3 The Strong Force

The full structure of the Standard Model will only become apparent in Section 5, after we understand the implications of parity violation. But, before we get there, there are two self-contained aspects of the theory that we can explore in some detail. These are the electromagnetic and strong forces.

We've already met the former in our first course on Quantum Field Theory. The action is

$$S = \int d^4x \, \left(-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + i\bar{\psi}\, \not{D}\psi - m\bar{\psi}\psi \right) \, . \tag{3.1}$$

Here $F_{\mu\nu}$ is the field strength of electromagnetism and it's excitations are photons. Meanwhile ψ is a Dirac spinor that describes the electron. We can always add further fields corresponding to any other electrically charged particles, like the muon. Upon quantisation, this theory is known as *quantum electrodynamics*, or *QED* for short.

For QED, what you see is what you get. You can stare at the action and, from your knowledge of perturbative quantum field theory, read off immediately that the theory describes a massless photon, coupled to a charged fermion of mass m. This, it turns out, is the only time we will be able to do this. The rest of the Standard Model is considerably more rich and interesting.

Our goal in this section is to describe the strong force. Remarkably, the action for the strong force is almost identical to that of QED. The only real difference is that the $U(1)$ group of electromagnetism is replaced by the gauge group

$$G = SU(3) \, . \tag{3.2}$$

The theory of the strong force is referred to as *quantum chromodynamics*, or *QCD* for short, and is given by

$$S = \int d^4x \, \left(-\frac{1}{2} \mathrm{Tr}\, G_{\mu\nu} G^{\mu\nu} + i \sum_i \bar{q}_i \, \not{D} q_i - m_i \bar{q}_i q_i \right) \, . \tag{3.3}$$

We'll explain what the various parts of this action mean, before we turn to quantum dynamics.

To avoid confusion with the photon, we denote the gauge field as G_μ. It is, like all Yang-Mills fields, Lie-algebra valued which means that we should think of each G_μ as a 3×3 Hermitian matrix. Replete with its gauge indices, we would write it as $(G_\mu)^a{}_b$ with $a, b = 1, 2, 3$. In the context of QCD, this additional index is referred to as *colour*[6]. The dimension of $SU(N)$ is $\dim SU(N) = N^2 - 1$ so there are 8 gauge bosons contained within the matrix G_μ. These are known, collectively, as *gluons*.

We can decompose G_μ into these gluon fields by writing $G_\mu = G_\mu^A T^A$ where T^A are generators of $SU(3)$ which we take to obey

$$\mathrm{Tr}(T^A T^B) = \frac{1}{2}\delta^{AB} \ . \tag{3.4}$$

A convenient basis is given by

$$T^A = \frac{1}{2}\lambda^A \ . \tag{3.5}$$

Here the λ^a the collection of 3×3 *Gell-Mann matrices*

$$\begin{gathered}
\lambda^1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \ , \quad \lambda^2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \ , \quad \lambda^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \ , \\
\lambda^4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \ , \quad \lambda^5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} \ , \quad \lambda^6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \ , \\
\lambda^7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \ , \quad \lambda^8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \ .
\end{gathered} \tag{3.6}$$

These are to $SU(3)$ what the Pauli matrices are to $SU(2)$. Indeed, you can see the Pauli matrices sitting in the top-left corner of λ^1, λ^2, and λ^3, reflecting the existence of an $SU(2)$ sub-group of $SU(3)$. Because $SU(3)$ has rank 2, there are two diagonal Gell-Mann matrices, λ^3 and λ^8. These span the Cartan sub-algebra.

We define the associated field strength

$$G_{\mu\nu} = \partial_\mu G_\nu - \partial_\nu G_\mu - i g_s [G_\mu, G_\nu] \ . \tag{3.7}$$

[6]Americans prefer to work with the convention $u = 1$.

This too is Lie-algebra valued. Note that the gauge potential and field strength are both called G and are distinguished only by the number of μ, ν spacetime indices that they carry. Buried within the field strength we have the strong coupling constant g_s. This is a dimensionless coupling that characterises the strength of the strong force. We will give its value shortly.

The gluons couple to quarks. These are Dirac spinors that we will call q_α where $\alpha = 1, 2, 3, 4$ is the usual spinor index that adorns a Dirac fermion. The quarks transform in the fundamental 3-dimensional representation of $SU(3)$. In group theoretic language, this is usually denoted as $\mathbf{3}$. This means that, in addition to the spinor index, the quarks also carry a colour index $a = 1, 2, 3$. We should think of this colour degree of freedom as a complex, normalised 3-vector that is rotated by $SU(3)$. To cheer us up, we sometimes refer to these three orthogonal states as red, green and blue. Needless to say, if you prefer to label them by your own favourite choice of colours then the physics remains unchanged.

The covariant derivative for each quark q is given by (now suppressing the spinor index)

$$\mathcal{D}_\mu q^a = \partial_\mu q^a - i g_s (G_\mu)^a{}_b q^b \ . \tag{3.8}$$

Here too we see the strong coupling constant g_s multiplying the interaction term.

Finally, the quarks also come with a *flavour index*, $i = 1, \ldots, N_f$ which simply tells us what kind of quark we're dealing with. The full theory of QCD comes with $N_f = 6$ flavours of quarks which, for reasons that will become clearer only in Section 5, we should think of as three pairs. They are down and up; strange and charm; and bottom and top. These quarks have masses

$$\begin{aligned} m_{\text{down}} &= 5 \text{ MeV} \quad &&\text{and} \quad &&m_{\text{up}} = 2 \text{ MeV} \\ m_{\text{strange}} &= 93 \text{ MeV} \quad &&\text{and} \quad &&m_{\text{charm}} = 1.3 \text{ GeV} \\ m_{\text{bottom}} &= 4.2 \text{ GeV} \quad &&\text{and} \quad &&m_{\text{top}} = 173 \text{ GeV} \ . \end{aligned} \tag{3.9}$$

The most striking aspect of these masses is that they span almost 5 orders of magnitude! In Section 5, we'll get a deeper understanding of how the masses arise from the condensation of the Higgs boson. But we won't get any deeper understanding of the particular values that the masses take: we only know these masses by measuring them experimentally.

The quarks also carry electric charge, and so the theory of QCD (3.3) should be augmented by coupling to electromagnetism. Here we will largely ignore the effects of electromagnetism in the dynamics because, as we will see, it is small compared to the strong force. It will, however, prove useful to just list the electric charges Q of various particles that we come across. For the first generation of quarks they are

$$Q_{\text{down}} = -\frac{1}{3}e \quad \text{and} \quad Q_{\text{up}} = \frac{2}{3}e \ . \tag{3.10}$$

Clearly, these are fractional charges relative to the electron. This pattern then repeats itself: the strange and bottom quark both have $Q = -\frac{1}{3}e$ while the charm and top both have $Q = +\frac{2}{3}e$. Note that, in this regard, the first generation of up and down quarks is the odd one out because the charge $\frac{2}{3}$ quark is lighter than the charge $-\frac{1}{3}$ quark.

This completes our discussion of the various elements in the QCD action 3.3. Now it's time to understand the physics.

3.1 Strong Coupling

If you look naively at the action (3.3), you would think that QCD is a theory of massless gluons interacting with quarks. But that's certainly not what we see in the world around us. Any massless gauge boson would mediate a long range force which drops off, like electromagnetism, as $1/r^2$. Yet we know that the effects of the strong force don't extend beyond the nucleus of the atom, which isn't particularly big. In addition, we don't see quarks wandering around freely. What we see are protons and neutrons. If the weak force didn't exist, these would be joined by light particles called pions. But not quarks.

All of which leads us to ask: why are the particles that we see in the world not directly related to the fields in the fundamental Lagrangian (3.3)?

3.1.1 Asymptotic Freedom

The answer to this question starts with the observation that the coupling constant of the strong force is not at all constant. Like all parameters in quantum field theory, its value depends on the distance scale, or equivalently energy scale, at which you look. This is the essence of renormalisation.

To illustrate the physics, we will briefly step back from QCD and consider the more general theory with $G = SU(N_c)$ gauge group, coupled to N_f *massless* quarks. Hence, N_c is the number of colours, and N_f the number of flavours. The gauge coupling g_s^2

depends on the energy scale μ at which the theory is probed and, at one-loop, is given by

$$\frac{1}{g_s^2(\mu)} = \frac{1}{g_0^2} - \frac{b_0}{(4\pi)^2} \log \frac{\Lambda_{UV}^2}{\mu^2} \ . \tag{3.11}$$

Here g_0^2 is the bare coupling that sits in the Lagrangian. It can be thought of as the coupling evaluated at the cut-off scale Λ_{UV} since $g_s^2(\Lambda_{UV}) = g_0^2$. The coefficient b_0 is given by

$$b_0 = \frac{11}{3} N_c - \frac{2}{3} N_f \tag{3.12}$$

A derivation of this result can be found in the lectures on Gauge Theory.

The running of the coupling constant is often summarised in terms of the one-loop beta function

$$\beta(g) \equiv \mu \frac{dg_s}{d\mu} = -\frac{b_0}{(4\pi)^2} g_s^3 \tag{3.13}$$

whose solution gives the logarithmic behaviour (3.11).

The all-important feature of the beta function is the overall minus sign. The flow of the coupling means that the theory is weakly coupled at high energies, a phenomenon known as *asymptotic freedom*. Conversely, it means that the theory is strongly coupled at low energies. From (3.12), we see that asymptotic freedom persists only if the number of flavours is sufficiently small

$$N_f < \frac{11}{2} N_c \ . \tag{3.14}$$

Clearly this is satisfied by QCD with $N_c = 3$ and $N_f = 6$.

Asymptotic freedom is rare in $d = 3 + 1$ dimensions. In fact, it only happens for non-Abelian gauge theories. Coupling constants in any theory run with scale but all of them – the QED fine structure constant, Yukawa couplings, self-interactions of scalars – get bigger as you go to high energies. It is only non-Abelian gauge theories where the coupling gets bigger as you go to low energies.

The comparison to QED is useful. At distances larger than $r \geq 10^{-12}$ m (which is the Compton wavelength of the lightest charged particle, namely the electron) the fine structure constant stops running and plateaus to the familiar value of $\alpha \approx 1/137$. But as you go to higher energies, or shorter distances, the fine structure constant increases. For example, at $r \approx 10^{-17}$ m, which corresponds to $E \approx 100$ GeV, we have $\alpha(\mu) \approx 1/127$.

Asymptotic freedom means that Yang-Mills theory is simple to understand at high energies, or short distance scales. Here it is a theory of massless, interacting gluon fields whose dynamics are well described by the classical equations of motion, together with quantum corrections which can be computed using perturbation methods. However, it becomes much harder to understand what is going on at large distances where the coupling gets strong. Indeed, the beta function (3.13) itself was computed in perturbation theory and is valid only when $g_s^2(\mu) \ll 1$. This equation therefore predicts its own demise at large distance scales.

We can estimate the distance scale at which we think we will run into trouble. Taking the one-loop beta function at face value, we can ask: at what scale does $g_s^2(\mu)$ diverge? This happens at a finite energy

$$\Lambda_{\text{QCD}} = \mu \exp\left(-\frac{8\pi^2}{b_0 g_s^2(\mu)}\right) \tag{3.15}$$

This is known as the *strong coupling scale*, or just the *QCD scale*. It has the property that $d\Lambda/d\mu = 0$. In other words, it is an RG invariant. This is the scale at which the gauge coupling becomes strong.

Viewed naively, there's something very surprising about the emergence of the scale Λ_{QCD}. This is because the classical theory has no dimensionful parameter. Yet the quantum theory has a physical scale, Λ_{QCD}. It seems that the quantum theory has generated a scale out of thin air, a phenomenon which goes by the name of *dimensional transmutation*. In fact, as the definition (3.15) makes clear, there is no mystery about this. Quantum field theories are not defined only by their classical action alone, but also by the cut-off Λ_{UV}. Although we might like to think of this cut-off as merely a crutch, and not something physical, this is misleading. It is not something we can do without. And it is this cut-off which evolves to the physical scale Λ_{QCD}.

$$\Lambda_{\text{QCD}} = \Lambda_{UV}\, e^{-8\pi^2/b_0 g_0^2} \ . \tag{3.16}$$

This means that if the bare coupling is small, $g_0 \ll 1$, as it should be then the physical scale Λ_{QCD} is exponentially suppressed relative to the UV cut-off: $\Lambda_{\text{QCD}} \ll \Lambda_{UV}$. It's a beautiful example of how a low-energy scale can be naturally generated from a high energy scale. (A similar mechanism can be seen in other contexts, including the BCS theory of superconductivity and the Kondo effect.)

The QCD Scale for QCD

So far, our discussion has been for the general theory of $SU(N_c)$ with N_f flavours of massless quarks. What happens for actual QCD?

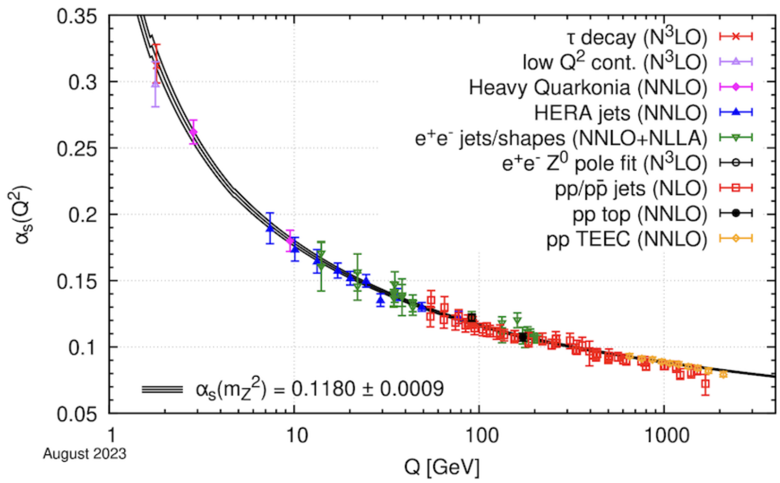


Figure 8. The running of the strong coupling coupling constant $\alpha_s = g_s^2/4\pi$ in terms of energy which is denoted Q in the plot. This is taken from the particle data group's review of QCD.

There is one important modification which is needed because the quarks in QCD are most certainly not massless. This is easy to accommodate. A quark of mass m contributes to the beta function as if it were massless for scales $\mu \gg m$. And it decouples from the physics for scales $\mu \ll m$. For scales $\mu \sim m$ you need to be more careful, but we'll simply duck the issue.

Revisiting the quarks masses in (3.9), we see that the beta function acts as if it has $N_f = 6$ massless quarks for $\mu \gg 173$ GeV. And for 4.2 GeV $\ll \mu \ll 173$ GeV, it acts as if it has $N_f = 5$ massless quarks, and so on. The combined experimental data for the running of $\alpha_s = g_s^2/4\pi$ is shown in Figure 8.

The most important question is: what is the strong coupling scale $\Lambda_{\rm QCD}$. As we will see, this determines the scale at which the interesting physics happens. For the strong force it lies around

$$\Lambda_{\rm QCD} \approx 200 \text{ MeV} . \tag{3.17}$$

This definition isn't precise and you'll also see statements that it is closer to 300 MeV. This could be due to different regularisation schemes, or whether you choose the definition of this scale to be $\alpha_s(\Lambda_{\rm QCD}) = \infty$ or $\alpha_s(\Lambda_{\rm QCD}) = 1$ (which doesn't change things too much). There's no right or wrong answer. As we will see, the point of $\Lambda_{\rm QCD}$ is to give a ballpark energy scale at which much of the physics of QCD takes place.

To give a value for the strength of the coupling g_s itself, we need to specify the energy scale at which we do the measurement. A useful benchmark is the mass of the Z-boson, $M_Z \approx 90$ GeV. Here the strong coupling constant has been measured remarkably accurately

$$\alpha_s(M_Z) = \frac{g_s^2(M_Z)}{4\pi} = 0.1184 \pm 0.0007 \; . \tag{3.18}$$

This is small enough to trust perturbation theory at these scales.

3.1.2 Anti-Screening and Paramagnetism

It's useful to have some intuition for why non-Abelian gauge theories exhibit asymptotic freedom, with a negative beta function, while all other quantum field theories do not. Ultimately, to see this result you just have to roll up your sleeves and do the calculation (and an opportunity will be offered in the sister course on AQFT). Here we give a nice, but slightly handwaving, analogy from condensed matter.

In condensed matter physics, materials are not boring passive objects. They contain mobile electrons, and atoms with a flexible structure, both of which can respond to any external perturbation such as applied electric or magnetic fields. One consequence of this is an effect known as *screening*. In an insulator, screening occurs because an applied electric field will polarise the atoms which, in turn, generate a counteracting electric field. One usually describes this by introducing the electric displacement $\mathbf{D}$, related to the electric field through

$$\mathbf{D} = \epsilon \mathbf{E} \tag{3.19}$$

where the permittivity $\epsilon = \epsilon_0(1 + \chi_e)$ with χ_e the electrical susceptibility. For all materials, $\chi_e > 0$. This ensures that the effect of the polarisation is always to reduce the electric field, never to enhance it. You can read more about this in Section 7 of the lecture notes on Electromagnetism.

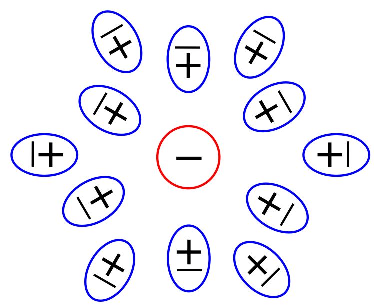
(As an aside: In a metal, with mobile electrons, there is a much stronger screening effect which turns the Coulomb force into an exponentially suppressed Debye-Hückel, or Yukawa, force. This was described in the final section of the notes on Electromagnetism, but is not the relevant effect here.)

What does this have to do with quantum field theory? In quantum field theory, the vacuum is not a passive boring object. It contains quantum fields which can respond to any external perturbation. In this way, quantum field theories are very much like condensed matter systems. A good example comes from QED. There the one-loop beta function is positive and, at distances smaller than the Compton wavelength of the electron, the gauge coupling runs as

$$\frac{1}{e^2(\mu)} = \frac{1}{e_0^2} + \frac{1}{12\pi^2}\log\left(\frac{\Lambda_{UV}^2}{\mu^2}\right) \ . \tag{3.20}$$

This tells us that the charge of the electron gets effectively smaller as we look at larger distance scales, a phenomenon that is understood in very much the same spirit as condensed matter systems. In the presence of an external charge, electron-positron pairs will polarize the vacuum, as shown in the figure, with the positive charges clustering closer to the external charge. This cloud of electron-positron pairs shields the original charge, so that it appears reduced to someone sitting far away.

The screening story above makes sense for QED. But what about QCD? The negative beta function tells us that the effective charge is now getting larger at long distances, rather than smaller. In other words, the Yang-Mills vacuum does not screen charge: it anti-screens. From a condensed matter perspective, this is weird. As we mentioned above, materials always have $\chi_e > 0$ ensuring that the electric field is screened, rather than anti-screened.



However, there's another way to view the underlying physics. We can instead think about magnetic screening. Recall that in a material, an applied magnetic field induces dipole moments and these, in turn, give rise to a magnetisation. The resulting magnetising field $\mathbf{H}$ is defined in terms of the applied magnetic field as

$$\mathbf{B} = \mu\mathbf{H} \tag{3.21}$$

with the permeability $\mu = \mu_0(1+\chi_m)$. Here χ_m is the magnetic susceptibility and, in contrast to the electric susceptibility, can take either sign. The sign of χ_m determines the magnetisation of the material, which is given by $\mathbf{M} = \chi_m\mathbf{H}$. For $-1 < \chi_m < 0$, the magnetisation points in the opposite direction to the applied magnetic field. Such materials are called *diamagnets*. (A perfect diamagnet has $\chi_m = -1$. This is what happens in a superconductor.) In contrast, when $\chi_m > 1$, the magnetisation points in the same direction as the applied magnetic field. Such materials are called *paramagnets*.

In quantum field theory, polarisation effects can also make the vacuum either diamagnetic or paramagnetic. Except now there is a new ingredient which does not show up in real world materials discussed above: relativity! This means that the product must be

$$\epsilon\mu = 1$$

because "1" is the speed of light. In other words, a relativistic diamagnetic material will have $\mu < 1$ and $\epsilon > 1$ and so exhibit screening. But a relativistic paramagnetic material will have $\mu > 1$ and $\epsilon < 1$ and so exhibit anti-screening. Phrased in this way, the existence of an anti-screening vacuum is much less surprising: it follows simply from paramagnetism combined with relativity.

For free, non-relativistic fermions, we calculated the magnetic susceptibility in the lectures on Statistical Physics when we discussed Fermi surfaces. In that context, we found two distinct contributions to the magnetisation. Landau diamagnetism arose because electrons form Landau levels. Meanwhile, Pauli paramagnetism is due to the spin of the electron. These two effects have the same scaling but different numerical coefficients.

When you dissect the computation of the one-loop beta function in Yang-Mills theory, you can see that the gluons also give two distinct contributions: one diamagnetic, and one paramagnetic. And the paramagnetic contribution wins. Viewed in this light, asymptotic freedom can be traced to the paramagnetic contribution from the gluon spins.

3.1.3 The Mass Gap

When the coupling is small, quantum field theories look similar to their classical counterparts. For example, classical Maxwell theory provides a decent guide to what you might expect from QED. In contrast, when the coupling is large, all bets are off. The quantum theory and classical theory may be completely different. Yang-Mills and QCD provide the archetypal example.

We will start our discussion by ignoring the quarks completely and look just at Yang-Mills theory,

$$S = \int d^4x \; -\frac{1}{2}\mathrm{Tr}\, G_{\mu\nu}G^{\mu\nu} \; . \tag{3.22}$$

For QCD we take gauge group $G = SU(3)$, but everything we're about to say holds for any simple, compact Lie group.

Classically, Yang-Mills describes massless, interacting spin 1 fields. Its solutions include, among other things, waves that propagate at the speed of light. The question that we want to ask is: what is the physics of the quantum theory?

Because the coupling is strong at low energies, we can't answer this question using the traditional perturbative techniques that we learned in our first course on Quantum Field Theory. In fact, if we rely purely on analytic methods we can't answer this question at all! Instead, we rely on numerical simulation and experiment, together with some heuristic ideas and a number of solvable toy models which give us intuition for what quantum field theories can do. But we do have a robust, clear answer:

Quantum Yang-Mills is not a theory of massless particles, Instead, the lightest particle has a mass of $m \sim \Lambda_{\rm QCD}$. This particle is called a *glueball*. We say that the theory is *gapped* which means that there is a gap between the ground state and the first excited state with energy $E = mc^2$. These glueballs also exist in our world, although they mix strongly with various neutral meson states and so don't have a very clean experimental signature.

We don't currently have the ability to prove that Yang-Mills is gapped from first principles. It is generally considered one of the most important and challenging open problems in mathematical physics.

3.1.4 A Short Distance Coulomb Force

The existence of a mass gap goes hand in hand with another phenomenon: this is *confinement*.

To highlight the physics, it's best if we again look at the slightly more general case of $G = SU(N)$ gauge theory. We can ask the kind of questions that we studied in our first course on Electromagnetism. Suppose that you take two test particles, a quark in the fundamental representation $\mathbf{N}$ and an anti-quark in the anti-fundamental $\overline{\mathbf{N}}$. What force do they feel?

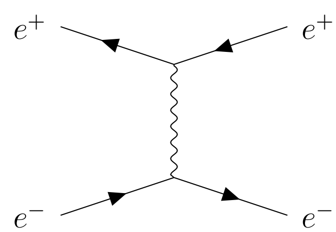
There are two different answers to this question, depending on the separation r between the particles. If they are separated just a short distance $r \ll \Lambda_{\rm QCD}^{-1} \approx 5 \times 10^{-15}$ m, then the coupling g_s^2 is small and we can trust the classical result. However, if the particles are separated by a large distance $r \gg \Lambda_{\rm QCD}^{-1}$, then we're firmly in the regime of strongly coupled physics and we might expect that the classical result is not a good guide.

Here we start by considering the short-distance regime $r \ll \Lambda_{\rm QCD}^{-1}$. The Compton wavelength of a particle of mass m is $\lambda \sim 1/m$ and it only makes sense to talk about separating two quantum particles a distance r if $r \gg \lambda$. This means that to talk about the short-distance force experienced by two quarks, the quarks must have mass $m \gg \Lambda_{\rm QCD}$. In the context of QCD, that means that the analysis below is valid only for charm, bottom and top quarks.

Let's remind ourselves of the story in QED. In electromagnetism, two particles of equal and opposite charges $\pm e$, separated by a distance r, experience an attractive Coulomb force, described by the potential energy $V(r)$,

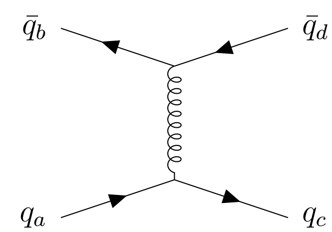
$$V(r) = -\frac{e^2}{4\pi r} \ . \tag{3.23}$$

In the framework of QED, we can reproduce this from the the tree-level exchange of a single photon (where time should be viewed as flowing left-to-right in this diagram)



This computation can be found in the lectures on Quantum Field Theory.

Now we want to do the same calculation in QCD. The diagram is the same, but with a gluon, rather than a photon, as the intermediary. The only difference lies in the fact that quarks carry colour indices, which are the $a, b, c, d = 1, \ldots, N$ indices in the Feynman diagram below



Using the Feynman rules for QCD, the tree level potential between the quarks is given by the same Coulomb force law, dressed with the group theoretic factor

$$V(r) = \frac{g_s^2}{4\pi r} T^A_{ca} T^{\star A}_{db} \ . \tag{3.24}$$

We've still got those colour indices to deal with. At first glance, it looks like there's N^2 different possibilities for the states of the ingoing particles ($a, b = 1, \ldots, N$) and a further N^2 different possibilities for the states of the outgoing particles ($c, d = 1, \ldots, N$). Happily, all of this boils down to some simple group theory. In the present case, we have the tensor product of representations

$$\mathbf{N} \otimes \overline{\mathbf{N}} = \mathbf{1} \oplus \text{adj} \tag{3.25}$$

where the adjoint representation has dimension $N^2 - 1$. The object $T^A T^{A\,\dagger}$, viewed as a $N^2 \times N^2$ dimensional matrix, will then have two different eigenvalues, one for each of these representations. This will lead to two different coefficients for the forces. Our goal is to determine them. Here we give the general result:

Claim: Suppose that we have two particles in representations $\mathbf{R}_1$ and $\mathbf{R}_2$. For each representation $\mathbf{R} \subset \mathbf{R}_1 \otimes \mathbf{R}_2$, the force experienced by the two particles will be proportional to

$$C(R) - C(R_1) - C(R_2) \tag{3.26}$$

where $C(R)$ is a number that characterises each representation, known as the *quadratic Casimir*, defined as

$$T^A(R) T^A(R) = C(R)\, \mathbb{1}\ . \tag{3.27}$$

Proof: Gluon exchange will result in a Coulomb-like force law (3.24), but with the group theoretic factor $T^A(R_1) T^A(R_2)$. (For $\mathbf{R}_1 = \mathbf{N}$ and $\mathbf{R}_2 = \overline{\mathbf{N}}$, this coincides with the result (3.24).) Consider the operator

$$S^A = T^A(R_1) \otimes \mathbf{1}\ +\ \mathbf{1} \otimes T^A(R_2)\ . \tag{3.28}$$

Squaring and rearranging, we have

$$T^A(R_1) \otimes T^A(R_2) = \frac{1}{2}\left[S^A S^A - T^A(R_1) T^A(R_1) \otimes \mathbf{1} - \mathbf{1} \otimes T^A(R_2) T^A(R_2)\right]\ . \tag{3.29}$$

(This is the same kind of calculation that one does in atomic physics when computing consequence of the spin orbit coupling $\mathbf{L}\cdot\mathbf{S}$. You can read more about this in the lectures on Topics in Quantum Mechanics.) Each of the final two terms on the right-hand side is a quadratic Casimir (3.27), while the first term decomposes into block diagonal matrices, with components labelled by the irreducible representations $\mathbf{R} \subset \mathbf{R}_1 \otimes \mathbf{R}_2$. We have

$$T^A(R_1) \otimes T^A(R_2)\Big|_R = \frac{1}{2}\left[C(R) - C(R_1) - C(R_2)\right] \tag{3.30}$$

as promised. □

The upshot is that to calculate the force between a quark and anti-quark (or, indeed, between any two representations) we just need to known the quadratic Casimirs. For $G = SU(N)$, the Casimirs for the fundamental, anti-fundamental and adjoint are

$$C(\mathbf{N}) = C(\overline{\mathbf{N}}) = \frac{N^2-1}{2N} \quad \text{and} \quad C(\text{adj}) = N \; . \tag{3.31}$$

We also have $C(\mathbf{1}) = 0$ for the singlet (trivial) representation. This means that a quark-anti-quark pair with their colour degrees of freedom entangled as a singlet experience a force proportional to

$$\frac{1}{2}\Big[C(\mathbf{1}) - C(\mathbf{N}) - C(\overline{\mathbf{N}})\Big] = -\frac{N^2-1}{2N} \; . \tag{3.32}$$

The minus sign means that this force is attractive. This is what we would have expected from our classical intuition. However, when the quarks sit in the adjoint channel, we have

$$\frac{1}{2}\Big[C(\text{adj}) - C(\mathbf{N}) - C(\overline{\mathbf{N}})\Big] = \frac{1}{2N} \; . \tag{3.33}$$

Perhaps surprisingly, this is a repulsive force.

We can do the same analysis if we have two quarks, rather than a quark and anti-quark. Now the group theoretic decomposition is

$$\mathbf{N} \otimes \mathbf{N} = \square\!\square \oplus \begin{smallmatrix}\square\\\square\end{smallmatrix}$$

where $\square\!\square$ is the Young tableaux representation for the symmetric representation, with $\dim(\square\!\square) = \frac{1}{2}N(N+1)$ while $\begin{smallmatrix}\square\\\square\end{smallmatrix}$ means the anti-symmetric representation with $\dim(\begin{smallmatrix}\square\\\square\end{smallmatrix}) = \frac{1}{2}N(N-1)$. The relevant Casimirs are

$$C\left(\square\!\square\right) = \frac{(N-1)(N+2)}{N} \quad \text{and} \quad C\left(\begin{smallmatrix}\square\\\square\end{smallmatrix}\right) = \frac{(N-2)(N+1)}{N}$$

From this we learn that two quarks which sit in the symmetric channel classically repel each other, since

$$\frac{1}{2}\left[C\left(\square\!\square\right) - C(\mathbf{N}) - C(\mathbf{N})\right] = \frac{N-1}{2N} \; . \tag{3.34}$$

Meanwhile, two quarks that sit in the anti-symmetric channel feel a classical attractive force,

$$\frac{1}{2}\left[C\left(\begin{smallmatrix}\square\\\square\end{smallmatrix}\right) - C(\mathbf{N}) - C(\mathbf{N})\right] = -\frac{N+1}{2N} \; . \tag{3.35}$$

Ultimately, our interest lies in QCD with $G = SU(3)$. Here there's a group theoretic novelty because the anti-symmetric representation is actually the same as the anti-fundamental,

$$\mathbf{3} \otimes \mathbf{3} = \overline{\mathbf{3}} \oplus \mathbf{6} \, . \tag{3.36}$$

This means that two quarks will attract in the anti-symmetric $\bar{\mathbf{3}}$ channel. But we could then add a third quark and, from (3.32), this too will feel an attractive force if all three sit in the singlet. We see that three quarks can feel a mutually attractive force in QCD. Of course, this force is computed classically and it falls off with a $1/r$ potential, just like the Coulomb force of electromagnetism. Nonetheless, this is the first time that we see why it might be energetically preferable for three quarks to form colour singlet bound states.

3.1.5 A Long Distance Confining Force

The analysis above was only for particles separated by very short distances $r \ll \Lambda_{\rm QCD}^{-1} \approx 5 \times 10^{-15}$ m. But our real interest is in what happens at large distance scales where the Yang-Mills coupling becomes strong.

Previously, we stated (but didn't prove!) that Yang-Mills has a mass gap. This means that, at distances $\gg 1/\Lambda_{QCD}$, the force will be due to the exchange of massive particles rather than massless particles. In many situations, the exchange of massive particles results in an exponentially suppressed Yukawa force, of the form $V(r) \sim e^{-mr}/r$, and you might have reasonably thought this would be the case for Yang-Mills. You would have been wrong.

Let's again consider a quark and an anti-quark, in the $\mathbf{N}$ and $\overline{\mathbf{N}}$ representations respectively. At large distances, the potential energy between the two turns out to grow linearly with distance

$$V(r) = \sigma r \tag{3.37}$$

for some value σ that has dimensions of energy per length. For reasons that we will explain shortly, it is often referred to as the *string tension*. On dimensional grounds, we must have $\sigma \sim \Lambda_{QCD}^2$ since there is no other dimensionful parameter in the game. The force law (3.37) is, to put it mildly, a dramatic departure from what we're used to. The potential energy now *increases* with separation. Indeed, it costs an infinite amount of energy to pull the quark-anti-quark pair to infinity.

For two quarks, both in the fundamental representation, the result is even more dramatic. Now the tensor product of the two representations does not include a singlet (at least this is true for $SU(N)$ with $N \geq 3$). The energy of the two quarks turns out to be infinite. This is a general property of quantum Yang-Mills: the only finite energy states are gauge singlets. The theory is said to be *confining*, meaning that an individual quark cannot survive on its own, but is forced to enjoy the company of friends.

The phenomenon of confinement is, like the mass gap, something that we can't prove from first principles. Once again, however, there is clear numerical evidence together with a plethora of heuristic explanations.

In Section 3.3, we'll look more closely at how quarks and anti-quarks bind together in QCD. Roughly speaking, there are two possibilities. First a quark and anti-quark can bind together to form a colour singlet. The resulting particle is known as a *meson*. But, alternatively, three quarks can bind together to form a colour singlet by dint of the invariant tensor ϵ^{abc} of $SU(3)$. The resulting particle is called a *baryon*, with the proton and neutron being the most obvious examples.

Note that if the strong force was described by $SU(N)$, with $N \neq 3$, then mesons would always be quark-anti-quark pairs and, hence, are always bosons. In contrast, baryons in $SU(N)$ contain N quarks and hence are fermions when N is odd and bosons when N is even.

The QCD Flux Tube

We've already seen an example of a confining potential (3.37) in Section 2.3 when discussing superconductivity. In that context, magnetic monopoles experience a confining force, and the reason was clear: the Meissner effect means that it's energetically preferable for the magnetic field lines to form flux tubes.

No such simple explanation is known for confinement in QCD, but it's clear from numerical simulations that a similar flux tube, or string, does form, now comprised of chromoelectric field lines. Two examples are shown in Figure 9, where we see flux tubes between the quark-anti-quark that form a meson and also between three quarks that form a baryon. In fact, some of the original studies of string theory were motivated by understanding the dynamics of these flux tubes.

However, in contrast to the the Higgs phase of a superconductor, it doesn't make sense to search for a classical solution to the equations of motion that describes the QCD flux tube. Instead the QCD flux tube is very much a quantum effect, arising only after performing the path integral, which involves summing over many different

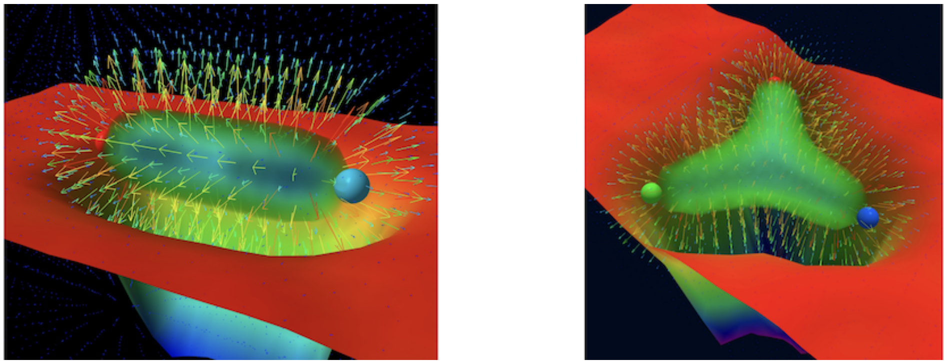


Figure 9. The chromoelectric flux tube between a quark and anti-quark in a meson state, on the left, and between three quarks in a baryon state on the right. From the QCD simulations of Derek Leinweber.

field configurations. To emphasise the physics, it's best to work with the alternative rescaling of the Yang-Mills action (1.103) in which the gauge coupling sits as an overall coefficient, so the path integral over the gauge field takes the schematic form

$$Z = \int \mathcal{D}G_\mu \; \exp\left(-\frac{i}{2g_s^2} \int d^4x \; \mathrm{Tr}\, G_{\mu\nu} G^{\mu\nu} \right) \; . \tag{3.38}$$

At weak coupling, we have $g_s^2 \ll 1$ and we may use saddle-point techniques to show that the path integral is dominated by solutions to the classical equations of motion. But at strong coupling, we have $g_s^2 \to \infty$ which, roughly speaking, is telling us that there's no suppression to the path integral at all. All field configurations, regardless of how wildly they oscillate, contribute equal weight. Among the infinity of different field configurations, those that look like a flux tube seem to dominate. But we don't know why.

Perhaps the best explanation of confinement (although one that falls well short of a proof) comes from an approach that discretises Yang-Mills theory known as *lattice gauge theory*. In that context, you can show that if you naively sum over all field configurations without any weighting, then you do indeed reproduce the confining behaviour. You can find details of this calculation, together with an explanation of why the calculation is not really performed in the physical regime, in the lectures on Gauge Theory.

It's tempting to push the superconductivity analogy further. In a superconductor, electrically charged particles condense (the Cooper pairs) and the result is that magnetic charges confine. Flipping this on its head, if magnetically charged particles were to condense, then electric charges would be confined. This idea goes by the name of the *dual Meissner effect*. It seems right, but it's hard to make it concrete. What are these mysterious chromomagnetic charges that condense in QCD causing quarks to confine? We don't know. However, there are other 4d gauge theories where we can prove confinement analytically and it does happen through the condensation of monopoles. (This is what happens in the famous Seiberg-Witten solution of $\mathcal{N} = 2$ supersymmetric gauge theories.)

The Effect of Light Quarks

As if the problem of confinement wasn't difficult enough, things are actually more complicated than I've sketched above. This is because, in real world QCD, the simple force formula (3.37) that designates a confining theory, simply isn't true!

Here's the deal. Suppose that we have pure Yang-Mills theory. Then, for any choice of non-Abelian gauge group, including $G = SU(3)$, the theory is strongly believed to have a mass gap, determined by its strong coupling scale $\Lambda_{\rm QCD}$, and confine. Here "confinement" means that if you introduce two test particles into the theory – a quark and anti-quark – then the long-distance force law between them will exhibit the linear behaviour (3.37).

Now suppose that you have Yang-Mills theory coupled to a single dynamical quark that has mass $m \gg \Lambda_{\rm QCD}$. For example, you could think of the artificial world in which there is only a charm quark and nothing else. We can again ask what the energy is between two test particles that we take to be a quark-anti-quark pair. At large distances $r \gg \Lambda_{\rm QCD}^{-1}$, we have a confining potential

$$V(r) = \sigma r \ . \tag{3.39}$$

But, this time, it doesn't persist for all r. This is because once we stretch the particles past the point $\sigma r > 2m$, then you can lower the energy of the state by creating a quark-anti-quark pair from the vacuum. The $q\bar{q}$ pair will break the string and you'll be left with two meson-like states, in which your original quark-anti-quark test particles are now bound to the dynamical quarks of the theory.

This means that the regime of the confining force (3.39) is limited. It happens only for long distances, but not too long distances. Using the fact that the string tension scales as $\sigma \sim \Lambda_{\rm QCD}^2$, we see that quarks experience the confining force only in a region

$$\frac{1}{\Lambda_{\rm QCD}} \ll r \ll \frac{m}{\Lambda_{\rm QCD}^2} \ . \tag{3.40}$$

Nonetheless, if we only have dynamical quarks with mass $m \gg \Lambda_{\rm QCD}$, then there's still a window in which we see the confining behaviour.

However, for real world QCD, there is no such window! The lightest quark has mass $m \ll \Lambda_{\rm QCD}$. If you like, the string breaks through the pair creation of up and down $q\bar{q}$ pairs before we even get to the confining regime $r \gg \Lambda_{\rm QCD}^{-1}$. This means that thinking about the confining nature of real world QCD in terms of the linear potential (3.39) is a useful, but not entirely accurate, fiction.

What does survive, however, is the statement that all finite energy states in QCD are necessarily colour singlets. That is the key takeaway that we will need when discussing the observed particle spectrum in Section 3.3.

3.2 Chiral Symmetry Breaking

Here's a general piece of advice. If you want to understand the dynamics of a quantum field theory, first understand the symmetries. They dictate how the dynamics is organised and will often contain clues about the nature of the low-energy physics.

So what are the symmetries of QCD? Well, obviously the theory is based on a $G = SU(3)$ gauge group but, as we've stressed previously, that's really a redundancy rather than a symmetry. Here we are interested in global symmetries.

The actual symmetry group of the QCD action (3.3) is $U(1)^{N_f}$, which rotates the phase of each individual Dirac quark field. That alone doesn't give us much insight. However, there is a much larger *approximate* symmetry of the theory. This emerges if we pretend that the quarks are massless.

First, we should ask: why are we allowed to pretend that quarks are massless? The reason is that QCD comes with its own dynamical scale $\Lambda_{\rm QCD}$. This is the scale at which all the interesting physics happens. This means that if we have any quark with a mass $m \ll \Lambda_{\rm QCD}$, then it's appropriate to first understand the dynamics of the gauge fields in the massless limit, and subsequently figure out how the presence of the mass changes things as corrections of order $m/\Lambda_{\rm QCD}$.

As we've seen, we have $\Lambda_{\rm QCD} \approx 200$ MeV, while the masses of the quarks are

$$\begin{aligned} m_{\rm down} &= 5 \text{ MeV} \quad \text{and} \quad m_{\rm up} = 2 \text{ MeV} \\ m_{\rm strange} &= 93 \text{ MeV} \quad \text{and} \quad m_{\rm charm} = 1.3 \text{ GeV} \\ m_{\rm bottom} &= 4.2 \text{ GeV} \quad \text{and} \quad m_{\rm top} = 173 \text{ GeV} \; . \end{aligned} \tag{3.41}$$

Clearly there's no sense in which the charm, bottom and top quarks are light. In fact, they're so much heavier than the QCD scale that they effectively just decouple from the low-energy dynamics and, for the story that we're about to tell, we can just ignore them. (We'll revisit these heavy quarks in Section 3.3 when we look more closely at the kinds of mesons and baryons that we can form.)

At the other end, no one's going to argue against the statement that $m_{\rm up}, m_{\rm down} \ll \Lambda_{\rm QCD}$ and it's an excellent approximation to treat these as massless and then see how the very small mass changes things. That leaves us with the strange quark. While it's certainly true that $m_{\rm strange} < \Lambda_{\rm QCD}$, you might reasonably complain that it's a bit of stretch to replace $<$ with $\ll$. All of which means that it will certainly be useful to pretend that there are two massless quarks, and it's probably worth seeing what happens if we're more optimistic and pretend that there are three massless quarks.

At this stage, we don't need to commit to the number of massless quarks, and we can work in generality. In fact, we don't even need to commit to the number of colours. Consider $G = SU(N_c)$ Yang-Mills, coupled to N_f flavours of massless fundamental fermions that we will continue to refer to as "quarks".

The additional symmetry comes from the realisation that each 4-component Dirac spinor q decomposes into two 2-component Weyl spinors, as in (1.48),

$$q = \begin{pmatrix} q_L \\ q_R \end{pmatrix} \ . \tag{3.42}$$

Each of the Weyl spinors q_L and q_R carries a colour index that runs over $1, \ldots, N_c$ and a flavour index $i = 1, \ldots, N_f$, as well as it's 2-component spinor index. Written in terms of these Weyl fermions, our generalised but massless, QCD action (3.3) becomes

$$S = \int d^4x \ \left(-\frac{1}{2} {\rm Tr}\, G_{\mu\nu} G^{\mu\nu} + i \sum_{i=1}^{N_f} \bar{q}_{L\,i} \bar{\sigma}^\mu \mathcal{D}_\mu q_{L\,i} + \bar{q}_{R\,i} \sigma^\mu \mathcal{D}_\mu q_{R\,i} \right) \ . \tag{3.43}$$

where we've suppressed both colour and spinor indices in this expression. Written in this way, we see that the classical Lagrangian has a global symmetry

$$G_F = U(N_f)_L \times U(N_f)_R \tag{3.44}$$

which acts on the flavour indices as

$$U(N_f)_L : q_{L\,i} \mapsto L_{ij} q_{L\,j} \quad \text{and} \quad U(N_f)_R : q_{R\,i} \mapsto R_{ij} q_{R\,j} \tag{3.45}$$

where both $L, R \in U(N_f)$. This is known as a *chiral* symmetry because it acts differently on left-handed and right-handed Weyl spinors. This chiral symmetry is a symmetry only of the theory with massless fermions because as soon as we add a mass term like $\bar{q}_L q_R$, it breaks the chiral symmetry to its diagonal subgroup.

As we will see, in the quantum theory different parts of the symmetry group G_F suffer different fates. Perhaps the least interesting is the overall $U(1)_V$, under which both q_L and q_R transform in the same way: $q_{L\,i} \to e^{i\alpha} q_{L\,i}$ and $q_{R\,i} \to e^{i\alpha} q_{R\,i}$. This symmetry survives in the quantum theory and the associated conserved quantity counts the number of quark particles of either handedness. In the context of QCD, this is referred to as *baryon number*, because it counts baryons, but not mesons which have a quark-anti-quark pair.

The other Abelian symmetry is the axial symmetry, $U(1)_A$. Under this, the left-handed and right-handed fermions transform with an opposite phase: $q_{L\,i} \to e^{i\beta} q_{L\,i}$ and $q_{R\,i} \to e^{-i\beta} q_{R\,i}$. This is more subtle. It turns out that although this is a symmetry of the classical Lagrangian, it is *not* a symmetry of the full quantum theory due to a phenomenon known as the *anomaly*. We will explain this in Section 4. For now, you will have to just trust me when I say that $U(1)_A$ is not actually a symmetry and we will not discuss it for the rest of this section.

This means that the global symmetry group of the quantum theory is

$$G_F = U(1)_V \times SU(N_f)_L \times SU(N_f)_R \tag{3.46}$$

The two non-Abelian symmetries act as (3.45), but where L and R are now each elements of $SU(N_f)$ rather than $U(N_f)$. The question that we want to ask is: what becomes of this chiral symmetry?

3.2.1 The Quark Condensate

There are two striking phenomena in QCD-like theories. The first is confinement. The second, which at first glance seems less dramatic, is the formation of a *quark condensate*, also known as a *chiral condensate*.

The quark condensate is a vacuum expectation value of the composite operators $\bar{q}_{L\,i}(x) q_{R\,j}(x)$. (As usual in quantum field theory, one has to regulate coincident operators of this type to remove any UV divergences). It turns out that the strong coupling dynamics of non-Abelian gauge theories gives rise to an expectation value of the form

$$\langle \bar{q}_{L\,i} q_{R\,j} \rangle = -\sigma \delta_{ij} \tag{3.47}$$

Here σ is a constant which has dimension of $[\text{Mass}]^3$ because a free fermion in $d = 3+1$ has dimension $[\psi] = \frac{3}{2}$. (An aside: in Section 3.1 we referred to the string tension as σ; it's not the same object that appears here.) The only dimensionful parameter in our theory is the strong coupling scale Λ_{QCD}, so we expect that parameterically $\sigma \sim \Lambda_{QCD}^3$, although they differ by some order 1 number.

The first question to ask is: why does the condensate (3.47) form? The honest answer is: we don't know. It is, like confinement and many other properties of strongly coupled gauge theories, an open question. It turns out that the formation of the condensate is implied by confinement, a statement that we will prove in Section 4.3. We will also give some very heuristic and hand-waving intuition for the formation of the condensate shortly.

Of more immediate concern are the consequences of the condensate (3.47). This is surprisingly easy to answer because as we now explain, everything is entirely determined by symmetry.

The key point is that, while our theory enjoys the full symmetry group (3.46), the vacuum does not. This is because, under G_F, the condensate (3.47) transforms as

$$\langle \bar{q}_{L\,i} q_{R\,j} \rangle \to -\sigma (L^\dagger R)_{ij}$$

This means that massless QCD exhibits a dynamical spontaneous symmetry breaking which, in the present context, is known as *chiral symmetry breaking* (sometimes shortened to χSB). We see that the condensate remains untouched only when $L = R$. This tells us that the symmetry breaking pattern is

$$G_F = U(1)_V \times SU(N_f)_L \times SU(N_f)_R \;\to\; U(1)_V \times SU(N_f)_V \tag{3.48}$$

where $SU(N_f)_V$ is the diagonal subgroup of $SU(N_f)_L \times SU(N_f)_R$.

At this stage, a large part of the physics follows from our general discussion of symmetry breaking in Section 2.2. There will necessarily be a manifold of ground states (2.61), given by the coset

$$\mathcal{M}_0 = \left[SU(N_f)_L \times SU(N_f)_R\right]/SU(N_f)_V \ . \tag{3.49}$$

The number of massless Goldstone bosons is given by the dimension

$$\dim \mathcal{M}_0 = N_f^2 - 1 \ . \tag{3.50}$$

This means that, if we pretend that we have $N_f = 2$ massless quarks (up and down), then we should find 3 massless Goldstone bosons in our world. We will soon identify these with light mesons known as *pions*. If we're happy to be bold and think that there are really $N_f = 3$ (up, down, and strange), then we should find 8 massless Goldstone bosons. These additional Goldstone bosons are not-so-light mesons called *kaons* and the *eta*.

In our world the pions are not massless. But this is because the constituent quarks are not exactly massless so the chiral symmetry is not exact. Nonetheless, the chiral symmetry is an approximate symmetry which, in turn, means that the would-be Goldstone bosons are light, but not exactly massless. Indeed, the pions are notably lighter than all other hadrons in QCD. We'll look more closely at the details as this section proceeds.

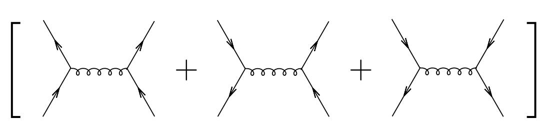
At a more theoretical level, we learn something interesting. Yang-Mills theory has a mass gap. But massless QCD, at least for $N_f \geq 2$ where there is a non-Abelian global symmetry, does not. Even if the theory confines, giving massive baryons and glueballs, chiral symmetry breaking means that there are massless Goldstone bosons.

How to Think About the Quark Condensate

The existence of a quark condensate (3.47) is telling us that the vacuum of space is populated by quark-anti-quark pairs. Again, there is an analogy with superconductivity, albeit with the part of superconductivity that we did not discuss in Section 2.3.2. In a superconductor, the Cooper pairing means that the vacuum is populated by electron pairs. Importantly, these are really electron pairs, rather than electron-hole pairs, which is responsible for the breaking of $U(1)_{\rm em}$. In contrast, the QCD vacuum contains quark-anti-quark pairs so the overall $U(1)_V$ survives, and it's the chiral symmetry that is broken.

In a superconductor, the instability to formation of an electron condensate is a result of the existence of a Fermi surface, together with a weak attractive force mediated by phonons. In the vacuum of space, however, things are not so easy. The formation of a quark condensate does not occur in weakly coupled theory. Indeed, this follows on dimensional grounds because, as we mentioned above, the only relevant scale in the game is Λ_{QCD}.

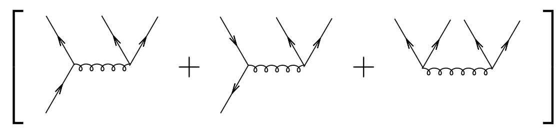
To gain some intuition for why a condensate might form, let's look at what happens at weak coupling $g_s^2 \ll 1$. Here we can work perturbatively and see how the gluons change the quark Hamiltonian. There are two, qualitatively different effects. The first is the kind that we already met in Section 3.1.4; a tree level exchange of gluons gives rise to a force between quarks. This takes the form

$\Delta H_1 = g_s^2$ 

As we saw in Section 3.1.4, the upshot of these diagrams is to provide a repulsive force between two quarks in the symmetric channel, and an attractive force in the anti-symmetric channel. Similarly, a quark-anti-quark pair attract when they form a colour singlet and repel when they form a colour adjoint.

The second term is more interesting for us. The relevant diagrams take the form

$$\Delta H_2 = g_s^2$$



The novelty of these terms is that they provide matrix elements which mix the empty vacuum with a state containing a quark-anti-quark pair. In doing so, they change the total number of quarks + anti-quarks.

The existence of the quark condensate (3.47) is telling us that, in the strong coupling regime, terms like ΔH_2 dominate. The resulting ground state has an indefinite number of quark-anti-quark pairs. It is perhaps surprising that we can have a vacuum filled with quark-anti-quark pairs while still preserving Lorentz invariance. To do this, the quark pairs must have opposite quantum numbers for both momentum and angular momentum. Furthermore, we expect the condensate to form in the attractive colour singlet channel, rather than the repulsive adjoint.

The handwaving remarks above fall well short of demonstrating the existence the quark condensate. So how do we know that it actually forms? Historically, it was first realised from experimental considerations since it explains the spectrum of light mesons; we will describe this in some detail in Section 3.3. At the theoretical level, the most compelling argument comes from numerical simulations on the lattice. However, a full analytic calculation of the condensate is not yet possible. (For what it's worth, the situation is somewhat better in certain supersymmetric non-Abelian gauge theories where one has more control over the dynamics and objects like quark condensates can be computed exactly.) Finally, there is a beautiful, but rather indirect, argument which tells us that the condensate (3.47) must form whenever the theory confines. We will give this argument in Section 4.3.

3.2.2 The Chiral Lagrangian

Chiral symmetry breaking implies the existence of Goldstone bosons. Our next task is to construct the theory that describes these massless particles. This too is dictated entirely by the symmetry structure of the theory.

As we've seen, in any theory with a spontaneously broken continuous symmetry, there is a manifold of ground states $\mathcal{M}_0$ which, for us, is given by (3.49). The different points in $\mathcal{M}_0$ are parameterised by the condensate which, in general, takes the form

$$\langle \bar{q}_{L\,i} q_{R\,j} \rangle = -\sigma\, U_{ij}$$

where $U = L^\dagger R \in SU(N_f)$. The Goldstone bosons are long-wavelength ripples of the condensate where it's value now varies in space and time: $U = U(x)$. As we've seen, there are $N_f^2 - 1$ such Goldstone bosons, one for each broken generator in (3.48). We parameterise these excitations by writing

$$U(x) = \exp\left(\frac{2i}{f_\pi}\,\pi(x)\right) \quad \text{with} \;\; \pi(x) = \pi^a(x)\, T^a \ . \tag{3.51}$$

Here $\pi(x)$ is valued in the Lie algebra $su(N_f)$. The matrices T^a_{ij} are the generators of the $su(N_f)$. (Note: we've changed notation here: previously we denoted Lie algebra generators as T^A, with a capital A index. But having capital letters as indices is offensive and this particular index will proliferate. Hence the change. To make things worse, in other chapter the index a was used to denote colour. Not so here.)

We will collectively refer to the component fields $\pi^a(x)$, labelled by $a = 1, \ldots, N_f^2 - 1$ as *pions*, although strictly this terminology is only accurate for $N_f = 2$. Indeed, in the case of $N_f = 2$, we can expand the field π in generators of $SU(2)$ and write

$$\pi = \frac{1}{2}\begin{pmatrix} \pi^0 & \sqrt{2}\pi^- \\ \sqrt{2}\pi^+ & -\pi^0 \end{pmatrix} \ . \tag{3.52}$$

We will later identify the field π^0 with the neutral pion, and $\pi^\pm$ with charged pions. (We'll give the extension to $N_f = 3$, for which the Goldstone bosons are pions, kaons, and a meson called the eta, in Section 3.3.)

We have also introduced a constant f_π in the definition (3.51) with mass dimension $[f_\pi] = 1$. For now, this ensures that the pions have canonical dimensions for scalar fields in four dimensions, $[\pi] = 1$. It is called the *pion decay constant*, although this name makes very little sense purely in the context of QCD because the pions are stable excitations and don't decay. We'll see where the name comes from in Section 5 when we look at the weak force. On general grounds, we expect $f_\pi \sim \Lambda_{\text{QCD}}$. In fact, it is measured to be around $f_\pi \approx 130$ MeV.

The Low-Energy Effective Action

We want to construct a theory that governs the Goldstone bosons U. We will require that our theory is invariant under the full global chiral symmetry $G_F = U(1)_V \times SU(N_f)_L \times SU(N_f)_R$, under which

$$U(x) \to L^\dagger U(x) R \ . \tag{3.53}$$

What kind of terms can we add to the action consistent with this symmetry? The obvious term is $\mathrm{tr}\, U^\dagger U$ but this doesn't work because $U \in SU(N_f)$ and so $\mathrm{tr}\, U^\dagger U = 1$. (Here we've denoted the trace over the N_f flavour indices as tr to distinguish from the trace Tr over colour indices that we used in the action (3.43).) Happily, this is consistent with the fact that U is a massless Goldstone field.

Next, we can look at kinetic terms. At first glance, it looks as if there are three different candidates:

$$(\mathrm{tr}\, U^\dagger \partial_\mu U)^2 \ , \quad \mathrm{tr}\,(\partial^\mu U^\dagger \partial_\mu U) \ , \quad \mathrm{tr}\,(U^\dagger \partial_\mu U)^2 \ . \tag{3.54}$$

The first term in (3.54) vanishes because $U^\dagger \partial U$ is an $su(N)$ generator and, hence, traceless. Furthermore, we can use the fact that $U^\dagger \partial U = -(\partial U^\dagger)U$ to write the third term in terms of the second. This means that there is a unique two-derivative Lagrangian that describes the dynamics of pions,

$$\mathcal{L}_{\rm pion} = \frac{f_\pi^2}{4}\, \mathrm{tr}\,(\partial^\mu U^\dagger\, \partial_\mu U) \ . \tag{3.55}$$

This is the *chiral Lagrangian*. Although the Lagrangian is very simple, this is not a free theory because U is valued in $SU(N_f)$. This is a kind of non-linear sigma model of the kind we met in Section 2.2. Indeed, this is really the original non-linear sigma model, first introduced by Gell-Mann and Lévy in 1960.

We've constructed our sigma-model to have both $SU(N_f)_L \times SU(N_f)_R$, given in (3.53), as symmetries. But because U is valued in $SU(N_f)$, we cannot just set $U = 0$. Indeed, our sigma-model describes a degeneracy of ground states, but in each of them $U \neq 0$. This ensures that the chiral Lagrangian spontaneously breaks the $SU(N_f)_L \times SU(N_f)_R$ symmetry, as it must. The field U itself is the Goldstone boson associated to this symmetry breaking.

Pion Scattering

The beauty of the chiral Lagrangian is that it contains an infinite number of interaction terms, packaged in a simple form by the demands of symmetry. To see these interactions more explictly, we rewrite the chiral Lagrangian in terms of the pion fields defined in (3.51). Keeping only terms quadratic and quartic, the chiral Lagrangian $\mathcal{L}_{\rm pion}$ becomes

$$\mathcal{L}_{\rm pion} = \mathrm{tr}\,(\partial_\mu \pi)^2 - \frac{2}{3f_\pi^2} \mathrm{tr}\, \left(\pi^2 (\partial_\mu \pi)^2 - (\pi \partial_\mu \pi)^2\right) + \ldots \tag{3.56}$$

Note that if we use $\mathrm{tr}\, T^a T^b = \frac{1}{2}\delta^{ab}$ for $su(N_f)$ generators, then the kinetic term has the standard normalisation for each pion field: $\mathrm{tr}\,(\partial_\mu \pi)^2 = \frac{1}{2}\, \partial^\mu \pi^a \partial_\mu \pi^a$.

For concreteness, we work with $N_f = 2$ and take the $su(2)$ generators to be proportional to the Pauli matrices: $T^a = \frac{1}{2}\sigma^a$. The quartic interaction terms then read

$$\mathcal{L}_{\rm int} = -\frac{1}{6f_\pi^2}\left(\pi^a\pi^a\partial\pi^b\partial\pi^b - \pi^a\partial\pi^a\pi^b\partial\pi^b\right) \ . \tag{3.57}$$

From this we can read off the tree-level $\pi\pi \to \pi\pi$ scattering amplitude using the techniques that we described in the Quantum Field Theory lectures. We label the two incoming momenta as p_a and p_b and the two outgoing momenta as p_c and p_d. The amplitude is

$$\begin{aligned} i\mathcal{A}^{abcd} = \frac{i}{6f_\pi^2}\Big[\ \ & \delta^{ab}\delta^{cd}\Big(4(p_a\cdot p_b + p_c\cdot p_d) + 2(p_a\cdot p_c + p_a\cdot p_d + p_b\cdot p_c + p_b\cdot p_d)\Big) \\ & + (b\leftrightarrow c) + (b\leftrightarrow d)\Big] \ . \end{aligned} \tag{3.58}$$

Momentum conservation, $p_a + p_b = p_c + p_d$, ensures that some of these terms cancel. This is perhaps simplest to see using Mandelstam variables which, because all particles are massless, are defined as

$$\begin{aligned} s &= (p_a+p_b)^2 = 2p_a\cdot p_b = 2p_c\cdot p_d \\ t &= (p_a-p_c)^2 = -2p_a\cdot p_c = -2p_b\cdot p_d \\ u &= (p_a-p_d)^2 = -2p_a\cdot p_d = -2p_b\cdot p_c \ . \end{aligned} \tag{3.59}$$

Using the relation $s+t+u=0$, the amplitude takes the particularly simple form,

$$i\mathcal{A}^{abcd} = \frac{i}{f_\pi^2}\Big[\delta^{ab}\delta^{cd}s + \delta^{ac}\delta^{bd}t + \delta^{ad}\delta^{bc}u\Big] \ . \tag{3.60}$$

There are various ways in which we could improve the description of pion scattering. First, we could include higher loop corrections to the amplitude above. The non-linear sigma model is non-renormalisable which means that we need an infinite number of counterterms to regulate divergences. However, this shouldn't be viewed as any kind of obstacle; the theory is designed only to make sense up to a UV cut-off of order f_π. As long as we restrict our attention to low-energies, the theory is fully predictive.

In addition, we could think about adding higher derivative terms to the chiral Lagrangian. These are corrections that are suppressed by E/f_π where E is the energy of the scattering process. At the next order in the derivative expansion, there are three independent terms:

$$\begin{aligned} \mathcal{L}_4 = & \ a_1\left(\mathrm{tr}\,\partial^\mu U^\dagger\,\partial_\mu U\right)^2 + a_2\left(\mathrm{tr}\,\partial_\mu U^\dagger\,\partial_\nu U\right)\left(\mathrm{tr}\,\partial^\mu U^\dagger\,\partial^\nu U\right) \\ & + a_3\mathrm{tr}\left(\partial_\mu U^\dagger\,\partial^\mu U\partial_\nu U^\dagger\,\partial^\nu U\right) \end{aligned} \tag{3.61}$$

Here a_i are dimensionless coupling constants. There is one further, very important term, known as the Wess-Zumino-Witten (WZW) term that appears at the same order, but can't be written in terms of a 4d action. This is the start of a long and gorgeous story that we won't have time to discuss in these lectures. You can read more about it in the lectures on Gauge Theory.

Currents

We started with quarks and gluons in (3.43) and, at low energies, end up with a very different looking theory of pions (3.55). It's interesting to ask how operators get mapped from one theory to the other. This is particularly straightforward when the operators in question are the currents associated to the $SU(N_f)_L \times SU(N_f)_R$ chiral symmetry.

In the microscopic theory, we have flavour currents for $SU(N_f)_L$ and $SU(N_f)_R$, given by

$$J_L^{a\,\mu} = \bar{q}_{L\,i}\bar{\sigma}^\mu T^a_{ij} q_{L\,j} \quad \text{and} \quad J_R^{a\,\mu} = \bar{q}_{R\,i}\sigma^\mu T^a_{ij} q_{R\,j} \tag{3.62}$$

where T^a_{ij} are $su(N_f)$ generators and the colour and spinor indices have been suppressed. If we write these in terms of the vector and axial combinations: $J_V^{a\,\mu} = J_L^{a\,\mu} + J_R^{a\,\mu}$ and $J_A^{a\,\mu} = J_L^{a\,\mu} - J_R^{a\,\mu}$ then we get the familiar expressions

$$J_V^{a\,\mu} = \bar{q}_i T^a_{ij}\gamma^\mu q_j \quad \text{and} \quad J_A^{a\,\mu} = \bar{q}_i T^a_{ij}\gamma^\mu\gamma^5 q_j \ . \tag{3.63}$$

Now we can ask: what are the analogous expressions for $J_L^{a\,\mu}$ and $J_R^{a\,\mu}$ in the chiral Lagrangian?

To answer this, let's start with $SU(N_f)_L$. Consider the infinitesimal transformation

$$L = e^{i\alpha^a T^a} \approx 1 + i\alpha^a T^a$$

Under this $SU(N_f)_L$, we have $U \to L^\dagger U$ so, infinitesimally,

$$\delta_L U = -i\alpha^a T^a U \ . \tag{3.64}$$

We can now compute the current using the standard trick: elevate $\alpha^a \to \alpha^a(x)$. The Lagrangian is no longer invariant but instead transforms as $\delta\mathcal{L} = \partial_\mu\alpha^a J^a_{L\,\mu}$ and the function $J^a_{L\,\mu}$ is the current that we're looking for. Implementing this, we find

$$J^a_{L\,\mu} = \frac{if_\pi^2}{4}\,\mathrm{tr}\Big(U^\dagger T^a\partial_\mu U - (\partial_\mu U^\dagger)T^a U\Big) \ . \tag{3.65}$$

We can also expand this in pion fields (3.51). To leading order we have simply

$$J^a_{L\,\mu} \approx -\frac{f_\pi}{2}\partial_\mu \pi^a \ . \tag{3.66}$$

Similarly, under $SU(N_f)_R$, we have $\delta U = i\alpha^a U T^a$ and

$$J^a_{R\,\mu} = \frac{if_\pi^2}{4}\Big(- T^a U^\dagger \partial_\mu U + (\partial_\mu U^\dagger) U T^a \Big) \approx +\frac{f_\pi}{2}\partial_\mu \pi^a \ . \tag{3.67}$$

Both currents have non-vanishing matrix elements between the vacuum $|0\rangle$ and a one-particle pion state $|\pi^a(p)\rangle$ that carries momentum p. For example

$$\langle 0|J^a_{L\,\mu}(x)|\pi^b(p)\rangle = -\frac{i}{2} f_\pi \delta^{ab}\, p_\mu e^{-ix\cdot p} \ . \tag{3.68}$$

This tallies with our general discussion of symmetry breaking in (2.2) where we saw that the Goldstone bosons are created by acting with the broken symmetry generators on the vacuum (2.75).

Because the Goldstone bosons are associated to the broken symmetry generators for axial current $J^a_{A_\mu}$, which is a pseudovector, the pions must also be pseudoscalars, meaning that they are odd under parity. We'll look more closely at the quark content of the pions in Section 3.3.

Historically, the approach to thinking of chiral symmetry breaking in terms of currents was known as *current algebra*, and predates our understanding of quarks. The equation (3.68) played a starring role in this story. It is telling us that the chiral $SU(N_f)_L \times SU(N_f)_R$ is spontaneously broken, and acting on the vacuum gives rise to the particles that we call pions. In the language of current algebra, we see that the diagonal combination $SU(N_f)_V$ survives since $\langle 0|J^a_{V\,\mu}|\pi^b\rangle = \langle 0|J^a_{L\,\mu} + J^a_{R\,\mu}|\pi^b\rangle = 0$.

Adding Masses

Our discussion so far has been for massless quarks. That's not particularly realistic. Nonetheless, as we stressed in the introduction to this section, there is reason to expect that the massless limit provides a good jumping off point to understand the physics of light quarks. Our next task is to understand how to incorporate masses.

The QCD action is

$$S = \int d^4x \ \left(-\frac{1}{2}\mathrm{Tr}\, G_{\mu\nu}G^{\mu\nu} + \sum_{i=1}^{N_f} \left(i\bar{q}_i \not{D} q_i - m_i \bar{q}_i q_i \right) \right) \ . \tag{3.69}$$

If the masses are large compared to Λ_{QCD}, then the quarks play no role in the low-energy physics. This is the case for the charm, bottom, and top quarks and we continue to ignore them in what follows.

But for the up, down and (optimistically!) strange quarks, we may assume that the quark condensate (3.47)

$$\langle \bar{q}_{L\,i} q_{R\,j} \rangle \approx -\sigma\, U_{ij} \tag{3.70}$$

continues to form at the scale $\sigma \sim \Lambda^3_{\rm QCD}$, with the masses giving small corrections. We can then incorporate the masses in the chiral Lagrangian by introducing the $N_f \times N_f$ mass matrix,

$$M = \text{diag}(m_1, \ldots, m_{N_f}) \ . \tag{3.71}$$

Because we're now dealing with a low-energy effective theory, the masses that appear here should be the renormalised masses, rather than the bare quark masses quoted earlier in (3.41). In the presence of masses, the leading order chiral Lagrangian is then

$$\mathcal{L}_{\rm pion} = \frac{f_\pi^2}{4}\,\text{tr}\,(\partial^\mu U^\dagger\, \partial_\mu U) + \frac{\sigma}{2}\text{tr}\,\left(MU + U^\dagger M^\dagger\right) \ . \tag{3.72}$$

This lifts the vacuum manifold of the theory. It can be thought of as adding a potential to the vacuum moduli space $\mathcal{M}_0$, resulting in a unique ground state. To see the effect in terms of pion fields, we can again expand $U = e^{2i\pi/f_\pi}$, to find

$$\mathcal{L}_2 = \text{tr}\,(\partial\pi)^2 - \frac{\sigma}{f_\pi^2}\,\text{tr}(M + M^\dagger)\pi^2 + \ldots \tag{3.73}$$

and we see that we get a mass term for the pions as expected. These almost-Goldstone bosons are sometimes referred to as *pseudo-Goldstone bosons*.

For example, if we restrict to $N_f = 2$, we have $M = \text{diag}(m_d, m_u)$. Then, expanding the matrix π in terms of the component fields (3.52),

$$\pi = \frac{1}{2}\begin{pmatrix} \pi^0 & \sqrt{2}\pi^- \\ \sqrt{2}\pi^+ & -\pi^0 \end{pmatrix} \ . \tag{3.74}$$

the quadratic terms in (3.73) become

$$\mathcal{L}_2 = \frac{1}{2}\partial_\mu\pi^0\partial^\mu\pi^0 + \partial_\mu\pi^+\partial^\mu\pi^- - \frac{\sigma}{2f_\pi^2}\,(m_d + m_u)\left((\pi^0)^2 + 2\pi^+\pi^-\right) \ . \tag{3.75}$$

We see that all three pions get an equal mass, given by

$$m_\pi^2 = \frac{\sigma}{f_\pi^2}(m_u + m_d) \ . \tag{3.76}$$

We learn that the square of the pion mass scales linearly with the quark masses. This is known as the *Gell-Mann-Oakes-Renner relation*. The proportionality constant is the (so-far undetermined) ratio σ/f_π^2.

3.2.3 Phases of Massless QCD

Throughout this section, we've couched our discussion in the broader context of a gauge theory with $G = SU(N_c)$ Yang-Mills, coupled to N_f flavours of massless quarks. Obviously, if our interest is in the real world then we can focus on $N_c = 3$ and $N_f = 2$ or 3, depending on taste. But there's a broader theoretical question that we could ask which is: what is the low-energy physics of the theory with general N_c and N_f?

In this section, we take a quick detour to explain what's known. As we will see, there are a number of open questions.

We start with low N_f:

- When $N_f = 0$, we have pure Yang-Mills. The theory sits in the confining phase, with a mass gap.
- When $N_f = 1$, there is no chiral symmetry group (3.46) and so no chiral symmetry breaking. The theory is again thought to have a mass gap, with quarks bound in mesons and baryons.
- When $2 \leq N_f \leq N^\star$ the theory confines and exhibits chiral symmetry breaking. This means that the low energy theory consists of freely interacting Goldstone bosons, parameterising the moduli space (3.49).

  The big question here is: what is the maximum value $N^\star$ for which chiral symmetry breaking occurs? We don't know the answer to this. Various approaches, including numerics, suggest that it is somewhere around

  $$N^\star \approx 4N_c$$

  This means that, for the $N_f = 2$ or 3 of QCD, we are firmly in the chiral symmetry breaking regime. But, in general, our lack of knowledge of this simple question highlights just how poorly we understand strongly interacting field theories.

Now let's jump to high values of N_f and we'll then try to fill in the details in the middle.

- When $N_f \geq \frac{11}{2}N_c$, the beta function is positive. You can see this from the general expression for the beta function (3.12),

  $$b_0 = \frac{11}{3}N_c - \frac{2}{3}N_f \ . \qquad (3.77)$$

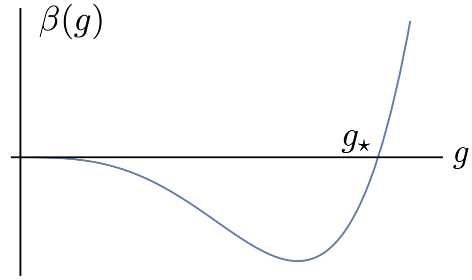


Figure 10. The beta function for N_f slightly below the asymptotic freedom bound has a zero which indicates the existence of an interacting conformal field theory.

This means that the theory is weakly coupled in the infra-red: the low-energy physics consists of massless gluons, weakly interacting with massless quarks. As we go to smaller and smaller energies, the interactions become weaker and weaker. Strictly speaking, in the far IR, the physics is free.

On the flip side, these theories become arbitrarily strongly coupled in the UV, with the gauge coupling diverging at some very high scale. This doesn't mean that we should discard them, but they don't make sense at arbitrarily high energy scales. Said another way, we can't take the UV cut-off Λ_{UV} to infinity while keeping any low-energy interactions. Nonetheless, it's quite possible that these theories may arise as the low-energy limit of some other theory.

That leaves us with the physics in the middle region. We'll keep working down from the asymptotic freedom bound $11N_c/2$.

- When $N^{\star\star} < N_f < \frac{11}{2}N_c$, things are more interesting. To see what happens, we need the two-loop beta function

$$\beta(g) = -\frac{b_0}{(4\pi)^2}g^3 - \frac{b_1}{(4\pi)^4}g^5 + \ldots \tag{3.78}$$

with the one-loop coefficient b_0 given in (3.77) and the two-loop coefficient

$$b_1 = \frac{34N_c^2}{3} - \frac{N_f(N_c^2-1)}{N_c} - \frac{10N_fN_c}{3}\ . \tag{3.79}$$

In the window of interest, $b_0 > 0$ and $b_1 < 0$, so we can play the one-loop contribution against the two-loop contribution to find a zero of the beta function

$$g_\star^2 = -(4\pi)^2\frac{b_0}{b_1}\ . \tag{3.80}$$

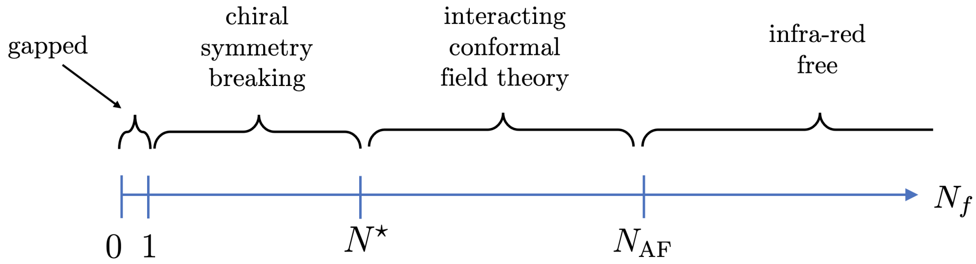


Figure 11. The expected phases of massless QCD. The asymptotic freedom bound is $N_f = \frac{11}{2}N_c$. The lower edge of the conformal window is not known but is expected to be somewhere around $N_f \approx 4N_c$.

with $\beta(g_\star) = 0$. The beta function is shown in Figure 10. The existence of such a fixed point is telling us that we have an interacting conformal field theory: there are massless modes, but they are no longer free in the infra-red. This is known as the *Banks-Zaks fixed point*.

Importantly, when N_f lies just below the asymptotic freedom bound, so $N_f/N_c = 11/2-\epsilon$, this fixed point lies at $g_\star \ll 1$ which means that we can trust the analysis without having to worry about higher order corrections. Moreover, because $g_\star$ is small we can use perturbation theory to calculate anything that we want.

However, as N_f decreases, the value of the fixed point $g_\star$ increases until we can no longer trust the analysis above. The expectation is that we get a conformal field theory only for some range of N_f, lying within $N^{\star\star} < N_f < \frac{11}{2}N_c$. This is known as the *conformal window*. We don't currently know the value of $N^{\star\star}$.

That leaves us with understanding what happens in the middle when $N^\star < N_f \leq N^{\star\star}$. Our best guess is that there is no such regime, and the upper edge of the chiral symmetry breaking phase coincides with the lower edge of the conformal window,

$$N^{\star\star} = N^\star$$

This guess is motivated partly by numerics and partly by a lack of any compelling alternative. For us, the lesson to take away is that strongly interacting quantum field theories are hard and even the most basic questions are beyond our current abilities. A summary of the expected behaviour of massless QCD is shown in Figure 11.

Quark	Charge	Mass (in MeV)
d = down	-1/3	5
u = up	+2/3	2
s = strange	-1/3	93
c = charm	+2/3	1270
b = bottom	-1/3	4200
t = top	+2/3	170,000

Table 3. The quarks

3.3 Hadrons

Confinement means that quarks are bound into colour singlets. There are two group-theoretic possibilities: quark-anti-quark pairs, known as *mesons*, or a collection of three quarks known as *baryons*. Collectively these particles are called *hadrons*[7].

Much of hadron physics is messy and complicated. Some balm comes, once again, from symmetries. Recall that, if we assume that quarks are massless, then the global symmetry exhibits the symmetry breaking pattern

$$U(1)_V \times SU(N_f)_L \times SU(N_f)_R \rightarrow U(1)_V \times SU(N_f)_V \ . \tag{3.81}$$

The broken generators give rise to pions and other Goldstone bosons, and we'll see how these arise in terms of quarks shortly. But, for now, our interest lies in the surviving $SU(N_f)_V$ symmetry. This is what we will use to organise the spectrum of hadrons.

We don't need the quarks to be massless to get an $SU(N_f)$ symmetry: we just need their masses to all be equal. Their masses, together with their electric charges, are presented in Table 3.

It seems very reasonable to view $m_{\rm up} \approx m_{\rm down}$, at least to a first approximation. (Remember that we're comparing these values against $\Lambda_{\rm QCD} \approx 200$ MeV.) And, indeed, we will see that there is a clear $SU(2)_V$ symmetry in the hadronic spectrum. This was first identified by Heisenberg, who noted that the proton and neutron have almost identical interactions with the strong force, and is known as *isospin*. (Not a great name as it has nothing to do with "spin".)

[7]I strongly recommend that you take a look, even a brief one, at the booklet published by the Particle Data Group to get a sense for the hadronic world that lies beneath you.

Meanwhile, despite the obvious difference in the strange quark mass, there's also a very visible, albeit approximate, $SU(3)_V$ symmetry in the hadronic spectrum. This was observed, independently, by Gell-Mann and Ne'eman in 1961 and is known as the *eightfold way*. (Because $\dim SU(3) = 8$.) Note that this $SU(3)_V$ has nothing to do with the gauge group $SU(3)$ of QCD. It is an entirely different (and approximate) global $SU(3)_V$ that rotates the different flavours of light quarks.

There are other symmetries of QCD that we can use to assign quantum numbers to particles. These are rotations, corresponding to angular momentum or spin of the particle J, parity, and charge conjugation, both of which are symmetries of QCD, albeit not of the full Standard Model. Particles often come with a label J^{PC}, where $P = \pm$ denotes that the state is even or odd parity and $C = \pm$ denotes even or odd under charge conjugation, which is typically called C-parity in this context.

(As an aside: if you look through the particle data book, you'll sometimes see the additional quantum numbers I^G. Here I is the I_3 eigenvalue of isospin. So for example, particles come in $I = \pm\frac{1}{2}$ pairs if they sit in a double of isospin. Meanwhile G stands for G-parity which is the combination $G = Ce^{i\pi I_2}$ where the isospin rotation is designed to send $I_3 \mapsto -I_3$.)

In the rest of this section, we will describe the hadrons that contain up, down, and strange quarks, and see how they furnish representations of the $SU(3)_V$ flavour symmetry. We then finish by looking at the kinds of particles we can make with heavy charm, bottom, and top quarks.

3.3.1 Mesons

Many hundreds of mesons are observed in nature. A simple model of a meson views it as a bound state of a quark and an anti-quark, or some linear combination of these states. .Each quark is a fermion, so mesons are bosons and, as such, have integer spin. Here we will describe some of the lightest mesons with spin 0 and 1, containing only up, down and strange quarks.

Our three flavours of quarks (d, u, s) transform in the $\mathbf{3}$ of $SU(3)_V$. A little group theory tells us that quark and anti-quark must then transform in

$$\mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{1} \oplus \mathbf{8} \ . \tag{3.82}$$

So we expect mesons to sit in two representations of $SU(3)_V$: the singlet $\mathbf{1}$ and the adjoint $\mathbf{8}$.

Meson	Quark Content	Mass (in MeV)	Lifetime (in s)
pion π^+	$u\bar{d}$	140	10^{-8}
pion π^0	$\frac{1}{\sqrt{2}}(u\bar{u} - d\bar{d})$	135	10^{-16}
eta η	$\frac{1}{\sqrt{6}}(u\bar{u} + d\bar{d} - 2s\bar{s})$	548	10^{-19}
eta Prime η'	$\frac{1}{\sqrt{3}}(u\bar{u} + d\bar{d} + s\bar{s})$	958	10^{-21}
kaon K^+	$u\bar{s}$	494	10^{-8}
kaon K^0	$d\bar{s}$	498	$10^{-8} - 10^{-11}$

Table 4. The pseudoscalar mesons

Pseudoscalar Mesons

We first look at the lowest mass mesons with spin 0. We get total spin zero if the individual spins of the quarks are anti-aligned, and the particles have zero orbital angular momentum. We saw in Section 1.4 that if a fermion has parity +1 then the anti-fermion has parity −1, which means that the spin 0 meson has odd parity. We write $J^{PC} = 0^{-+}$.

We first give the experimental data for these mesons, and we will then see how they fit into what we know. The names, quark content, masses, and lifetime of the lightest pseudoscalar mesons are shown in Table 4. The $\pm$ and 0 superscripts tell us the electromagnetic charge of the meson. The charged mesons, π^+ and K^+ both have anti-particles, π^- and K^- respectively. The neutral mesons π^0, η and η' are all their own anti-particles; each is described by a real scalar field. Finally, the neutral K^0 is described by a complex scalar field and its anti-particle is denoted $\bar{K}^0$. This means that there are 9 different meson states in total, in agreement with our simple expectation (3.82).

First, an obvious comment: the masses of the mesons are not equal to the sum of the masses of their constituent quarks! We already anticipated this from our analysis of the chiral Lagrangian and the Gell-Mann-Oakes-Renner relation (3.76). This gets to the heart of what it means to be a strongly coupled quantum field theory. The mesons – and, indeed the baryons – are complicated objects, consisting of a bubbling sea of gluons, quarks and anti-quarks. This is what gives mesons and baryons mass, and also makes these particles hard to understand.

The nine different meson states can be decomposed into the $\mathbf{1} \oplus \mathbf{8}$ multiplets by writing

$$\begin{pmatrix} u \\ d \\ s \end{pmatrix} \otimes (u, d, s) = \begin{pmatrix} u\bar{u} & u\bar{d} & u\bar{s} \\ d\bar{u} & d\bar{d} & d\bar{s} \\ s\bar{u} & s\bar{d} & s\bar{s} \end{pmatrix} = \mu_0 \mathbb{1} + \sum_{a=1}^{8} \mu_a \lambda_a \ . \tag{3.83}$$

Here λ_a are the Gell-Mann matrices (3.6), now in their role as the generators of $SU(3)_V$. We'll ignore the singlet μ_0 for now and focus on the mesons that sit in the $\mathbf{8}$. These are precisely the would-be Goldstone bosons that we met previously. The various fields μ_a naturally rearrange themselves into two real and three complex fields that we call *pions*, *kaons*, and the *eta meson*,

$$\pi^0 = \mu_3 \ , \quad \pi^\pm = \frac{1}{\sqrt{2}}(\mu_1 \mp i\mu_2) \tag{3.84}$$

$$K^0 = \frac{1}{\sqrt{2}}(\mu_6 - i\mu_7) \ , \quad K^\pm = \frac{1}{\sqrt{2}}(\mu_4 \mp i\mu^5) \ , \quad \eta = \mu_8 \ .$$

The matrix (3.83) is identified with the Goldstone boson matrix that we met in the previous section. We previously wrote this in (3.52) for $N_f = 2$ quarks. The extension to $N_f = 3$ quarks is

$$\pi = \frac{1}{2}\sum_{a=1}^{8} \mu_a \lambda^a = \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{\pi^0}{\sqrt{2}} + \frac{\eta}{\sqrt{6}} & \pi^+ & K^+ \\ \pi^- & -\frac{\pi^0}{\sqrt{2}} + \frac{\eta}{\sqrt{6}} & K^0 \\ K^- & \bar{K}^0 & -\frac{2\eta}{\sqrt{6}} \end{pmatrix} \tag{3.85}$$

You can check that this reproduces the quark content shown in Table 4. If the masses of the three quarks were equal, then these 8 particles would all have the same mass.

The group theoretic underpinnings of these mesons encourages us to draw them on an $SU(3)$ weight diagram, as shown in Figure 12. The charges under the two $U(1)^2 \subset SU(3)_V$ Cartan elements are also shown. These are taken to be isospin $I_3 \subset SU(2)_V \subset SU(3)_V$ and "strangeness" S which effectively counts the number of strange quarks in the meson. A suitable combination, shown on the diagonal, gives the electric charge Q. These are exact quantum numbers in QCD (but not when we include weak interactions) and, historically, it was by observing their conservation in dynamical processes, such as particle decays, that the pattern above was identified.

If we compare pions to kaons, we see from the data that the addition of a strange quark adds about 350 MeV to the mass of a meson. That's significantly more than the

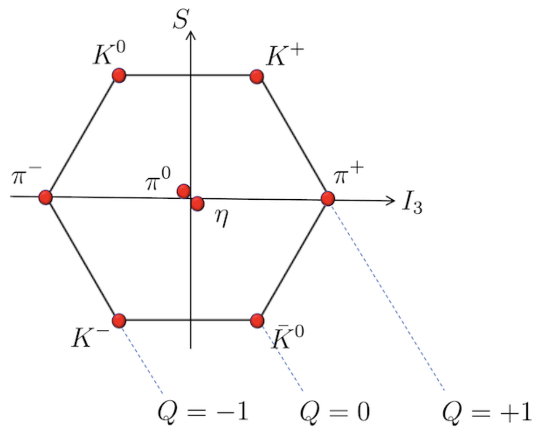


Figure 12. The eightfold way for pseudoscalar (and pseudo-Goldstone) mesons.

bare mass of ~ 100 MeV of a strange quark. Again, this highlights the difficulty of strongly interacting field theories: you don't just read off the physics from the classical Lagrangian.

We can make some progress by looking at the mesons through the lens of the chiral Lagrangian. We return to the massive Lagrangian (3.73), now with the mass matrix $M = \mathrm{diag}(m_u, m_d, m_s)$. Again, I stress that these should be renormalised masses, not bare masses. Expanding out the action using (3.85), we find the masses

$$\mathcal{L}_{\mathrm{mass}} = \frac{-\sigma}{f_\pi^2}\left[\frac{1}{2}(m_u+m_d)\left((\pi^0)^2+2\pi^+\pi^-\right)+(m_u+m_s)K^-K^+\right. \tag{3.86}$$
$$\left.+(m_d+m_s)\bar{K}^0K^0+\frac{1}{2}\left(\frac{m_u}{3}+\frac{m_d}{3}+\frac{4m_s}{3}\right)\eta^2+\frac{1}{\sqrt{3}}(m_u-m_d)\pi^0\eta\right] .$$

This generalises our previous result (3.75). Note that there is mixing between π^0 and η, albeit one that disappears when $m_u = m_d$ so that isospin is restored. By taking ratios, we can eliminate the overall scale σ/f_π^2 and relate meson and quark masses directly. For example, we have

$$\frac{m_{K^+}^2 - m_{K^0}^2}{m_\pi^2} = \frac{m_u - m_d}{m_u + m_d} \tag{3.87}$$

We can also derive expected relationships between the meson masses. For example, we have $3m_\eta^2 + m_\pi^2 = \frac{2\sigma}{f_\pi^2}\left(2(m_u+m_d)+4m_s\right)$. If we accept that $m_u \approx m_d$, then we get the relation

$$4m_K^2 \approx 3m_\eta^2 + m_\pi^2\ . \tag{3.88}$$

This is known as the *Gell-Mann-Okubo relation*. Comparing against the experimentally measured masses, we have $\frac{1}{2}\sqrt{3m_\eta^2 + m_\pi^2} \approx 480$ MeV, which is not far off the measured value of $m_K \approx 495$ MeV.

So far, there is one scalar meson that we've not yet discussed. This is the singlet in the decomposition $\mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{1} \oplus \mathbf{8}$, associated to the field μ_0 in (3.83). This field corresponds to the meson η', pronounced *eta-prime*,

$$\eta' = \frac{1}{\sqrt{3}}(u\bar{u} + d\bar{d} + s\bar{s}) \ . \tag{3.89}$$

From Table 4, we see that this is by far the heaviest of the scalar mesons. This is because, in contrast to the other mesons, it is not a pseudo-Goldstone boson: if you sent the quark masses to zero, then the pions and kaons and eta all become massless. The eta-prime remains massive.

In fact, there's more to the story of the eta-prime. Recall that back in Section 3.2, we mentioned that the classical Lagrangian of massless QCD also has an axial $U(1)_A$ symmetry. Naively, it appears as if this too is spontaneously broken by the condensate (3.47) . If this were true, the eta-prime meson would be the corresponding pseudo-Goldstone boson, in which case we have a puzzle on our hands because it seems too heavy to be Goldstonesque.

The answer to this puzzle will be presented in Section 4 where we'll see that $U(1)_A$, while a symmetry of the classical action, is not a symmetry of the quantum theory because it suffers something called an anomaly. The fact that the eta-prime is inordinately heavy is one consequence of this.

Pseudovector Mesons

This same pattern of $\mathbf{1} \oplus \mathbf{8}$ repeats many more times in excited meson states, in which the spins of the quarks are aligned (rather than anti-aligned) or the quarks have some additional relative orbital angular momentum L. The total parity of these excited meson states is $P = (-1)^{L+1}$.

The first such collection occurs when the spins are aligned, but $L = 0$, giving a collection of 9 pseudovector mesons with $J^{PC} = 1^{--}$, as listed in Table 5. The lightest of these spin 1 mesons are the rhos, $\rho^\pm$ and ρ^0, which can be viewed as excited pions. The heaviest is the phi meson, which is again the singlet $\mathbf{1}$. Note that by the time we get to the excited kaons, some naming exhaustion has set in, and the fact that these are excited states is denoted merely by the addition of a star.

Meson	Quark Content	Mass (in MeV)	Lifetime (in s)
rho ρ^+	$u\bar{d}$	770	10^{-24}
rho ρ^0	$\frac{1}{\sqrt{2}}(u\bar{u} - d\bar{d})$	770	10^{-24}
omega ω	$\frac{1}{\sqrt{2}}(u\bar{u} + d\bar{d})$	780	10^{-22}
phi ϕ	$s\bar{s}$	1020	10^{-22}
kaon $K^{+\star}$	$u\bar{s}$	890	10^{-24}
kaon $K^{0\star}$	$d\bar{s}$	890	10^{-24}

Table 5. The pseudovector mesons

If you look closely at the quark content of the scalar and vector mesons, you'll see that the analogy between them isn't quite perfect. In particular, the excited versions of the η and η' are the ω and ϕ. But the quark content of the pseudoscalar mesons is

$$\eta : \ \frac{1}{\sqrt{6}}(u\bar{u} + d\bar{d} - 2s\bar{s}) \quad \text{and} \quad \eta' : \ \frac{1}{\sqrt{3}}(u\bar{u} + d\bar{d} + s\bar{s}) \tag{3.90}$$

while the quark content of the pseudovector mesons is:

$$\omega : \ \frac{1}{\sqrt{2}}(u\bar{u} + d\bar{d}) \quad \text{and} \quad \phi : \ s\bar{s} \ . \tag{3.91}$$

What's going on? Why are these so different?

This is an issue of particle mixing, something that we will see more of when we come to discuss the weak force and neutrinos. First note that the quantum numbers of η and η' are the same (in particular, $I_3 = S = 0$ and hence $Q = 0$ for both). Similarly for the ω and ϕ. In any quantum mechanical system, if you have states with the same quantum numbers then you have to diagonalise the Hamiltonian to find the energy (or in this case, mass) eigenstates. That can lead to linear superpositions of the original states.

That's what's going on here. There are two competing aspects at play. One is the $SU(3)_V$ flavour symmetry that pushes the energy eigenstates to form as $\mathbf{1} \oplus \mathbf{8}$ multiplets, which results in the quark content seen in the pseudoscalars (3.90). The other is the bare mass terms of the quarks, that prefers the energy eigenstates to be the more straightforward $q\bar{q}$. For both pseudoscalar and pseudovector mesons there is some competition between these, meaning that neither (3.90) nor (3.91) is entirely correct. Instead, the honest answer is that the quark content is some linear combination of the two results in both cases, but the group theory dominates for the pseudoscalars, while the mass difference of the strange quark dominates for the pseudovectors.

Of course, this still begs the question of why scalar mesons fall one way, and vectors the other. This is, like many things in QCD, complicated, but it boils down to the fact that the scalar mesons are would-be Goldstone bosons.

Note that masses don't entirely get their own way for the vector mesons. The ρ^0 and ω have constituents $u\bar{u} \pm d\bar{d}$, rather than $u\bar{u}$ and $d\bar{d}$, so the $SU(2)_V$ isospin symmetry is still powerful enough to hold sway over the up/down mass difference.

If you flip through the particle data group booklet, you will find further collections of excitations with $J^{PC} = 0^{++}$ around 1150 MeV. These have orbital angular momentum $L = 1$ and spin $S = 1$ and are given catchy names like a_0, a_1, etc. Then there are states with with $J^{PC} = 1^{+-}$ at around 1250 MeV that have $L = 1$ and $S = 0$. These have equally catchy names $b_0, b_1, \ldots$. And so it continues.

3.3.2 Lifetimes

So far we've not said anything about the lifetime of mesons, which we also listed in Tables 4 and 5. This is largely because many of these lifetimes are dictated by the weak force that we haven't yet described. Nonetheless, there are a few straightforward comments that we can make here.

The first is that there is a very wide range of lifetimes exhibited by mesons, from the charged pions and kaons which decay in 10^{-8} seconds to the rho which decays in 10^{-24} seconds. This reflects the different ways in which these particles can decay.

For example, despite their similar masses, the neutral and charged pions have rather different lifetimes. The neutral pion decays through the electromagnetic force to two photons

$$\pi^0 \to \gamma + \gamma \; . \tag{3.92}$$

It has a lifetime of around 10^{-16} seconds. In contrast, the charged pions π^+ and π^- decay only through the weak force. We'll see in Section 5 that they typically decay to a muon and a neutrino

$$\pi^+ \to \mu^+ + \nu_\mu \quad \text{and} \quad \pi^- \to \mu^- + \bar{\nu}_\mu \; . \tag{3.93}$$

They live for 10^{-8} seconds, an eternity in the subatomic world and much longer than any of the other hadrons, except for the proton and neutron.

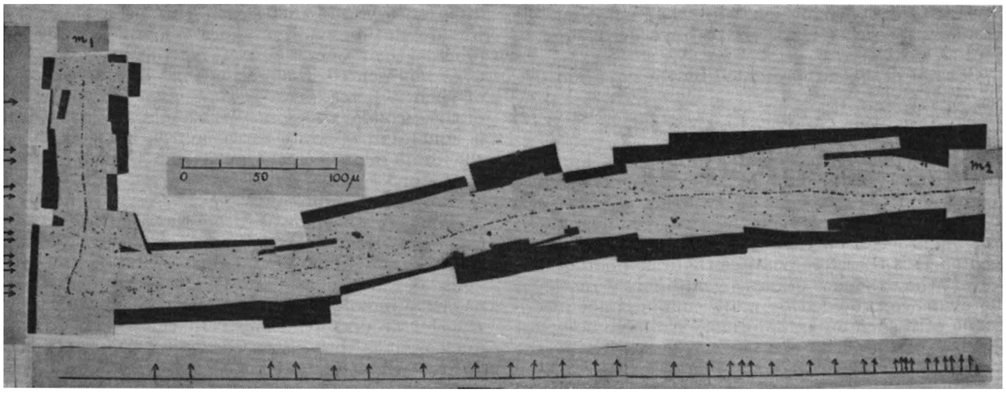


Figure 13. The discovery of the charged pion in 1947. The pion enters in the top left (labelled m_1), slows in the bromide and comes to rest, before decaying into a muon that flies off to the right (labelled m_2) and an anti-neutrino which is invisible in the picture

As a general rule of thumb, each force comes with a characteristic time scale that determines the lifetime of the hadron:

- Strong decay: $\sim 10^{-22}$ to 10^{-24} seconds.
- Electromagnetic decay: $\sim 10^{-16}$ to 10^{-21} seconds.
- Weak decay: $\sim 10^{-7}$ to 10^{-13} seconds.

Where you sit within each range depends on other factors, such as the relative masses of the parent and daughter particles.

In a world with just the strong force, all the pseudoscalar mesons listed in Table 4 would be stable and, despite the fact that some can disappear in 10^{-20} seconds or so, physicists continue to refer to them as stable. In contrast, anything that decays via the strong force is said to be a *resonance*, rather than a particle. All of the vector mesons listed in Table 5 are resonances. For example, the rho decays via the strong force to (predominantly) two pions. If you look through the particle data book, you'll find that resonances are always listed with their mass in brackets. So, for example, you will find $\rho(770)$ in the book but, just above it, η with no brackets.

You'll often find lifetimes quoted in terms of the width, which is an energy scale, rather than a time. The conversion factor is

$$100 \text{ MeV} \approx 10^{-23} \text{ s}^{-1} \ . \tag{3.94}$$

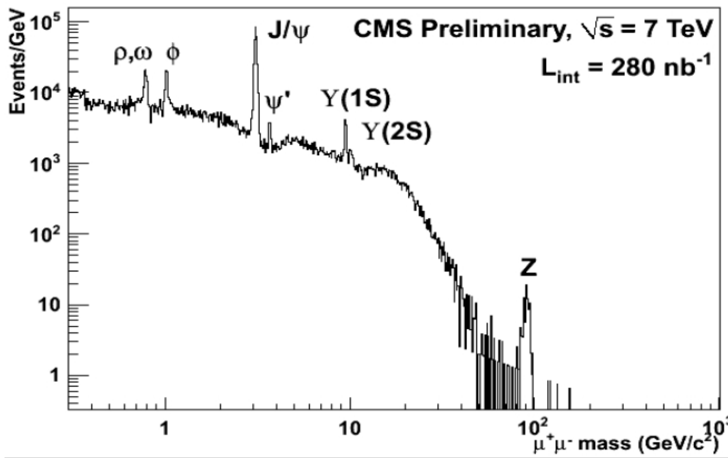


Figure 14. The centre-of-mass energy of $\mu^+\mu^-$ pairs reveals a zoo of mesonic resonances at low energies, with the Z-boson sitting at high energies. This is a plot from 2010 made by the CMS collaboration.

This coincides with what we saw above. The relevant energy scale of the strong force is somewhere around $\Lambda_{\rm QCD} \sim 100$ish MeV and if the strong force does something (like enable a decay), then is typically takes around $T_{\rm QCD} \sim 10^{-23}$ seconds to do it.

Of course, our world has more than the strong force and that means that there's nothing qualitatively different between a particle like the pion and a resonance like the rho. Both will decay in less than the blink of an eye. But it does make a difference for experiments. If something lasts for 10^{-10} seconds then, with good technology, you can take a photograph of the particle's track in a cloud chamber or bubble chamber. For example, the discovery photo of the pion is shown in Figure 13. When a particle leaves such a vivid trace, it's hard to deny its existence. In contrast, we're never going to take a photograph of something that lasts 10^{-20} seconds. But that doesn't mean that it's any less real! It just leaves its signature in more subtle ways, typically as a bump in the cross-section for some process. (See, for example, the chapter on scattering theory in the lectures on Topics in Quantum Mechanics for a discussion of how this comes about.) The glorious plot shown in Figure 14 shows bumps in the number of back-to-back $\mu^+\mu^-$ pairs that were seen in the CMS detector in the early days of the LHC. The resonances start, on the far left, with the ρ, ω and ϕ but then, as the energy increases, there are clear peaks for the J/ψ, which is a charmed meson, the upsilon Υ which is a bottom meson and, far off the right, the Z-boson which is one of the gauge bosons for the weak force.

Finally, hiding within the data are some interesting stories that we will meet again later. For example, the decay of the neutral pion $\pi^0 \to \gamma + \gamma$ is closely tied to the anomaly, and we will revisit this in Section 4.

The lifetime of the neutral kaons also holds an important lesson. Curiously they appear to have two different lifetimes, either 10^{-7} seconds or 10^{-10} seconds, depending on how you count! That's kind of weird. It turns out to be a manifestation of the fact that the weak force violates time-reversal! We will discuss this in Section 5.

The Elusive Sigma

There is one light scalar meson listed in the particle data book that I have not yet mentioned. It has $J^{PC} = 0^{++}$ and goes by the catchy name of $f_0(500)$ and has a mass which is listed as somewhere between 400 - 550 MeV. The reason that it's so difficult to pin down is that it decays very quickly – via the strong force rather than weak force – to two pions and so has a large width. Moreover, it has vanishing quantum numbers (angular momentum, parity, isospin and strangeness are all zero).

Experimentally, its probably best not to refer to this resonance as a particle at all. However, theoretically it has played a very important role, for this is the "sigma" after which the sigma-model is named. It can be thought of as the excitation that arises from ripples in the value of the quark condensate, $\sigma = \bar{\psi}\psi$, rather than rotations in the quark condensate U.

3.3.3 Baryons

Three quarks can form a gauge singlet by anti-symmetrising over their colour indices $a = 1, 2, 3$ to form a baryon,

$$\mathcal{B} = \epsilon^{abc} q_a q_b q_c \; . \tag{3.95}$$

For baryons constructed of light d, u, and s quarks, these too sit in representations of the $SU(3)_V$ flavour symmetry.

We can again do a little group theory. For two quarks we have

$$\mathbf{3} \otimes \mathbf{3} = \bar{\mathbf{3}} \oplus \mathbf{6} \; . \tag{3.96}$$

Adding the third quark, we have

$$\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = (\bar{\mathbf{3}} \otimes \mathbf{3}) \oplus (\mathbf{6} \otimes \mathbf{3}) = \mathbf{1} \oplus \mathbf{8} \oplus \mathbf{8}' \oplus \mathbf{10} \; . \tag{3.97}$$

Importantly, we want to think of these as representations of the $SU(3)_V$ flavour symmetry rather than the $SU(3)$ gauge symmetry. This tells us that we expect baryons to sit in one of the representations above.

Baryon	Quark Content	Mass (in MeV)	Lifetime (in s)
proton p	uud	938	stable
neutron n	udd	940	10^3
lambda Λ^0	uds	1115	10^{-10}
sigma Σ^+	uus	1189	10^{-10}
sigma Σ^0	uds	1193	10^{-19}
sigma Σ^-	dds	1197	10^{-10}
cascade Ξ^0	uss	1315	10^{-10}
cascade Ξ^-	dss	1321	10^{-10}

Table 6. The octet of spin $\frac{1}{2}$ baryons.

At this point, we have to remember that quarks are fermions and, as such, obey the Pauli exclusion principle. We can look at each of the possibilities above in turn:

- The singlet $\mathbf{1}$ is fully anti-symmetrised in flavour indices. But any baryon is necessarily fully anti-symmetrised in colour indices, as shown in (3.95), and the Pauli exclusion principle says that the state must be anti-symmetrised overall. We still have the spin degree of freedom to play with, but it's not possible to fully anti-symmetrise in spin so this baryon must have some orbital angular momentum to satisfy Pauli. That makes it heavy and messy. Candidates exist but we won't discuss them.

- At the other end, the decuplet $\mathbf{10}$ is fully symmetrised in flavour indices and so we can satisfy Pauli by symmetrising over spin degrees of freedom. This means that the decuplet of baryons should have spin $\frac{3}{2}$.

- The $\mathbf{8}$ and $\mathbf{8}'$ are a bit more tricky: one is anti-symmetrised only in the first two indices, the other symmetrised in the first two indices, so we have to work a little harder. But it turns out that we can take a suitable linear combination of them that gives a fully anti-symmetrised wavefunction (including colour) when the quarks have total spin $\frac{1}{2}$.

The octet contains the two most famous baryons: protons and neutrons. Collectively, these are called *nucleons*. Others in this multiplet have a mass that differs by about 30% from that of the nucleons. The Σ baryons contain a single strange quark while the Ξ baryons, known either as *xi* or, with a rhetorical flourish, *cascades*, contain two

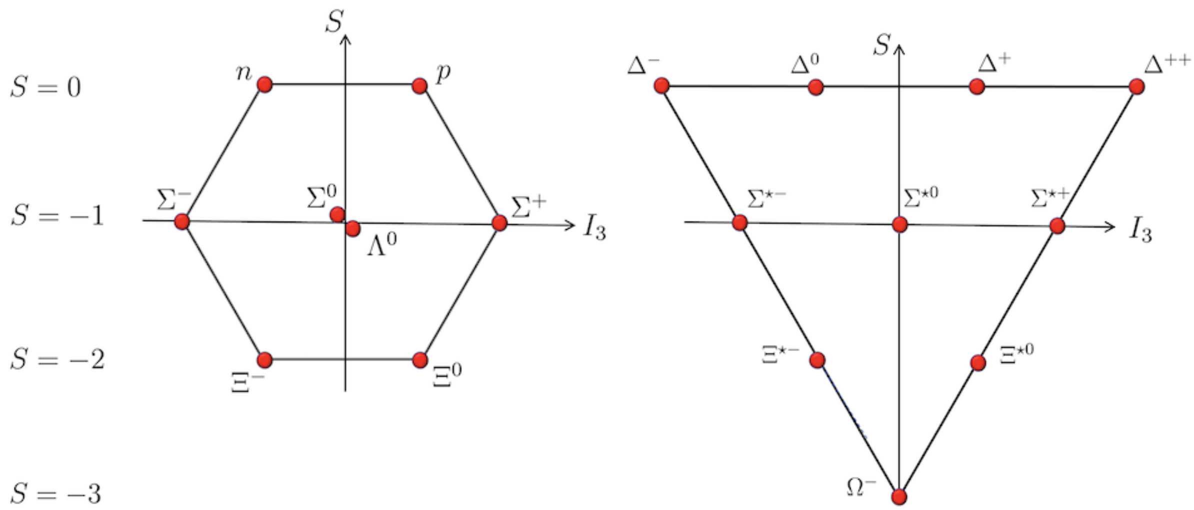


Figure 15. The octet and decuplet of baryons.

strange quarks. The full collection of eight spin $\frac{1}{2}$ baryons are shown in Table 6, and in an $SU(3)$ weight diagram, reflecting their group theoretic origins, in Figure 15.

We saw previously that the octet of pseudoscalar mesons have an interpretation as almost-Goldstone modes. That means, in particular, that if the quarks were massless, then the pions, kaons and eta would all be massless as well. What is the analogous story for the baryons?

Here there is a surprise. If the up and down quark were massless, the mass of the proton and neutron would be more or less unchanged from the values we measure! The mass of the baryons – at least those comprised of light quarks – is not driven by the bare quark mass. Instead, it's driven by the strong coupling scale Λ_{QCD}. In fact, on general grounds one can argue that the mass of baryons in $SU(N_c)$ QCD scales as $N_c\Lambda_{QCD}$.

That's not to say that the mass of the quarks is entirely unimportant. Crucially, the fact that the down quark is heavier than the up quark is the reason why the neutron is heavier than the proton. If this weren't true, the weak force would allow the proton to decay into the neutron, rather than the other way around, and it's hard to see how atoms and chemistry and physicists could exist.

Similarly, the strange baryons are heavier than the proton and neutron. You can see from the data that each strange quark adds about 140±10 MeV to the baryon mass.

That's smaller than the corresponding amount for mesons, but still bigger than the bare mass $m_s \approx 93$ MeV.

You may have heard it said that the Higgs is responsible for all the mass in the universe. This is a blatant lie. In Section 5, we will see that the Higgs is responsible for the mass of all *elementary particles*, meaning the leptons and quarks. But the overwhelming majority of mass in atoms is contained in the protons and neutrons that make up the nucleus, and this mass has nothing to do with the Higgs boson. It is entirely due to the urgent thrashing of strongly interacting quantum fields.

While we're talking about fairytales that we were subjected to when we were young, here's another one: we are usually told that the strong force is what keeps the nucleus together in the atom. This one is kind of true, but only in an indirect way. The strong force binds quarks together into baryons, which are fermions, and into mesons, which are bosons. But, as described in the lectures on Quantum Field Theory, scalar particles mediate forces. In particular, the pions mediate a force of a Yukawa type, with potential

$$V(r) \sim -\frac{e^{-m_\pi r}}{r} \; . \tag{3.98}$$

This is what binds the protons and neutrons together in the nucleus.

We refer to this force mediated by pions as the strong nuclear force, but it would be better to give it a different name — say "mesonic force", or "Yukawa force" — to highlight the fact that it is really a residual, secondary effect. The upshot is that there are two layers to the strong force: we start with one force and a set of matter particles — gluons interacting with quarks — and end up with a very different force and a new set of matter particles — the mesonic force interacting with protons and neutrons. In this sense, both the particles in the nucleus, and the force that holds them together, are *emergent* phenomena, arising from something more fundamental underneath.

Finally, we briefly look at the spin $\frac{3}{2}$ baryons, that sit in the flavour decuplet. They go by the names Δ (with charges 0, ±1 and 2), $\Sigma^\star$ (with charges 0 and ±1), $\Xi^\star$ (with charges -1 and 0) and Ω^- with charge -1. The full list of particles is given in Table 7 and the weight diagram shown in Figure 15.

The real novelties among these baryons are the three outliers, in which all quarks are the same. The Δ^{++} played an important historic role because it was the first particle to be found with charge $+2$ as opposed to 0 or ±1 and helped enormously in piecing

Baryon	Quark Content	Mass (in MeV)	Lifetime (in s)
Δ^{++}	uuu	1232	10^{-24}
Δ^{+}	uud	1232	10^{-24}
Δ^{0}	udd	1232	10^{-24}
Δ^{-}	ddd	1232	10^{-24}
$\Sigma^{\star -}$	dds	1383	10^{-23}
$\Sigma^{\star 0}$	dus	1384	10^{-23}
$\Sigma^{\star +}$	uus	1387	10^{-23}
$\Xi^{\star -}$	dss	1535	10^{-23}
$\Xi^{\star 0}$	uss	1532	10^{-23}
Ω^{-}	sss	1672	10^{-11}

Table 7. The decuplet of spin $\frac{3}{2}$ baryons.

together the story of the underlying quarks. The Ω^- baryon, meanwhile, holds a special place in the history of science because Gell-Mann used the simple quark model described above to predict its mass and properties before it was discovered experimentally. In that way, he followed Mendeleev and Dirac in predicting the existence of a "fundamental" particle of nature (where, as should by now be clear, the meaning of the word "fundamental" is time-dependent).

One of the lessons to take away from this section is that QCD is complicated. We can make some progress by using symmetries (or approximate symmetries) as organising principles, but that only takes us so far. It is natural to wonder how much of the results above we can calculate from first principles, starting from the Lagrangian of QCD.

If your first principles involve only pen and paper, then the answer is: not much. QCD is hard. But if you extend your first principles to embrace numerical simulations which, in this context, go by the name of lattice QCD, then you can do pretty well. After many decades of work, much of the spectrum described above can be computed to within, say, 5% accuracy. There is now no doubt that the complexity seen in the hadron spectrum can be entirely explained by the dynamics of QCD.

3.3.4 Heavy Quarks

So far, we've only discussed the hadrons constructed from the three lightest quarks. We've still to discuss the heavy ones.

It turns out that there are no hadrons comprised of the top quark. Its extreme high mass means that the top quark decays with a lifetime of around 10^{-25} seconds, which is faster than the characteristic timescale $T_{\rm QCD} \approx 10^{-23}$ seconds of the strong force. This means that such "top hadrons" decay before they even form. Needless to say, none have been observed.

That still leaves us with the charm and bottom. The masses of hadrons containing these quarks are determined more by the bare quark mass than by $\Lambda_{\rm QCD}$. Two sets of these mesons deserve a special mention. The first is *charmonium*, a bound state of charm and anti-charm quark. It also goes by the dual name J-psi (J/ψ),

$$J/\psi\ (\bar{c}c) \quad m \approx 3.1 \text{ GeV} . \tag{3.99}$$

Its lifetime is around 10^{-21} seconds. The discovery of this particle in 1974, showing up as a very sharp resonance similar to what is seen in Figure 14, was the first glimpse of the charm quark and played a key role in cementing the Standard Model.

There are a collection of lighter mesons that contain just a single charm quark. These are called (somewhat peculiarly) *D-mesons*. The lightest are:

$$\begin{aligned} &D^0\ (c\bar{u}) \qquad m \approx 1865 \text{ MeV} \\ &D^+\ (c\bar{d}) \qquad m \approx 1869 \text{ MeV} . \end{aligned} \tag{3.100}$$

These are remarkably long lived particles, with the D^+ surviving a whopping 10^{-12} seconds, and the D^0 about half this time. The long lifetime is because these particles decay only through a somewhat subtle property of the weak force. We will learn more about this in Section 5.

Similarly, the bottom quark was first discovered in *bottomonium*, also known as the upsilon (Υ)

$$\Upsilon\ (\bar{b}b) \quad m \approx 9.5 \text{ GeV} . \tag{3.101}$$

This has a lifetime of 10^{-20} seconds. Once again, it is neither the lightest nor the longest lived meson containing a b-quark. The lightest *B-mesons* are

$$B^+\ (u\bar{b}) \text{ and } B^0\ (d\bar{b}) \quad m \approx 5280 \text{ MeV} . \tag{3.102}$$

Despite being significantly heavier, they actually live (very) slightly longer than the D-mesons, with a lifetime of around 1.5×10^{-12} seconds. It's worth stressing how astonishing this is: the ratio of the mass to the width of the B-meson is $m_B/\Gamma_B \sim 10^{13}$ You can compare this to the common or garden mesons, like the ρ, which has $m_\rho/\Gamma_\rho \sim$ 4!. Again, this is down to intricacies of the weak force.

A small comment on terminology. The third generation of quarks was originally termed *beauty* and *truth*. (What can I say? It was the 70s.) Eventually, out of a due sense of embarrassment, these names were phased out in preference for the more boring "bottom" and "top". This has persisted for the top quark, but the term "beauty" lingers. For example, the important experiment LHCb which investigates B-mesons, prefers to be thought of, for obvious reasons, as focussing on the study of beauty, rather than the study of bottoms.

There are also baryons containing charm and bottom quarks. Here the names become increasingly unimaginative, with subscripts c and b denoting the quark content. For example, in addition to the Σ^+, comprised of uus, there is also a Σ_c^+ comprised of uuc and Σ_b^+ comprised of uub, and similar stories for cascades. There are also excited states of all these baryons, in which the quarks orbit each other, not dissimilar to the way in which the electrons orbit the proton in the excited states of the hydrogen atom.

3.4 The Theta Term

For QCD, we've seen that the action is gloriously simple:

$$S = \int d^4x \left(-\frac{1}{2}\mathrm{Tr}\, G_{\mu\nu}G^{\mu\nu} + \sum_{i=1}^{N_f} \left(i\bar{q}_i \not{D} q_i - m_i \bar{q}_i q_i \right) \right) \ . \tag{3.103}$$

The question that we would like to pose is: are there any other interaction terms that we could write down that we've missed.

The answer is that there is one, but that it's rather subtle. This is known as the Yang-Mills theta term,

$$S_\theta = \frac{\theta g_s^2}{16\pi^2} \int d^4x \ \mathrm{Tr}\, G_{\mu\nu}{}^\star G^{\mu\nu} \tag{3.104}$$

where ${}^\star G^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}G_{\rho\sigma}$. Here θ is the eponymous theta angle, and should be viewed as an additional parameter of QCD.

Before we get to the theory underlying the theta term, let me first give some commentary on why we haven't mentioned this term until now. The reason is that, as far as we can tell from experiment, the theta parameter takes the value $\theta = 0$. Said more precisely, we can bound the theta parameter to be

$$\theta < 10^{-10} \ . \tag{3.105}$$

So why should we care about something that doesn't exist? The reason is that zero is a number too! The game that we play in the Standard Model is the same as for all other quantum field theories: after you've figured out what fields you're dealing with, you then write down all possible relevant and marginal interactions that could change the low energy physics. Each of these terms typically comes with a parameter that we have to determine by experiment. These parameters are things like the masses of particles (or, more precisely, Yukawa couplings as we'll see in Section 5.) Out of all these parameters, θ is special because it's the only one that appears to vanish. And that's crying out for an explanation.

What would the consequences be if θ were not to vanish? The answer is pretty dramatic because, in contrast to all other terms in the QCD action (3.103), the theta term violates various discrete symmetries. Written in terms of the chromoelectric and chromomagnetic fields, it takes the form

$$G_{\mu\nu}{}^{\star}G^{\mu\nu} \sim \mathbf{E}\cdot\mathbf{B}\ . \tag{3.106}$$

We've seen in Section 1.4 that, under parity P, charge conjugation C, and time reversal T, the electric and magnetic fields transform as

$$\begin{aligned} P &: \mathbf{E} \mapsto -\mathbf{E} \quad \text{and} \quad P : \mathbf{B} \mapsto +\mathbf{B} \\ C &: \mathbf{E} \mapsto -\mathbf{E} \quad \text{and} \quad C : \mathbf{B} \mapsto -\mathbf{B} \\ T &: \mathbf{E} \mapsto +\mathbf{E} \quad \text{and} \quad C : \mathbf{B} \mapsto -\mathbf{B} \end{aligned} \tag{3.107}$$

This means that the theta term breaks both P and CP or, equivalently, T. As we saw previously, a consequence of CP violation is that particles are endowed with an electric dipole moment. The most precise experimental tests are for the neutron which, experimentally, is found to have an electric dipole moment d_n bounded by

$$d_n < 10^{-26}\ \mathrm{e\,cm}\ . \tag{3.108}$$

This, ultimately, translates into the bound (3.105). (For what it's worth the CP violation in the weak sector is predicted to give the neutron a dipole moment around $d_n \approx 10^{-30}$ e cm, somewhat below current experimental bounds.)

So why do we have $\theta = 0$? The answer is: we don't know. One might want to state by fiat that QCD should be invariant under P and CP and that's why the theta term is disallowed. That's a reasonable argument in the context of stand-alone QCD, but not when viewed within the broader framework of the Standard Model which, as we will see, is invariant under neither P nor CP. (Indeed, the fuller story is that the QCD theta term is infected by various other terms in the Standard Model Lagrangian and somehow they collectively conspire to ensure that $\theta = 0$.) The question of why $\theta = 0$ is known as the *strong CP problem*. It is surely one of the most important clues for what lies beyond the Standard Model.

3.4.1 Topological Sectors

The theta term is also special for other reasons. Indeed, of all the terms that we could write down in the Standard Model, it is by far the most subtle. In this sense, it's something of a shame that it vanishes!

We can discuss the physics for a general gauge group G, rather than restricting to QCD and, for that reason, we will revert to the notation of Section 1.3 and refer to the Yang-Mills gauge field as A_μ and the field strength as $F_{\mu\nu}$ (rather than G_μ and $G_{\mu\nu}$ for QCD).

The first important property of the theta term is that it's a total derivative. You can show that

$$S_\theta = \frac{\theta g_s^2}{8\pi^2} \int d^4x\ \partial_\mu K^\mu \quad \text{with} \quad K^\mu = \epsilon^{\mu\nu\rho\sigma} \mathrm{Tr}\left(A_\nu \partial_\rho A_\sigma - \frac{2i}{3} A_\nu A_\rho A_\sigma \right) \ . \tag{3.109}$$

This means that it does not affect the classical equations of motion. Nonetheless, it can affect the quantum dynamics of gauge theories. This arises because the path integral receives contributions from field configurations that have something interesting going on at infinity so that the boundary term S_θ is non-vanishing. This something interesting can be found in the topology of the gauge group.

To explain this, we first Wick rotate so that we work in Euclidean spacetime $\mathbb{R}^4$. Configurations that have a finite action from the Yang-Mills term must asymptote to pure gauge,

$$A_\mu \to \frac{i}{g} \Omega \partial_\mu \Omega^{-1} \quad \text{as } x \to \infty \tag{3.110}$$

with $\Omega \in G$. This means that finite action, Euclidean field configurations involve a map

$$\Omega(x) : \mathbf{S}^3_\infty \mapsto G\ . \tag{3.111}$$

with $\mathbf{S}^3_\infty = \partial \mathbb{R}^4$. Maps of this kind fall into disjoint classes. These arise because the gauge transformations can "wind" around the spatial $\mathbf{S}^3$ in such a way that one gauge transformation cannot be continuously transformed into another. Such winding is characterised by *homotopy theory*. In the present case, the maps are labelled by an element of the homotopy group which, for all simple, compact Lie groups G, is given by

$$\Pi_3(G) = \mathbb{Z}\ . \tag{3.112}$$

This means that the winding of gauge transformations (3.110) at infinity is classified by an integer n.

This statement is most intuitive for $G = SU(2)$ since, viewed as a manifold, $SU(2) \cong \mathbf{S}^3$ and the homotopy group counts the winding from one $\mathbf{S}^3$ to another. For higher dimensional groups, including $G = SU(3)$ relevant for QCD, it turns out that it's sufficient to pick an $SU(2)$ subgroup of G and consider maps which wind within that. You then need to check that these maps cannot be unwound within the larger G.

It can be shown that, in general, the winding $n \in \mathbb{Z}$ is computed by

$$n(\Omega) = \frac{1}{24\pi^2} \int_{\mathbf{S}^3_\infty} d^3S\ \epsilon^{ijk} \mathrm{Tr}\,(\Omega\partial_i\Omega^{-1})(\Omega\partial_j\Omega^{-1})(\Omega\partial_k\Omega^{-1}) \tag{3.113}$$

Evaluated on any configuration that asymptotes to (3.110), the theta term gives

$$S_\theta = \theta n \quad \text{with} \quad n \in \mathbb{Z}\ . \tag{3.114}$$

It is the contribution from configurations with $n \neq 0$ in the path integral that means that observables in quantum gauge theories can depend on θ. In general, all observables are thought to depend on the value of θ. For example, it's expected that the masses of particles in Yang-Mills theory, or indeed, in QCD, depend on θ. (The "expected" in that sentence is because it's very hard to know for sure, largely because it's very difficult to do numerical simulations of these theories when $\theta \neq 0$.)

When exponentiated in the path integral, the theta term contributes to the Euclidean action as $e^{iS_\theta} = e^{i\theta n}$. Importantly, it is a complex phase. The fact that it is complex can be traced to the $\epsilon^{\mu\nu\rho\sigma}$ tensor in S_θ. This means that S_θ contains a single time derivative and so, upon Wick rotation, still sits in the path integral as e^{iS_θ} rather than e^{-S_θ}. The fact that $n \in \mathbb{Z}$ means that θ is a periodic variable, with

$$\theta \in [0, 2\pi)\ . \tag{3.115}$$

For this reason, it's often called the *theta angle*. We see that the role of the theta term is to weight different topological sectors in the path integral with different phases $e^{i\theta n}$.

3.4.2 Instantons

We can say more if we work in a regime in which the theory is weakly coupled. Here the path integral is dominated by the saddle points, which are solutions to the classical equations of motion. This means that any θ dependence should come from field equations that wind at infinity, so $n \neq 0$, and solve the classical equations of motion,

$$\mathcal{D}_\mu F^{\mu\nu} = 0 \tag{3.116}$$

There is a cute way of finding solutions to this equation. The Yang-Mills action is

$$S_{YM} = \frac{1}{2g^2} \int d^4x \ \mathrm{Tr}\, F_{\mu\nu} F^{\mu\nu} \ . \tag{3.117}$$

Note that in Euclidean space, the action comes with a + sign. (This is to be contrasted with the Minkowski space action which comes with a minus sign.) We can write the Euclidean action as

$$S_{YM} = \frac{1}{4g^2} \int d^4x \ \mathrm{Tr}\,(F_{\mu\nu} \mp {}^\star F_{\mu\nu})^2 \pm \frac{1}{2g^2} \int d^4x \ \mathrm{Tr}\, F_{\mu\nu} {}^\star F^{\mu\nu} \geq \frac{8\pi^2}{g^2}|n| \ . \tag{3.118}$$

where, in the last inequality, we've used the result (3.114). We learn that in the sector with winding n, the Yang-Mills action is bounded by $8\pi^2 n/g^2$. The action is minimised when the bound is saturated. This occurs when

$$F_{\mu\nu} = \pm^\star F_{\mu\nu} \ . \tag{3.119}$$

These are the (anti)-self-dual Yang-Mills equations. The argument above shows that solutions to these first order (anti)-self-dual equations necessarily minimise the action in a given topological sector and so must solve the equations of motion (3.116). In fact, it's straightforward to see that this is the case since it follows immediately from the Bianchi identity $\mathcal{D}_\mu {}^\star F^{\mu\nu} = 0$.

Solutions to the (anti)-self-dual Yang-Mills equations (3.119) have finite action, which means that any deviation from the vacuum must occur only in localised regions of Euclidean spacetime. In other words, these solutions correspond to point-like objects in Euclidean spacetime $\mathbb{R}^4$. Because they occur for just an "instant of time" they are known as *instantons*. They are very much analogous to the classical tunnelling solutions for the quantum mechanical double well potential that we met in Section 2.1.

There is much to say about instantons. You can read about the role they play in quantum Yang-Mills in the lectures on Gauge Theory and more about the structure of the solutions to (3.119) in the lectures on Solitons. For our purposes, it will suffice to point out that the contributions of instantons to any quantity comes with the characteristic factor

$$e^{-S_{\text{instanton}}} = e^{-8\pi^2|n|/g^2} e^{i\theta n} \ . \tag{3.120}$$

Famously, the function $e^{-8\pi^2/g^2}$ has vanishing Taylor expansion about the origin $g^2 = 0$. This is telling us that effects due to instantons are smaller than any perturbative contribution, which takes the form g^{2n}. Nonetheless, that doesn't mean that instantons are useless since they can contribute to quantities that apparently vanish in perturbation theory.

Instantons are usually referred to as *non-perturbative* effects. This is a little bit of a misnomer. The use of instantons requires weak coupling $g^2 \ll 1$, so in this sense they are just as perturbative as usual perturbation theory. The name *non-perturbative* really means "not perturbative around the vacuum". Instead, the perturbation theory occurs around the instanton solution.

An Example: An Instanton in $SU(2)$

It is fairly straightforward to write down the instanton solutions with winding $n = 1$. For $SU(2)$, such a configuration is given by

$$A_\mu = \frac{1}{x^2+\rho^2}\,\eta^a_{\mu\nu} x^\nu \sigma^a \tag{3.121}$$

Here ρ is a parameter whose role we will describe shortly. The $\eta^a_{\mu\nu}$ are usually referred to as 't Hooft matrices. They are three 4×4 matrices which provide an irreducible representation of the $su(2)$ Lie algebra. They are given by

$$\eta^1_{\mu\nu} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \ , \ \ \eta^2_{\mu\nu} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \ , \ \ \eta^3_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \ . \tag{3.122}$$

These matrices are self-dual: they obey $\frac{1}{2}\epsilon_{\mu\nu\rho\sigma}\eta^i_{\rho\sigma} = \eta^i_{\mu\nu}$. (Note that we're not being careful about indices up vs down as we are in Euclidean space with no troublesome minus signs.) In the solution (3.121), the 't Hooft matrices intertwine the $su(2)$ group index $a = 1, 2, 3$ with the spacetime index μ and this implements the asymptotic winding of the gauge fields.

The associated field strength is given by

$$F_{\mu\nu} = -\frac{2\rho^2}{(x^2+\rho^2)^2}\eta^a_{\mu\nu}\sigma^a \ . \tag{3.123}$$

This inherits its self-duality from the 't Hooft matrices: $F_{\mu\nu} = {}^\star F_{\mu\nu}$ and therefore solves the Yang-Mills equations of motion, $\mathcal{D}_\mu F_{\mu\nu} = 0$.

We can get some sense of the form of this solution. First, the non-zero field strength is localised around the origin $x = 0$. (By translational invariance, we can shift $x^\mu \to x^\mu - X^\mu$ to construct a solution localised at any other point X^μ.) The solution depends on a parameter ρ which can be thought of as the size of the instanton lump. The fact that the instanton has an arbitrary size follows from the classical conformal invariance of the Yang-Mills action.