the total cross section with an intercept of $\alpha(0) = 1.08$ and a slope $\alpha' = 0.25$ as mentioned in Sect. 12.6.2. This means that when a hard scale is involved the energy dependence is steeper relative to soft diffraction (see Eq. 12.18). Taking diffractive vector meson production at HERA as an example: the energy dependence of the γ^*p cross section for J/ψ production corresponds to an intercept $\alpha(0) \backsim 1.4$ (see Refs. [4019,4020]). Such an intercept of the hard Pomeron, represented by a two gluon state, agrees with what has been calculated in perturbative QCD by re-summation of the leading logarithms. The small slope α' of the "hard" Pomeron is also reproduced in perturbative QCD calculations [4021].

Hard diffraction has extensively been studied at HERA in $\gamma^* p$ processes and the results have been interpreted in terms of diffractive Parton Distribution Functions of the Pomeron and a Pomeron flux factor based upon Regge theory [3629]. It was shown within QCD that factorization is valid for diffractive hard scattering in $\gamma^* p$ processes [4022]. However, using the same formalism and using the DPDF's determined at HERA from $\gamma^* p$ processes applied to $\bar{p}p$ processes at the Tevatron gives about an order of magnitude too high cross section for QCD jet production (see e.g. [4023]). At the Tevatron the process is completely hadronic, and the reduction of the cross section is thought to be due to the fact that in 90% of the cases the rapidity gap is filled or partially filled by hadron remnants which are not present in $\gamma^* p$ processes. In this case, the factorisation using diffractive DPDF's suggested by the HERA data breaks down.

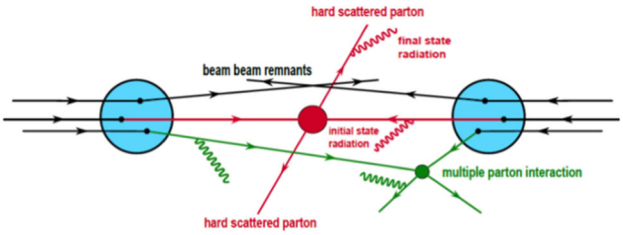


Fig. 338 Schematic drawing illustrating a typical Underlying Event. Figure taken from ATLAS slides ATL-PHYS-SLIDE-2013-330

12.6.5 The underlying event

The underlying event is not to be confused with Minimum Bias Events. As the name indicates, minimum bias events are events collected with as little bias as possible. The concept of Underlying Event (UE) is different. Here one refers to events that contain a hard parton–parton interaction and the term underlying event refers to all the activity that accompanies the hard scatter but is not a part of it. The Underlying Event has several different components. There are contributions from initial and final-state radiation but also particles from the proton break-up so called beam–beam remnants contribute. An important part of the Underlying Events consists of Multiple Parton Interactions (MPI) i.e. one or more soft interactions together with a hard interaction within the same pp interaction. In Fig. 338 a typical UE is shown in schematized way.

A good description of the UE is needed to extract the relevant signals from the hard scatter and rely upon MC generators which are based upon different phenomenological approaches. Many input parameters in the MC generators parameters have to be tuned with data. To get information relevant for the UE the event is often divided into different regions of the phase space as indicated in Fig. 339. Normally a "transverse" region is defined relative to the azimuthal angle of the leading p_T particle. This region is then taken as the reference region for the underlying event. In Fig. 340 is shown an example of the mean charged-particle multiplicity as a function of the leading p_T for the different regions around the leading particle [4024]. All regions exhibit a fast rise at low p_T up to a p_T of about 5 GeV. Here there are no real "hard" processes present. However at higher p_T, hard processes start to dominate and the transverse region which is decoupled from the hard scatter reach a plateau.

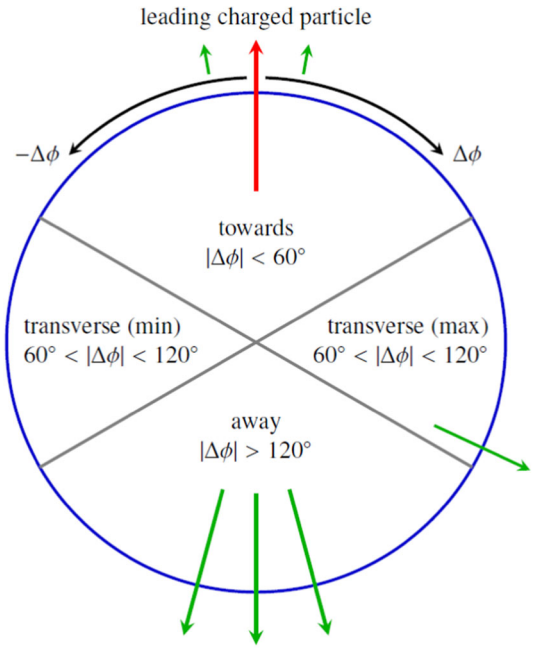


Fig. 339 Definition of regions in the azimuthal angle with respect to the leading (highest-p_T) charged particle, with arrows representing particles associated with the hard scattering process and the leading charged particle highlighted in red. Conceptually, the presence of a hard-scatter particle on the right-hand side of the transverse region, increasing its Σp_T, typically leads to that side being identified as the "trans-max" and hence the left-hand side as the "trans-min", with maximum sensitivity to the UE. Figure taken from Ref. [4024]

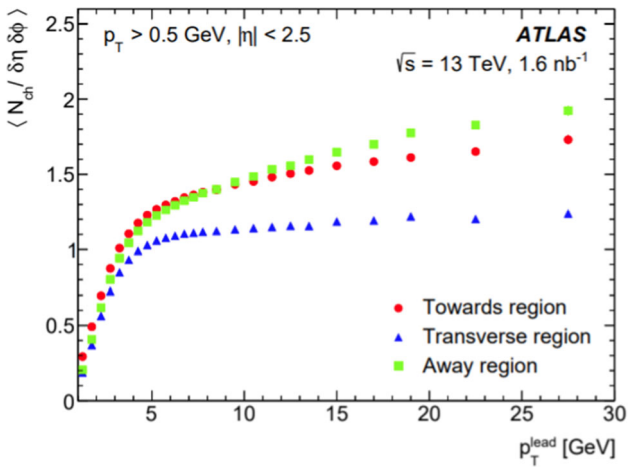


Fig. 340 Mean $\eta - \phi$ densities of charged-particle multiplicities as a function of the transverse momentum of the leading charged particle in the transverse, towards, and away azimuthal regions. The error bars, which are mostly hidden by the data markers, represent combined statistical and systematic uncertainty. Figure taken from Ref. [4024]

12.6.6 Charged particle density

The charged particle density as a function of rapidity is an important observable in pp collisions. The measurement covering the largest rapidity interval has been done by a combination of CMS and TOTEM at the LHC [4025,4026]. The result of their measurement is shown in Fig. 341.

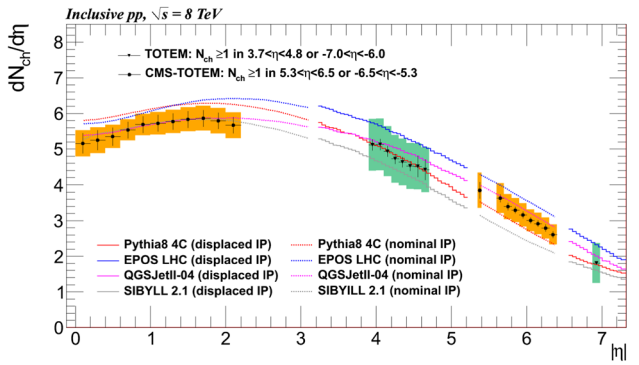


Fig. 341 Charged particle pseudorapidity distributions obtained in pp collisions at $\sqrt{s} = 8$ TeV for inelastic events as measured by the CMS and TOTEM experiments. The colored bands show the combined systematic and statistical uncertainties and the error bars represent the η uncorrelated uncertainties. The colored lines represents different model predictions. Figure taken from Ref. [4025]

To describe the entire rapidity interval models must be able to combine and connect perturbative QCD with non-perturbative approaches. The experimental points are compared to a number of different models which are available. The approaches are different but there are also several common elements in the models. As can be seen the gross features of the distribution are reasonably well described by the models.

The density at $\eta = 0$ as a function of the centre of mass energy has been plotted in Fig. 342 using data from the Sp$\bar{p}$S collider and the Tevatron in addition to the LHC data [4026]. The data points have been fitted with a power law. It is interesting to note that the increase of the density at $\eta = 0$ is faster than the increase of the total cross section with energy. This can be understood in terms of Pomeron interaction. To calculate an inclusive cross section like the density at $\eta = 0$ it is enough to use a one-pomeron exchange diagram. For the total cross section on the other hand one has to take into account multi-pomeron exchanges which tames the rise of the total cross section.

12.6.7 Conclusion

As mentioned in the introduction "soft" processes cover a large part of the total cross section. Collider experiments, at HERA and the Tevatron and now also at LHC, have produced a large amount of measurements related to low p_T reactions. The large rapidity coverage of the LHC detectors, and dedicated small angle experiments such as TOTEM, have offered new possibilities and there is still more to come. Moreover, the high center-of-mass energy of the LHC means that kinematically a larger rapidity range is available which opens up a window of studies where a separa-