Introduction to Elementary Particle Phenomenology

(student handout version)

Philip G. Ratcliffe
(philip.ratcliffe@uninsubria.it)

Dipartimento di Scienza e Alta Tecnologia
Università degli Studi dell'Insubria in Como
via Valleggio 11, 22100 Como (CO), Italy

(last revised 2nd October 2017)

Preface

The present is a written version of lecture notes for the introductory course on the *Phenomenology of Elementary Particles*, held within the framework of the Master's equivalent degree course in Physics at Insubria University in Como. The lectures were first delivered in the academic year 2005/06 and have since undergone some evolution. The notes have been augmented and edited with the aim of being as self-contained as is reasonably possible and therefore of more general utility. They are thus primarily intended for use by students with a basic knowledge of classical electromagnetism, quantum mechanics and special relativity but not necessarily, for example, of quantum field theory. However, this volume should also hopefully represent a useful reference text and study aid for other similar courses.

Contents

Chapter 1

Introduction

1.1 Supplementary reading

A (very short) list of suggested supplementary reading material follows. The books by Perkins and by Martin and Shaw are particularly recommended for their style, clarity and completeness and their relevance to the subject matter of this course. In addition, each chapter contains a more-or-less comprehensive list of cited works (both original papers and more general pedagogical review articles), which may be consulted for further study.

We should finally mention the Particle Data Group (PDG) Review of Particle Physics (Patrignani *et al.*, 2016), which is updated every two years and, besides providing a compendium of high-energy particle-physics data, contains many clear and concise review articles on various aspects of particle physics, including the theory behind the standard model (SM).

Reading list

Braibant, S., Giacomelli, G. and Spurio, M. (2012), *Particelle e Interazioni Fondamentali* (Springer-Verlag Italia), 2nd. edition.

Cahn, R.N. and Goldhaber, G. (1989), *The Experimental Foundations of Particle Physics* (Cambridge U. Press).

Close, F.E. (1979), *An Introduction to Quarks and Partons* (Academic Press).

Halzen, F. and Martin, A.D. (1984), *Quarks and Leptons* (John Wiley & Sons).

Martin, B.R. and Shaw, G. (1997), *Particle Physics* (John Wiley & Sons), 2nd. edition.

Patrignani, C. *et al.*, Particle Data Group (2016), *Chin. Phys.* **C40** [10], 100001.

Perkins, D.H. (2000), *Introduction to High Energy Physics* (Cambridge U. Press), 4th. edition.

Povh, B., Rith, K., Scholz, C. and Zetsche, F. (1995), *Particles and Nuclei* (Springer–Verlag).

Quinn, H.R. *et al.*, Fundamental Particles and Interactions Chart Committee (1989), preprint SLAC-PUB-4879.

1.2 Aims and philosophy

The specific topic of this lecture series is the phenomenology of particle physics at both high and low energies. While a certain basic knowledge of the theoretical foundations (in particular, notions of quantum mechanics and special relativity) is assumed and is necessary for a full appreciation of the subjects treated, the main emphasis is placed on the phenomenological aspects and on the experimental manifestation of the underlying dynamics and symmetries. In particular, the role of symmetries (exact and approximate) and their violation is central to many of the discussions. In this context, I can also recommend, as general supplementary reading, the Dirac Memorial Lecture delivered by Weinberg (1987).

While the formulation and physical significance of the Dirac equation (1928) is briefly outlined, no attempt is made to enter into the realms of quantum field theory. The formalism of Dirac spinors and their relation to the symmetries of spin, parity and charge-conjugation is, however, necessary. Moreover, in order to describe some of the phenomena that have played important roles in the growth of our knowledge of particle interactions, the notion of Feynman diagrams (1949) is introduced in an intuitive manner.

1.3 Bibliography

Dirac, P.A.M. (1928), *Proc. Royal Soc. (London)* **A117**, 610.

Feynman, R.P. (1949), *Phys. Rev.* **76**, 769.

Weinberg, S. (1987), in *Elementary Particles and the Laws of Physics*, eds. R. MacKenzie and P. Doust (Cambridge U. Press), p. 61.

Chapter 2

Symmetries (Discrete and Continuous)

In this chapter we shall mainly be concerned with discrete symmetries, or more precisely: parity, charge-conjugation and time-reversal invariances and also their *violation*. However, since we shall be forced to examine the weak interaction in some depth it is natural that some discussion be presented here of the flavour symmetries present in the quark sector. In particular, in order to explain CP violation, it will be necessary to examine the question of *quark mixing* and the Cabibbo model (1963) with its extension to the full three-generation picture due to Kobayashi and Maskawa (1973). However, the principal topics here will remain the symmetries of C, P and T.

2.1 Parity violation in weak interactions

Within the realm of particle physics perhaps one of the first and most notable phenomenological manifestations of symmetry (or lack thereof) is related to the role of parity in the weak interaction. Of course, symmetry and its role in general was recognised much earlier: Einstein's development of the theory of relativity (both special and general) rests on notions of symmetry (with respect to the choice of reference frames) while Noether's seminal work (1918) on the relationship between continuous symmetries and conserved quantities stands at the very foundation of all modern theory, both classical and quantum.

Parity, however, holds a special place as the first, simplest and previously unquestioned symmetry to be found violated in Nature (Wu *et al.*, 1957).* This

* It is often recounted that when Abdus Salam, as a young researcher, proposed a theory involving parity violation to Wolfgang Pauli, he was unceremoniously dismissed with the remark: "This young man does not realise the sanctity of parity!"

opens the window onto a completely new perspective: the *breaking* of symmetries. Just over half a century later it is now quite normal to seek violation of symmetries and indeed to use the natural violations that can occur at the quantum level to explain, at least in part, the phenomenology of the particles populating the world we see and experience.

2.1.1 The $\tau-\theta$ puzzle

In the early fifties a puzzle arose (Dalitz, 1953) involving two new subatomic particles, then called τ and θ. Both were members of the newly found family of so-called *strange* particles, relatively long-lived objects that were being produced in the new accelerator experiments. The long lifetimes of these particles suggested that, although the final states often only contained strongly interacting particles, they did not decay via the strong interaction; they were thus deemed strange. The τ^+ and the θ^+ are now known to be one and the same particle, the charged kaon or K^+.

In fact, the τ^+ and θ^+ were found to be identical in terms of their mass, charge and other properties (within experimental precision) and were only distinguished by their decay modes.* The two particles decayed quite differently and hence their being considered distinct. The τ^+ decayed into three pions ($\pi^+\pi^+\pi^-$ or $\pi^+\pi^0\pi^0$) while the θ^+ produced only two ($\pi^+\pi^0$). Indeed, it was *only* the different decay modes that distinguished them and suggested the need for two separate particles. The fact that their lifetimes were also very similar (identical within errors) rendered the idea of two distinct yet almost identical states (one might say "twins") very puzzling—no other known pair of particles displayed such striking similarity.

The intriguing necessity for two distinct particles arose from the realisation that the parity of the two final states must be different (Dalitz, 1953). In 1956, based on observations of the charged-kaon decays, Lee and Yang were led to make the (then extravagant) proposition that parity conservation might be violated.† They further suggested that if the answer to the τ–θ puzzle were indeed parity violation, then such an effect might also be observed in the spatial distribution of the β-decay of polarised nuclei. In essence, they proposed measuring a pseudo-scalar quantity such as $\boldsymbol{p}\cdot\boldsymbol{s}$, where $\boldsymbol{p}$ is, say, the final electron momentum and $\boldsymbol{s}$, say, the spin of the decaying nucleus.

Let us first examine why the parity assignments must differ (Dalitz, 1953). Experiments showed that the spin of both objects was zero; thus, the question of

* Naturally, their *charge-conjugate* versions or *antiparticles*, with opposite charge, also exist and exhibit identical behaviour.

† The 1957 Nobel Prize for Physics was awarded equally to Chen Ning Yang and Tsung-Dao Lee for "their penetrating investigation of the so-called parity laws which has led to important discoveries regarding the elementary particles."

angular-momentum conservation is rather simple since pions too have spin zero: the total orbital angular momentum of the final state must then also be zero. With zero orbital contribution to the total angular momentum of the system, the overall parity is just the product of the intrinsic parities of the final-state particles—recall that the parity of the spatial part of the wave-function is simply $(-1)^L$. This is trivial in the two-pion case; for three pions, while they may individually have non-zero orbital angular momentum, the overall total cancellation required implies overall even parity of the spatial part. The parities of the final states are then determined by the intrinsic parities of the pions they contain.

2.1.2 Intrinsic parity and its measurement

All particles either naturally possess or may be assigned an *intrinsic parity*. In the case of fermions a consequence of the Dirac equation is that any given fermion and antifermion have *opposite* parities. Since fermion number is conserved (only fermion–antifermion pairs may be created or annihilated), the absolute value of, say, the electron parity is undetectable and irrelevant. By convention, the parity of fermions (antifermions) is taken as positive (negative) although no physical significance may be attached to either separately. However, the intrinsic parity of a fermion–antifermion pair is meaningful and is thus predicted to be *negative*.

Recall that the parity of a compound state is just the product of the parities of the parts. The parity of a $q\bar{q}$ meson is thus the product of the intrinsic parities of the quarks of which it is composed and the parity of the spatial wave-function describing their relative orbital motion. In other words, $P_\pi = P_q P_{\bar{q}} (-1)^L = (-1)^{L+1}$, where L is the orbital quantum number. This has the immediate consequence that a pion, being the lowest-mass, zero-spin $q\bar{q}$ state (and therefore presumably with zero internal orbital angular momentum and with the quark and antiquark spins antiparallel), should have *negative* intrinsic parity. This is indeed experimentally verified; *i.e.* it is pseudoscalar.

The measurement of $P_\pi = -1$ is conceptually rather simple: consider associated production (via the parity-conserving strong interaction) of a neutron pair from a low-energy, negatively charged pion incident on a deuteron (Chinowsky *et al.*, 1954):

$$\pi^- d \rightarrow nn. \tag{2.1.1}$$

The deuteron is a spin-one nucleus ($L_{pn} = 0$ and $S_{pn} = 1$) of *positive* parity and the pion has zero spin, while the proton and neutron both have spin one-half. In any case, the two nucleons are, with respect to parity, merely spectators in this process. What are important are the relative initial- and final-state orbital angular momenta. Now, the process is actually that of K-capture: the pion is initially trapped, forming an excited pionic atom, and then rapidly cascades down

to the lowest Bohr orbit, *i.e.* an s-wave, from which it then interacts with the deuteron. The total angular momentum of the initial state is thus one, but with $L_{\pi d}=0$ and therefore positive spatial parity. The parity of the initial and (so too) the final states is therefore precisely the pion parity: $P_\pi=(-1)^{L_{nn}}$.

The orbital angular momentum of the final state could, in principle, be measured by studying (statistically) the angular distribution of the neutrons produced. However, it is easier to appeal to the Fermi–Dirac statistics obeyed by a system of two *identical* fermions. The final nn system must have unit total angular momentum and this can be constructed from either a sum of the two neutron spins giving zero (singlet) and orbital motion $L=1$ or a sum of spins giving one (triplet) and orbital motion $L=0$, 1 or 2.* For the spin-singlet state, we have the following spin wave-function:

$$\tfrac{1}{\sqrt{2}}\left(|{\uparrow\downarrow}\rangle - |{\downarrow\uparrow}\rangle\right). \tag{2.1.2a}$$

Note, in particular, that it is antisymmetric under interchange of the neutrons. Since they are identical fermions, the spatial part must then be symmetric and thus L even. The total spin can then only be even and therefore one is excluded. For the triplet state, on the other hand, we have the following three possible spin wave-functions:

$$|{\uparrow\uparrow}\rangle, \qquad \tfrac{1}{\sqrt{2}}\left(|{\uparrow\downarrow}\rangle + |{\downarrow\uparrow}\rangle\right), \qquad |{\downarrow\downarrow}\rangle. \tag{2.1.2b}$$

These are all symmetric under interchange and thus the spatial part must be antisymmetric, giving L odd, which will only accommodate $L=1$, combining with $S=1$ to provide total spin one. Therefore, we must have $S=1$, $L=1$, negative spatial parity and $P_\pi=-1$.

Exercise 2.1.1. *By considering the intrinsic parities of the proton and neutron, together with their known orbital and spin alignments inside the deuteron, show that indeed we do expect $P_d=+1$.*

2.1.3 The physical consequences of parity violation

As mentioned earlier, the idea then was to measure the dependence of some decay rate on a *pseudo-scalar* quantity such as $\boldsymbol{p}\cdot\boldsymbol{s}$. The reason for this is quite simple. Let us suppose that the transition matrix element or quantum amplitude for some decay rate (or, of course, interaction cross-section) takes the general form

$$\mathcal{M} \propto A + B\,\boldsymbol{p}\cdot\boldsymbol{s}, \tag{2.1.3}$$

with A and B scalar quantities that do *not* depend (linearly) on either $\boldsymbol{p}$ or $\boldsymbol{s}$. Then, since $\boldsymbol{p}$ is a polar vector, it changes sign under the action of parity inversion

* Recall that angular momenta are to be added vectorially.

while $\boldsymbol{s}$, a pseudovector or axial vector, does not. The B term will thus change sign with respect to the A term and so

$$\left|\mathcal{M}^{\mathcal{P}}\right| \neq \left|\mathcal{M}\right|, \tag{2.1.4}$$

where $\mathcal{M}^{\mathcal{P}}$ stands for the corresponding matrix element under parity inversion. Such a difference would be precisely a manifestation of the parity violation that Lee and Yang sought.

Now, although it is not possible to apply $\mathcal{P}$ as such experimentally, the presence of a parity-violating term may be detected from the spatial dependence it implies. For example, for a fixed spin $\boldsymbol{s}$, we also have that

$$\left|\mathcal{M}(-\boldsymbol{p})\right| \neq \left|\mathcal{M}(+\boldsymbol{p})\right|. \tag{2.1.5a}$$

And so it is sufficient to simply examine, *e.g.*, the decay rate for final-state electron momenta parallel and antiparallel to the nucleon spin axis. Note that the two statements are entirely equivalent. A final-state electron observed moving, say, parallel to the nucleon spin axis, moves in the opposite direction in the parity-inverted experiment. Note also that analogously, for fixed $\boldsymbol{p}$, we have

$$\left|\mathcal{M}(-\boldsymbol{s})\right| \neq \left|\mathcal{M}(+\boldsymbol{s})\right|. \tag{2.1.5b}$$

2.1.4 Parity violation in β-decay

The invitation to perform an experiment to detect such parity violation was taken up shortly after in 1957 by Wu *et al.*[*†] of Columbia University in New York. At that time Wu was considered a world expert on β-decay and Lee also worked at Columbia.

The basic requirement of polarising the initial-state decaying nucleus then led Wu *et al.* to select ^{60}Co as it has a nuclear spin of $J=5$ in natural units (with a large magnetic moment too). The experiment was by no means simple; in order to substantially polarise the cobalt specimen and avoid depolarisation by thermal motion, exceedingly low temperatures and thus new refrigeration techniques were necessary. To this end Wu enlisted the help of Ernest Ambler, a cryogenics expert.

[*] Why, after all this time, some texts still refer to the leader of the group as *Madame* Wu is a mystery. In a modern age reference to the gender of a research scientist (or indeed of anyone when gender is irrelevant) is, at best, belittling. Indeed, if one wished to attach a title to the name of, say, Feynman or Fermi, it would be Prof. and certainly not *Monsieur*.

[†] Although, perhaps rather surprisingly, Wu did not receive the Noble prize for this discovery, her contribution was recognised via many other honours and awards: she was the first female president of the American Physical Society and the first winner of the Wolf Prize in Physics, to mention just two.

Indeed, so convinced was Wu of the importance of such an experiment that she forwent a long-awaited trip to her native China and immediately set about preparing the necessary equipment at the National Bureau of Standards' headquarters in Maryland, where important facilities were available. And so, with group members sleeping in the laboratory in order to be ready when the required temperatures were reached, just a few months after the discussions with Lee and Yang the experiments were performed.

The measurement of any variations with respect to $\boldsymbol{p}$ would require either two identical detectors or two independent runs with a single detector being placed first above and then below the decaying specimen. However, the difficulties in control over systematics would render any difference found highly suspect. The observation exhibited in Eq. (2.1.5b) was thus exploited and the polarisation axis flipped by inverting the applied (vertical) polarising magnetic field. Still, some control over systematics was desirable and so advantage was taken of the full decay chain. The basic β-decay process is

$$^{60}\mathrm{Co} \;\rightarrow\; ^{60}\mathrm{Ni}^{*} + e^{-} + \bar{\nu}_{e}, \tag{2.1.6}$$

where the nickel daughter ($J=4$) is left in an excited state. The subsequent transition to the ^{60}Ni ground state is a two-step process involving double γ emission. Both transitions are of the electric-quadrupole (E4) type with its characteristic double-lobe angular distribution. The presence of nuclear polarisation may thus be checked by observation of the relative γ-ray intensities in the polar (*i.e.* vertical) and equatorial directions (*i.e.* horizontal); the emitted photons prefer to lie in the equatorial plane rather than along the polarisation axis.

The procedure thus essentially consisted of cooling the specimen to $\sim 0.01\,\mathrm{K}$ and applying a magnetic field along, say, the z direction. The refrigeration system was then switched off and continuous read-outs were taken of both the polar β flux and the polar and equatorial γ intensities. This was then repeated with the polarising magnetic field inverted. The results are displayed in Fig. 2.1 and may be summarised via the following angular decay distribution for the electrons with respect to the polarisation axis of the ^{60}Co nucleus (as suggested by Lee and Yang):

$$I(\theta) \;\propto\; 1 + \alpha\,\frac{v}{c}\cos\theta, \tag{2.1.7}$$

where, given that $\cos\theta = \pm 1$ here, the asymmetry parameter α is effectively measured directly. Wu *et al.* found α negative and gave a lower-limit estimate of $|\alpha| \gtrsim 0.7$. In fact, α is negative for electrons and positive for positrons. More precise measurements later showed that, in general, $|\alpha| = 1$; *i.e.* parity is *maximally* violated in β-decay.

Since the energies are so low that orbital angular momentum can play no role,

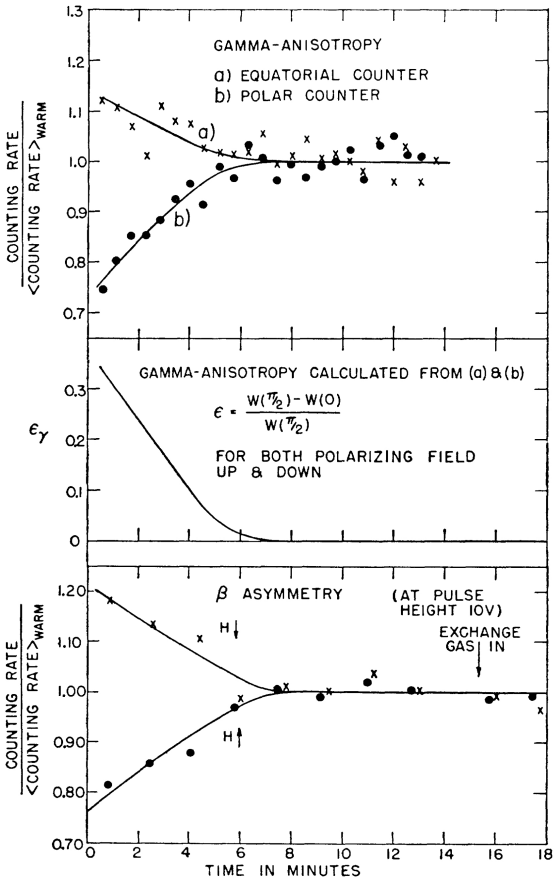


Figure 2.1: Results of the parity violation experiment performed by Wu *et al.* (1957). The γ anisotropy and the β asymmetry with the polarising magnetic field pointing up and down as functions of the time after switching off the cryostat.

one can make an argument to infer the electron spin by noting that $\Delta J = -1$ for the nuclei in this decay and therefore the electron–neutrino pair must carry off one unit of spin aligned in the positive z direction, *i.e.* both spins must be aligned along the same positive z direction. Since the electron tends to move along the negative z direction, it must have helicity -1;* in other words, it is *left handed*. Likewise, in β-plus decays the positron is found to be always right handed.

> "The sudden liberation of our thinking on the very structure of the physical world was overwhelming."
>
> Chien-Shiung Wu

* Recall that helicity is defined as $h := \hat{\boldsymbol{p}} \cdot \hat{\boldsymbol{s}}$ and therefore $-1 < h < +1$.

2.1.5 The helicity of the neutrino

The measurements performed by Wu *et al.*, while unequivocally indicating that parity is indeed violated and that the electrons (positrons) emerging are left (right) handed, did not provide any indication as to the relative spin alignment of the (undetected) neutrinos. Just a few months later Goldhaber *et al.* (1958) thus set out to measure the helicity of neutrinos produced in β-decay. The method devised is ingenious (see Fig. 2.2), combining as it does a number of non-trivial physical

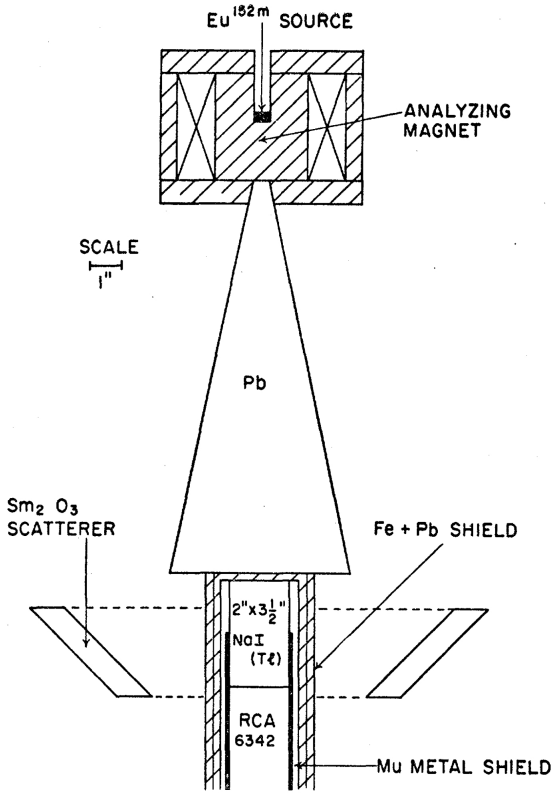


Figure 2.2: Apparatus for analysing the circular polarisation of γ-rays in the experiment to measure neutrino helicity performed by Goldhaber *et al.* (1958).

effects and phenomena.

The process used was K-capture:

$$\begin{array}{rcccc} & & ^{152}\mathrm{Eu} + e^- & \rightarrow & ^{152}\mathrm{Sm}^* + \nu_e \\ & & & & \hookrightarrow\ ^{152}\mathrm{Sm} + \gamma, \\ \mathrm{J} & = & 0 & & 1 \qquad 0 \end{array} \tag{2.1.8}$$

in which the samarium daughter nucleus subsequently de-excites via γ-emission.

The spin actually measured was that of the emitted γ. The analysing magnet and block of magnetised iron surrounding the source served to filter out one or other of the two possible γ helicities: a photon may be absorbed by an atomic electron if and only if their spins are opposite (the electron then flips its spin to conserve angular momentum). Conversely, a photon with spin parallel to the electrons in the block of iron will pass relatively unhindered.

$$\begin{array}{ccccc} \longrightarrow & & \overset{\gamma}{\Rightarrow} + \overset{e^-}{\leftarrow} & & \overset{\gamma}{\Rightarrow} + \overset{e^-}{\rightarrow} \quad \ldots \\ \text{photon direction} & & \text{absorbed} & & \text{unabsorbed} \end{array} \tag{2.1.9}$$

Since the γ-emission process involves a $J=1$ nucleus decaying into $J=0$, the photon evidently carries the same spin as the original nucleus, which in turn must be opposite to that of the neutrino.

In order to extract the neutrino helicity, information is also needed on its direction of motion (without directly detecting it). The lower samarium *rescatterer* is only effective for those photons with energy *exactly* corresponding to the first excited state (from which the photons were originally produced). Now, the effect of the nuclear recoil in such processes results in emitted photons with slightly *less* than the excitation energy while absorbed photons need slightly *more*. Therefore, were both the emitting and absorbing samarium nuclei both stationary, there could be *no* absorbtion. However, in the K-capture process, for which the captured electron is essentially at rest while the neutrino is emitted with a non-negligible energy, the daughter nucleus recoils. Such a recoil may be sufficient, if it is in the right direction, to provide the necessary extra energy to the subsequently emitted photon. Thus, only a photon produced from a nucleus moving in the same direction (and therefore opposite to the neutrino) may be rescattered and finally detected.

To recap, the photon and neutrino move in opposite directions and also have opposite spin projections; they therefore have the same helicities. By comparing the counting rate with the magnetic field applied in the positive and negative vertical directions, Goldhaber *et al.* were thus able to infer the neutrino helicity or handedness. The results demonstrate that the neutrino too is always left handed. Similar experiments on antineutrino emission (using, *e.g.*, ^{203}Hg) show that the antineutrino is instead right handed (as too is the positron emitted in β^+-decay).

2.1.6 Muon-decay experiments

The findings of Wu *et al.* were confirmed by parallel experiments on muon decay carried out at almost the same time by Garwin, Lederman and Weinrich (1957). For completeness, we shall just briefly review the basic concepts. Once again, the idea came from Lee and Yang: they suggested studying pion decay into muons

and the subsequent in-flight muon decay, thus

$$\begin{aligned} \pi^+ &\rightarrow \mu^+ + \nu_\mu \\ &\quad \hookrightarrow e^+ + \nu_e + \bar{\nu}_\mu. \end{aligned} \tag{2.1.10}$$

In the first decay, thanks to parity violation, the muon spin is predominantly aligned along its direction of motion. The muons are then stopped in a carbon target and are assumed to lose little of their polarisation in the stopping process. If this is the case, then again, by virtue of parity violation, the decay electrons will emerge with an angular distribution with respect to the muon direction of the form

$$I(\theta) \propto 1 + \alpha \cos\theta. \tag{2.1.11}$$

In fact, rather than moving the detectors, the experimenters exploited the magnetic moment of the muon to precess its spin. The measured distribution suggested a value $\alpha = 0.33 \pm 0.03$. The fact that here $|\alpha| \neq 1$ can be explained by the non-trivial composite nature of the pion.

2.1.7 Interpretation

While one might be led to attribute the existence of parity violation to the neutrino itself, a closer inspection reveals this to be wrong. The reason that suspicion falls on the neutrino has to do with its mass, or rather lack thereof. A left-handed electron appears so only in certain reference frames. However, if the observer is boosted to a velocity exceeding that of the electron (which is always possible for a massive particle), then it will now appear to have the opposite velocity while maintaining the same spin projection and will thus have effectively flipped its helicity to become right handed. There can therefore be nothing intrinsically special about a right- or left-handed electron, or indeed any massive fermion. In contrast, a massless neutrino always travels at the speed of light and so a neutrino that is left handed in some given reference frame appears left handed in *any* frame. It is thus tempting to attribute parity violation to the non-existence of the right-handed neutrino and left-handed antineutrino.

What has not been mentioned here yet is that there are also a large number of weak decays that do not involve neutrinos (so-called *non-leptonic* decays) but in which parity is still violated. A simple example is furnished by the decay channels

$$\Lambda^0 \rightarrow p\pi^-,\ n\pi^0. \tag{2.1.12}$$

By virtue of a non-zero magnetic moment, the Λ^0 may easily be polarised (this is indeed how its magnetic moment is measured) and thus one may measure the

correlation between the Λ^0 spin direction and say the momentum direction of the emitted nucleon (the nucleon and pion naturally emerge back-to-back in the Λ^0 rest frame). Again, such a quantity, depending on a scalar product $\boldsymbol{p}\cdot\boldsymbol{s}$, violates parity. In this case the asymmetry parameter is $\alpha \approx 0.64$. Again, the reason $|\alpha| \neq 1$ has to do with the non-trivial internal structure of baryons.

Consequently, we must ascribe the violation of parity to the nature of the weak interaction itself. This will be discussed fully in the next section; suffice it to say here that the nature of the weak force is such that it only couples left-handed fermions and right-handed antifermions. In other words, at this level it is entirely irrelevant whether or not the right-handed neutrino exists: it is simply non-interacting or *sterile*.

It should be pointed out that the effects examined above are all examples of *direct* parity violation; *i.e.*, they are directly caused by parity violation in the interaction involved in the process observed. However, the presence of a parity violating interaction can induce parity mixing in, *e.g.*, bound states (nucleons, nuclei and even atoms), which may then manifest itself via the apparent violation of parity in some process (usually a decay). Such an effect is then said to be *indirect* parity violation.

2.1.8 Closing remarks

> "Discoveries in physics often depend on looking toward a new direction, quite often with the very latest detector technology. Parity non-conservation is an exception. The reason it was not discovered [earlier] was not because it was at the margin of detector technology, but simply because people did not look for it."
>
> Tsung-Dao Lee

Indeed, such an experiment could easily have been performed thirty years earlier and it is probable that the effect had been seen as early as the twenties but not recognised. In fact, it was not until the thirties and the work of Wigner that the role of parity was at all appreciated.

2.2 $V-A$ formulation of the weak currents

As we have seen in the previous section, parity violation, at least in the case of leptonic interactions, is maximal. The questions now raised are how this may be explained and what implications there might be.

2.2.1 The Dirac equation

As mentioned in the previous section, the properties of fermions with respect to spatial inversion are dictated by the form of the Dirac* equation (1928). This is connected to the question of spin and the fact that the Dirac equation relates particle to antiparticle and automatically includes spin states. We shall now provide a very brief description of the Dirac equation and its main properties.

In quantum mechanics the wave equations governing particle motion may be deduced from the classical laws via substitution of dynamical variables such as momentum and energy with operators, thus: $E \to \mathrm{i}\partial/\partial t$ and $\boldsymbol{p} \to -\mathrm{i}\boldsymbol{\nabla}$ or, for a relativistically covariant version, $p^\mu \to \mathrm{i}\partial^\mu$.† The free-particle Schrödinger equation is then derived from

$$E = \frac{\boldsymbol{p}^2}{2m}. \tag{2.2.1}$$

The Dirac equation is the result of an attempt to formulate a Lorentz covariant wave equation that avoids certain problems of negative energies. If we attempt to replace (2.2.1) with Einstein's version

$$p^2 = m^2, \tag{2.2.2}$$

where p is now a four-vector, we obtain the Klein–Gordon equation (Klein, 1927 and Gordon, 1926). However, owing to the presence of E^2 in Einstein's equation, apparently unphysical negative-energy solutions naturally arise.

Dirac's idea was effectively to take the square root of this equation and seek to adopt something of the form

$$p^\mu \sim m. \tag{2.2.3}$$

Now, since p^μ transforms as a vector while m is a scalar quantity, we evidently need to saturate the index μ:

$$\gamma^\mu p_\mu = m, \tag{2.2.4}$$

where γ^μ, a new object introduced *ad hoc*, must evidently have special properties. Indeed, this equation should still agree with the Einstein relation (2.2.2) and so, squaring, we require that

$$\gamma^\mu p_\mu \gamma^\nu p_\nu = p^2 \tag{2.2.5}$$

to be completely equivalent to (2.2.2). That is,

$$\tfrac{1}{2}\{\gamma^\mu, \gamma^\nu\} p_\mu p_\nu = \gamma^\mu p_\mu \gamma^\nu p_\nu = p^2 = p_\mu p_\mu = g^{\mu\nu} p_\mu p_\nu, \tag{2.2.6}$$

* The 1976 Nobel Prize for physics was awarded equally to Erwin Schrödinger and Paul Adrien Maurice Dirac for "the discovery of new productive forms of atomic theory."

† We shall immediately adopt the "natural" units of the high-energy physicist, in which both c and $\hbar$ are set equal to unity.

where the first step is made by assuming the coefficient vectors γ^μ to commute with p^μ, as they must to avoid spurious forces, and then symmetrising in the indices μ,ν (since $p_\mu p_\nu$ is symmetric). The above requirement is then satisfied if

$$\{\gamma^\mu, \gamma^\nu\} \;=\; 2g^{\mu\nu}\,\mathbb{1}, \tag{2.2.7}$$

where the $\mathbb{1}$ is the identity in the space spanned by the γ^μ. This is just the Clifford algebra, which may be rendered with a matrix representation.

The standard 4×4 representation is that of Dirac, given in the appendices:

$$\gamma^0 \;=\; \begin{pmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\gamma} \;=\; \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ -\boldsymbol{\sigma} & 0 \end{pmatrix}, \tag{A.1.11}$$

where a 2×2 block notation has been used and the $\boldsymbol{\sigma}$ are just the usual Pauli matrices. The Dirac equation is then

$$\mathrm{i}\gamma^\mu \partial_\mu \psi \;=\; m\psi, \tag{2.2.8}$$

where ψ must now be a four-component *spinor*. The role of the four components is uncovered by coupling the equation to a (classical) electromagnetic potential A^μ via the usual procedure of *minimal coupling*: $p^\mu \to p^\mu - eA^\mu$ or equivalently $\mathrm{i}\partial^\mu \to \mathrm{i}\,\partial^\mu - eA^\mu$. Consideration of a classical electric (scalar) potential A^0 and the form of γ^0 reveals that, if the upper two components of ψ are to describe an electron of charge $-|e|$, then the lower two describe an object of charge $+|e|$, or the positron. Moreover, consideration of the magnetic part in the low-energy limit reveals the spin content and the prediction of a gyromagnetic ratio equal to two.

We further find that the various transformations of spatial inversion (or parity), time reversal and charge conjugation are obtained via multiplication by suitable combinations of the γ^μ (together with any other necessary transformations, *e.g.* $\boldsymbol{x} \to -\boldsymbol{x}$ or $t \to -t$). In particular, the parity operation, besides sending $\boldsymbol{x} \to -\boldsymbol{x}$ requires the spinors to be multiplied by γ^0. The sign difference between the upper and lower blocks of this matrix leads to the opposite parity assignment for fermion and antifermion.

If we wish to have a natural (positive definite) object for a probability density, such as $\psi^\dagger\psi$, then we are led to define the current corresponding to the Dirac equation as

$$j^\mu \;\equiv\; \overline{\psi}\,\gamma^\mu\,\psi, \tag{2.2.9}$$

where the natural "conjugate" spinor is $\overline{\psi} \equiv \psi^\dagger\gamma^0$. In electromagnetism the coupling between the electron and the electromagnetic field is then perfectly well described by an interaction of the form $j\cdot A$. Taking this as a template for particle interactions, one immediately realises it is not unique; the most general form for a

"current" is $\bar{\psi}\Gamma\psi$, where Γ can be any one of a number of matrices spanning the Dirac spinor space.

Briefly, as the reader may easily verify, the free-particle plane-wave solutions to the Dirac equation take the following form:

$$\psi(x) \;=\; w(p,s)\,\mathrm{e}^{-\mathrm{i}\epsilon\, p\cdot x}, \tag{2.2.10}$$

where the sign $\epsilon = \pm$ will be explained shortly, $w(p,s)$ is a constant, four-component spinor, containing both energy–momentum (p) and spin (s) information and which satisfies

$$\epsilon\,\gamma^{\mu} p_{\mu}\, w(p,s) \;=\; m\, w(p,s), \tag{2.2.11}$$

in which p is no longer an operator. The sign of the exponent determines two types of solutions. In fact, the Dirac equation does not eliminate the negative-energy solutions but gives them a meaning: antimatter. Thus, according as to whether $\epsilon = \pm$ the spinor $w(p,s)$ takes on different forms.

Exercise 2.2.1. *Verify that the free-particle plane-wave solutions to the Dirac equation do indeed take the form shown in Eq. (2.2.10), with the spinor $w(p,s)$ satisfying Eq. (2.2.11).*

The general form of a free-particle state is then

$$\psi(x) \;=\; w(\boldsymbol{p})\,\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}\epsilon\, p\cdot x}, \tag{2.2.12}$$

where the spinor $w(\boldsymbol{p})$ is

$$w(\boldsymbol{p}) \;=\; \begin{cases} u(\boldsymbol{p}) \;=\; \begin{pmatrix} \sqrt{E+m}\;\mathbb{1} \\ \sqrt{E-m}\;\boldsymbol{\sigma}\cdot\hat{\boldsymbol{p}} \end{pmatrix} \otimes \chi \quad \text{(positive-energy)}, \\[2ex] v(\boldsymbol{p}) \;=\; \begin{pmatrix} \sqrt{E-m}\;\boldsymbol{\sigma}\cdot\hat{\boldsymbol{p}} \\ \sqrt{E+m}\;\mathbb{1} \end{pmatrix} \otimes \chi \quad \text{(negative-energy)}, \end{cases} \tag{2.2.13}$$

and the two-component spinor χ is $\binom{1}{0}$ and $\binom{0}{1}$.

2.2.2 Relativistic currents

Thus far we have two possibilities: $\mathbb{1}$ and γ^{μ}. However, this sixteen-dimensional Dirac space offers three other different possibilities. First of all, let us construct the special matrix $\gamma_5 \equiv \mathrm{i}\gamma^0\gamma^1\gamma^2\gamma^3$, which immediately provides five more independent pieces: γ_5 and $\gamma_5\gamma^{\mu}$.* Finally, it is conventional to include the antisymmetric

* Note that, as defined, in four dimensions we have $\{\gamma_5, \gamma^{\mu}\} = 0$.

product (with six independent components) $\sigma^{\mu\nu} \equiv \frac{\mathrm{i}}{2}[\gamma^\mu, \gamma^\nu]$. There are then five distinct types of currents:

$$\bar{\psi}\,\mathbb{1}\,\psi, \quad \bar{\psi}\,\gamma_5\,\psi, \quad \bar{\psi}\,\gamma^\mu\,\psi, \quad \bar{\psi}\,\gamma_5\gamma^\mu\,\psi, \quad \bar{\psi}\,\sigma^{\mu\nu}\,\psi, \tag{2.2.14}$$

which go under the names of **s**calar, **p**seudoscalar, (polar) **v**ector, **a**xial (or pseudo-) vector and **t**ensor, respectively. These names reflect their properties under Lorentz transformations and also under spatial inversion. Each current also has specific properties under temporal inversion and charge conjugation.

From the preceding discussion one sees that the vector current is related to the momentum of a particle while examination of the role of γ_5 and $\gamma_5\gamma^\mu$ reveals that the axial-vector current is related to intrinsic spin. The others, however, have no simple physical interpretation, while the form of vector and axial-vector currents suggests how to proceed in order to obtain parity-violating matrix elements. Given the above form of the spinor solutions, one finds that $\bar{\psi}\gamma_5\gamma^\mu\psi$ measures just the spin s^μ of a particle.

Exercise 2.2.2. *Verify that $\bar{\psi}\gamma_5\gamma^\mu\psi$ does indeed give precisely the spin s^μ of the particle and find the constant of proportionality.*

2.2.3 Fermi theory

The basis for the Fermi theory of nuclear β-decay (1934)* rests on the *ansatz* of a four-body point-like interaction:

$$\mathcal{M}_{fi} \;\propto\; G_{\mathrm{F}} \int \mathrm{d}^3\boldsymbol{x}\, \psi^*_{A'}(\boldsymbol{x})\, \psi^*_e(\boldsymbol{x})\, \psi^*_\nu(\boldsymbol{x})\, \psi_A(\boldsymbol{x}). \tag{2.2.15}$$

The (trivial) spin structure implicit in this form for the matrix element naturally leads to a description of the so-called Fermi transitions, in which the e–ν pair carries zero total angular momentum. Including the two-component spinorial forms due to Pauli allows for a description of the so-called Gamow–Teller transitions, in which the e–ν pair have their spins aligned and so carry one unit of angular momentum. However, there is no apparent relation between the two types of transitions. Phenomenologically, even taking into account a factor three for the triplet final state, a further (phenomenological) factor $\sim 5/4$ is needed if the same constant G_{F} is to describe both.†

* The 1938 Nobel Prize for Physics was awarded to Enrico Fermi for "his demonstrations of the existence of new radioactive elements produced by neutron irradiation, and for his related discovery of nuclear reactions brought about by slow neutrons."

† To be precise, a complete description of the substructure in terms of quarks is required, but we shall see how this works later.

One of the great assets of Dirac's relativistic formulation is that it places severe constraints on the different components, linking spin-dependent and -independent matrix elements. In order to replace the simple non-relativistic Fermi interaction above with a relativistic version, the concept of current–current interaction must be introduced, which is borrowed from the only complete theory we have: namely, quantum electrodynamics (QED). Thus, it is natural to write the possible interaction terms schematically in the form

$$\propto G_{\mathrm{F}} \int \mathrm{d}^3\boldsymbol{x}\, j_1^\dagger(\boldsymbol{x}) \cdot j_2(\boldsymbol{x}), \tag{2.2.16}$$

where $j_{1,2}$ may be any combination of the above five currents above, provided that the indices may be suitably saturated. This is the most general form compatible with the Dirac construction. It turns out that each current leads to different parity and angular-momentum selection rules and also to different angular distributions. One may thus experimentally identify which is the correct form. It is found that only the vector, $\bar{\psi}\gamma^\mu\psi$, and axial-vector, $\bar{\psi}\gamma_5\gamma^\mu\psi$, forms (describing Fermi and Gamow–Teller transitions, respectively) contribute to the weak interaction.

Now, the parity assignments are precisely those of a vector (negative) and an axial-vector (positive). However, since the decay distribution is determined by the modulus squared of the transition matrix element, neither taken separately can provoke parity violation. Therefore, a combination must be taken. The two possible extremes are $V \pm A$ or

$$j^\mu_{\mathrm{W}} \;=\; \bar{\psi}\,(\gamma^\mu \pm \gamma^\mu\gamma_5)\,\psi. \tag{2.2.17}$$

The ordering $\gamma^\mu\gamma_5$ is chosen so that the factorised form $\gamma^\mu(\mathbb{1}\pm\gamma_5)$ makes explicit the natural projective property.

Exercise 2.2.3. *Using the Dirac matrix algebra and the spinor structure given, show that the operators* $P_{R/L} := \frac{1}{2}(\mathbb{1}\pm\gamma_5)$ *project onto right- and left-handed helicities respectively.*

It turns out that the signs of the asymmetries found in the experiments by Wu *et al.*, Goldhaber *et al.* and Garwin *et al.* (and many others) are consistent with just one specific choice: $V-A$ (Sudarshan *et al.*, 1958 and Feynman *et al.*, 1958). That is, the weak interaction is *maximally* parity violating and involves only *left-handed* currents. A simple and experimentally verifiable consequence of this form is that parity will only be violated in those nuclear β-decays in which *both* Fermi and Gamow–Teller transitions are possible and play a role. Indeed, parity violation is seen to be the result of an *interference* effect and, as such, is evidently a *quantum-mechanical* phenomenon—it cannot be accommodated by classical physics.

As mentioned above, in other than purely leptonic decays, when hadrons take part, this maximal violation is not always apparent owing to the complex internal structure of hadronic matter (in particular, that of the baryons). However, if hadrons are described as bound states of quarks and proper account is made for the quark-spin symmetry (see the quark model later), then one finds that again the violation is maximal at the purely quark level and the weak current is still precisely $V-A$. Let us again emphasise that the $V-A$ form factorises as follows:

$$\bar{\psi}\,(\gamma^\mu - \gamma^\mu\gamma_5)\,\psi \;=\; 2\bar{\psi}\gamma^\mu\tfrac{1}{2}\,(\mathbb{1} - \gamma_5)\,\psi \;=\; 2\bar{\psi}_{\rm L}\gamma^\mu\psi_{\rm L}, \tag{2.2.18}$$

where $\psi_{\rm L} \equiv \frac{1}{2}(\mathbb{1}-\gamma_5)\psi$. This demonstrates that only left-handed (elementary) fermions interact weakly although, of course, the existence of right-handed fermions is not explicitly precluded.

It should also be appreciated that implicit in this description of parity (P) violation is the violation of charge conjugation (C) symmetry: all fermions that interact weakly are left handed while all antifermions are right handed. However, since the two violations go perfectly hand-in-hand the combined symmetry CP is conserved. In other words, while comparisons of either "mirror" or "antiparticle" experiments reveal differences, the translation to "mirror–antiparticle" experiments returns the original experimental observations, *e.g.*, β^+-decay viewed in a mirror is indistinguishable from β^--decay. Later on we shall examine very special systems in which even CP is violated.

Exercise 2.2.4. *Show how the transformation properties derived earlier lead to a form $\gamma^\mu(1+\gamma_5)$ for antifermion currents.*

2.3 Cabibbo theory

2.3.1 Universality

The Fermi description of the weak interaction requires a new coupling constant: $G_{\rm F}$. The β-decay matrix element is proportional to $G_{\rm F}$ and thus the decay rates are proportional to $G_{\rm F}^2$. In muon decay all participants are elementary and so there are no other unknown ingredients. Knowledge of the muon lifetime thus translates directly into a measurement of $G_{\rm F}$. We have (Patrignani *et al.*)

$$\tau_\mu \;=\; 2.19703 \pm 0.00004 \times 10^{-6}\ \text{s}, \tag{2.3.1}$$

which (in natural units) leads to

$$G_{\rm F} \;=\; 1.16637 \pm 0.00001 \times 10^{-6}\ \text{GeV}^{-2}. \tag{2.3.2}$$

Note that to arrive at this result, the full electroweak theory (see later) has been used. Now, already the formulation in terms of Dirac spinors has unified the description of Fermi and Gamow–Teller transitions. However, there still remains the question of the relationship between hadronic and leptonic weak couplings, and even between different hadrons. That is, how does the above value for $G_{\rm F}$ compare with that deduced, for example, from neutron β-decay or from electron–neutrino scattering? In other words, is it *universal*?

In the case of other purely leptonic processes the answer is simple, the agreement is perfect within errors. Here, though, the theory is rather simple: for purely leptonic processes, the particles participating are all elementary. When dealing with semi-leptonic and even more so for non-leptonic processes, the details of the hadronic bound state (baryon or meson) cloud the issue somewhat. Let us examine in more detail the case of neutron and also nuclear β-decay.

When Cabibbo was developing his theory of hadronic weak couplings in 1963 there was, as yet, no notion of the quark substructure and so the description was constructed purely in terms of the physical baryon or meson states. And this is how we shall proceed for the time being. As already noted, there are two general types of nuclear transitions (Fermi and Gamow–Teller), which we have discovered correspond to vector and axial-vector couplings respectively. In analogy with electromagnetism, it is natural to assign charges g_V and g_A respectively. As has already been hinted, the ratio g_A/g_V is not unity, as might be hoped by appealing to universality, but is nearer to 5/4.* However, given that hadrons are not point-like objects (evidently having non-trivial substructures) and that the stringent relations on particle spins provided by the Dirac formulation only apply to point-like elementary particles (*e.g.* the electron), it is not surprising that g_V and g_A do not appear to have a strict relationship.†

The vector charge g_V, by analogy with electromagnetism, is thus given quite simply as a sort of weak *charge*, which may be set to unity via the normalisation and definition of $G_{\rm F}$. On the other hand, g_A may be thought of as a sort of weak *magnetic moment* and, as such, for composite objects cannot be fixed by any normalisation. This means that in order to measure $G_{\rm F}$, processes are needed in which the transition is purely of the Fermi type. By noting that the Gamow–Teller transitions involve a spin-one electron–neutrino pair, we see that it is sufficient to consider transitions in which this is prohibited. In general, even if the variation of total nuclear spin ΔJ is zero, since angular momenta add vectorially, it is still possible to have $L_{e\nu} \neq 0$. The exception is the case in which $J = 0$ for both the

* The ratio may be determined by measuring various angular and/or spin correlations. The present PDG value is rather precise: $g_A/g_V = 1.2701 \pm 0.0025$ (PDG-2016 – Patrignani *et al.*, 2016). Note that the overall sign is a matter of convention.

† We shall see later, however, that the quark model of hadronic structure, with its symmetries, does actually salvage even this aspect.

initial and final states. There are just 9 such decays that have been measured; these are the so-called *superallowed* $J^P = 0^+$ transitions. Examples are:

$$^{10}\mathrm{C}(\beta^+)^{10}\mathrm{B}^* \quad \text{and} \quad ^{14}\mathrm{O}(\beta^+)^{14}\mathrm{N}^*. \tag{2.3.3}$$

The others are the β^+-decays of ^{26}Al, ^{34}AC, ^{38}K, ^{42}Sc, ^{46}V, ^{50}Mn and ^{54}Co. In such decays, the initial and final nuclear wave-function overlap integrals are also essentially free of nuclear uncertainties, since the wave-functions are almost identical. They thus provide a second, self-consistent and accurate determination of G_{F}.

However, the result of comparing the two determinations of G_{F} (*i.e.* from μ-decay and super-allowed nuclear decays) leads to a small discrepancy. Interestingly, attempts to reconcile the two numbers first looked to improved calculations and the quantum corrections: in particular, that due to the Coulomb repulsion (these decays are all β^+). However, such more accurate studies actually worsened the situation. The final discrepancy, although small in magnitude, was many standard deviations. Numerically, at the level of decay rates Γ, it was found that the nuclear decays were approximately $4-5\,\%$ smaller. In terms of the extracted nuclear G_{F} (since $\Gamma \propto G_{\mathrm{F}}^2$) this implies approximately $2\,\%$ weaker.

2.3.2 The mixing of weak and mass eigenstate

Attempting to resolve the issue, Cabibbo (1963) examined other known β-decays: namely, those of the so-called *strange* particle Λ^0. In this case it turns out that the discrepancy is enormous. Using these decay rates to extract the Fermi coupling, the valued obtained is approximate $22\,\%$ of the muon value. Now, $0.22^2 \sim 0.05$; this was interpreted by Cabibbo as the missing $5\,\%$. The idea then was to invoke the well-known quantum phenomenon of *mixing*. The physical particles actually detected experimentally correspond to mass eigenstates (technically *asymptotic* states), *i.e.* those that propagate in space–time, and these do not necessarily coincide wth the eigenstates of the weak interaction.

To understand this, recall, for example, the effect of an electromagnetic field applied to the hydrogen atom: the perturbing interaction mixes the usual hydrogen levels. In other words, the corresponding new eigenstates are different to those of the free hydrogen atom, being superpositions of these.

The strange β-decay considered is $\Lambda^0 \to p e^- \bar{\nu}$, where the final state is identical to that of neutron β-decay. Therefore, if the states are properly normalised, it is natural to describe such mixing, which here then is between just two states (n and Λ^0), via an angle:

$$n_{\mathrm{W}} = \cos\theta_{\mathrm{C}}\, n + \sin\theta_{\mathrm{C}}\, \Lambda^0, \tag{2.3.4a}$$

$$\Lambda^0_{\rm W} \;=\; -\sin\theta_{\rm C}\, n + \cos\theta_{\rm C}\, \Lambda^0, \tag{2.3.4b}$$

where $n_{\rm W}$ and $\Lambda^0_{\rm W}$ are weak eigenstates, n and Λ^0 being the usual mass eigenstates; $\theta_{\rm C}$ is known as the *Cabibbo angle*. The transitions must now be described by matrix elements involving these new states, that is, we must use so-called *weak currents*:

$$\mathcal{M} \;\propto\; G_{\rm F} \int {\rm d}^3\boldsymbol{x}\, j_h^{W\,\dagger}(\boldsymbol{x}) \cdot j_l^{W}(\boldsymbol{x}), \tag{2.3.5}$$

where $j^W_{h,l}$ describe weak hadronic and leptonic currents respectively and are thus of the form

$$j^\mu_{\rm W} \;:=\; \bar\psi^{\rm W}_f \left(g_V \gamma^\mu + g_A \gamma^\mu \gamma_5\right) \psi^{\rm W}_i. \tag{2.3.6}$$

In this theory the leptonic states do not suffer mixing (as we shall see later, this is to do with the fact that neutrinos are massless).*

With the mixing structure describe above, it is evident that matrix elements between the real physical states will pick up a factor $\sin\theta_{\rm C}$ or $\cos\theta_{\rm C}$ depending on whether the initial hadronic state is strange or not. That is, for the hadronic current, we now have

$$\begin{aligned} j^\mu_{{\rm W},h} \;&=\; \bar\psi_{{\rm W},p} \left(V^\mu - A^\mu\right) \psi_{{\rm W},n} \\ &=\; \cos\theta_{\rm C}\, \bar\psi_p \left(V^\mu - A^\mu\right) \psi_n + \sin\theta_{\rm C}\, \bar\psi_p \left(V^\mu - A^\mu\right) \psi_\Lambda. \end{aligned} \tag{2.3.7}$$

Taking $\sin\theta_{\rm C} \simeq 0.22$, both decays are then well described by one and the same coupling $G_{\rm F}$. This picture is well corroborated by the description it provides of other transitions, for example, the analogous pair of decays $\pi^- \to \pi^0 e^- \bar\nu$ and $K^- \to \pi^0 e^- \bar\nu$. In fact, there are many similar β-like transitions to which the theory may be applied. Note, however, that quite why the *same* Cabibbo angle should work for both baryons and mesons is evidently a mystery until we move over to the quark description of hadrons.

2.4 The GIM mechanism

2.4.1 A brief introduction to the quark model

Around the same time the *quark model* of Gell-Mann (1962)† was starting to take shape and, although originally only intended as a mathematical expression

* In recent years we have learnt that neutrinos are not, in fact, massless and so there can indeed be the same type of mixing in the leptonic sector too—we shall discuss this later.

† For a more complete review, see for example Gell-Mann and Ne'eman (1964) or Sec. 3.2 in the present notes.

of the underlying *flavour* symmetry, many started to consider the possibility that quarks were, in fact, real physical entities. Among those who saw early on the possibility that this could explain other puzzling experimental observations were Glashow, Iliopoulos and Maiani (1970). The problem they addressed was that of certain weak-decay channels, which were *not* observed, despite there being no apparent reason for any suppression. As we shall now see, once again quantum mechanics plays an important role, in this case via the phenomenon of *interference*. The particular processes under consideration were possible purely leptonic decays of the neutral kaon and its antiparticle.

Let us first rephrase the Cabibbo picture in terms of quarks. We now know that the baryons (among which we find the spin-half nucleons, the proton and the neutron, together with the Λ^0 already mentioned) are all composed of three quarks in particular spin configurations, which for the moment are inessential, while the mesons (*e.g.*, $\pi^{0,\pm}$, $K^{0,\pm}$ *etc.*) are simply quark–antiquark pairs. The neutron and proton are thus

$$|p\rangle \;=\; |uud\rangle \quad \text{and} \quad |n\rangle \;=\; |udd\rangle, \tag{2.4.1}$$

where the two quarks u and d are known as *up* and *down* respectively. The Λ^0 hyperon is similar but contains a strange quark s:

$$|\Lambda^0\rangle \;=\; |uds\rangle. \tag{2.4.2}$$

The β-decay process is then seen to be the transformation of either a d or an s into a u quark, accompanied by the emission of an $e^-\bar{\nu}_e$ pair. The remaining quarks are considered to be mere *spectators* and play no role. With no change in the reasoning, the Cabibbo mixing (2.3.4) can then be simply rewritten as

$$|d\rangle_{\rm W} \;=\; \quad \cos\theta_{\rm C}\,|d\rangle + \sin\theta_{\rm C}\,|s\rangle, \tag{2.4.3a}$$

$$|s\rangle_{\rm W} \;=\; -\sin\theta_{\rm C}\,|d\rangle + \cos\theta_{\rm C}\,|s\rangle. \tag{2.4.3b}$$

The case of charged pions and kaons is similar:

$$|\pi^-\rangle \;=\; |d\bar{u}\rangle \quad \text{and} \quad |K^-\rangle \;=\; |s\bar{u}\rangle. \tag{2.4.4}$$

And again, in the decays $\pi^-(K^-)\to\pi^0 e^-\bar{\nu}_e$ it is just a d (s) that decays into a u quark, accompanied by an electron–neutrino pair.

2.4.2 Feynman diagrams

By way of introduction to the concept, let us represent these decays with the use of so-called Feynman diagrams.* The idea is to assign each particle participating in a given process a *line* (internal lines represent propagation, external lines the initial- and final-state wave-functions) while the interactions between particles are indicated by *vertices*. Thus, the diagram describing neutron β-decay is shown in Fig. 2.3. In the diagram time flows from left to right while the vertical axis rep-

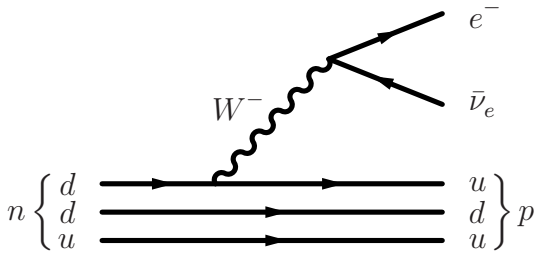


Figure 2.3: The Feynman diagram describing the neutron β-decay process in terms of the quark-level weak interaction.

resents generic spatial position. The arrows on the fermion lines do *not* represent direction of mechanical motion, but the flow of particle quantum numbers; thus, an antiparticle has an arrow directed backwards (in time).

The true significance of such a representation is that each element corresponds to a mathematical object, which then goes to make up the matrix element or scattering amplitude. We shall not labour this point here, but merely note that an important aspect of the Feynman rules is that all possible (different) diagrams connecting the same initial and final states must be *added* together to provide the total amplitude, which is then squared and integrated to give the final rate. This opens the way to quantum-interference effects: if two similar diagrams contribute with similar magnitude but *opposite sign* then they cancel each other in the sum (at least partially) and can thus lead to vanishing (or suppressed) rates even though there may be no real obstacle or selection rule forbidding the process.

2.4.3 Unobserved neutral-kaon decays

Consider then the two neutral pseudoscalar strange mesons K^0 and $\overline{K}^0$, the quark content of these two hadrons is, recall, $K^0 = d\bar{s}$ and $\overline{K}^0 = \bar{d}s$.† As we shall see later

* The 1965 Nobel Prize for physics was awarded equally to Sin-Itiro Tomonaga, Julian Schwinger and Richard P. Feynman for "their fundamental work in quantum electrodynamics, with deep-ploughing consequences for the physics of elementary particles."

† The assignment *particle–antiparticle* in the case of neutral mesons is dictated by the up–down quark content. The K^0 thus contains a d (or matter) while $\overline{K}^0$ contains $\bar{d}$ (antimatter).

this pair of particles has a rather complicated dynamics and CP is violated in their decays. However, for the present purposes they may simply be considered as the quark composites just defined. Of course, these states are meant to represent the mass eigenstates, or at least the quark mass eigenstates. Therefore, to describe possible weak decays, we shall need the rotated states unveiled in the previous section. Since the weak interaction then couples *both* the d and s quarks to the u quark, we can envisage and indeed calculate the rate for the decay $K^0 \to e^+e^-$, depicted in Fig. 2.4. This process is, however, *not* observed at the estimated rate.

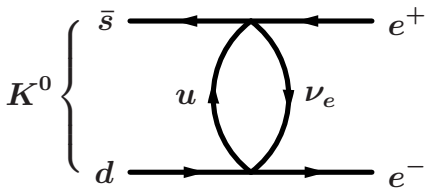


Figure 2.4: A possible process in which a K^0 might decay via internal annihilation into an electron–positron pair (recall that time runs from left to right).

Note that by exchanging the roles of the electron and the neutrino, we have a process in which a kaon apparently disappears into thin air—this too is not seen.

The way out of this *impasse* was provided by Glashow, Iliopoulos and Maiani (1970). Since the weak state d^W couples to the u quark, it is tempting to hypothesise that the orthogonal combination of mass eigenstates in s^W couples to some new state, the *charm* or c quark. If this is the case, then there is a second diagram, similar to that of Fig. 2.4, in which the u quark is replaced by c. Let us now draw the diagrams also including the $\cos\theta_{\rm C}$ and $\sin\theta_{\rm C}$ vertex factors with their relative *signs*, as given in Eq. (2.4.3) and shown in Fig. 2.5, we then immediately

Figure 2.5: The *destructively* interfering diagrams contributing to K^0 decay into an electron–positron pair.

see that there is a cancellation (or destructive interference) owing to the overall sign difference. Of course, the u and c quarks in the internal propagators do not have the same mass ($m_c \gg m_u$) and so the cancellation is not perfect, as it would be were the masses equal. This simply means that the non-observation of such a channel at some level provides an *upper limit* on the mass of the c quark.

Exercise 2.4.1. *One might imagine that a similar cancellation should apply to the decay* $K^0 \to \pi^+\pi^-$*. By considering the quark diagrams responsible for this channel, show that this is not the case.*

The Glashow–Iliopoulos–Maiani (GIM) mechanism was proposed in 1970 and the predicted new quark (with a mass that could not be much more than about 1.5 GeV) was discovered in e^+e^- collisions four years later by two independent groups (Aubert *et al.*, 1974 and Augustin *et al.*, 1974, see Sec. 4.2.2). This discovery then gave rise to the concept of quark (and lepton) *family*. At the quark level, we place the states into pairs of up-type and down-type, paralleling the first known pair u–d; the corresponding leptons are the e–ν_e and μ–ν_μ pairs. Although it must be recognised that Kobayashi and Maskawa (1973) had already contemplated such a structure (and more) and its consequences, as we shall now see.

2.5 The CKM matrix and CP violation

The following questions now naturally arise:

- Why have we taken the rather one-sided position of rotating (or labelling as specifically weak) the down-type states and not those of the up-type?*
- Since the state functions are naturally complex in quantum mechanics, the rotation matrix, in general, could contain complex phases – why then is the parametrisation in terms of only a single real parameter (an Euler angle)?
- What would happen if there were more quark states or families in the game?
- Since this happens for the weak interaction in the quark or hadronic sector, should there not be an analogous mixing among leptons?

All but the last (with which we shall deal immediately) will be treated in the following sections, the answer to the first will be answered at the end of the next subsection while we shall tackle the other two shortly.

The case of leptons is conceptually rather simple: in general, mixing occurs because there are two well-defined, unambiguous and distinct bases: the mass eigenstates (*i.e.* those that propagate and which therefore correspond to experimentally detected particles) and those of the weak interaction (*i.e.* those that are produced, decay or anyway interact). However, even assuming that the weak-interaction basis also remains well-defined for the leptons, the mass eigenstates of the neutrinos are ambiguous in the standard model (SM) since they are all considered *massless*. That is, they are *only* distinguishable by virtue of their weak interactions. Of course, in the wake of the recent experimental and theoretical developments concerning solar and atmospheric neutrinos, we now believe their

* Naturally, Cabibbo, not having the c quark, only had the possibility he actually adopted.

masses to be non-zero and, indeed, work is already under way on mapping out the leptonic mixing matrix.

2.5.1 *CPT*

In order to motivate the following discussion, we need to understand a little more of the discrete symmetries C and T, mentioned earlier, and their relationship through the so-called CPT theorem. A rather self-contained and sufficiently comprehensive discussion is provided for the unfamiliar reader in App. A.2—it is highly recommended to consult this before continuing.

2.5.2 The Kobayashi–Maskawa extension

We now turn to the other two intimately related questions posed: what happens in the case of more families and why is there no complex phase in the Cabibbo description? The *naïve* answer to the first of these is that the mixing matrix simply grows to be $n\times n$ for n families, but, as we shall see there is more to it than this. The second is a loaded question: a complex phase would allow for the violation of time-reversal invariance, which, if CPT is conserved, would be equivalent to and would imply CP non-conservation, as discussed in the appendix.

Let us attack the phase problem first. Once known that CP was *not* an exact symmetry of nature (Christenson *et al.*, 1964, see Sec. 2.5.6), the quest began for its origins and the possibility of introducing a complex phase, via the weak-interaction mixing matrix, thus became highly relevant. However, Cabibbo quite rightly described the two-component mixing in terms of a single, real, Euler angle. The reason is that there are the various constraints that must be imposed and also a certain phase freedom.

Firstly, the 2×2 matrix, let us call it $V_{\rm C}$, must be unitary ($V_{\rm C}^\dagger V_{\rm C}=\mathbb{1}$) in order to respect the ortho-normality of the bases involved. This implies four constraints on the possible four real amplitudes and four complex phases. Matrix theory tells that the constraints are actually divided up into three on the amplitudes and one on the phases. The matrix is thus already reduced to just the one real amplitude (or Euler angle) of Cabibbo. Secondly, the number of free phases is further reduced when one considers that each of the four quantum-mechanical quark states multiplied by the matrix possesses an arbitrary (unphysical and unmeasurable) intrinsic phase. All except one, which must serve as a reference point, may therefore be rotated so as to absorb phases in the matrix. The three remaining phases may thus all be absorbed into redefinitions of the quark states. Finally then, one is left with a real, one-parameter matrix, as proposed by Cabibbo.

This exercise naturally leads to investigation of the matrix in the case of more

than two families (Kobayashi and Maskawa, 1973).* A general, complex, $n\times n$, matrix contains n^2 real amplitudes and n^2 complex phases. Again, matrix theory reveals that unitarity imposes n^2 constraints: $\frac{1}{2}n(n+1)$ on the real amplitudes and $\frac{1}{2}n(n-1)$ on the imaginary phases. The $2n$ "external" quark states allow arbitrary phase rotations and thus absorption of a further $2n-1$ phases. A little arithmetic then reveals that the most general rotation matrix for n families may be described in terms of

$$\tfrac{1}{2}n(n-1) \qquad \textit{real} \text{ Euler angles} \tag{2.5.1a}$$

and

$$\tfrac{1}{2}(n-1)(n-2) \quad \text{complex phases.} \tag{2.5.1b}$$

The case $n=2$ just confirms what we have already described while $n=3$ leads to the possibility of three Euler angles and precisely one complex phase. Thus, the important finding of Kobayashi and Maskawa is that the mixing induced by the weak interaction in the case of three or more families is sufficient to accommodate a T- or CP-violating phase.

Note that, historically, at that time, a year before the discovery of charm, even the second family of leptons was still incomplete and it would not be until 1975 that evidence of a third family would emerge via the discovery of τ lepton by Perl *et al.*, with the b quark being discovered by Herb *et al.* in 1977.

Now, we still have the first question to answer: the justification of limiting the mixing to down-type quarks. Let us carefully examine how the currents should be defined in the presence of the most general possible mixing. *A priori*, we should define *both* up-type ψ_U^W *and* down-type ψ_D^W weak eigenstates, as distinct from their mass-eigenstate counterparts. There should thus be two $n\times n$ mixing matrices, $\mathcal{U}$ and $\mathcal{D}$ say:

$$\psi_U^W \;=\; \mathcal{U}\,\psi_U \;\text{ and }\; \psi_D^W \;=\; \mathcal{D}\,\psi_D, \tag{2.5.2}$$

where ψ_U and ψ_D now represent n-component spinors in *flavour* space:

$$\psi_U \;=\; \begin{pmatrix} u \\ c \\ t \\ \vdots \end{pmatrix} \quad \text{and} \quad \psi_D \;=\; \begin{pmatrix} d \\ s \\ b \\ \vdots \end{pmatrix}. \tag{2.5.3}$$

The natural extension of Eq. (2.3.7) is then

$$j^\mu_{\rm W} \;=\; \bar{\psi}_U^W (V^\mu - A^\mu)\,\psi_D^W$$

* One half of the 2008 Nobel Prize for Physics was awarded to Yoichiro Nambu for "the discovery of the mechanism of spontaneous broken symmetry in subatomic physics" and one quarter each to Makoto Kobayashi and Toshihide Maskawa for "the discovery of the origin of the broken symmetry which predicts the existence of at least three families of quarks in nature."

$$
\begin{aligned}
&= \mathcal{U}\overline{\psi}_U \,(V^\mu - A^\mu)\,\mathcal{D}\psi_D \\
&= \bar{\psi}_U \,\mathcal{U}^\dagger \,(V^\mu - A^\mu)\,\mathcal{D}\,\psi_D \\
&= \bar{\psi}_U \,(V^\mu - A^\mu)\,\mathcal{U}^\dagger\mathcal{D}\,\psi_D \;=:\; \bar{\psi}_U \,(V^\mu - A^\mu)\,V_{\rm CKM}\,\psi_D,
\end{aligned}
\tag{2.5.4}
$$

where the various steps are made possible owing to the fact that, for example, the Dirac matrices (γ) and the flavour matrices ($\mathcal{U}$ and $\mathcal{D}$) commute since they act on different spaces. One thus finds that only the combination $V_{\rm CKM} = \mathcal{U}^\dagger\mathcal{D}$ counts and, moreover, that this can be seen as acting *either* to the right and mixing the down-type quarks, *or* to the left and mixing the up-type quarks. Both interpretations are physically equivalent. It thus becomes a matter of mere convention that we describe this situation in terms of the mixing of down-type quarks.

The above discussion on the most general parametrisation and the possible introduction of a complex phase must now naturally be applied unaltered to the full Cabibbo–Kobayashi–Maskawa (CKM) matrix $V_{\rm CKM}$ with $n = 3$. This means that the matrix can, as anticipated, admit just one complex phase, with the consequent possible violation of time-reversal invariance. Given the CPT theorem, this translates into the parallel and consequent violation of CP, to which we shall now turn our attention.

2.5.3 The neutral-kaon system

We now come to what is possibly one of the richest systems in particle physics: the K^0–$\overline{K}^0$ pair. We recall that these two particles have the following quark composition:

$$
K^0 \;=\; d\bar{s} \quad \text{and} \quad \overline{K}^0 \;=\; \bar{d}s \tag{2.5.5}
$$

and thus, although neutral, they are *distinct* particles. However, they do have in common several decay modes: as with the charged kaons they can decay into two or three pions (with total charge zero). This fact has the consequence that they can *oscillate*, *i.e.* each may transform spontaneously, via a virtual two- or three-pion intermediate state, into the opposite particle or antiparticle state. Historically, however, the first problem arose with regard to their CP assignment.*

The first question is then: what is the CP signature of the two- or three-pion final states in neutral-kaon decays? First of all, recall that it is always true that, for a C eigenstate, $C = \pm 1$ (since $C^2 = 1$). Recall too that both neutral and charged pions have negative intrinsic parity.

* The phenomena described here are not to be confused with the case of the τ–θ puzzle described earlier, which involves the *charged* kaons and does *not* imply CP violation.

CP of the two-pion final state

As already discussed, the kaon and pion have spin zero and therefore angular momentum conservation forces the two-pion final state to have $L=0$. The overall parity of this state is then simply

$$P_{2\pi} \;=\; P_{\pi}^{2} \;=\; +1. \tag{2.5.6}$$

To discuss the signature under charge conjugation, we must distinguish between the two possibilities $\pi^0\pi^0$ and $\pi^+\pi^-$. In the case of a charged-pion pair, the operation of $\mathcal{C}$ interchanges the two and therefore introduces a factor $(-1)^L$ owing to the spatial wave-function, for an s-wave we thus have $+1$. The properties of a fermion–antifermion pair (such as go to make up a neutral pion) under $\mathcal{C}$ are such that $C_{\pi^0}=+1$. This is indeed confirmed experimentally by the observation of the principal decay mode

$$\pi^0 \;\rightarrow\; \gamma+\gamma \tag{2.5.7a}$$

and the *non*-observation of

$$\pi^0 \;\rightarrow\; \gamma+\gamma+\gamma. \tag{2.5.7b}$$

Putting all this together, we find that the two-pion final state in neutral-kaon decay must have $CP=+1$.

CP of the three-pion final state

There are again two possibilities to consider: $\pi^0\pi^0\pi^0$ or $\pi^0\pi^+\pi^-$. The presence of an extra particle complicates the discussion of both the spatial-inversion and charge-conjugation properties. Any pair may now have non-zero orbital motion with respect to the remaining pion, taken as a reference point. However, since the total must still be zero, the two must have identical L, with equal and opposite L_z. Thus, the final state with three neutral pions has spatial-parity signature $P=(-1)^{2L}=+1$ and therefore $CP=(+1)(-1)^3=-1$. The $\pi^0\pi^+\pi^-$ case is a little more complex as the charge-conjugation signature depends on the relative orbital angular momentum of the charged pair, which may be odd. However, studies of the decay angular distribution indicate $L_{\pi^+\pi^-}=0$, as might be deduced from the very low Q-value of this decay. The three-pion final state thus always has $CP=-1$.

Although the problem of parity violation was already understood and it was accepted that the weak two- and three-pion decays of the kaons violate P, the product symmetry CP was still believed to hold. Indeed, for example, the neutral pion, which is its own antiparticle, displays no evidence of CP violation. The problem was elegantly solved by Gell-Mann and Pais (1955). While the neutral pion is its own antiparticle, the same is not true for the neutral kaons and thus neither K^0 nor $\overline{K}^0$ is an eigenstate of $\mathcal{C}$ and therefore certainly not of $\mathcal{CP}$. However,

since $\mathcal{C}$ transforms K^0 into $\overline{K}^0$ and *vice versa*, the following linear combinations are easily seen to be eigenstates not only of $\mathcal{C}$ but also of $\mathcal{CP}$:

$$|K_1^0\rangle \;:=\; \tfrac{1}{\sqrt{2}}\left(|K^0\rangle + |\overline{K}^0\rangle\right) \tag{2.5.8a}$$

and

$$|K_2^0\rangle \;:=\; \tfrac{1}{\sqrt{2}}\left(|K^0\rangle - |\overline{K}^0\rangle\right). \tag{2.5.8b}$$

Note that the standard phase convention for the action of $\mathcal{C}$ sets

$$\mathcal{C}\,|K^0\rangle \;=\; -|\overline{K}^0\rangle \quad \text{and} \quad \mathcal{C}\,|\overline{K}^0\rangle \;=\; -|K^0\rangle. \tag{2.5.9}$$

As always, the intrinsic phase of a transformation such as $\mathcal{C}$ may be altered by redefining (or rotating) the phases of one or other of the states involved. Thus, the minus sign is merely conventional and has no effect on any physical results. With this definition and, of course,

$$\mathcal{P}\,|K^0\rangle \;=\; -|K^0\rangle \quad \text{and} \quad \mathcal{P}\,|\overline{K}^0\rangle \;=\; -|\overline{K}^0\rangle \tag{2.5.10}$$

we have

$$\mathcal{CP}\,|K^0\rangle \;=\; +|\overline{K}^0\rangle \quad \text{and} \quad \mathcal{CP}\,|\overline{K}^0\rangle \;=\; +|K^0\rangle, \tag{2.5.11}$$

which for our new superposition states $|K_{1,2}^0\rangle$ implies

$$\mathcal{CP}\,|K_1^0\rangle \;=\; +|K_1^0\rangle \quad \text{and} \quad \mathcal{CP}\,|K_2^0\rangle \;=\; -|K_2^0\rangle. \tag{2.5.12}$$

It is now obvious that the decays may be explained by associating the initial state K_1^0 (K_2^0) with the final state containing two (three) pions. In fact, since the two-pion decay mode has a shorter lifetime (by a factor of order 600) the two states are then identified as "K-short" ($K_{\rm S}^0 \to 2\pi$) and "K-long" ($K_{\rm L}^0 \to 3\pi$):

$$|K_{\rm S}^0\rangle \;:=\; \tfrac{1}{\sqrt{2}}\left(|K^0\rangle + |\overline{K}^0\rangle\right) \tag{2.5.13a}$$

and

$$|K_{\rm L}^0\rangle \;:=\; \tfrac{1}{\sqrt{2}}\left(|K^0\rangle - |\overline{K}^0\rangle\right). \tag{2.5.13b}$$

The phenomenology is thus perfectly well explained and CP is *not* violated here: in some given production process one might imagine that an s quark is created and encounters a $\bar{d}$ (*e.g.* from a virtual $d\bar{d}$ pair) to then form a $\overline{K}^0$. Now, such a state may be rewritten, by inverting the above relations, as an equal mixture of $K_{\rm S}^0$ and $K_{\rm L}^0$, which will then decay according to their natural probabilities into either two or three pions. The mean lifetimes are (PDG-2016 – Patrignani *et al.*,

2016):

$$\tau_{\rm S} \;=\; (0.8954 \pm 0.0004) \times 10^{-10}\ {\rm s} \quad \text{and} \quad \tau_{\rm L} \;=\; (0.5116 \pm 0.0021) \times 10^{-7}\ {\rm s}, \tag{2.5.14}$$

with $K^0_{\rm S}$ decaying predominantly into two pions (99.9 %) while $K^0_{\rm L}$ has a 68 % semileptonic branching ratio and only 32 % into three pions.

2.5.4 Regeneration

Thanks to the peculiar form of the neutral-kaon eigenstates, various interesting quantum-mechanical phenomena become possible. One such is that known as *regeneration*, suggested by Pais and Piccioni (1955). For the purposes of this discussion we may ignore any possible effects of CP violation (which we shall discuss shortly*) and take the $K^0_{\rm S}$ and $K^0_{\rm L}$ as being pure K^0_1 and K^0_2 states respectively. As already noted, in the generation of strange particles one typically produces either a strange quark or antiquark and therefore either a pure $\overline{K}^0$ or K^0 respectively. For definiteness, let us assume that predominantly K^0 is being produced (as is typically the case since matter contains many more d quarks than $\bar{d}$). That is, the beam created is initially an equal mixture or *superposition* of $K^0_{\rm S}$ and $K^0_{\rm L}$. Likewise, a beam of pure $K^0_{\rm L}$, say (as will always be the case after a period of time that is long with respect to the $K^0_{\rm S}$ lifetime), can be viewed as an equal mixture of $\overline{K}^0$ and K^0. The question now arises as to how such states evolve in time.

Let us first recall that the two states $\overline{K}^0$ and K^0 are very different with respect to their content in terms of ordinary matter (by which we mean up and down quarks): the first contains $\bar{d}$ while the second contains d. This means that the first may undergo strong interactions in which the $\bar{d}$ annihilates with a d found in matter while, since ordinary matter does not contain $\bar{d}$, the second may not. Therefore, while the $\overline{K}^0$ is very likely to decay or effectively *disappear* on contact with matter, K^0 is not as it may only interact weakly or electromagnetically. The *background* to this disappearance is just their normal weak decays, which are no competition for the strong interaction. It is thus expected that on passage through matter a $\overline{K}^0$ beam should be subject to severe attenuation, while a K^0 beam should survive much longer.

More formally, after some time we may say that fractions f and $\bar{f}$ of initially pure K^0 and $\overline{K}^0$ beams will survive, with $f \gg \bar{f}$. Thus, an initially pure $\overline{K}^0$ state will evolve (in the vacuum) into a pure state of $K^0_{\rm L}$ (that is, an equal mixture of K^0 and $\overline{K}^0$), which on passing through ordinary matter will become

$$\tfrac{1}{\sqrt{2}} \left\{ f|K^0\rangle - \bar{f}|\overline{K}^0\rangle \right\} \;=\; \tfrac{1}{2} \left\{ (f-\bar{f})|K^0_{\rm S}\rangle + (f+\bar{f})|K^0_{\rm L}\rangle \right\}. \tag{2.5.15a}$$

* The phenomenon we are about to describe will actually turn out to be a source of background for CP-violation measurements.

Now, since f and $\bar{f}$ are different or rather $(f-\bar{f})\neq 0$, we have the *reappearance* of $K^0_{\rm S}$. In fact, since $f \gg \bar{f}$, then to a good approximation we may write the new state as

$$\simeq \tfrac{1}{2}\left\{f|K^0_{\rm S}\rangle + f|K^0_{\rm L}\rangle\right\}, \qquad (2.5.15b)$$

or roughly equal populations. This can, of course, be easily tested by, *e.g.*, the observation of two-pion decays immediately after passage through matter, where immediately prior there were none. This phenomenon is known as *regeneration*.* The first experimental demonstrations were performed by Good *et al.* (1961).

2.5.5 Quantum oscillation

A related, but more subtle, effect is that known as *oscillation*, in which states effectively transform back and forth between K^0 and $\overline{K}^0$. The phenomenon is mathematically the same as the effect known as *beating* in wave mechanics, or more simply, as is the case here, that seen in a system of two *weakly* coupled oscillators. The central point in such phenomena is the presence of two slightly different natural frequencies in the system—in the case of weakly coupled oscillators, if the individual natural frequencies are identical, then the weak coupling induces a splitting between the two lowest possible coupled modes (typically *in* and *out of phase*); the *in-phase* mode usually has the lowest fundamental frequency while the *out-of-phase* mode is slightly higher. This difference results in *beats*: namely, if the starting condition has only one of the two oscillators in motion, then the subsequent evolution will see the other begin to move while the first comes to a stop and *vice versa*. The frequency of these beats is just the frequency difference between the two lowest modes.

Our oscillating system is just that of the particle states themselves: according to quantum mechanics, for an energy eigenstate we have

$$\phi(t,\boldsymbol{x}) \;=\; \phi_0(\boldsymbol{x})\,\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}Et}, \qquad (2.5.16)$$

where, for our purposes here, the spatial part $\phi_0(\boldsymbol{x})$ is irrelevant. Taking into account special relativity, the energy of a physical particle state must include its rest mass, $E=mc^2$. If we thus neglect the kinetic energy of the particles involved, we may substitute E above with mc^2. If then the particle (or antiparticle) state under consideration is described as an equal superposition of two states of different

* Naturally then, any attempts at measuring CP violation though detection of two-pion decays must avoid all regeneration effects, which could literally swamp the tiny CP-violating effect.

masses $m_{1,2}$ (with $m_2 > m_1$ say), at time $t = 0$ we have

$$|a\rangle \;=\; \tfrac{1}{\sqrt{2}}\left(|1\rangle + |2\rangle\right) \tag{2.5.17a}$$

and

$$|\bar{a}\rangle \;=\; \tfrac{1}{\sqrt{2}}\left(|1\rangle - |2\rangle\right), \tag{2.5.17b}$$

where, for clarity, we suppress any (irrelevant) spatial dependence. The inverse relations are

$$|1\rangle \;=\; \tfrac{1}{\sqrt{2}}\left(|a\rangle + |\bar{a}\rangle\right) \tag{2.5.18a}$$

and

$$|2\rangle \;=\; \tfrac{1}{\sqrt{2}}\left(|a\rangle - |\bar{a}\rangle\right). \tag{2.5.18b}$$

At a later time t, for an initially pure $|a\rangle$ state we have (also suppressing now the factors c^2 and $\hbar$)

$$|a,t\rangle \;=\; \tfrac{1}{\sqrt{2}}\left[\mathrm{e}^{-\mathrm{i}m_1 t}\,|1\rangle + \mathrm{e}^{-\mathrm{i}m_2 t}\,|2\rangle\right], \tag{2.5.19}$$

We then re-express this in terms of the particle–antiparticle states:

$$\begin{aligned} |a,t\rangle \;&=\; \tfrac{1}{2}\left[\mathrm{e}^{-\mathrm{i}m_1 t}\left(|a\rangle + |\bar{a}\rangle\right) + \mathrm{e}^{-\mathrm{i}m_2 t}\left(|a\rangle - |\bar{a}\rangle\right)\right] \\ &=\; \tfrac{1}{2}\left[\left(\mathrm{e}^{-\mathrm{i}m_1 t} + \mathrm{e}^{-\mathrm{i}m_2 t}\right)|a\rangle + \left(\mathrm{e}^{-\mathrm{i}m_1 t} - \mathrm{e}^{-\mathrm{i}m_2 t}\right)|\bar{a}\rangle\right] \\ &=\; \mathrm{e}^{-\mathrm{i}\overline{m}\,t}\left[\cos(\tfrac{1}{2}\Delta m\, t)|a\rangle + \mathrm{i}\,\sin(\tfrac{1}{2}\Delta m\, t)|\bar{a}\rangle\right], \end{aligned} \tag{2.5.20}$$

where $\overline{m} := \frac{1}{2}(m_1 + m_2)$ and $\Delta m := m_2 - m_1$.

We thus see that the particle–antiparticle content oscillates: the sine and cosine coefficients giving particle and antiparticle content $\cos^2(\frac{1}{2}\Delta m t) = \frac{1}{2}(1 + \cos\Delta m t)$ and $\sin^2(\frac{1}{2}\Delta m t) = \frac{1}{2}(1 - \cos\Delta m t)$ respectively. As already noted in the classical case, the oscillation frequency is then given by the energy (or mass) difference:

$$\omega_{\mathrm{osc}} \;=\; \frac{|m_2 - m_1|c^2}{\hbar}. \tag{2.5.21}$$

The physical meaning should be evident: a beam initially containing, say, only K^0 will at a later time actually contain some (oscillating) fraction of $\overline{K}^0$. This can be verified experimentally by examining the decays: K^0 (containing $d\bar{s}$) can decay into $\pi^- e^+ \nu$ while $\overline{K}^0$ (containing $\bar{d}s$) decays into $\pi^+ e^- \nu$. The measured lepton charge asymmetry as a function of time (or distance travelled) should thus oscillate.

The real experimental situation is a little more complicated owing to the finite and, indeed, rather short lifetimes of the particles involved. So, we must now

included the effects of decay into the above description. A state with a finite lifetime may be described by:

$$\phi(t) \;=\; \phi_0\, \mathrm{e}^{-\mathrm{i}Et}\, \mathrm{e}^{-\frac{1}{2}\Gamma t}, \tag{2.5.22}$$

where Γ is just the decay rate. This can be seen by considering the number density:

$$|\phi(t)|^2 \;=\; |\phi_0|^2\, \mathrm{e}^{-\Gamma t}, \tag{2.5.23}$$

which satisfies the standard exponential decay-law equation

$$\frac{\mathrm{d}}{\mathrm{d}t}|\phi(t)|^2 \;=\; -\Gamma\, |\phi(t)|^2. \tag{2.5.24}$$

The previous temporal evolution equations are then modified as follows

$$\begin{aligned} |a,t\rangle \;&=\; \tfrac{1}{2}\left[\mathrm{e}^{-\mathrm{i}m_1 t}\, \mathrm{e}^{-\frac{1}{2}\Gamma_1 t}\left(|a\rangle + |\bar{a}\rangle\right) + \mathrm{e}^{-\mathrm{i}m_2 t}\, \mathrm{e}^{-\frac{1}{2}\Gamma_1 t}\left(|a\rangle - |\bar{a}\rangle\right)\right] \\ &=\; \tfrac{1}{2}\left[\left(\mathrm{e}^{-\mathrm{i}m_1 t}\, \mathrm{e}^{-\frac{1}{2}\Gamma_1 t} + \mathrm{e}^{-\mathrm{i}m_2 t}\, \mathrm{e}^{-\frac{1}{2}\Gamma_2 t}\right)|a\rangle \right. \\ &\qquad\qquad \left. + \left(\mathrm{e}^{-\mathrm{i}m_1 t}\, \mathrm{e}^{-\frac{1}{2}\Gamma_1 t} - \mathrm{e}^{-\mathrm{i}m_2 t}\, \mathrm{e}^{-\frac{1}{2}\Gamma_2 t}\right)|\bar{a}\rangle\right], \end{aligned}$$

which we shall write as

$$=\; f(t)|a\rangle + \bar{f}(t)|\bar{a}\rangle. \tag{2.5.25}$$

Thus, the particle (*vis à vis* antiparticle) content of the beam is given by

$$\begin{aligned} \left|f(t)\right|^2 \;&=\; \tfrac{1}{4}\left|\mathrm{e}^{-\mathrm{i}m_1 t}\, \mathrm{e}^{-\frac{1}{2}\Gamma_1 t} + \mathrm{e}^{-\mathrm{i}m_2 t}\, \mathrm{e}^{-\frac{1}{2}\Gamma_2 t}\right|^2 \\ &=\; \tfrac{1}{4}\left[\mathrm{e}^{-\Gamma_1 t} + \mathrm{e}^{-\Gamma_2 t} + 2\, \mathrm{e}^{-\overline{\Gamma} t}\cos(\Delta m t)\right], \end{aligned} \tag{2.5.26}$$

where $\overline{\Gamma} := \frac{1}{2}(\Gamma_1 + \Gamma_2)$. Similarly one can calculate the antiparticle fraction. Since $\Gamma_{\mathrm{L}} \ll \Gamma_{\mathrm{S}}$, the only terms that survive in the large-t limit are $\frac{1}{4}\mathrm{e}^{-\Gamma_L t}$, which merely implies the expected survival (up to its own decay) of the part of the initial beam corresponding to the K^0_{L} state. A graphical representation of the fractional intensities (*i.e.*, normalised to the overall $\mathrm{e}^{-\Gamma_L t}$ decay behaviour) is displayed in Fig. 2.6.

Experimentally, as seen in Fig. 2.6, $\Delta m \tau_{\mathrm{S}} \sim 1\!/_2$. This is fortunate as it allows just *one* oscillation before the decay process kills the signal. Had the value turned out much smaller, *no* oscillation would have been observable. Equally, had it turned out much larger then the risk would have been that the rapid oscillations

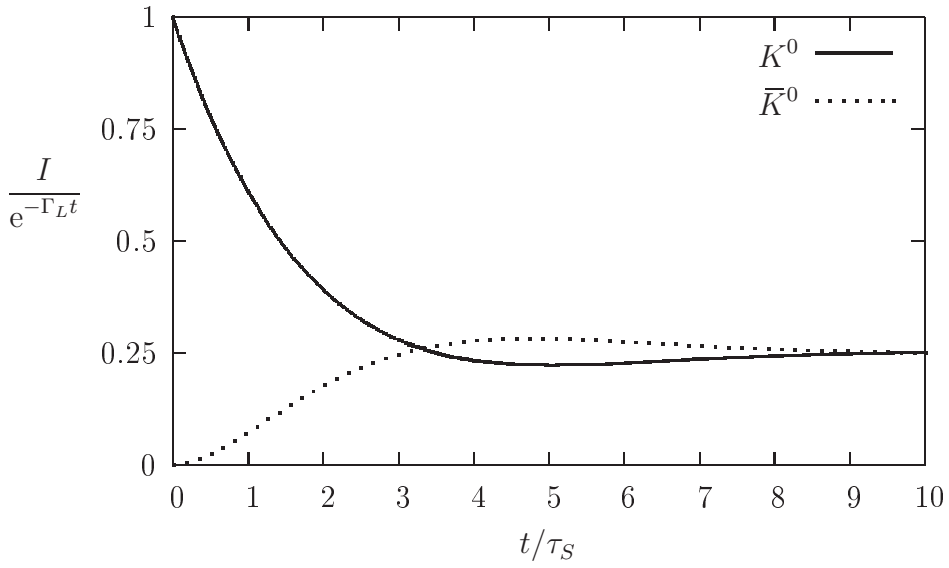


Figure 2.6: A graphical representation of the time-dependent K^0 and $\overline{K}^0$ intensities (normalised to the overall $e^{-\Gamma_L t}$ decay behaviour).

might have washed themselves out. The precise measured value is

$$\Delta m\,\tau_{\rm S} \;=\; 0.474 \pm 0.002, \tag{2.5.27}$$

which, given the value of $\tau_{\rm S}$, implies (see PDG-2016 – Patrignani *et al.*, 2016)

$$\Delta m \;\sim\; 3.5 \times 10^{-6}\ \text{eV}, \tag{2.5.28}$$

or

$$\frac{\Delta m}{m} \;\sim\; 0.7 \times 10^{-14}. \tag{2.5.29}$$

Experimentally, it is also found that $m_{\rm L} > m_{\rm S}$. The first such mass-difference measurements were performed by Good *et al.* (1961) and Fit.

As remarked earlier, this phenomenon is rather general and more specifically may also occur in the leptonic sector once non-zero neutrino masses are established. Indeed, there is already some understanding of the form of the lepton mixing matrix, which, in order to be non-trivial, requires the neutrino masses to be different. We might, however, remark here that technically the situation for oscillation is slightly different. The neutral-kaon mass difference is exceedingly small as compared to their rather large mass. In the neutrino case the masses are very small (especially with respect to any kinetic energy they might possess) while the differences are comparable. In particular, this means that it is not so much the mass differences themselves that are felt but differences in the corresponding

neutrino energies for a given momentum. The formalism thus changes somewhat.

2.5.6 *CP* violation

The enormous difference in decay rates suggests a possible method to search for CP violation (Christenson *et al.*[*], 1964). In short, if we wait for long enough (but not too long) all the $K^0_{\rm S}$ in an initially purely K^0 or $\overline{K}^0$ sample will have decayed and only the $K^0_{\rm L}$ component will have survived. Since the initial populations are equal, the ratio at some later instant t will be

$$\frac{N_{\rm S}}{N_{\rm L}} = \frac{{\rm e}^{-t/\tau_{\rm S}}}{{\rm e}^{-t/\tau_{\rm L}}}, \tag{2.5.30}$$

where, recall, $\tau_{\rm S}/\tau_{\rm L} \sim 1/600$. For $\tau_{\rm S} \ll t \lesssim \tau_{\rm L}$ this ratio is very small indeed: we thus have $N_{\rm S}/N_{\rm L} \sim O({\rm e}^{-600}) \sim O(10^{-260})$ and we should therefore no longer see any two-pion decays *at all*.

Exercise 2.5.1. *Ignoring relativistic time-dilation effects, calculate the mean distances that K^0_S and a K^0_L mesons moving at roughly the speed of light will travel before decaying.*

For their experiment Christenson *et al.* (1964) used the Brookhaven alternating gradient synchrotron (AGS) 30 GeV[†] proton beam, incident on a beryllium target, to produce a *secondary* beam of neutral kaons. The detector (shown in Fig. 2.7) was placed a little over 17 m away. An initial lead collimator and suitable magnetic fields ensured a relatively pure[‡] kaon beam while a final collimator guaranteed the direction of motion (important for reconstructing the kinematics). In order to exclude three-pion events in which one pion goes undetected, final pion pairs were selected by requiring their invariant mass to be near that of the K^0 (about 498 MeV);

Put simply, the measured branching ratio was found to be

$$\frac{K^0_{\rm L} \to \pi^+\pi^-}{K^0_{\rm L} \to \text{ all charged modes}} = (2.0 \pm 0.4) \times 10^{-3}. \tag{2.5.31}$$

This was based on a two-pion sample of 45 ± 10 events.

[*] The 1980 Nobel Prize for Physics was awarded jointly to James Watson Cronin and Val Logsdon Fitch for "the discovery of violations of fundamental symmetry principles in the decay of neutral K-mesons."

[†] In their paper Christenson *et al.* (1964) use the notation BeV, standing for billion electron-volts; the present-day accepted form is GeV.

[‡] Note that typical contaminants, such as photons and neutral pions cannot create the sought after two-pion final state.

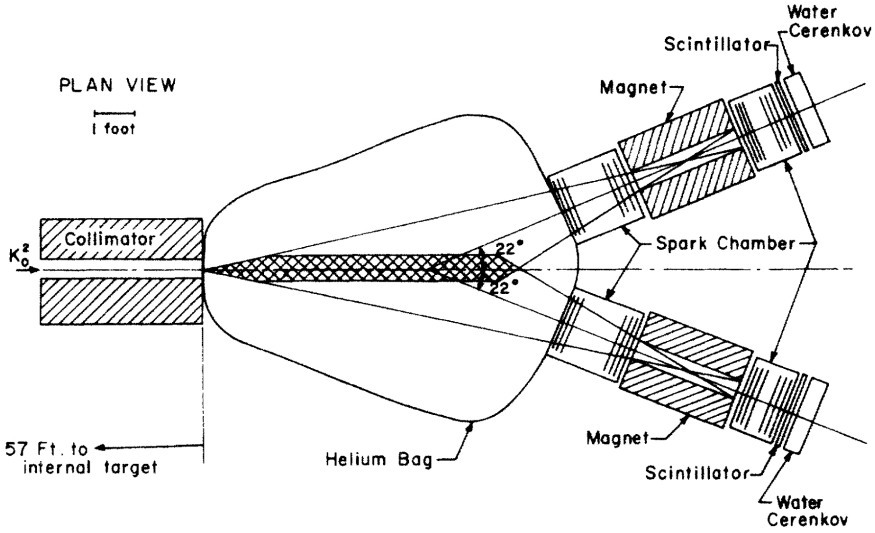


Figure 2.7: The apparatus used by Christenson *et al.* (1964) to detect two-pion decays of the $K_{\rm L}^0$ at a little over 17 m from the production point. Note that 1 foot $\simeq$ 30 cm.

The interpretation is that the $K_{\rm L}^0$ state is not purely K_2^0, but instead contains a small admixture of K_1^0:

$$|K_{\rm L}^0\rangle \;:=\; \tfrac{1}{\sqrt{1+|\epsilon|^2}}\left(|K_2^0\rangle + \epsilon|K_1^0\rangle\right), \tag{2.5.32a}$$

where the measured branching ratio implies $|\epsilon| \simeq 2.3\times10^{-3}$. The corresponding $K_{\rm S}^0$ state is

$$|K_{\rm S}^0\rangle \;:=\; \tfrac{1}{\sqrt{1+|\epsilon|^2}}\left(|K_1^0\rangle + \epsilon|K_2^0\rangle\right). \tag{2.5.32b}$$

We must, however, be careful in the explanation of the *observed* CP violation we have just described since there are, in fact, two distinct possible underlying phenomena:

Direct CP violation

The introduction of a complex phase into the CKM matrix allows for CP-violation at an elementary *interaction* level. For example, the decays $K_1^0 \to 3\pi$ and $K_2^0 \to 2\pi$ are thus possible. This then is a *direct* effect.

Indirect CP violation

It turns out that the K^0–$\overline{K}^0$ case just illustrated is an *indirect* consequence of CP violation inasmuch as it derives from mixing of CP eigenstates, which in turn is,

of course, due to CP violation in the interaction.*

Further measurements must be made in order to disentangle the two phenomena. Suffice it to note here that experimental results demonstrate that it is the *indirect* effect that dominates in neutral-kaon decays.

* By the same token, parity violation is observed in atomic physics, due to the effect of mixing of parity eigenstates—in this case such indirect effects are very small indeed.

2.6 Bibliography

??, to appear.

Aubert, J.J. *et al.* (1974), *Phys. Rev. Lett.* **33**, 1404.

Augustin, J.-E. *et al.* (1974), *Phys. Rev. Lett.* **33**, 1406.

Cabibbo, N. (1963), *Phys. Rev. Lett.* **10**, 531.

Chinowsky, W. and Steinberger, J. (1954), *Phys. Rev.* **95**, 1561.

Christenson, J.H., Cronin, J.W., Fitch, V.L. and Turlay, R. (1964), *Phys. Rev. Lett.* **13**, 138.

Dalitz, R.H. (1953), *Phil. Mag.* **44**, 1068.

Dirac, P.A.M. (1928), *Proc. Royal Soc. (London)* **A117**, 610.

Fermi, E. (1934), *Z. Phys.* **88**, 161; *transl.*, Wilson, F.D., *Am. J. Phys.* **36**, 1150.

Feynman, R.P. and Gell-Mann, M. (1958), *Phys. Rev.* **109**, 193.

Garwin, R.L., Lederman, L.M. and Weinrich, M. (1957), *Phys. Rev.* **105**, 1415.

Gell-Mann, M. (1962), *Phys. Rev.* **125**, 1067.

Gell-Mann, M. and Ne'eman, Yu., eds. (1964), *The Eightfold Way* (W.A. Benjamin).

Gell-Mann, M. and Pais, A. (1955), *Phys. Rev.* **97**, 1387.

Glashow, S.L., Iliopoulos, J. and Maiani, L. (1970), *Phys. Rev.* **D2**, 1285.

Goldhaber, M., Grodzins, L. and Sunyar, A.W. (1958), *Phys. Rev.* **109**, 1015.

Good, R.H. *et al.* (1961), *Phys. Rev.* **124**, 1223.

Gordon, W. (1926), *Z. Phys.* **40**, 117.

Herb, S.W. *et al.* (1977), *Phys. Rev. Lett.* **39**, 252.

Klein, O. (1927), *Z. Phys.* **41**, 407.

Kobayashi, M. and Maskawa, T. (1973), *Prog. Theor. Phys.* **49**, 652.

Lee, T.-D. and Yang, C.-N. (1956), *Phys. Rev.* **104**, 254; *erratum*, *ibid.* **106**, 1371.

Noether, E. (1918), *Nachr. v. d. Ges. d. Wiss. zu Göttingen*, 235; *transl.*, *Transport Theory and Statistical Mechanics* 183.

Pais, A. and Piccioni, O. (1955), *Phys. Rev.* **100**, 1487.

Patrignani, C. *et al.*, Particle Data Group (2016), *Chin. Phys.* **C40** [10], 100001.

Perl, M.L. *et al.* (1975), *Phys. Rev. Lett.* **35**, 1489.

Sudarshan, E.C.G. and Marshak, R.E. (1958), *Phys. Rev.* **109**, 1860.

Wu, C.-S., Ambler, E., Hayward, R.W., Hoppes, D.D. and Hudson, R.P. (1957), *Phys. Rev.* **105**, 1413.

Chapter 3

Hadronic Physics

In this chapter we shall discuss various aspects of the physics of the strongly interacting particles or *hadrons*. Hadrons are not elementary particles but are made up of combinations of quarks and antiquarks. The present description of interactions between quarks is constructed on the basis of a so-called *non-Abelian gauge theory* known as quantum chromodynamics (QCD). In particular then, we shall discuss the theoretical and experimental foundations of this theory, from the early developments of the quark–parton model, due to Gell-Mann and Feynman, up to the modern version of the theory of strong interactions.

3.1 Pre-history

3.1.1 The bootstrap model

Until as recently as the early seventies there was no truly fundamental theory of the strong interaction, but rather a model based mainly on ideas of mathematical self-consistency. Although QED provided an substantially complete quantum field theory for the electromagnetic forces, no such theory could be constructed for the strong interaction. First of all, this is not surprising since we know that the hadrons are not elementary particles and therefore cannot be expected to have point-like interactions. Moreover, if one insists on constructing a model in which, for example, the pion acts as the exchange field for the strong force, simple fits to say nucleon–nucleon scattering data indicate a strong fine-structure constant of order 10. This would render nonsense of a perturbative approach and such a model is thus almost useless.

Instead, theorists appealed to the general mathematical structure that typically emerges from quantum field theory approaches and attempted to directly construct the so-called S-matrix (or scattering matrix). The natural constraints to be applied

were then:

- analyticity,
- crossing,
- symmetry.

While the meaning of the last should be obvious, the other two (borrowed, so-to-speak, from field-theoretical descriptions) need a little explanation. At any rate, the idea was simply that, in the absence of a truly fundamental theory from which one could, in principle, have *calculated* the scattering matrix or S-matrix, the S-matrix should be *constructed* or rather *parametrised* from general principles and experimental data. The above requirements would then become constraints on such a construction.

Analyticity

This refers to a general mathematical property of scattering amplitudes: namely, that they should be analytic functions of the energies, momenta and other variables involved, which is actually a rather strong constraint. It means, in particular, that Cauchy's theorem applies and therefore the pole structure or spectrum of the theory determines to a large extent the general nature of the S-matrix.

Crossing

This indicates the requirement that amplitudes for processes differing only by the exchange of initial- and final-state particles should be obtainable simply via interchange of the relevant variables (*i.e.* four-momenta and possible spin variables). Thus, for example, the processes $\pi^0 p \to \pi^+ n$ and $\pi^- p \to \pi^0 n$ should be described by one and the same amplitude with the incoming and outgoing pion momenta exchanged. And by exchanging instead, say, the proton and π^+, the amplitude for $\pi^0 \pi^- \to \bar{p} n$ is obtained *etc.*

Symmetry

The concept of symmetry simply refers to the fact that the model must include or respect all known discrete and continuous symmetries of the strong interaction (such as C, P, T *etc.*) and all conservation laws (such as electric charge, strangeness, energy and momentum *etc.*). Moreover, as we shall now see, there are many more observed (at least approximate) symmetries, for which explicit account should be made.

Put together, these requirements place very stringent boundary conditions on the construction of possible scattering amplitudes. Such a picture was partially

justified by the development of a string theory* of hadronic interactions, which in turn was supported by Regge theory. However, as far as the *structure* of hadrons was concerned, one had a vision in which, so to speak, *everything was made of everything* and nothing was fundamental. This self-generating view of strong-interaction phenomenology led to the name "bootstrap".† While, to some extent, this permitted a self-consistent description and even some useful predictions, the overall agreement and predictive power were not acceptable for a complete theory or understanding of hadronic physics.

Before moving on, we should mention that while such an approach could never provide a fundamental description of hadronic interactions, nevertheless it has some use. In particular, it does not rely on perturbative techniques and can thus provide important information in those situations where standard perturbative methods fail. For a detailed discussion of the S-matrix, the reader is referred to the classic text by Eden *et al.* (1966) although the book by Collins and Martin (1984) provides more insight to modern applications.

3.1.2 The birth of quarks and partons

This evermore unsatisfactory situation spurred physicists on in their quest for newer and, in particular, more fundamental descriptions. Among these were Gell-Mann and Ne'eman (1964)‡ with the *quark*§ theory of the observed hadronic symmetries (see also Sakata, 1956, for early work in this direction) and Feynman (1969) with his point-like *parton* constituents of the proton. While these two roads were initially followed independently and were indeed somewhat orthogonal in inspiration, it soon became evident that the two pictures coincided and merely described two different aspects of the same fundamental objects: quark–partons. Mention should also be made of the role played by Bjorken (1969) in uniting the known symmetries with a high-energy (and therefore short-distance) view of particle interactions. This chapter describes then the unfolding of these two paths to their eventual unification and the successive development of the theory of the strong interaction now known as quantum chromodynamics (QCD).

---

* This is not to be confused with modern *superstring* theory, which is considered by many to represent a possible theory of *all* known particle interactions (including gravity).

† The origins of the expression "bootstrap", as employed here, are usually ascribed to Baron Münchhausen. The story attributes him with the claim that, finding himself once stuck in a swamp and unable to escape, he pulled himself out of the mud by his own bootstraps.

‡ The 1969 Nobel Prize for physics was awarded to Murray Gell-Mann for "his contributions and discoveries concerning the classification of elementary particles and their interactions."

§ Note that, according to Gell-Mann, the word "quark" rhymes with "walk" (*not* "park") and was *possibly* inspired by James Joyce's *Finnegan's Wake*.

3.2 Gell-Mann's flavour SU(3)

3.2.1 The eightfold way

The reader should already be familiar with the symmetry associated with what is known as *isotopic spin* or, more simply, *isospin*[*]; we shall now describe how such a picture is extended to include strangeness. Let us first remark that the necessity for introducing this new quantum number arose from observations such as the relatively long lifetime of the Λ^0 hyperon (and other so-called strange particles). Despite having a large enough mass to decay comfortably into $p\pi^-$ or $n\pi^0$ (*i.e.* with sufficient phase-space or Q-value so as not to be suppressed), the lifetime of the Λ^0 is $2.6{\times}10^{-10}$ s, far from that of a strong decay, for which typical lifetimes are of order 10^{-23} s. In addition, one notices that in *strong*-interaction processes certain particles (such as kaons and hyperons) are only ever produced in pairs, which on examination may be consistently assigned labels of strange or anti-strange.

First, let us note that, for historical reasons, the strangeness associated with the strange quark (and hence with baryons that contain one) is -1 while a strange antiquark has $S=+1$. Now, if one also defines a new quantum number B or baryon number, which is $+1$ for qqq baryons (-1 for antibaryons) and 0 for $q\bar{q}$ mesons, then by inspection one discovers the Gell-Mann–Nishijima relation (Gell-Mann, 1953; Nakano and Nishijima, 1953):

$$Q \;=\; I_3 + \tfrac{1}{2}(B+S) \;=\; I_3 + \tfrac{1}{2}Y, \tag{3.2.1}$$

where Q is the electric charge of the baryon in units of the proton charge and where we have also taken the liberty of introducing yet another quantum number,

$$Y \;:=\; B+S, \tag{3.2.2}$$

the hypercharge. This formula correctly reproduces the charges of *all* known hadrons (*i.e. both* baryons and *mesons*).

Combining isospin with hypercharge leads to a natural set of *periodic* tables for the baryons and mesons (see Fig. 3.1). The fact that both *baryons* and *mesons* fall into octets is an accidental property of SU(3). That they should also be the lowest mass states is, however, a property of the strong interaction itself. In Fig. 3.1 the masses of the particles are approximately the same along the rows (separately for each group, of course), with the exception of the isospin singlets: Λ^0 and η^0. The third component of isospin varies horizontally while the vertical axis represents the hypercharge (or strangeness since B is constant in any given diagram). The early identification of the pseudoscalar meson octet led Ohnuki already in 1960 to

[*] For a brief introduction to the concept of isospin see App. B.2.

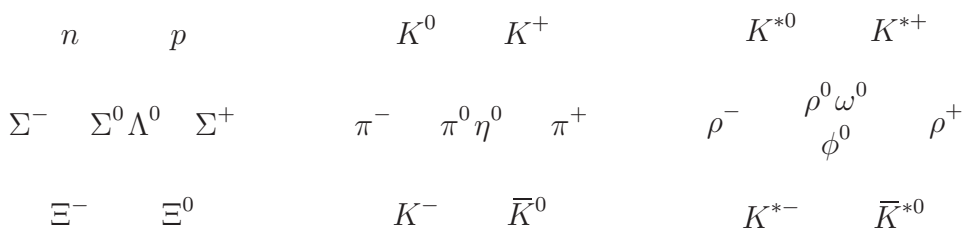


Figure 3.1: From left to right: the arrangement of the observed lowest-mass baryons, pseudoscalar mesons and vector mesons into three octets of *flavour* SU(3).

predict the existence of the (then unknown) η^0.

The rightmost group in Fig. 3.1 actually contains nine particles: along with the octet, there is an SU(3) would-be singlet (of larger mass). In fact, also in the pseudoscalar meson case one can identify a ninth (singlet) particle: namely, the $\eta'(958)$. In the case of the vector mesons the similarity of the masses favours strong mixing and thus the distinction between singlet and octet member loses any clear meaning. The baryon case is rather more complex: the colour and flavour wave-functions should be antisymmetric and thus zero orbital angular momentum and spin-½ are not possible if the wave-functions is to be overall antisymmetric as required by Fermi–Dirac statistics.

At any rate, such an arrangement in octets is readily explained via an underlying *flavour* SU(3) symmetry. In mathematical terms, baryons are products of three fundamental representations while mesons are constructed from one fundamental representation and one anti-fundamental representation. Such composite objects can the be decomposed into the following irreducible representations:

$$3 \otimes 3 \otimes 3 = 1 \oplus 8 \oplus 8 \oplus 10, \tag{3.2.3a}$$

$$3 \otimes \overline{3} = 1 \oplus 8. \tag{3.2.3b}$$

As noted, the baryon singlet is difficult to identify while the meson singlet is naturally associated with the η'. The extra baryon octet is observed as a set of so-called N^* resonances, with similar properties but heavier. The decuplet nicely accommodates the set of spin three-halves, isospin three-halves baryon resonances shown in Fig. 3.2. The question mark in the lowest entry of the table represents a particle that was unknown when the table was first laid down, but which was discovered shortly after (Barnes *et al.*, 1964): namely, the Ω^- an *sss* state (see Fig. 3.3).* Now, not only does the table evidently predict the existence of such

*It is interesting to note that in 1973 by reanalysing earlier cosmic-ray photographic-emulsion data Álvarez demonstrated that it had actually been unwittingly "seen" as early as 1954.

$$\Delta^- \quad \Delta^0 \quad \Delta^+ \quad \Delta^{++}$$

$$\Sigma^{*-} \quad \Sigma^{*0} \quad \Sigma^{*+}$$

$$\Xi^{*-} \quad \Xi^{*0}$$

$$?$$

Figure 3.2: The arrangement of the lowest-mass, spin-$^3/_2$, isospin-$^3/_2$, baryons into a decuplet of *flavour* SU(3).

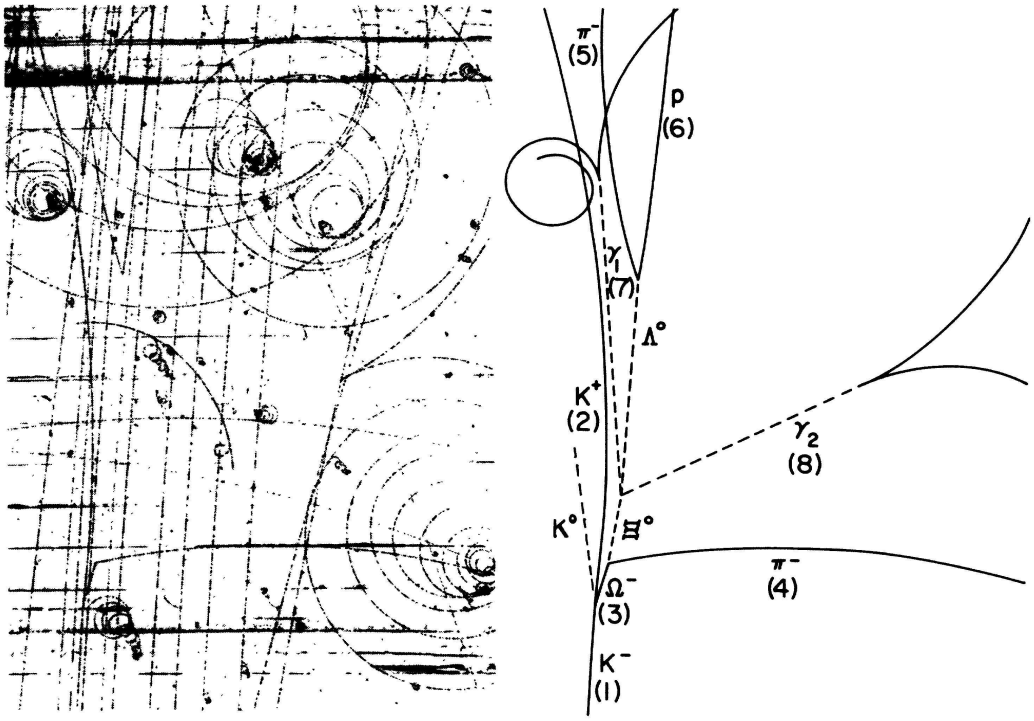


Figure 3.3: The bubble-chamber image in which the Ω^- was discovered. A K^- strikes a proton, producing $\Omega^- K^0 K^+$. These unstable particles then all decay. The $\Omega^- K^+$ charged decay products are registered. The dashed lines indicate neutrals, which do not produce tracks. The figure is taken from Barnes *et al.* (1964).

a particle, but also its mass. Indeed, SU(3) symmetry in this case leads to the simple prediction that the mass spacing between rows of the table is constant.

3.2.2 SU(3) mass relations

Now, the approximate SU(3) symmetry of the Gell-Mann model also leads to a surprisingly good description of the baryon-octet masses, in this case via just three parameters. Such detailed agreement is highly non-trivial: while the decuplet mass spacing is even, in the case of the octet it is not and, for example, the Λ^0–Σ^0 mass difference can neither be ignored nor simply ascribed to electromagnetic effects. The celebrated Gell-Mann–Okubo mass formulæ may thus be used, for example, to very successfully "predict" the mass of the Λ^0 hyperon, given the mean masses of the three isospin multiplets: n–p, $\Sigma^{-,0,+}$ and $\Xi^{-,0}$ (see also Coleman and Glashow, 1961, 1964).

To derive these formulæ, we need to examine the representations of SU(3). Recall that SU(2) is locally isomorphic to SO(3) (it actually provides a double covering for the rotation group) and has just three generators, which are conveniently represented by the three Pauli σ-matrices:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.2.4}$$

The SU(3) group, which incidentally is *not* related to any SO(N) group, has eight generators*, which can be constructed in a very similar fashion

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_2 = \begin{pmatrix} 0 & -\mathrm{i} & 0 \\ \mathrm{i} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \lambda_5 = \begin{pmatrix} 0 & 0 & -\mathrm{i} \\ 0 & 0 & 0 \\ \mathrm{i} & 0 & 0 \end{pmatrix}, \tag{3.2.5}$$

$$\lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \mathrm{i} \\ 0 & -\mathrm{i} & 0 \end{pmatrix}, \quad \lambda_8 = \tfrac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

The matrices in the first row are evidently a direct extension to 3×3 of the Pauli matrices (*i.e.* they connect u to d). On close inspection, one sees that also the pairs (λ_4,λ_5) and (λ_6,λ_7) play a role similar to that of (λ_1,λ_2), connecting u to s and d to s respectively. Indeed, the usual ladder operators to raise or lower the third

* In general, the group SU(N) has N^2-1 generators, which form an adjoint representation, while the fundamental representation is N dimensional.

component of isospin (or I-spin), so-called V-spin and U-spin may be constructed as follows:

$$\tfrac{1}{2}(\lambda_1 \pm \mathrm{i}\lambda_2), \qquad \tfrac{1}{2}(\lambda_4 \pm \mathrm{i}\lambda_5), \qquad \tfrac{1}{2}(\lambda_6 \pm \mathrm{i}\lambda_7). \tag{3.2.6}$$

Finally, the eigenvalues of the matrix λ_8 correspond to hypercharge.

To describe (though not truly explain) the baryon mass spectrum, one would naturally wish to write a formula of the form $m_B = \langle B|\mathsf{M}|B\rangle$, where M represents the (unknown) mass operator. We thus seek an SU(3) representation of the baryons themselves. Without formally deriving such, let us simply state that the following does the job:

$$\begin{pmatrix} \frac{1}{\sqrt{2}}\Sigma^0 - \frac{1}{\sqrt{6}}\Lambda^0 & \Sigma^+ & p \\ \Sigma^- & -\frac{1}{\sqrt{2}}\Sigma^0 - \frac{1}{\sqrt{6}}\Lambda^0 & n \\ \Xi^- & \Xi^0 & \frac{2}{\sqrt{6}}\Lambda^0 \end{pmatrix}. \tag{3.2.7}$$

The interpretation of (3.2.7) is that the matrix used to represent any given baryon will have entries corresponding to the coefficients of that baryon in the above matrix. Thus, for example,

$$p = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \Lambda^0 = \frac{1}{\sqrt{6}} \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \qquad \text{etc.} \tag{3.2.8}$$

As far as the SU(3) dependence is concerned, the interaction is then constructed by simply multiplying the tensors representing the physical states together with the relevant interaction matrix and saturating the indices. For the two-dimensional representations adopted here, this simply means matrix multiplication and an overall trace.

If SU(3) were exact, then the interaction would have the unit-matrix form shown below as M_0 while a term that violates SU(3) via the strange-quark mass (a reasonable though not proven hypothesis) should take the form of $\delta\mathsf{M}$ below:

$$\mathsf{M}_0 = m_0 \mathbb{1} \quad \text{and} \quad \delta\mathsf{M} = \delta m\, \lambda_8. \tag{3.2.9}$$

We now simply need to evaluate traces of products of the matrices involved.* The SU(3)-symmetric piece leads to something like $\mathrm{Tr}[\bar{B}\mathsf{M}_0 B]$, where $\bar{B}$ implies the transpose of the corresponding matrix. Since M_0 is proportional to the unit matrix this reduces trivially to $\mathrm{Tr}[\bar{B}B] = 1$. The SU(3)-breaking term is a little more difficult since λ_8 does not necessarily commute with B and thus *a priori* there are two possible inequivalent orderings. The best we can do is associate each

* Recall that traces are invariants and therefore depend only on the relevant symmetry-group structure and not on the representation adopted.

with a new parameter, as follows:

$$m_B \;=\; \langle B|\mathsf{M}|B\rangle \;=\; m_0 + \delta m_1\,\mathrm{Tr}[\bar{B}B\lambda_8] + \delta m_2\,\mathrm{Tr}[B\bar{B}\lambda_8]. \tag{3.2.10}$$

The final form is thus a three-parameter expression for the masses of the eight baryons.* The term m_0 corresponds to the value all the baryon masses would have for exact SU(3) symmetry while the two terms in $\delta m_{1,2}$ describe the symmetry breaking.

Exercise 3.2.1. *Evaluate the two coefficients of $\delta m_{1,2}$ in each of the four independent cases: N, Λ, Σ and Ξ.*

Now, since we have not considered isospin breaking, we have already effectively set $m_p = m_n$, the masses of the three Σs equal and also the pair of Ξs to have equal masses. Therefore, there are actually only four independent quantities to consider. However, this still leaves room for a prediction (or rather "post-diction"): for example, the Λ^0 mass is completely determined by the others. The result depends on how one weights the individual contributions in the three separate isospin multiplets and is also affected by the presence of electromagnetic corrections; nevertheless, the mass so obtained is in excellent agreement with the experimental value. Alternatively, one may use the four equations obtained from Eq. (3.2.10) to eliminate the three unknown mass parameters, this leads to

$$3m_\Lambda + m_\Sigma \;=\; 2(m_N + m_\Xi). \tag{3.2.11}$$

This is just one of many mass formulæ obtainable assuming an approximate (but broken) SU(3) flavour symmetry.

Exercise 3.2.2. *Using the coefficients derived in the previous exercise, verify the above mass relation.*

Finally, we note that one can also consider taking into account the small SU(2) or isospin breaking (leading to the proton–neutron mass difference). This certainly has at least two origins: the differing charges of the u and d quarks and the differing masses of the same. In such a simple picture (we have no dynamics here) the two effects cannot be separated, but may both be included in the formulæ via the inclusion of another breaking term:

$$\delta\mathsf{M}' \;=\; \delta m\,\lambda_3. \tag{3.2.12}$$

Two more parameters are necessary and thus one then has a five-parameter formula, to describe though the *eight* independent baryon masses. Indeed, another

* It is not difficult to show that any choice for the matrix $\delta\mathsf{M}$ that does not violate SU(2) (*i.e.* treats u and d quarks equally) would lead to an equivalent formula.

relation, due to Coleman and Glashow (1961), deals precisely with these so-called electromagnetic corrections:

$$(m_p - m_n) - (m_{\Sigma^+} - m_{\Sigma^-}) + (m_{\Xi^0} - m_{\Xi^-}) \;=\; 0. \tag{3.2.13}$$

Exercise 3.2.3. *Introduce $SU(2)$ breaking as described above and thus add two new coefficients, say $\delta m_{3,4}$. Calculate the four coefficients of δm_{1-4} and determine the parameters using the five independent cases of say p, n, Λ^0 and $\Xi^{0,-}$. Using the values thus found, predict the $\Sigma^{0,\pm}$ masses.*

Exercise 3.2.4. *As a final check, insert the known values into the previous mass formulæ and examine how closely they are actually satisfied.*

3.2.3 The nature of quarks

At this point it is perhaps relevant to note that although it is often stated that Gell-Mann himself did not believe in quarks as real physical objects inside hadrons, he has more recently claimed:

> "I always believed they were real—I just said that they had such strange properties that they were better stuck away where they can't be seen. But I didn't know that one could find them inside particles."

At any rate, many did begin to believe in the physical reality of quarks and indeed, independently, Feynman was already working towards a description of the possible constituents of hadrons or, more precisely, of the way they might reveal their presence through interaction with an external probe at very high energies (one should not forget here the important contributions of Bjorken).

If we do take the physical reality of quarks seriously, then assuming the lowest-mass baryons (the spin-$\frac{1}{2}$ octet and the spin-$\frac{3}{2}$ decuplet already mentioned) to all be composed of three quarks in an s-wave state, one evidently requires the quarks themselves to be spin-$\frac{1}{2}$. Their electric charges are easily determined from, say, the p–n system:

$$2Q_u + Q_d \;=\; 1 \quad \text{and} \quad Q_u + 2Q_d \;=\; 0, \tag{3.2.14}$$

leading to

$$Q_u \;=\; \tfrac{2}{3} \quad \text{and} \quad Q_d \;=\; -\tfrac{1}{3}. \tag{3.2.15}$$

The same conclusions are reached by considering the Δ quadruplet. Indeed, these charge and spin assignments, together with $Q_s = -\frac{1}{3}$ correctly reproduce the charges, spins and parities (taking into account the orbital angular momentum assignments) of all known hadrons, both baryons and mesons (not yet including, of course, charm or beauty). Finally, taking the individual quark magnetic moments as free parameters, it is possible to obtain similar formulæ for the baryon magnetic

moments; the agreement here is not quite so striking, but is nevertheless another success for the theory.

3.3 Feynman's parton model

3.3.1 High-energy electron–proton scattering

At this point in history (the late 1960's) the experimental capabilities became the determining factor in progress. At the Stanford Linear Accelerator Center (SLAC) the machine then in operation was capable of delivering an intense electron beam with an energy of around 2 GeV. It was being used to study the internal structure of the proton in much the same way that Rutherford and collaborators had used α-particles to study the internal structure of the atom (Geiger and Marsden, 1909; Rutherford, 1911). The correct energy to use is, of course, a question of the length scale one wishes to resolve. An α-particle with a kinetic energy of 5 MeV has momentum

$$p_\alpha \;=\; \sqrt{2\,m_\alpha c^2\,E_\alpha} \;\simeq\; \sqrt{2\times 4000\times 5}\ \text{MeV/c} \;=\; 200\ \text{MeV/c}. \tag{3.3.1}$$

Using the scale set by $\hbar c \sim 200$ MeV/c, we immediately see that the best resolution attainable is of the order of 1 fm (in practice it will always be rather poorer), which is what would be necessary to approach the nuclear size.*

If we now wish to look *inside* the proton, we evidently need a resolution roughly an order of magnitude better, which would mean a factor of 100 in the α-particle energy. However, in order to avoid confusing signals, we also require a probe that does *not* partake in the strong interaction and which thus avoids a convolution of the non-trivial structure of the projectile and that of the target. The obvious choice then, as indeed already much used in nuclear physics, is the electron. At such energies it is already highly relativistic, which actually simplifies the calculations since we have $pc \approx E$. And so an electron of energy 2 GeV has a resolving power of order 0.1 fm.

3.3.2 The parton model

In order to appreciate how and why Feynman was led to develop a picture of hard, point-like constituents inside the proton (for an interesting account of the model, see Feynman, 1972), we need to take a few steps back to the work done in nuclear physics. We wish to describe e–p interactions at *very high* energies, where

* Rutherford and collaborators were indeed able to provide an estimate for the size of a nucleus based on the observed deviations from the simple Rutherford formula.

the possibility arises to "shatter" the proton into numerous (hadronic) fragments (see Fig. 3.4). The initial and final electron four-momenta are $\ell^\mu = (E, \boldsymbol{\ell})$ and

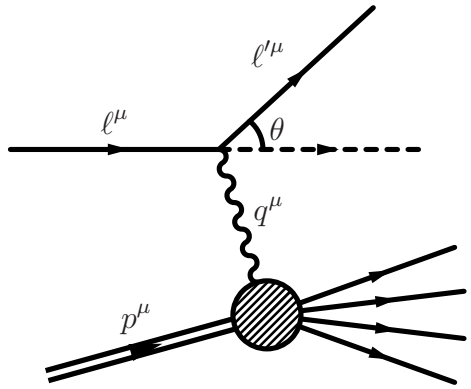


Figure 3.4: The Feynman diagram describing the electron–proton deeply inelastic scattering process, θ denotes the laboratory-frame electron scattering angle.

$\ell'^\mu = (E', \boldsymbol{\ell}')$ while θ denotes the angle between $\boldsymbol{\ell}$ and $\boldsymbol{\ell}'$ in the laboratory frame. In the energy regime of interest the energy transferred by the photon is

$$\nu \equiv E - E' \gg M, \tag{3.3.2}$$

where M is the nucleon mass.* We thus require a formalism capable of dealing with *inelastic* processes, which, by virtue of the short wavelengths (high energies) involved, take place in very small regions of space and thus probe the small-scale structure of the targets. Let us once again stress that, particularly in the field of particle physics, *high-energy* and *short-distance* are considered synonymous, as too are *low-energy* and *long-distance*—the dimensional translation factor is, as always, $\hbar c \sim 200\,\mathrm{MeV\,fm}$.

3.3.3 High-energy elastic *ep* scattering

The starting point will be the simplest form of e–p interaction: namely, elastic scattering. The low-energy case is dealt with in App. C.1.2, where we present the effective Mott formula, applicable when nuclear recoil may be neglected. At higher energies the recoil of the target nucleon is no longer negligible and the reduced Mott formula Eq. (C.1.6) must be replaced by the full form (C.1.8):

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}^{\mathrm{Mott}} = \frac{E'}{E}\frac{\mathrm{d}\tilde{\sigma}}{\mathrm{d}\Omega}^{\mathrm{Mott}}, \tag{3.3.3}$$

* What actually interests us is that this is more–or–less the same as the condition that the effective wavelength of the photon be much smaller than the radius of the proton.

where the factor $\frac{E'}{E}$ accounts for the (now non-negligible) recoil effect of the target nucleon. The four-momentum squared of the exchange photon assumes a certain importance and in the high-energy limit, where we may safely neglect the electron mass, we have

$$q^2 \;=\; (\ell - \ell')^2 \;\simeq\; -4EE' \sin^2 \tfrac{\theta}{2}. \tag{3.3.4}$$

Since this expression is evidently always negative, it is traditional to introduce the positive variable $Q^2 := -q^2$. This quantity appears in the propagator for the photon and therefore naturally sets the scale for the process. We shall see that in the energy regime of interest *both* ν and Q^2 become large.

As the electron energy increases and its motion becomes ultra-relativistic, it also becomes necessary to include the magnetic interaction (normally suppressed by a factor v/c). For a point-like particle having gyromagnetic ratio exactly two the full elastic cross-section takes on the form

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}^{\text{Dirac}} \;=\; \left[1 + 2\tau \tan^2 \tfrac{\theta}{2}\right] \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}^{\text{Mott}}, \tag{3.3.5}$$

where the suffix "Dirac" indicates a point-like cross-section for a spin-half object and the variable $\tau := \frac{Q^2}{4M^2}$. Note that the new term, proportional to $\tan^2(\theta/2)$, disappears for $\theta = 0^\circ$, reflecting the spin-*flip* nature of the magnetic interaction together with the usual requirement of angular-momentum conservation, coupled to electron-helicity conservation due to the vector nature of the interaction.

However, neither the proton nor the neutron is point-like. Indeed, even the neutron has an appreciable magnetic moment and can therefore scatter high-energy electrons with a cross-section comparable to that of the proton. Moreover, since we are now moving into a regime where the substructure becomes apparent, we must also take into account both the charge and magnetic-moment *distributions* inside the nucleons. As discussed in App. C.2, this simply requires the inclusion of form factors, which are nothing other than Fourier transforms of the distributions in question. The cross-section for elastic electron–nucleon scattering thus takes on the Rosenbluth form (1950):

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}^{\text{Rosen.}} \;=\; \left[\frac{G_E^2(Q^2) + \tau\, G_M^2(Q^2)}{1+\tau} + 2\tau\, G_M^2(Q^2) \tan^2 \tfrac{\theta}{2}\right] \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}^{\text{Mott}}. \tag{3.3.6}$$

In the limit $Q^2 \to 0$, where the photon wavelength becomes infinite, the scattering becomes effectively point-like and the electric and magnetic form factors $G_E(Q^2)$ and $G_M(Q^2)$ take on the their so-called "static" values:

$$G_E^p(0) \;=\; 1, \qquad G_M^p(0) \;=\; \;\; 2.79, \tag{3.3.7a}$$

$$G_E^n(0) \;=\; 0, \qquad G_M^n(0) \;=\; -1.91, \tag{3.3.7b}$$

that is, at zero momentum transfer the electric form factor measures the total charge while the magnetic form factor measures the magnetic moment (in units of the *nuclear* magneton). In the absence of a theory for these form factors, the Q^2 dependence must simply be measured experimentally. The dependence on θ allows a two-dimensional plot (Q^2 also depends on E and so is an independent variable), from which the functions $G(Q^2)$ may be extracted separately. For example, note that for Q^2 constant, the ratio $\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}^{\text{Rosen.}}/\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}^{\text{Mott}}$ is linear in $\tan^2\frac{\theta}{2}$, with slope $2\tau G_M^2(Q^2)$.

Performing such measurements, one finds that the three form factors with a non-zero limiting value all have a dipole-like behaviour (of the form mentioned in App. C.2) while the neutron electric form factor is more difficult both to measure and categorise:

$$G_E^p(Q^2) \;=\; \frac{G_M^p(Q^2)}{2.79} \;=\; \frac{G_M^n(Q^2)}{-1.91} \;=\; \left[1+\frac{Q^2}{M_V^2}\right]^{-2}, \tag{3.3.8}$$

where the single, phenomenological, mass parameter is $M_V \simeq 0.84\,\mathrm{GeV}$. It can be shown that such a dipole form corresponds to an exponentially decaying charge density:

$$\rho(r) \;=\; \rho(0)\,\mathrm{e}^{-ar} \quad \text{with} \quad a \;=\; 4.2\ \mathrm{fm}^{-1}. \tag{3.3.9}$$

Taking the $Q^2 \to 0$ limit of the experimentally measured slope, one deduces a typical root-mean-square radius for the nucleon of approximately 0.8 fm.

3.3.4 Deeply inelastic scattering

As the energy transfer increases, processes other than elastic scattering become possible and inelastic scattering sets in. Since the strict one-to-one constraint of the relation between the outgoing electron energy and scattering angle is then lost, for a fixed detector (or spectrometer) angle a broad spectrum of energies will be observed. The upper limit is obviously the standard elastic scattering final-state energy but many events are seen for energies below this. The first important structure one observes is due to the quadruplet of Δ resonances, with masses around 1230 MeV. It is convenient to introduce a new variable W, the final hadronic-state invariant mass. The four-momentum of the hadronic state emerging after photon absorbtion (assuming nothing is emitted) is simply $p^\mu + q^\mu$ and the invariant mass we seek is thus

$$W^2 \;\equiv\; (p+q)^2 \;=\; M^2 + 2M\nu - Q^2, \tag{3.3.10}$$

which, recalling that for fixed beam energy and scattering angle,

$$q^2 \;=\; -4EE'\sin^2\tfrac{\theta}{2}, \tag{3.3.11}$$

is linear in E'. Recall too that, as defined*, ν is actually implicitly a Lorentz invariant, as is of course Q^2. Figure 3.5 displays a typical cross-section or spectrum for $ep\to eX$ near the Δ resonance mass peak. Allowing for the underlying inelastic

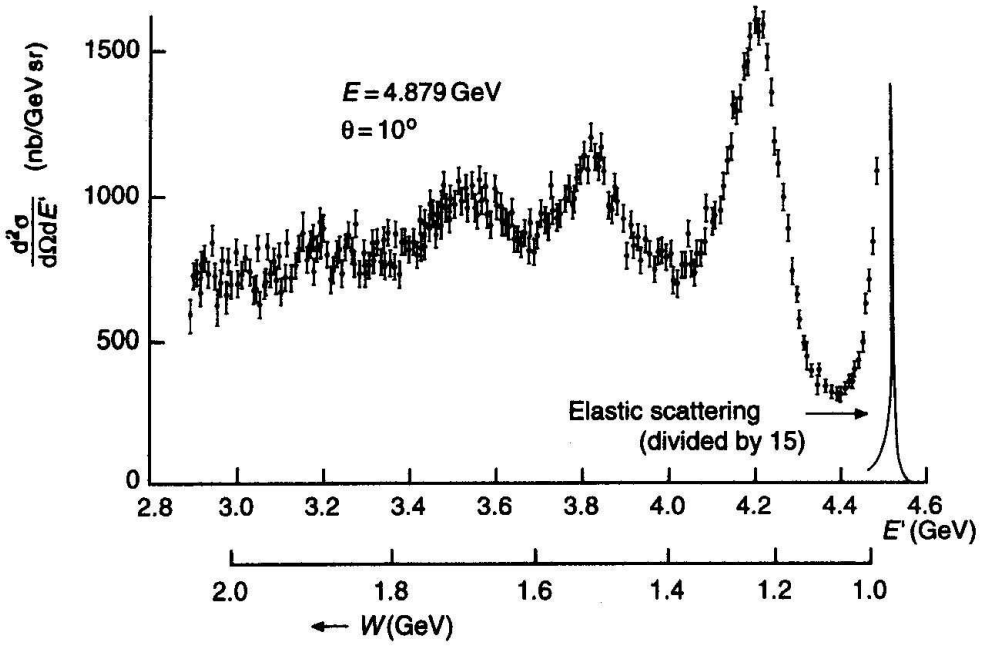


Figure 3.5: The cross-section for $ep\to eX$ for electron beam energy $E=4.9\,\mathrm{GeV}$ and scattering angle $\theta=10^\circ$. The figure is taken from Bartel *et al.* (1968).

events (*e.g.* pion production), the shape of the resonance is a classic Breit–Wigner (BW)†, from which one can deduce a mass of a little over 1200 MeV and a width of around 100 MeV. Up to around 2 GeV in W it is possible to identify other resonances, after which the continuum production of multiparticle states starts to dominate.

Exercise 3.3.1. *Calculate the maximum possible resolution, i.e. the shortest photon wavelength, obtainable with an electron beam energy of* 4.88 GeV. *For the same beam energy, calculate the maximum value of* W.

* Defined, that is, not specifically as the energy transfer in the rest frame, but as $M\nu = p\cdot q$.

† For a detailed discussion, the reader is referred to App. A.5

The question now is what will happen as we move to higher energies? The situation is very similar to that of Rutherford in the early 1900's: here the proton has taken over the role of the atom (and the electron that of the α-particle). So, if the proton is just a diffuse sphere of charge, again the large-angle cross-section is expected to fall off very rapidly. Instead the remarkable behaviour found in deeply inelastic scattering (DIS) at the end of the sixties (Bloom *et al.*, 1969; Breidenbach *et al.*, 1969) was that, far from dying away as Q^2 increased, as the above form factors would predict, the cross-sections remained large and (up to an overall dimensional scale factor) were independent of Q^2 for fixed ν/Q^2 ratio (see Fig. 3.6).* Note that such *scaling* behaviour was already observed from about $W = 2\,\mathrm{GeV}$ onwards and is therefore often termed "precocious". An early review may be found in Friedman and Kendall (1972).

Before continuing, let us try to understand the *expected* decrease in physical terms and thus better realise the implications of these findings. As the energy of the probe increases, the wavelength of the exchange photon decreases. At the moment of the interaction the photon effectively coherently "sees" only a small volume inside the nucleon, given roughly by the photon wavelength. Therefore, if the charge of the nucleon is distributed more–or–less uniformly throughout its volume, the interaction strength is expected to decrease rapidly with increasing energy. Since this does not happen, we are forced (just as was Rutherford) to entertain the idea that there are small, dense (or rather point-like) objects inside the nucleon, where its charge is then concentrated, thus avoiding the decreasing cross-section.

3.3.5 Bjorken scaling

For inelastic scattering there are two independent variables, which historically were taken as Q^2 and ν, and in place of the Rosenbluth formula (3.3.6) one writes

$$\begin{aligned}\frac{\mathrm{d}^2\sigma}{\mathrm{d}\Omega \mathrm{d}E'} &= \left[W_2(Q^2,\nu) + 2W_1(Q^2,\nu)\tan^2\tfrac{\theta}{2}\right]\frac{\mathrm{d}\tilde{\sigma}}{\mathrm{d}\Omega}^{\mathrm{Mott}} \\ &= \frac{4\alpha^2 E'^2}{(Q^2)^2}\left[W_2(Q^2,\nu)\cos^2\tfrac{\theta}{2} + 2W_1(Q^2,\nu)\sin^2\tfrac{\theta}{2}\right]. \qquad (3.3.12)\end{aligned}$$

However, as noted, the data are well described as a function of a single variable. This had already been largely foreseen by Bjorken (1969) essentially via dimensional analysis. He argued that since at high energies (and it must be admitted

* The 1990 Nobel Prize for physics was awarded equally to Jerome I. Friedman, Henry W. Kendall and Richard E. Taylor for "their pioneering investigations concerning deep inelastic scattering of electrons on protons and bound neutrons, which have been of essential importance for the development of the quark model in particle physics."

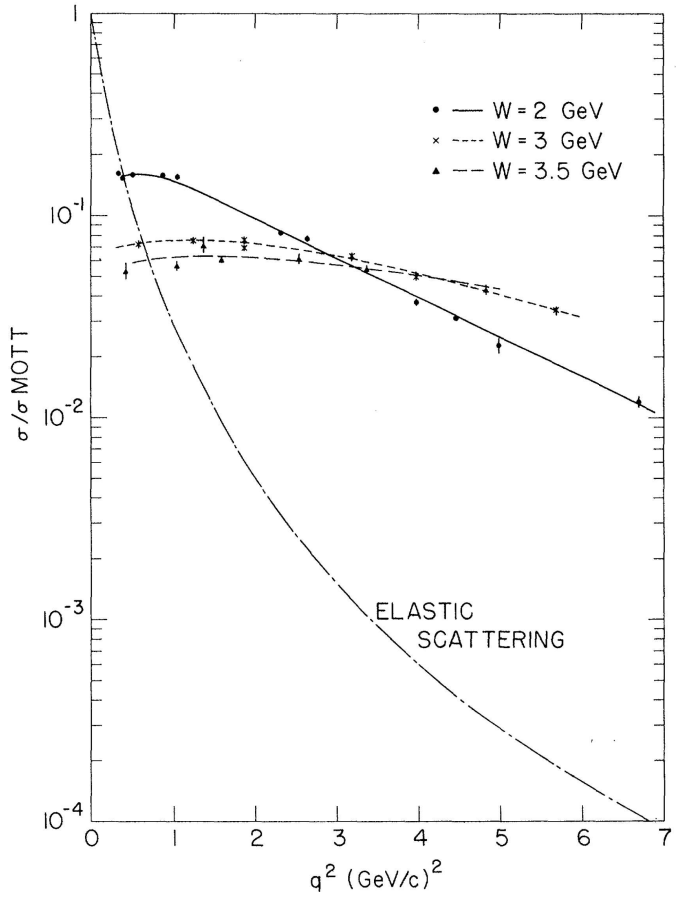


Figure 3.6: The ratio $(\mathrm{d}^2\sigma/\mathrm{d}\Omega\mathrm{d}E')/\sigma^{\mathrm{Mott}}$ as a function of q^2 for $W = 2$, 3 and 3.5 GeV in units of GeV^{-1}. The lines drawn through the data are meant to guide the eye. Also shown is the cross-section for elastic e–p scattering divided by σ^{Mott}, $(\mathrm{d}\sigma/\mathrm{d}\Omega)/\sigma^{\mathrm{Mott}}$ calculated for $\theta = 10^\circ$ using the dipole form factor. The relatively slow variation of the inelastic cross-section with q^2, as compared to the elastic cross-section, is clearly seen. The figure is taken from Breidenbach *et al.* (1969).

that he had rather more the just a few GeV in mind) all relatively small masses and energy-like parameters could be neglected. Therefore, once the *naïve* dimensions of, say, a cross-section (*e.g.* the E'^2/Q^4 pre-factor in the above formula) had been factored out, the remaining *dimensionless* form factors could only depend on *dimensionless* variables. Now, in DIS at high energies, for example, unless there is some new scale due to some new physics or dynamics, only two large quantities with dimensions of energy or mass remain important: namely, Q^2 and ν. One can only construct the adimensional, so-called, Bjorken scaling variable x_B, which

satisfies the following kinematical constraints:

$$0 \leq x_B := \frac{Q^2}{2M\nu} \leq 1. \tag{3.3.13}$$

It is thus an ideal candidate as the variable against which to plot the data.* We shall soon see that in the Feynman picture it also has a very special meaning. In 1969 Bjorken showed that the correct scaling behaviour is then obtained via the following substitutions:

$$MW_1(Q^2, \nu) \rightarrow F_1(x_B), \tag{3.3.14a}$$

$$\nu W_2(Q^2, \nu) \rightarrow F_2(x_B). \tag{3.3.14b}$$

3.3.6 The Feynman picture

It was, however, Feynman (1969) who gave more specific meaning to x_B and the form factors $F_{1,2}$ or *structure functions* (as they are now known). Assuming that there were point-like spin-half objects (which he called *partons*) inside the nucleon and that it was with these that the high-energy electromagnetic probe interacted, Feynman calculated the resulting cross-section, much in the same fashion as in App. C.3 for *quasi-elastic* scattering (see Fig. 3.7). The parton approach also

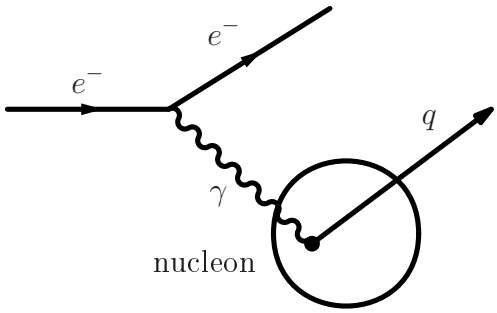


Figure 3.7: The process of electron–nucleon deeply inelastic scattering, according to Feynman's parton-model picture (the small blob inside the circle represents the struck parton bound inside the nucleon).

leads to a simplification: the two *a priori* unrelated structure functions $F_{1,2}$ turn out to be a single function (because the partons are treated as elementary Dirac particles with gyromagnetic ratio 2). Indeed, Callan and Gross (1969) quickly

* It might be mentioned that Bjorken himself used $x \equiv \frac{Q^2}{M\nu}$, which differs by a factor two. Indeed, in the early literature one can also find the inverse $\omega \equiv \frac{M\nu}{Q^2}$ while many authors (including Bjorken himself) still often simply used ν as the independent kinematic variable.

showed the following relation to hold in the high-energy, so-called, *scaling limit*:

$$F_2(x) \;=\; 2xF_1(x). \tag{3.3.15}$$

It is important to realise that this and indeed the whole picture is to be taken as an approximation that would be precise only in the limit $Q^2 \to \infty$. In fact, it turns out that higher-order quantum corrections also spoil the simple picture although they can be incorporated in a systematic way to provide precise and experimentally verifiable numerical predictions.

We can now try to understand the physical meaning of x_B and the structure functions that depend on it. It is natural (following the quasi-elastic scattering analogy) to attempt to describe the process as a collision between a virtual photon and a parton bound inside a nucleon. Indeed, one considers the lepton–photon vertex as so thoroughly understood as to be of no interest here. The only obstacle is that the partons are considered massless (or as having negligibly small masses) and thus their rest-frame is ill-defined. It is therefore necessary to work in a relativistically boosted frame. There are many possible choices of, so-called, *infinite-momentum* frames, but that which lends itself best to the present purpose is the Breit (or so-called "brick-wall") frame, in which the struck parton turns through 180° while retaining the same energy and absolute momentum, *i.e.* it simply reverses its momentum, as if it had indeed collided with a brick wall. In this frame the photon evidently carries zero energy and q^μ is therefore purely space-like. We shall also take the z-axis as the direction of the proton in this frame (the photon momentum is thus negative). Now, since $Q^2/2p{\cdot}q = x_B$ and $q^0 = 0$, then $q_z = -2x_B p_z$. To reverse the direction of the incoming parton (with momentum $\boldsymbol{k}$, say), we evidently require $q_z = -2k_z$. We thus finally obtain

$$x_B \;=\; k_z/p_z. \tag{3.3.16}$$

This is then the famous parton-model relation: x_B is just the fraction of the proton momentum carried by the parton as seen in an infinite-momentum frame.

Exercise 3.3.2. *Derive the above relation explicitly and thus demonstrate that the adimensional Bjorken scaling variable x_B, as defined in Eq. (3.3.13), is bounded to lie in the range* $[0,1]$.

At this point we have a picture in which a parton carrying a fraction x_B of the parent hadron momentum collides with an electron at very high energy and with very large momentum transfer. It may thus be reasonable (though more on this later) to avail ourselves of the impulse approximation already used in describing quasi-elastic scattering. In such an approach one can calculate the deeply inelastic scattering e–p cross-section for a given x_B as simply the product of the probability

of finding a parton with that momentum fraction and the cross-section for its *elastic* scattering with the incoming electron. Comparison of the two expressions immediately leads to the identification of the structure functions with probability distributions or densities for partons:*

$$F_2(x) \;=\; 2xF_1(x) \;=\; \sum_i Q_i^2 \left[x f_i(x) + x \bar{f}_i(x) \right], \tag{3.3.17}$$

where the sum runs over the different types or flavours of partons that might be found inside a hadron, Q_i is the charge (in units of the proton charge) of the i-th. parton type, f_i its probability distribution or density with respect to the momentum fraction x and $\bar{f}_i$ that of the corresponding antiparton.†

We have included antiquarks in the sum over parton types since we know that, via spontaneous quark–antiquark production (predicted by quantum field theory), at any given instant in time a hadron will also contain some (albeit small) fraction of antiquarks. Note that the cross-section is only sensitive to the charge squared Q_i^2 and therefore the contributions of partons and antipartons are indistinguishable. We may as well now start calling Feynman's partons *quarks*. We shall at times still continue to use the term *parton* since it may be taken to refer to *any* constituent of the proton, neutron or other hadrons and, as we shall see shortly, there are also the gluons to consider.

3.3.7 Difficulties with the Feynman approach

A number of (actually rather deep) questions now arise, which absolutely beg clarification:

1. parton transverse-momentum effects have been ignored,
2. we have no theory of the distributions $f(x)$,
3. gluons have not been included,
4. the destiny of the struck quark is not specified,
5. binding-energy effects have been ignored.

We shall deal with the second in some detail later. We shall also try to provide some understanding of the fourth and fifth shortly; they are related but are also somewhat more complex and profound issues. Let us begin, however, with a few brief comments.

* From now on, for simplicity of notation and according to accepted convention, we shall usually drop the suffix B on the Bjorken variable x_B.

† It should perhaps be stressed that x here has nothing whatsoever to do with position or configuration space.

Parton transverse-momentum effects

Let us then first comment on the momentum components of the quark in the plane transverse with respect to the z-axis (defined by the proton–photon directions in their centre–of–mass frame). In the above treatment they were totally ignored, this may be justified by assuming that they are due to Fermi motion of the quarks inside the nucleon, which is restricted to low momenta. According to the familiar Heisenberg principle, the mean (internal) momentum will be of the order of the inverse size of the nucleon in which they are bound and thus presumably of order 200 MeV. Moreover, transverse components are unaffected by longitudinal boosts. As a first approximation this is just fine although there are certain circumstances where the role of transverse momenta is non-negligible.

Parton distributions

As to a true theory for the quark–parton distributions $f(x)$, it simply does *not exist* as yet. The bound-state problem in QCD is still far from being solved and although various theoretical techniques and models have been developed, none provides truly satisfactory solutions for the bound states of three quarks or quark–antiquark pairs and certainly no approach is sufficiently advanced to provide complete *ab initio* calculations of the parton densities.

Gluons

As far as gluons are concerned, since they carry no electric charge, they cannot contribute to DIS in the Born approximation. However, at higher order in perturbation theory, a gluon inside the proton may spontaneously split into a quark–antiquark pair, of which one or other may then interact with the photon. Such effects can be calculated and included systematically into phenomenological descriptions. In fact, it is found that approximately only half the proton momentum is carried by quarks and antiquarks (these last carry only some 6 % or so), the other half being associated then with gluons.

Moreover, as we shall soon see, gluons carry the *colour* charge of QCD and therefore interact not only with quarks but also with each other. This means that in hadron–hadron collisions we may have a parton-level process in which two gluons collide and fuse to produce, *e.g.* a quark–antiquark pair. Indeed, at high energies such gluon–gluon fusion processes can even dominate the scattering cross-section. Yet another possibility is a gluon–quark Compton-like process.

Confinement and asymptotic freedom

The remaining two points are conceptually much tougher, especially in view of the fact that, no matter how high we go in energy experimentally, it has (so far) proved *impossible* to liberate a quark from inside its host hadron. The problem of binding leads us to two concepts that are central to the theory of strong interactions: namely *confinement* and *asymptotic freedom*. Indeed, so important are they that they deserve a dedicated section. However, this will be postponed until we have described at least a little of the nature of the interaction involved: namely, QCD.

3.4 Quantum chromodynamics

The attempt to construct a fundamental theory took as its starting point the theory known as QED. This is the field-quantised version of the classical field theory of electrodynamics, a gauge theory (the photon is rigorously massless) and as such possessing important symmetries, which guarantee suitable high-energy behaviour under quantisation. However, QCD differs from QED in that the gauge structure has a non-Abelian symmetry, *i.e.* it is of the Yang–Mills type (Yang and Mills, 1954). Apart from various theoretical complications this leads to a very important difference: there is not just one single charge but three. By analogy with the primary colours in optics, these are traditionally identified as red, blue and green. A quark may thus carry any one of these three charges while an antiquark carries an anti-colour charge and, as we shall explain, the (massless) gluon is also coloured.

3.4.1 Motivation for colour SU(3)

Let us first examine how such a theory came into being. The choice of $\mathrm{SU}(3)_{\mathrm{col}}$* as the strong-interaction gauge group (Greenberg, 1964; Han and Nambu, 1965) is uniquely determined by a number of phenomenological and theoretical observations (see Muta, 1998, for example). Note that the following do not constitute a requirement regarding the interaction, but merely indicate the nature of the symmetry group. However, it is natural, following in the footsteps of the highly successful theory of QED, to extend it to a local gauge symmetry and thus introduce a very desirable interaction. For a variety of reasons (which we shall shortly discuss), it is necessary to enlarge the symmetry group beyond the simple single parameter space of the QED U(1) to $\mathrm{SU}(3)_{\mathrm{col}}$

* There must be no confusion between $\mathrm{SU}(3)_{\mathrm{col}}$ and $\mathrm{SU}(3)_{\mathrm{flav}}$ – the first refers to the *local* gauge symmetry of the strong interaction (*i.e.* the gluon dynamics) while the second is a *global* symmetry of the quark fields irrespective of their interactions and contains *no* real dynamics.

(a) The group must admit a totally antisymmetric colour-singlet ("white") baryon composed of three quarks, qqq. Note that states with, *e.g.*, four quarks have never been observed. From the study of hadron spectroscopy it is known that the lowest-mass baryons, the spin-$\frac{1}{2}$ octet and the spin-$\frac{3}{2}$ decuplet of $\mathrm{SU}(3)_{\mathrm{flav}}$ (the approximate *flavour* symmetry that rotates the three light quarks u, d and s), are composed of three quarks in what are assumed to be colour-singlet states. Indeed, the qqq wave-function must be antisymmetric in colour, in order to satisfy Fermi–Dirac statistics. Consider, for example, a Δ^{++} with spin-z component $+\frac{3}{2}$: this has the form $|u^{\Uparrow}u^{\Uparrow}u^{\Uparrow}\rangle$ in an s-wave (likewise, Ω^- should be $|s^{\Uparrow}s^{\Uparrow}s^{\Uparrow}\rangle$ with $L=0$), *i.e.* three identical fermions in the same state. In space, spin and flavour the wave-function is thus totally symmetric and hence antisymmetry in colour is required for overall antisymmetry. This requirement is neatly satisfied by $\mathrm{SU}(3)_{\mathrm{col}}$ and the natural construct $\epsilon_{abc}q^aq^bq^c$, where a, b and c are $\mathrm{SU}(3)_{\mathrm{col}}$ indices.

(b) The group structure must admit complex representations in order to distinguish between quarks and antiquarks. In fact, there exist $q\bar{q}$ mesonic states while no analogous qq bound states are known. Among the simple groups, this restricts the choice to E(6), SU(N) with $N\geq 2$ and SO($4N+2$) with $N\geq 2$, taking into account that SO(6) has the same algebra as SU(4).

(c) The choice of the gauge group $\mathrm{SU}(N_{\mathrm{c}}=3)_{\mathrm{col}}$ is also confirmed *a posteriori* by many processes that directly or indirectly *measure* N_{c}. We shall now present some important examples.

The hadron-production rate in e^+e^- annihilation

The e^+e^- annihilation process proceeds via the production of a virtual intermediate neutral boson (γ or Z^0), which then "decays" into a fermion–antifermion pair (see Fig. 3.8). These may be charged leptons or quarks (neutrinos are also possible,

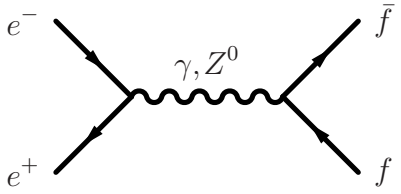


Figure 3.8: The e^+e^- annihilation process into a fermion–antifermion pair, via the intermediate production of a virtual photon or Z^0 boson.

though undetectable, in the case of Z^0). The overall rate for any given channel is proportional to the charge squared of the final-state fermions. Final states containing only hadrons are assumed to have their origins in a $q\bar{q}$ pair. Counting a

separate contribution for each quark colour, the total rate for hadron production in e^+e^- annihilation is thus proportional to N_c:*

$$R_{e^+e^-} \equiv \frac{\sigma(e^+e^- \to \text{hadrons})}{\sigma_{\text{point}}(e^+e^- \to \mu^+\mu^-)} = N_c \sum_f Q_f^2 \qquad (\text{for } 2m_f < E_{\text{CM}}),$$

where the sum runs over individual contributions (weighted by Q_f^2, the quark electric charge squared) from accessible $q_f\bar{q}_f$ final states. Above the $b\bar{b}$ threshold but well below m_Z we have $q_f = u$, c, d, s and b (t is, of course, too heavy):

$$R_{e^+e^-} \approx \left[2\times(\tfrac{2}{3})^2 + 3\times(-\tfrac{1}{3})^2\right] N_c = \tfrac{11}{9} N_c. \qquad (3.4.1)$$

The data nicely indicate $N_c = 3$, as seen from Fig. 3.9 (PDG-2016 – Patrignani *et al.*, 2016). Note that the cross-section excess in the data of a few percent with respect

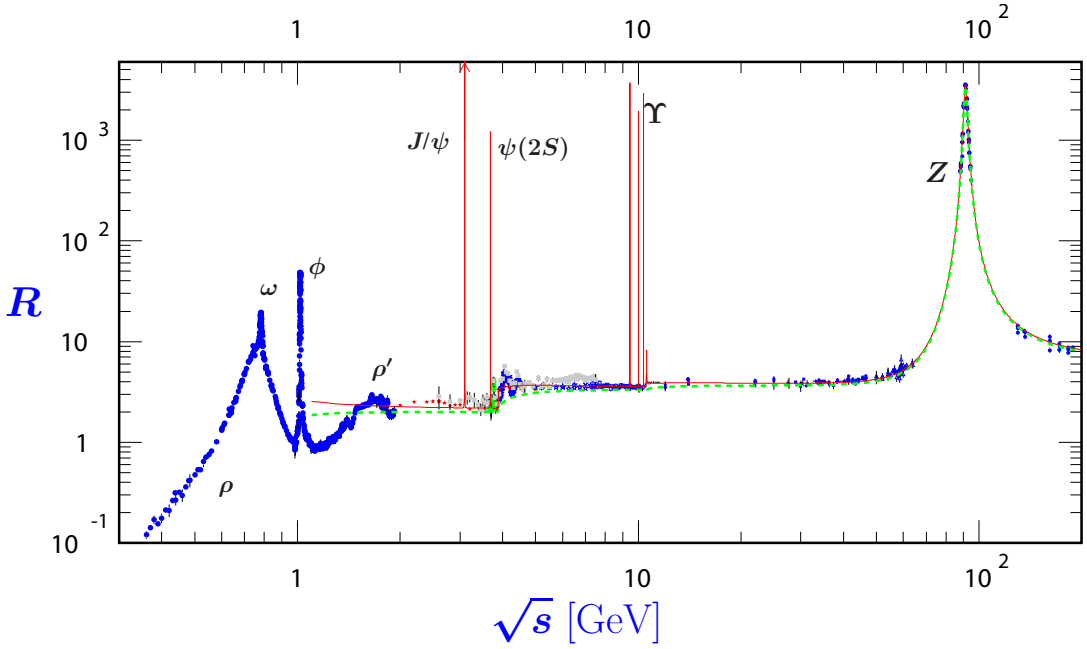


Figure 3.9: $R_{e^+e^-}$ as a function of total centre–of–mass energy. The figure is taken from PDG-2016 (Patrignani *et al.*, 2016).

to the value 11/3 can be accounted for by QCD radiative (or quantum) corrections.

* Important, known, quantum corrections have been neglected here (they will be discussed later).

The branching ratio $B(W^- \to e^- \bar{\nu}_e)$

A similar example is provided by the W^- decay rate (see Fig. 3.10). Again, in

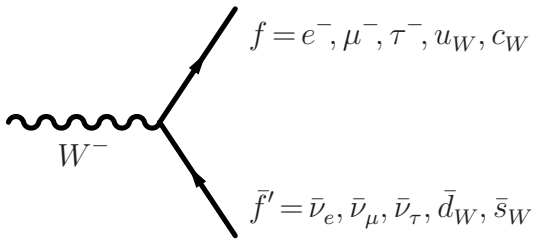


Figure 3.10: The general $W^- \to f\bar{f}'$ decay process (the natural charge-conjugate channels also exist for the W^+).

the Born approximation, the allowed fermion–antifermion ($f\bar{f}'$) final states in W^- decay are $e^-\bar{\nu}_e$, $\mu^-\bar{\nu}_\mu$, $\tau^-\bar{\nu}_\tau$, $(d\bar{u})_W$ and $(s\bar{c})_W$; the final state $(b\bar{t})_W$ is not possible since the top quark is too heavy to be produced. Each channel type (lepton or quark) contributes equally*, except that for quarks there are N_c colours and we must therefore include an extra weight factor $N_\mathrm{c} = 3$ for the quark channels:

$$B(W^- \to e^-\bar{\nu}_e) \equiv \frac{\Gamma(W^- \to e^-\bar{\nu}_e)}{\Gamma(W^- \to \text{all})} \approx \frac{1}{3 + 2N_\mathrm{c}}. \tag{3.4.2}$$

For $N_\mathrm{c} = 3$, $B = 11\,\%$ (it would be $20\,\%$ for $N_\mathrm{c} = 1$); to be compared with the experimental value $B = 10.75 \pm 0.13\,\%$.

The branching ratio $B(\tau^- \to e^- \bar{\nu}_e \nu_\tau)$

The τ lepton, having a mass of very nearly 1777 MeV, may decay into a number of final states, both leptonic and hadronic. The basic process $\tau \to f\bar{f}'\nu_\tau$, obviously an analogue of β-decay, is depicted in Fig. 3.11. Considering the energetically

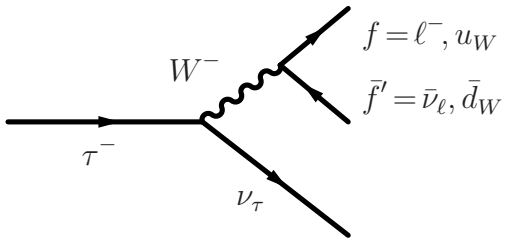


Figure 3.11: The general $\tau^- \to \nu_\tau f\bar{f}'$ decay process (ℓ stands for either of the two lighter charged leptons).

* Quark and lepton mass effects (m_c and m_τ) may be more-or-less neglected here.

available final-state channels, the $f\bar{f}'$ pair may be $e^-\bar{\nu}_e$, $\mu^-\bar{\nu}_\mu$, or $(d\bar{u})_W$. In principle, neglecting the small mass effects, each should contribute with equal weight. However, if the quarks are coloured, the number of $(d\bar{u})_W$ states available becomes $N_c = 3$. The branching ratio $B(\tau^- \to e^-\bar{\nu}_e\nu_\tau)$ is then

$$B(\tau^- \to e^-\bar{\nu}_e\nu_\tau) \equiv \frac{\Gamma(\tau^- \to e^-\bar{\nu}_e\nu_\tau)}{\Gamma(\tau^- \to \text{all})} \approx \frac{1}{2+N_c}. \tag{3.4.3}$$

For $N_c = 3$, $B = 20\,\%$ (it would be $33\,\%$ for $N_c = 1$) while the experimental number is $B = 17.84 \pm 0.05\,\%$ (the poorer agreement in this case is explained by the larger QCD radiative corrections since the mass of the τ^- is small and thus α_s large, see later).

The Drell–Yan process

In the Born approximation the rate for Drell–Yan processes (*e.g.*, $pp \to \mu^+\mu^- X$) is *inversely* proportional to N_c. Such a process proceeds via $q\bar{q}$ annihilation into a virtual (massive) photon, which subsequently decays into a $\mu^+\mu^-$ pair. Thus, for example, a quark of a given colour in one hadron must find an antiquark of the same colour in the other and hence only $1/N_c$ of the cases may actually proceed.

The rate $\Gamma(\pi^0 \to 2\gamma)$

The process $\Gamma(\pi^0 \to 2\gamma)$ is quadratic in N_c and is depicted in Fig. 3.12. The rate

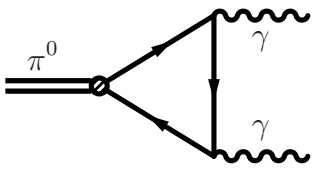


Figure 3.12: The quark description of decay $\pi^0 \to 2\gamma$. Given the natural mass scale involved (m_π), the internal fermion loop only involves the light quarks u, d and s.

can be reliably calculated via so-called soft-pion theorems and is related to the so-called *chiral anomaly*:

$$\Gamma(\pi^0 \to 2\gamma) \approx \left(\frac{N_c}{3}\right)^2 \frac{\alpha^2 m_{\pi^0}^3}{32\pi^3 f_\pi^2} = (7.73 \pm 0.04) \times \left(\frac{N_c}{3}\right)^2 \text{eV}, \tag{3.4.4}$$

where $f_\pi = (130.7 \pm 0.37)\,\text{MeV}$ is the charged-pion decay constant. The measured experimental value is $\Gamma = (7.7 \pm 0.5)\,\text{eV}$, in good agreement with $N_c = 3$.

Cancellation of the ABJ triangle anomaly

Another rather more technical problem, but related to the previous case, is that of the Adler–Bell–Jackiw (ABJ—Adler, 1969; Bell and Jackiw, 1969) so-called *triangle anomaly*. At a classical level electrodynamics has a U(1) vector symmetry, but the quantum field theory also a U(1) axial-vector symmetry—*i.e.* additional invariance with respect to transformations under γ_5.* However, the one-loop triangle graph with two vector vertices and one axial vertex breaks this symmetry—the axial current is therefore not conserved (see Fig. 3.13). The anomaly contribution

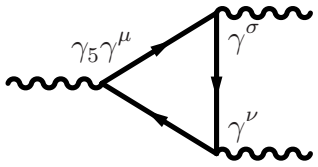


Figure 3.13: The triangle diagram contributing to the Adler–Bell–Jackiw anomaly; the loop contains all the leptons and all the quarks.

is proportional to the charge of the fermion circulating in the loop and is thus: -1 for the charged leptons, zero for neutrinos, $+\frac{2}{3}$ for up-type quarks and $-\frac{1}{3}$ for down-type quarks. Therefore, if and only if there are precisely three colours of quark, each quark–lepton family contributes with an overall coefficient

$$-1+0+3\times\left(\tfrac{2}{3}-\tfrac{1}{3}\right)\;=\;0. \tag{3.4.5}$$

3.4.2 Asymptotic freedom and confinement

That quarks (and gluons) appear to be inextricably bound inside hadrons (baryons and mesons), *i.e.* that they are not allowed to propagate in free space, is ascribed to the property (presumably of QCD) known as *confinement*. The behaviour observed in high-energy collisions, where the scattering processes between quarks, gluons and other particles occur as though the partons themselves were instead free, despite being bound, is referred to as *asymptotic freedom* (again, a property of QCD). While the latter can actually be demonstrated in perturbative approaches, the former is so far only a reasonable conjecture that may be partially understood and justified through arguments of plausibility.

The confinement problem is a serious obstacle to Feynman's picture: the impulse approximation is applicable in nuclear physics because the interaction time for the probe is much shorter than that of the nuclear motion; *i.e.* nucleons bound

* The associated axial current is analogous to that which contributes to the Gamow–Teller transitions in nuclear β-decay.

inside the nucleus do not actually *feel* the potential until they *touch* the boundaries (internally the potential is approximately constant). It is also true that the binding energy is much less than the potential-well depth; *i.e.* the binding is loose. Finally, the nucleons do actually emerge (and the energy difference due to the well depth is manifest in their spectrum); quarks, on the other hand, do not and thus would appear to have *infinite* binding energy. One might therefore presume the potential to be far from flat inside the nucleon. How then is it that they appear as though free?

The answer to this question was provided by Gross and Wilczek and, independently, by Politzer in 1973.* These three theoreticians examined the behaviour of coupling constants in general quantum field theories. In QED it had long been known that one effect of renormalisation is to *transmute* the coupling constant α into a function that varies with energy scale (one uses generically Q^2).

To understand this, let us imagine trying to measure the charge of an isolated electron, in a vacuum, by using another (infinitesimal) charge as a probe at some large distance r. Now, in quantum field theory the vacuum is *not* strictly empty; it is rather a sort of bubbling soup of virtual particle–antiparticle pairs being continually created and subsequently annihilating spontaneously. When such a pair is formed, the particle with positive charge is attracted to the electron that is the object of our measurement while the other is repelled. Consider then the Gaussian sphere at radius r: there will be a net movement of *neutralising* charge towards the electron under study. According to Gauss' theorem, this reduces the effective charge as measured by the probe, leading to a so-called "screening" effect. However, as the probe approaches the object the screening diminishes and the measured charge thus increases. A full calculation in QED reveals that there comes a point at high enough energies where the charge becomes effectively infinite while the large-distance (or $Q^2 \approx 0$) limit is well-defined and finite: its value is just the oft-quoted 1/137.

Fortunately, it turns out that the scale or Q^2 for which the charge would diverge (the so-called Landau pole) is actually far beyond the Planck mass and is thus of little physical relevance.† The *variation* of α with Q^2 is, in contrast, very physical and indeed observable. For example, at the Large Electron–Positron Collider (LEP) α has been measured for $Q^2 \approx M_Z^2$ and the value obtained is approximately 1/128, in perfect agreement with theory.

The above description is typical of Abelian gauge theories (such as QED) and

* The 2004 Nobel Prize for physics was awarded to David J. Gross, H. David Politzer and Frank Wilczek for "the discovery of asymptotic freedom in the theory of the strong interaction." The work of Gross and Wilczek (1973b) and, independently, of Politzer (1973) marks the effective birth of QCD as the theory of strong interactions.

† That is, until the role of gravity in quantum field theory is understood and/or becomes important phenomenologically, we need not (and indeed cannot) address such a problem.

indeed most quantum field theories, with the exception of non-Abelian theories (such as QCD and, by the way, the electroweak theory). In such theories the gauge fields themselves also carry the charge of the interaction and can thus interact, *even in the absence of matter fields*. It turns out that their contribution to the vacuum *colour* polarisation has the *opposite* effect to that of the fermions, *i.e.* it *antiscreens* an isolated charge.

The results of the calculations performed by Politzer, Gross and Wilczek may be summarised as follows. For convenience, as all variation is logarithmic in energy scale (Q^2), we introduce the so-called β-function as the logarithmic derivative of α with respect to the scale:

$$Q^2 \frac{\partial \alpha(Q^2)}{\partial Q^2} \;=\; \frac{\partial \alpha(t)}{\partial t} \;=\; \beta(\alpha(t))\,, \tag{3.4.6}$$

where we have defined

$$t \;=\; \ln \frac{Q^2}{\mu^2}, \tag{3.4.7}$$

with μ an arbitrary parameter—varying μ merely translates the t-axis, leaving derivatives unaffected. The β-function may be calculated perturbatively in quantum field theory and depends in an essential way on the type of theory. We thus make a power expansion in α:

$$\beta(\alpha) \;=\; -\alpha^2(b_0 + b_1\alpha + b_2\alpha^2 + \ldots). \tag{3.4.8}$$

Note that the first term turns out already $O(\alpha^2)$. The sign of the first coefficient is crucial in determining whether the coupling constant increases or decreases with growing energy scale (the overall minus sign is conventional).

In QED we find:

$$b_0^{\mathrm{QED}} \;=\; -\frac{1}{3\pi} \sum_f N_{\mathrm{c}f}\, Q_f^2, \tag{3.4.9a}$$

where $N_{\mathrm{c}f} = 3$ for quarks, 1 for leptons, and the sum runs over all fermions of charge Q_f that are active* at the chosen energy scale. In QCD, however,

$$b_0^{\mathrm{QCD}} \;=\; \frac{11N_{\mathrm{c}} - 2N_{\mathrm{f}}}{12\pi}, \tag{3.4.9b}$$

where, as usual, N_{f} is the number of active flavours of quarks. Therefore, provided

* By active we mean energetically accessible at the scale determined by Q^2.

$N_\mathrm{f} < 17$, the β-function in QCD is *negative*.* An important proven result is that, in four space–time dimensions, only non-Abelian gauge theories are asymptotically free (Gross and Wilczek, 1973a,b, 1974; Politzer, 1973, 1974).

If α is *small* enough for perturbation theory to be valid, defining $\alpha_0 \equiv \alpha(t)|_{t=0}$, the leading-order solutions to the differential equations are simply

$$\mathrm{QED}: \qquad \alpha(t) \simeq \frac{\alpha_0}{1 - |b_0|\alpha_0 t} \tag{3.4.10a}$$

and

$$\mathrm{QCD}: \qquad \alpha(t) \simeq \frac{\alpha_0}{1 + |b_0|\alpha_0 t}. \tag{3.4.10b}$$

A different and more transparent form may be adopted for QCD: by defining $\alpha_0^{-1} =: b_0 \ln(\mu^2/\Lambda_\mathrm{QCD}^2)$, for $b_0 > 0$, it may be rewritten as

$$\alpha(Q^2) \simeq \frac{1}{\frac{1}{\alpha_0} + b_0 t} = \frac{1}{b_0 \ln \frac{\mu^2}{\Lambda_\mathrm{QCD}^2} + b_0 \ln \frac{Q^2}{\mu^2}} = \frac{1}{b_0 \ln \frac{Q^2}{\Lambda_\mathrm{QCD}^2}}. \tag{3.4.11}$$

where a dimensional parameter Λ_QCD has been introduced to replace μ (and α_0). The logarithmic decrease of $\alpha(Q^2)$ with Q^2 is thus made manifest.

At this point we may consider Λ_QCD as an independent physical parameter, substituting the non-physical α_0 (the value of α for $Q^2 = \mu^2$). In quantum field theory jargon this parameter exchange is known as *dimensional transmutation*. The exact value extracted experimentally depends on the order of perturbation theory used and the energy scale (through the number of active flavours), but generally lies between 200 and 300 MeV.

In other words, so effective is the antiscreening of gluons that, unless there are more than 17 different quark types, the behaviour of the QCD coupling is the opposite of that in QED and the charge *decreases* with increasing Q^2. This leads to the notion of *asymptotic freedom*. With this concept in hand we can justify the apparent freedom of the quarks inside the proton: as long as they are probed at high enough energies, the effective interaction strength with the surrounding nucleon is small. To have some idea of this, for $Q^2 \simeq M_Z^2$ one finds (experimentally) $\alpha_s \simeq 0.11$, where we have used the standard notation of α_s to indicate the *strong* coupling constant (*i.e.* that of QCD). To be honest, at 1969 SLAC energies it was larger by about a factor 3 or 4; in other words, it was not really very small. In any case there are now many independent measurements of $\alpha(Q^2)$ and the agreement with perturbative QCD calculations is excellent (see Fig. 3.14).

How then, on the other hand, can such a picture be reconciled with the phenomenon of confinement? Here the discussion necessarily becomes less rigorous as

* In fact, b_1 and b_2 are known for QCD and have the same sign as b_0 for N_f not too large.

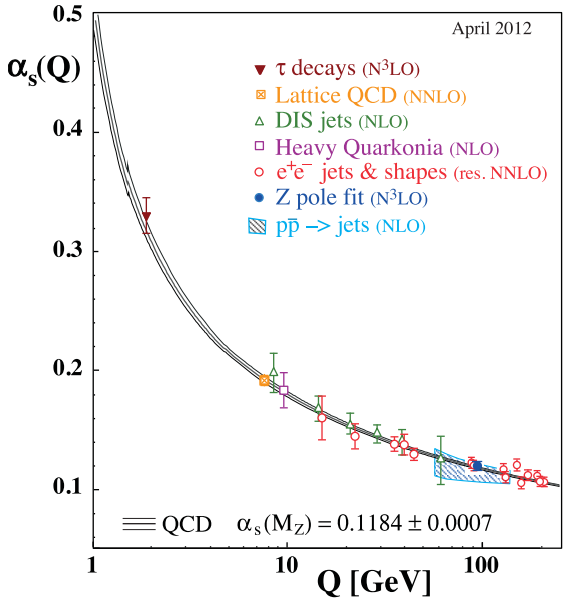


Figure 3.14: The running of $\alpha(\mu)$; the data points correspond (in increasing order of energy): τ-decay rate, Υ-decay rate, DIS scaling violation, and event shapes in e^+e^- annihilation to hadrons; figure taken from PDG-2016 (Patrignani *et al.*, 2016).

there is presently no way of performing complete and reliable *ab initio* calculations in the low-energy regime, where the coupling is strong and the non-trivial vacuum structure of QCD comes into play. Some very plausible arguments can, however, be made. Note first that the $1/r$ behaviour of the Coulomb potential is a consequence of living in *three* spatial dimensions: the flux lines are distributed over the surface of sphere and therefore the force decreases as $1/r^2$ (see Fig. 3.15a). In a spatially

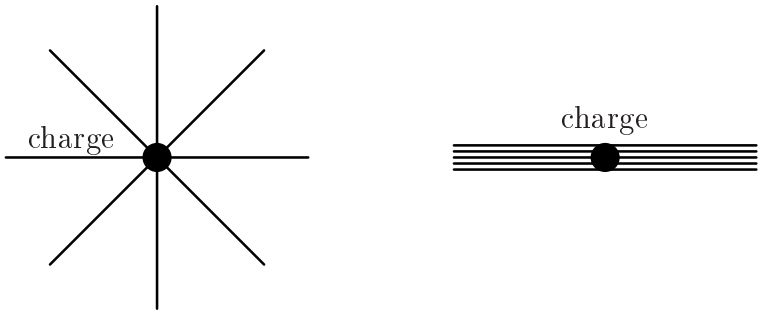


Figure 3.15: The force-field flux lines for spatially (a) three-dimensional and (b) one-dimensional field theories.

one-dimensional world the flux-line "density" is constant (as in Fig. 3.15b), so that

the force is constant and therefore the potential is proportional to r—evidently a *confining* potential. Theoretical work has been performed on such theories and much is known about them.

However, we do live in three dimensions. How then might the *effective* number of dimensions be reduced? Recall that the gluons themselves carry the colour charge and thus interact strongly with one another. One can show that a plausible effect of this is to *squeeze* the flux lines. If they are actually forced into a tube-like one-dimensional structure, then the force law indeed becomes string-like and the potential becomes linear in r (see Fig. 3.16).

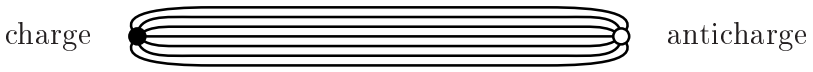


Figure 3.16: The force-field flux lines between a charge and anti-charge, assuming a one-dimensional tube-like behaviour.

One can imagine that such squeezing is only operative for large separations of two "isolated" colour charges and that at short distances the configuration returns to a three-dimensional Coulomb-type behaviour. The potential might then be described (approximately) by the form

$$V_{\mathrm{QCD}} \sim -\frac{a}{r} + b\,r. \tag{3.4.12}$$

The more usual Coulomb behaviour will only dominate in the short-distance regime while the string-like, linear, second term will naturally take over at sufficiently large distances. Indeed, a close inspection of the flux-tube picture shown in Fig. 3.16, in particular of the region very near to one of the charges, reveals that it depicts just such a situation. The actual scale at which the switch occurs depends on the coefficients a (coupling constant) and b (*string tension*).

In the light of the foregoing discussion, it is tempting to consider confinement and asymptotic freedom as being opposite sides of the same coin (the term *infrared slavery* is often used). However, this over-simplification is *misleading* and should be avoided. As we have seen, the same behaviour that leads to a diminishing coupling constant with growing energy also leads to an apparent divergence (the Landau pole) for low energies. It would be wrong, however, to associate confinement with this effect. At a formal level, at large distances the strong-coupling regime takes over and perturbative arguments no longer apply; *hic sunt leones*. More physically, as we have just discussed, confinement must be a result of the peculiar vacuum structure of QCD (which can have nothing to do with perturbation theory) and *may* arise owing to effective string-like forces between colour charges at large distances.

3.4.3 A brief survey of quark and gluon densities

Having, hopefully, convinced the reader of the validity of the picture presented by combining the ideas of Gell-Mann and Feynman, we must now demonstrate something of its utility. More to the point, as we shall see, there are various predictions of the model that, given their success, on the one hand lend strong support to the model and on the other provide useful information for both experimental analysis and planning.

The predictive power of the model, coupled with QCD, is twofold: firstly, the structure functions are universal and may be used to calculate cross-sections for processes other than DIS and, secondly, the scale variation is calculable, which means that relatively low-energy information, gathered early in history, may be exploited to make predictions for future high-energy experiments.

The intuitive picture we have derived for the DIS process is essentially that of a convolution of two basic ingredients: parton distributions or densities (which may be thought of as fluxes) and partonic cross-sections. It can be shown that for high-energy processes, where $\alpha_{\rm QCD}$ is small, many hadronic processes may described in a similar manner. From process to process, the partonic (hard or high-energy) scattering cross-sections will vary, but are *calculable*, while the (incalculable) parton densities are assumed to be the same, *i.e.* they are *universal* up to calculable scale variations. We may therefore use DIS, say, to *measure* them and then use the functions thus measured to make *predictions* for other processes.

Quark densities from electron scattering

Still today we have no reliable way to calculate the densities $f(x)$ from first principles and therefore do indeed need to measure them. Here we shall briefly review what is known about their general behaviour. In DIS at moderate energies (*e.g.* for $Q^2 \lesssim m_c^2$) only the three lightest quarks (u, d and s) contribute appreciably and so we may write

$$x^{-1}F_2^{ep}(x) \;=\; \tfrac{4}{9}\big[u_p(x)+\bar{u}_p(x)\big] + \tfrac{1}{9}\big[d_p(x)+\bar{d}_p(x)\big] + \tfrac{1}{9}\big[s_p(x)+\bar{s}_p(x)\big] \quad \text{(3.4.13a)}$$

and

$$x^{-1}F_2^{en}(x) \;=\; \tfrac{4}{9}[u_n(x)+\bar{u}_n(x)] + \tfrac{1}{9}\big[d_n(x)+\bar{d}_n(x)\big] + \tfrac{1}{9}[s_n(x)+\bar{s}_n(x)]\,, \quad \text{(3.4.13b)}$$

where the suffixes p and n on the quark densities indicate that they refer to a parent proton or neutron respectively. Now, assuming isospin to be a good symmetry, we expect $u_p = d_n$, $u_n = d_p$, $s_p = s_n$ *etc.* Exploiting this symmetry, the accepted convention is to drop the suffixes p and n and use densities that refer to the

proton. We thus write

$$x^{-1}F_2^{ep}(x) \;=\; \tfrac{4}{9}\left[u(x)+\bar{u}(x)\right]+\tfrac{1}{9}\left[d(x)+\bar{d}(x)\right]+\tfrac{1}{9}\left[s(x)+\bar{s}(x)\right] \qquad (3.4.14a)$$

and

$$x^{-1}F_2^{en}(x) \;=\; \tfrac{1}{9}\left[u(x)+\bar{u}(x)\right]+\tfrac{4}{9}\left[d(x)+\bar{d}(x)\right]+\tfrac{1}{9}\left[s(x)+\bar{s}(x)\right]. \qquad (3.4.14b)$$

Valence and sea quark separation

At this point we still have too many unknown functions to be able to determine or extract very much, but let us see what may be assumed and/or deduced. A first reasonable assumption is that, at least on average, the u and d antiquark and s quark and antiquark (or so-called *sea*-quark) densities should be suppressed with respect to those of the two *valence* quarks. In fact, we might decompose the u and d densities as follows:

$$q(x) \;=\; q_{\text{val}}(x)+q_{\text{sea}}(x) \qquad (q=u,d), \qquad (3.4.15)$$

where by *sea* we mean those quarks and antiquarks produced spontaneously while the *valence* quarks are those of Gell-Mann. We therefore have

$$\int_0^1 \mathrm{d}x\, u_{\text{val}}(x) \;=\; 2 \quad \text{and} \quad \int_0^1 \mathrm{d}x\, d_{\text{val}}(x) \;=\; 1. \qquad (3.4.16)$$

Moreover, since the sea quarks are always produced as quark–antiquark pairs, they must exist in equal numbers overall:

$$\int_0^1 \mathrm{d}x\, q_{\text{sea}}(x) \;=\; \int_0^1 \mathrm{d}x\, \bar{q}_{\text{sea}}(x) \qquad (q=u,d,s). \qquad (3.4.17)$$

One might hope that equality holds at each value of x, but there is no guarantee of this; the integrals, however, *must* be equal.

Let us subsume all the sea densities into one global function, say $\Sigma(x)$, thus

$$x^{-1}F_2^{ep}(x) \;=\; \tfrac{4}{9}u_{\text{val}}(x)+\tfrac{1}{9}d_{\text{val}}(x)+\Sigma(x) \qquad (3.4.18a)$$

and

$$x^{-1}F_2^{en}(x) \;=\; \tfrac{1}{9}u_{\text{val}}(x)+\tfrac{4}{9}d_{\text{val}}(x)+\Sigma(x). \qquad (3.4.18b)$$

Now, since the individual quantities on the right-hand side are all positive definite, then the ratio $F_2^{en}(x)/F_2^{ep}(x)$ is bounded to lie between $\frac{1}{4}$ and 4 (Nachtmann, 1972). It can only attain one or other bound if $\Sigma(x)$ is negligible. On the other hand, if the sea should dominate anywhere, then the ratio there will be approximately unity. The data (see Fig. 3.17) show that for x very small, the ratio does

indeed tend to unity while for large x it tends to the value $1/4$. This then is interpreted as implying that the valence quarks are important for large values of x while the sea grows as $x \to 0$. The actual large-x limit indicates that u dominates over d as $x \to 1$.

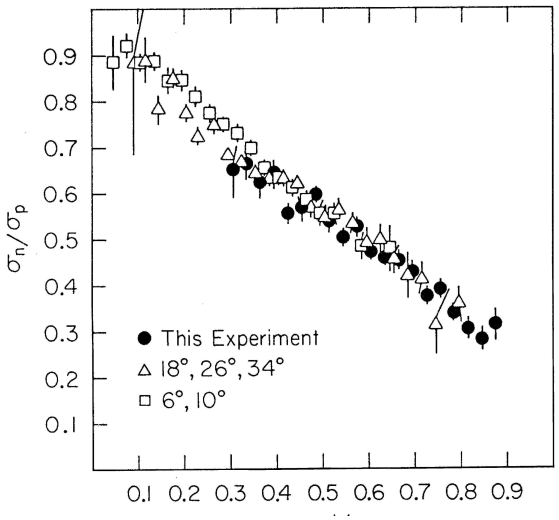


Figure 3.17: The ratio σ_n/σ_p for DIS as a function of x. The figure is taken from Bodek *et al.* (1974).

One may also consider the proton–neutron structure function difference:

$$x^{-1}\left[F_2^{ep}(x) - F_2^{en}(x)\right] \;=\; \tfrac{1}{3}\left[u_{\rm val}(x) - d_{\rm val}(x)\right]. \qquad (3.4.19)$$

The sea-quark contribution only cancels exactly in the limit of perfect SU(2) or isospin symmetry.* The result then is the difference between purely valence quarks. As a first approximation, one might make the assumption

$$u_{\rm val}(x) \;\simeq\; 2\,d_{\rm val}(x) \qquad \text{and, say,} \qquad d_{\rm val}(x) \;\simeq\; q_{\rm val}(x), \qquad (3.4.20)$$

where $q_{\rm val}(x)$ represents a generic single valence-quark density. The right-hand side of Eq. (3.4.19) is then roughly $\frac{1}{3}q_{\rm val}(x)$. With the availability of more precise and varied data, allowing the separation of $u_{\rm val}(x)$ and $d_{\rm val}(x)$, it was found that this is not a particularly good approximation, especially for large x, where the

* In fact, it is now known that, as far as the sea is concerned, SU(2) is broken quite strongly. This may be understood qualitatively by appealing to the Pauli exclusion principle: the proton contains more valence u quarks than d and thus the u–$\bar{u}$ content of the sea is suppressed, see Eq. (3.4.22). An experimentally observed consequence is a non-cancellation of the sea in the proton–neutron F_2 difference integral, Eq. (3.4.19).

following behaviour is typically found:

$$\frac{d_{\rm val}(x)}{u_{\rm val}(x)} \sim 1-x. \tag{3.4.21}$$

That is, the d quarks are relatively suppressed for $x\to 1$. A possible explanation for this may be found in the Pauli exclusion principle, which forces the more numerous u quarks to distribute themselves more evenly (*i.e.* to higher x). Unfortunately, with no real theory of the bound state in QCD, this remains merely a plausible conjecture. Some weight is, however, lent to such an argument by the observation of a surprisingly large difference between the antiquark distributions $\bar{u}(x)$ and $\bar{d}(x)$: experimentally one finds

$$\bar{u}(x) \simeq \tfrac{1}{2}\,\bar{d}(x). \tag{3.4.22}$$

Quark densities from neutrino scattering

In order to obtain more independent information we need new probes. That is, currents coupling differently to the various quarks. Deeply inelastic neutrino scattering ($\nu_e N\to e^- X$) via the weak interaction provides just such a case—the detected final state is no different to standard charged-lepton scattering ($e^- N\to e^- X$). However, since the weak and electric charges vary differently from quark to quark, such a process is sensitive to different combinations of the quark densities. Defining analogous structure functions for deeply inelastic neutrino scattering, we have

$$x^{-1}F_2^{\nu p}(x) = 2\,[d(x)+\bar{u}(x)] \tag{3.4.23a}$$

and

$$x^{-1}F_2^{\nu n}(x) = 2\left[u(x)+\bar{d}(x)\right]. \tag{3.4.23b}$$

Charge conservation requires a negatively charged quark in the neutrino–proton case (and then $d\leftrightarrow u$ for the neutron) while the strange and antistrange contributions are suppressed at low energies owing to the requirement that the final state be either a (heavy) c quark with a cross-section factor $\cos^2\theta_{\rm C}$ or a u quark with $\sin^2\theta_{\rm C}$. Similar formulæ apply to antineutrino scattering.

Consider now isoscalar targets (*i.e.* with equal numbers of protons and neutrons, such as the deuteron, ^{12}C or ^{20}Ca), which just average over the proton and neutron structure functions. In this case the electron-to-neutrino DIS ratio is

$$\frac{F_2^{ep}(x)+F_2^{en}(x)}{F_2^{\nu p}(x)+F_2^{\nu n}(x)} = \frac{\frac{5}{9}(u+\bar{u}+d+\bar{d})+\frac{2}{9}(s+\bar{s})}{2(u+\bar{u}+d+\bar{d})} \geq \frac{5}{18}. \tag{3.4.24}$$

This combination ratio is studied in Fig. 3.18. The fact that it saturates well for

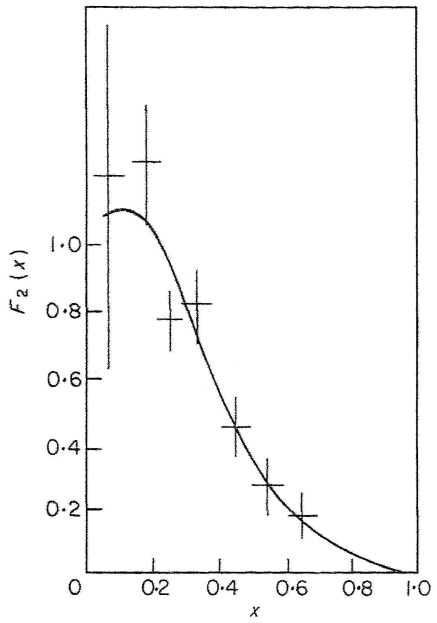


Figure 3.18: Data for the denominator in Eq. (3.4.24) compared with the curve of 18/5 times the numerator (from global experimental averages) as a function of x. The figure is taken from Perkins (1975).

$x \gtrsim 0.2$, but not below (not seen clearly in the figure owing to the large error bars), again suggests that the sea quarks are suppressed in this region, but are present for smaller values.

Sum rules for quark and gluon densities

We may integrate the quark (and gluon) densities with various weights and so relate them to known static properties of the nucleons. Evidently, the entire integrals cannot be determined purely from data since the regions $x \sim 0$ and $x \sim 1$ are experimentally inaccessible.* However, from theoretical arguments one can make reliable extrapolations to these limits.

First of all, the total charges of the proton and neutron are given by the sum over integrated quark densities weighted with their individual charges:

$$1 = \int_0^1 \mathrm{d}x \, \left[\tfrac{2}{3}(u - \bar{u}) - \tfrac{1}{3}(d - \bar{d})\right] \qquad (3.4.25a)$$

*Since Q^2 must be kept finite and not small, the limit $x \to 0$ requires an infinitely large beam energy, while for $x \to 1$ we find that the DIS cross-sections tend to zero and therefore data are limited by statistics in this region.

and

$$0 \;=\; \int_0^1 \mathrm{d}x \;\left[\tfrac{2}{3}(d-\bar{d}) - \tfrac{1}{3}(u-\bar{u})\right]. \tag{3.4.25b}$$

The s-quark contribution evidently vanishes in this sum rule since we must have zero overall strangeness or

$$0 \;=\; \int_0^1 \mathrm{d}x\,[s-\bar{s}]. \tag{3.4.26}$$

Rearranging, we obtain

$$2 \;=\; \int_0^1 \mathrm{d}x\,[u-\bar{u}] \;=\; \int_0^1 \mathrm{d}x\,u_{\mathrm{val}} \tag{3.4.27a}$$

and

$$1 \;=\; \int_0^1 \mathrm{d}x\,[d-\bar{d}] \;=\; \int_0^1 \mathrm{d}x\,d_{\mathrm{val}}, \tag{3.4.27b}$$

which is, of course, just what would be expected from a simple valence-quark model picture. Since experiments can never cover the complete x interval [0,1], an important use of these sum rules is to fix the overall normalisation of functional fits to data.

There are many such sum rules, but a last important one should be mentioned: by weighting the integral with x itself, one calculates the total fraction of the parent momentum carried by the quarks. If there were nothing other than quarks inside the nucleon then, summed over all quark types, this would give unity. In contrast, one finds experimentally

$$\int_0^1 \mathrm{d}x\,x\,[u+\bar{u}+d+\bar{d}+s+\bar{s}] \;\simeq\; 0.5, \tag{3.4.28}$$

for Q^2 in the few-GeV region. From this we may deduce that the gluon density is actually rather important, although invisible to DIS, and that

$$\int_0^1 \mathrm{d}x\,x\,g(x) \;\simeq\; 0.5, \tag{3.4.29}$$

where $g(x)$ is the probability of finding a gluon with momentum fraction x. This can be verified in other experiments; *i.e.* in hadron–hadron interactions (for example the Drell–Yan process mentioned earlier), where collisions between gluons and quarks (and even gluons and gluons) may contribute. The presence of gluons is also made manifest when considering higher-order corrections: a gluon inside the target hadron may *split* into a quark–antiquark pair, one of which may then interact with the virtual photon.

Exercise 3.4.1. *Show that the total quark momentum fraction may be obtained directly from the following combination of DIS structure function integrals:*

$$\int_0^1 \mathrm{d}x \, \left[\tfrac{9}{2}F_2^{ep+en} - \tfrac{3}{4}F_2^{\nu p+\nu n}\right]. \tag{3.4.30}$$

The shapes of parton distributions

As a first guess, assuming the quarks inside the nucleon to be non-interacting, for three valence quarks one would then expect the densities to be simple sharp spectral lines at precisely $x = \frac{1}{3}$ (see Fig. 3.19a). Treating them now as an interacting gas, the energy and momentum may be redistributed and one would expect a rather broader spectrum, still centred around $x = \frac{1}{3}$ (see Fig. 3.19b). Finally, allowing for gluon emission (or *Bremsstrahlung*) and quark–antiquark pair production one can imagine that the peaks should move down in x and the new sea and gluon distributions should be important for low energies and momenta (see Fig. 3.19c). A very schematic example of the experimentally measured quark distributions is shown in Fig. 3.20. One can clearly see all the above-mentioned features.

Other quark–parton model topics

In closing this chapter, we should at least make mention of certain topics not covered here, which are nonetheless rather important.

Meson and other parton densities: First of all, the entire exercise may now be carried out in parallel for mesons—the main difference is merely that there are then only two valence objects: the quark and antiquark—indeed, one might consider any hadron. Measurements can be performed by colliding, *e.g.* pion beams with standard hydrogen or other nuclear targets. Experimentally, only very limited information is available for the charged pions and almost nothing for other hadrons. Moreover, at sufficiently high-momentum transfers even a *real photon* may be viewed in the same manner—here, of course, there are *no* valence quarks and the entire partonic content is composed of sea quarks and gluons.

Spin correlations: There is another label that may be attached to the parton distributions: namely, spin. That is, one may ask how the spin or helicity of a quark or gluon is correlated to the spin or helicity of the parent hadron. Starting from the mid seventies (and with ever-increasing interest since the late eighties) a number of experimental groups have been active in this area. Again, one can find

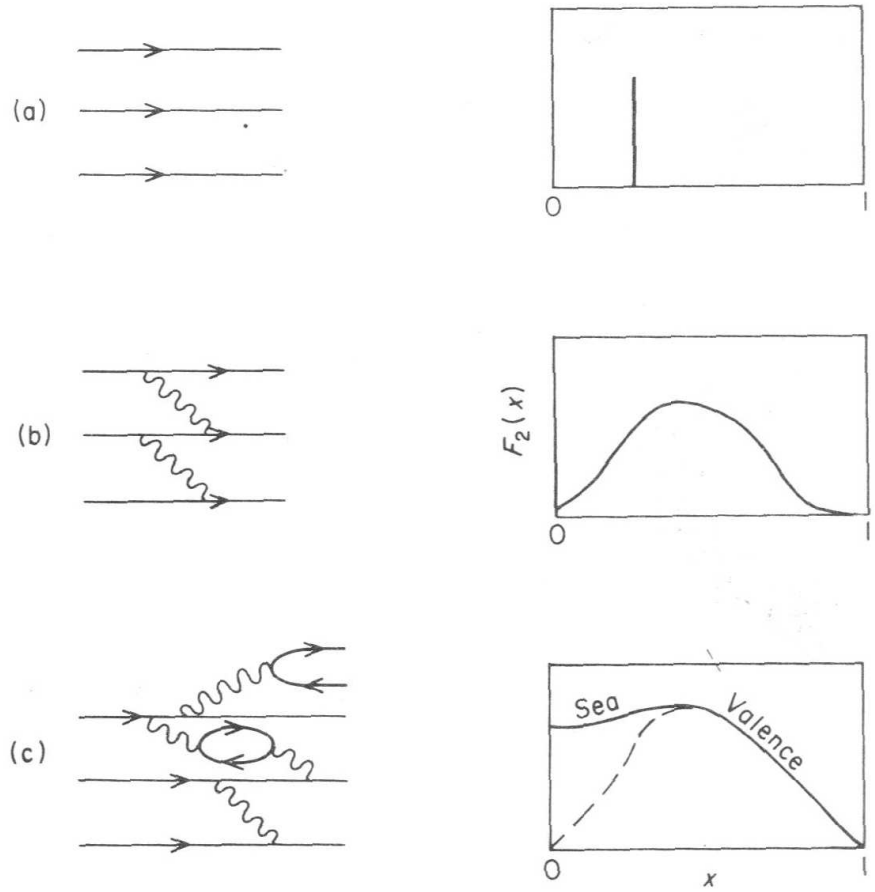


Figure 3.19: The DIS structure function $F_2(x)$ in the *naïve* and QCD-enhanced parton models: (a) The simple narrow spectral line expected for three static, non-interacting, valence quarks. (b) The spectral broadening effect of interactions between the three valence quarks. (c) The effect of spontaneous pair creation by the gluon field. The figure is taken from Close (1979).

constraining sum rules, in this case related to the axial-vector couplings of nucleon and hyperon β-decays.

Fragmentation functions: We must also mention that the, so-to-speak, *inverse* processes may be defined and studied. That is, we may ask the probability that a given quark emerging from the hard-scattering with the DIS photon materialise in the laboratory as a given hadron with a given fraction of the parent quark's energy. In such a way, it is natural to define so-called *fragmentation* functions, which can also be studied experimentally (especially in e^+e^- collisions). The situation is however rather more complex than for the case of distribution functions, from

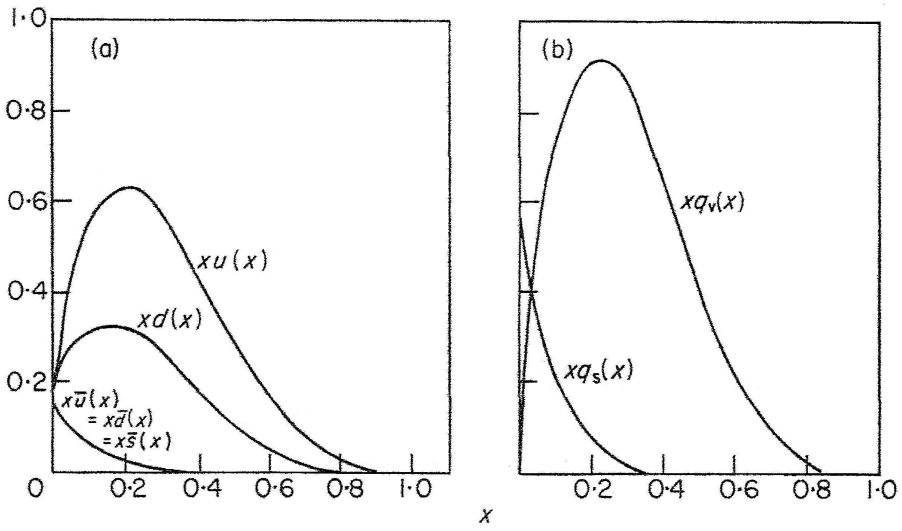


Figure 3.20: The general shape of the various quark densities. In the left panel they are shown for each quark and antiquark flavour, on the right the division is in terms of valence and sea.

both the theoretical and experimental points of view. We shall comment a little more on the process of fragmentation or *hadronisation* in the next chapter, where we shall present quark-model inspired pictures.

Fracture functions: A further development in recent years has been the study of hadrons emerging from the remnants of initial hadronic state. That is, those hadrons whose genealogies are not traceable to the struck quark, but rather to so-called *spectator* quarks inside the initial hadron. One can then define *fracture* functions. We shall, however, not delve further into this subject and simply refer the reader to the literature (see, *e.g.*, Trentadue and Veneziano, 1994).

Constituent *versus* current quarks: As a final note let us observe that while we have suggested here that there is a direct correspondence between the quarks of Gell-Mann and the partons of Feynman, the connection is not quite so simple. The objects described in Gell-Mann's theory have to do with the static properties of hadrons and certainly no account is given of the gluonic field binding them. Indeed, for such a picture to make sense, the quarks should be non-relativistic; they should, moreover, provide the mass of baryons in which they reside (and are in fact confined). One thus talks of *constituent* quarks with, for example, a constituent mass that depends on the specific environment and which for baryons must therefore be of the order of 300 MeV for the up and down quarks of the

nucleons (this is also consistent with the baryon magnetic moments). On the other hand, the infinite-momentum frame used by Feynman and also the kinematics of, say, DIS imply very light quarks with masses (deduced from other considerations) of just a few MeV. These are known as *current* quarks and are, for example, what should appear in the QCD Lagrangian and which, in the perturbative sense, know nothing of confinement.

3.5 Bibliography

Adler, S.L. (1969), *Phys. Rev.* **177**, 2426.

Álvarez, L.W. (1973), *Phys. Rev.* **D8**, 702.

Barnes, V.E. *et al.* (1964), *Phys. Rev. Lett.* **12**, 204.

Bartel, W. *et al.* (1968), *Phys. Lett.* **B28**, 148.

Bell, J.S. and Jackiw, R. (1969), *Nuovo Cim.* **A60**, 47.

Bjorken, J.D. (1969), *Phys. Rev.* **179**, 1547.

Bloom, E.D. *et al.* (1969), *Phys. Rev. Lett.* **23**, 930.

Bodek, A. *et al.* (1974), *Phys. Lett.* **B51**, 417.

Breidenbach, M. *et al.* (1969), *Phys. Rev. Lett.* **23**, 935.

Callan, C.G., Jr. and Gross, D.J. (1969), *Phys. Rev. Lett.* **22**, 156.

Close, F.E. (1979), *An Introduction to Quarks and Partons* (Academic Press).

Coleman, S.R. and Glashow, S.L. (1961), *Phys. Rev. Lett.* **6**, 423.

Coleman, S.R. and Glashow, S.L. (1964), *Phys. Rev.* **134**, B671.

Collins, P.D.B. and Martin, A.D. (1984), *Hadron Interactions* (Adam Hilger).

Eden, R.J., Landshoff, P.V., Olive, D.I. and Polkinghorne, J.C. (1966), *The Analytic S-Matrix* (Cambridge U. Press).

Feynman, R.P. (1969), *Phys. Rev. Lett.* **23**, 1415.

Feynman, R.P. (1972), *Photon–Hadron Interactions* (W.A. Benjamin).

Friedman, J.I. and Kendall, H.W. (1972), *Ann. Rev. Nucl. Part. Sci.* **22**, 203.

Geiger, H. and Marsden, E. (1909), *Proc. Royal Soc. (London)* **A82**, 495.

Gell-Mann, M. (1953), *Phys. Rev.* **92**, 833.

Gell-Mann, M. (1962), *Phys. Rev.* **125**, 1067.

Gell-Mann, M. and Ne'eman, Yu., eds. (1964), *The Eightfold Way* (W.A. Benjamin).

Greenberg, O.W. (1964), *Phys. Rev. Lett.* **13**, 598.

Gross, D.J. and Wilczek, F. (1973a), *Phys. Rev.* **D8**, 3633.

Gross, D.J. and Wilczek, F. (1973b), *Phys. Rev. Lett.* **30**, 1343.

Gross, D.J. and Wilczek, F. (1974), *Phys. Rev.* **D9**, 980.

Han, M.-Y. and Nambu, Y. (1965), *Phys. Rev.* **139**, B1006.

Muta, T. (1998), *Foundations of Quantum Chromodynamics: An Introduction to Perturbative Methods in Gauge Theories* (World Sci.), 2nd. edition.

Nachtmann, O. (1972), *Nucl. Phys.* **B38**, 397.

Nakano, T. and Nishijima, K. (1953), *Prog. Theor. Phys.* **10**, 581.

Ohnuki, Y. (1960), in proc. of the *Int. Conf. on High-Energy Physics* (Geneva, 1960), p. 843.

Okubo, S. (1962), *Prog. Theor. Phys.* **27**, 949.

Patrignani, C. *et al.*, Particle Data Group (2016), *Chin. Phys.* **C40** [10], 100001.

Perkins, D.H. (1975), *Contemp. Phys.* **16**, 173.

Politzer, H.D. (1973), *Phys. Rev. Lett.* **30**, 1346.

Politzer, H.D. (1974), *Phys. Rep.* **14**, 129.

Rosenbluth, M.N. (1950), *Phys. Rev.* **79**, 615.

Rutherford, E. (1911), *Phil. Mag.* **21**, 669.

Sakata, S. (1956), *Prog. Theor. Phys.* **16**, 686.

Trentadue, L. and Veneziano, G. (1994), *Phys. Lett.* **B323**, 201.

Yang, C.-N. and Mills, R.L. (1954), *Phys. Rev.* **96**, 191.

Chapter 4

The New Particles

In this chapter we shall trace the history of particle discovery, from the early experiments exploiting the first available source of energetic particles, namely so-called cosmic rays, right up to the most energetic particle colliders available today, in all their various forms: e^+e^-, ep, pp, $p\bar{p}$ *etc.* Now, although ep colliders do exist even with very high-energy beams, they are not normally considered as suited to particle discovery and here, after discussing the role of cosmic rays, we shall concentrate on e^+e^- and hadron–hadron machines.

Before discussing the various types of experiments in detail, let us briefly examine the basic requirements for the discovery of new particles. Firstly, of the various reasons that at some given point in time a particle has not previously been discovered, the most common is that its mass is larger than the available energies. Obviously, there are also other possible explanations: it interacts too weakly to be detected (*e.g.* the neutrino), there are conserved quantum numbers that suppress production processes (*e.g.* the strange particles) *etc.* And it is usually more a combination of such effects. However, the question of mass is fundamental: if a particle is too heavy it simply cannot be produced.

To make further discoveries then, one needs more control over the interacting system. In particular, we need to control (and raise) the energy. We also often need a *clean* initial system so that the details of the final state may emerge clearly and provide unambiguous indications of any new object produced. These requirements immediately suggest e^+e^- colliders as the prime candidate. The initial particles are point-like, well understood, can be produced with very precisely known energies and have no by-products that might pollute the final state. Having said that, hadron–hadron machines can reach much higher energies and have more channels open; in certain circumstances they must therefore be preferred.

4.1 Cosmic rays and the early discoveries

It might be said that the birth of particle physics lies in cosmic rays. Antimatter was first detected in cosmic rays, as were many of the first strange hadrons, not to mention the muon. Until the fifties cosmic rays were the only source of high-energy particles and still today they remain the source of the highest-energy particles that can be studied: single particles up to $O(10^9\,\mathrm{GeV})$ have been observed.

4.1.1 The positron

In 1930 Dirac proposed the particle–hole interpretation of the solutions to the relativistic wave equation he had himself derived earlier (Dirac, 1928). He had found that the equation not only correctly described the electron (including its gyromagnetic ratio of two), but also contained an object that was an exact copy of the electron with, however, the opposite charge: the antielectron or positron as it became known.

Dirac's hole theory suggested that a photon of energy slightly more than 1 MeV could, in principle, produce an electron–positron pair. However, kinematics does not permit direct production or conversion and more energy is required. Such energy is found naturally in cosmic rays. What are commonly called cosmic rays are, of course, only the by-products of extremely high-energy collisions between, typically, protons of cosmic origin and nuclei in the Earth's atmosphere. Such collisions, although totally uncontrollable, give rise to all energetically accessible states. The problem then lies in detecting the particles produced.

In 1932, Anderson* (see Anderson, 1933b) and, independently a few months later, Blackett and Occhialini (1933)† detected the passage of positively charged particles, similar in mass to the electron, using Wilson cloud chambers. While Anderson did not immediately connect his discovery to the prediction by Dirac, Blackett and Occhialini clearly recognised these particles as Dirac's positrons.

The idea behind the cloud chamber is that a charged particle passing through supersaturated water vapour provokes local condensation. The cloud chamber consists of a container, fitted with a piston, into which a saturated air–vapour mixture is injected. When the piston is moved suddenly to lower the pressure, the temperature also drops rapidly and the vapour passes into a supersaturated phase. Any charged particle traversing the chamber in this moment leaves a track of fine condensation droplets, which may be photographed (possibly from two different

* The 1936 Nobel Prize for physics was awarded equally to Victor Franz Hess for "his discovery of cosmic radiation" and to Carl David Anderson for "his discovery of the positron."

† The 1948 Nobel Prize for physics was awarded to Patrick Maynard Stuart Blackett for "his development of the Wilson cloud chamber method, and his discoveries therewith in the fields of nuclear physics and cosmic radiation."

angles so as to permit a stereo image). The presence of a magnetic field reveals the sign of the charge.

In Fig. 4.1 we see one of Anderson's positron events. The track enters from the

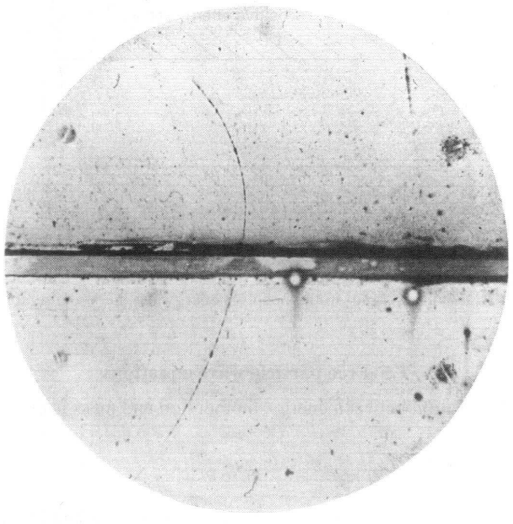


Figure 4.1: A Wilson cloud-chamber photograph showing the passage of a positron. The radius of curvature determines the momentum while the direction indicates the charge sign. The track length for such a relatively low-momentum particle allows one to distinguish between a positron and a proton. The figure is taken from Anderson (1933a).

bottom with very high energy (deduced from the large radius of curvature in the 1.5-T magnetic field used). It then passes through a 6-mm lead strip, which has the purpose of slowing the particles, and continues for nearly 3 cm before presumably annihilating with an atomic electron. The curvature of the upper track indicates a momentum of approximately 23 MeV. Were it a proton, this would correspond to a very low velocity and it is known that the range would then be only a few millimetres.*

4.1.2 The muon

The muon has a mass of 106 MeV and thus requires much higher energies to be produced in the laboratory. It was discovered in cosmic-ray experiments independently by Anderson and Neddermeyer (1936) and Street and Stevenson (1937).

Electrons and muons are produced copiously in high-energy collisions and while the electron is stable, the muon is not, decaying with a relatively long lifetime:

* Recall that low-energy cross-sections are typically *inversely* proportional to the velocity.

$\tau_\mu = 2.2 \times 10^{-6}$ s. Ignoring time-dilation effects, a highly relativistic muon therefore travel an average distance of nearly 0.7 km before decaying. On the other hand, the tau lepton is rather heavier and therefore less common, added to which its lifetime is $\tau_\tau = 2.9 \times 10^{-13}$ s. Thus, a τ produced by cosmic-ray interactions never reaches a laboratory on the ground. Moreover, its decay products are either a number of light hadrons or, in the leptonic mode, contain two neutrinos. It was not until 1975 and the availability of high-energy electron–positron colliders that the third of the charged leptons was discovered.

4.1.3 The pion

In the same year that Conversi, Pancini and Piccioni discovered that the muon was not Yukawa's mesotron (see App. B.1 for further details), the pion was discovered. This discovery was made, however, with photographic-emulsion detection methods at high altitude (typically used on mountain tops or even in aeroplanes). Recall that the neutral-pion lifetime is $(8.5 \pm 0.2) \times 10^{-17}$ s while that of the charged pions is 2.6×10^{-8} s. Moreover, the neutral-pion decays predominantly via the electromagnetic two-photon channel while the charged states undergo a weak β-type decay principally to $\mu\nu_\mu$.

4.1.4 The strange particles

There are, however, other particles having masses not too dissimilar to the muon and with very similar (weak) decay rates. They are the *kaons*: $K^\pm$, K^0 and $\overline{K}^0$. These then were ideal candidates for discovery in cosmic-ray experiments. Indeed, the strange baryons, Λ^0, $\Sigma^{0,\pm}$ and $\Xi^{0,-}$, having masses a little larger than the nucleons and lifetimes of the order of 10^{-10} s, were also soon discovered. Moreover, in the same year as the pion discovery the first "V" particles were detected too (Rochester and Butler, 1947), so-called owing to their distinctive two-pronged decay-state tracks, easily identifiable in cloud-chamber, emulsion and, later on, bubble-chamber experiments (see Figs. 4.2 and 4.3). Figure 4.3 shows both a photograph and a sketched version of an interesting event: a negatively charged pion enters from the left and strikes a proton, producing two uncharged particles (a neutral kaon and a lambda baryon):

$$\pi^- + p \rightarrow K^0 + \Lambda^0. \tag{4.1.1}$$

Note that this process does not violate any known conservation law and is therefore most likely a strong interaction. The kaon and lambda travel some distance (but,

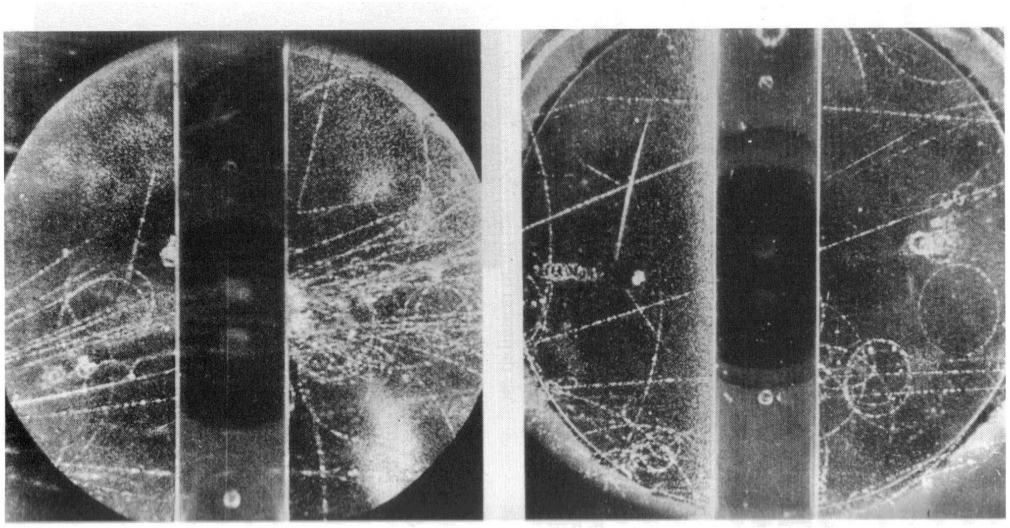


Figure 4.2: In the left-hand cloud-chamber photograph slightly to the right of the lead plate and above centre we see a typical "V" fork, probably due to the decay of a neutral (hence no incoming track) kaon into a $\pi^+\pi^-$ pair; the right-hand photograph shows a charged "V" (very open) top left, probably the signal of a charged kaon decaying into a muon plus neutrino. The figure is taken from Rochester and Butler (1947).

being neutral, leave no tracks) before decaying via the standard weak processes:

$$K^0 \rightarrow \pi^+ + \pi^- \quad \text{and} \quad \Lambda^0 \rightarrow p + \pi^-. \tag{4.1.2}$$

4.2 Electron–positron colliders

A particular limitation of cloud chambers, emulsions and bubble chambers is that the new particle produced must leave a detectable track. While this obviously severely limits their use in the case of neutrals, the most important consequence here is that they are also inadequate for very short-lived (intermediate) states such as the π^0 in the example above. One evidently needs an indirect method of detecting the presence of such an object. This is where the concept of a resonance *à la* Breit and Wigner (1936) arises.* Once a particle lifetime becomes too short to be measured directly (*e.g.* via the mean length of its cloud-chamber, emulsion or bubble-chamber tracks) one must move over to the energy counterpart, the decay

*For a detailed discussion, the reader is referred to App. A.5

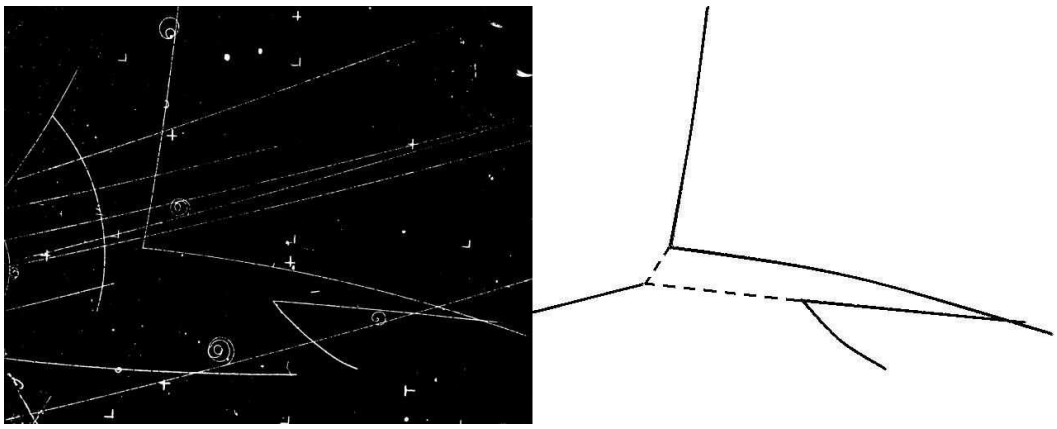


Figure 4.3: A negatively charged pion enters from the left and strikes a proton, producing two uncharged particles (a kaon and a lambda baryon) that leave no tracks until they too decay. The figure is taken from LBL NEWS Magazine, Vol. 6, 1981.

rate or *width*, related in the following way:

$$\Gamma = \hbar/\tau. \tag{4.2.1}$$

Thus, for shorter lifetimes we exploit the growing uncertainty in the mass of the state produced, which being (very) unstable is always an intermediate state in any given process. As the term "width" suggests, the uncertainty principle implies that given only a short time to determine the mass of a particle there will be some natural fluctuation around a central value compatible with the above expression; spectral lines are thus broadened and their width provides an indirect measure of the lifetime.

4.2.1 Resonance production

A first simple practical example in particle physics* is that of Δ production in pion–nucleon elastic scattering via, for example, the process

$$\pi^+ + p \rightarrow \Delta^{++} \rightarrow \pi^+ + p. \tag{4.2.2}$$

The nominal masses and widths of the Δ resonances are 1232 MeV and 120 MeV respectively. In this case the initial (and final) spins are 0 and $\frac{1}{2}$ while the intermediate Δ has spin $\frac{3}{2}$.

* The earliest applications of the partial-wave and Breit–Wigner approach are found in nuclear physics, where it is used to describe interactions that proceed via the formation of intermediate excited nuclear states or resonances.

Exercise 4.2.1. *Show that, for a charged-pion projectile incident on a fixed proton target, the resonance peak is attained for a pion kinetic energy just below* 200 MeV. *Calculate the centre–of–mass pion energy.*

With such energies, although the pion itself is quite relativistic in the laboratory frame, this is no longer true in the centre–of–mass frame and we are justified in using the non-relativistic Breit–Wigner (BW) form given in Eq. (A.5.15). Therefore, the maximum value of the cross-section is

$$\sigma_R^{\max} = \frac{4\pi\hbar^2}{p^2}\,\frac{4}{1\times 2} = \frac{8\pi\hbar^2}{p^2}. \tag{4.2.3}$$

Note that, owing to the rapidly falling underlying cross-section, the experimental peak lies a little below the true resonance energy, corresponding to the point where the experimental value coincides with the maximum-value curve (see Figs. 4.4). Note also the skew effect due to the variation of Γ over the width of the Δ reson-

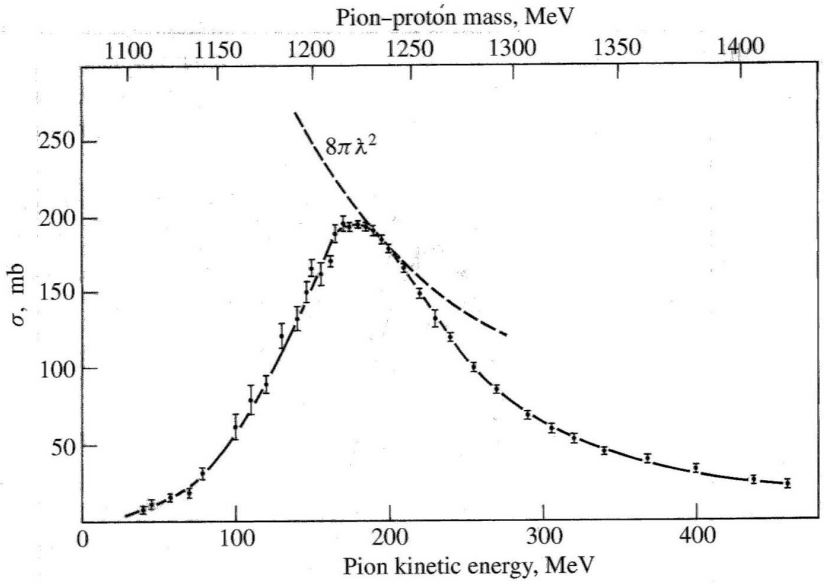


Figure 4.4: The $\pi^+ p$ elastic cross-section as a function of laboratory-frame pion kinetic energy. Note the skew effect due to the variation of Γ over the width of the Δ resonance.

ance. This is principally the effect of the variation of Γ in the numerator, which grows with increasing centre–of–mass energy. The high-energy tail is thus rather higher than the low-energy tail. Once out in the tails, the value of Γ used in the denominator usually has little influence.

A further way in which to search for resonances is to examine the invariant mass of, say, a $\pi^-\pi^+$ pair. A possible production mechanism is shown in Fig. 4.5.

Again, the form of the effective propagator for an intermediate resonance, having

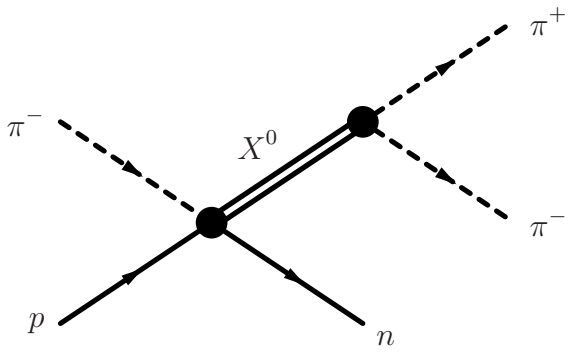


Figure 4.5: The production of an intermediate resonant state X^0 in the $\pi^+\pi^-$ channel of the process $\pi^- p \to \pi^- \pi^+ n$.

the same quantum numbers as two pions, leads to the classic BW shape in the invariant-mass distribution (see Fig. 4.6). In the figure one sees three clear peaks

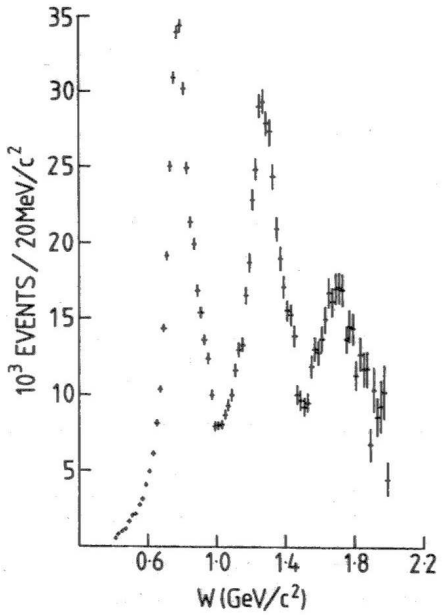


Figure 4.6: The $\pi^- p \to \pi^- \pi^+ n$ cross-section as a function of the $\pi^-\pi^+$ pair invariant mass W, for a pion beam momentum 17 GeV/c.

with widths of the order of $100 - 200$ MeV. These correspond to: $\rho^0(769)$ having $\Gamma \simeq 150$ MeV; $f_2^0(1275)$, $\Gamma \simeq 180$ MeV and $\rho^0(1700)$, $\Gamma \simeq 200$ MeV. Naturally, these peaks are superimposed on a background of non-resonant, continuum production of charged-pion pairs

In general, by considering a multiparticle state, we see that the invariant-mass distribution of the final composite system measures both the mass and width of any intermediate resonant state having the same quantum numbers as the detected system. To search for strange mesonic resonances, one might thus study the invariant mass of, say, a particular $K\pi$ final state.

One word of caution is in order. As we know from quantum mechanics, to calculate any particular process (scattering or decay), we must sum *all* amplitudes that can contribute. If the peaks are sufficiently narrow and/or well separated (as is the case above), then to a good approximation we may consider the contributions independently. However, it can happen that two or more resonances with the same quantum numbers have very similar masses and thus the BW peaks can overlap. In this case the phase variation and, in particular, the possibility of phase mismatch in the initial production process may mean that across the interval in which the two (or more) resonances contribute there are successive regions in which the two (or more) channels interfere destructively and constructively. These can distort the peaks in an essential way, leaving their appearance very misleading. We shall say more on this when we examine the very specific case of three-body decays.

4.2.2 Hadronic resonances in e^+e^- annihilation

As remarked earlier, one of the simplest imaginable situations is the search for and discovery of the neutral resonances that may be produced in e^+e^- annihilation. We have already seen the basic type of (resonance) diagram via which a fermion–antifermion pair may be produced (see Fig. 4.7). The process is very similar to that

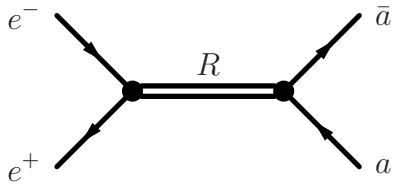


Figure 4.7: The e^+e^- annihilation process into a particle–antiparticle pair ($a\bar{a}$) via a direct intermediate resonant state R.

of e^+e^- annihilation through a virtual photon. Note that the quantum numbers and energy of the initial state are well determined.

The picture provided here is not yet complete. The initial-state e^+e^- pair cannot couple directly to a hadronic resonance and therefore the annihilation process proceeds via the formation of a virtual photon, which can then transform into any neutral resonance having the same J^{PC} quantum numbers. In such a way, whenever the centre–of–mass energy corresponds to the mass of a neutral, spin-one, flavourless hadronic resonance, we have the picture given in Fig. 4.8.

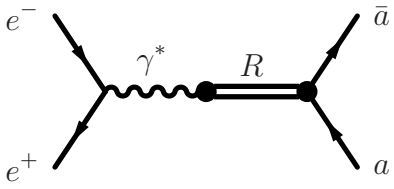


Figure 4.8: The e^+e^- annihilation process into a particle–antiparticle pair ($a\bar{a}$) via a photon coupling to an intermediate resonant state R.

The total angular momentum, P and C properties of the resonance so formed are thus predetermined. Indeed, the photon is characterised as having $J^{PC}=1^{--}$ and isospin $I=0,1$—the isospin is not completely determined as the electromagnetic interaction violates isospin (the up-quark and down-quark have different charges) and therefore the photon is not an eigenstate of isospin.

Now, it turns out to be more useful to examine the cross-section into hadrons rather than the elastic cross-section and we shall need to modify formula (A.5.15) slightly to take into account that the initial and final states are different (at higher energies many channels do indeed exist). The most general form of the BW approximation for such an *inelastic* process is then

$$\sigma_R \simeq \frac{4\pi\hbar^2}{p^2}\frac{(2J_R+1)}{(2s_1+1)(2s_2+1)}\frac{(\Gamma_i/2)(\Gamma_f/2)}{(E-E_0)^2+(\Gamma_{\rm tot}/2)^2}, \tag{4.2.4}$$

where the width $\Gamma_{\rm tot}$ appearing in the numerator is the *total* width of the resonance, while in the numerator we have the two *partial* widths Γ_i and Γ_f, corresponding to the decay channels into the actual initial and final states.

To understand how this comes about, let us examine the appropriate Feynman diagram for such a process (see Fig. 4.8). The approximation made is equivalent to considering this single amplitude as the product of two factorising sub-amplitudes: namely, $e^+e^-\to R$ and $R\to a\bar{a}$. The process is thus seen in two distinct and independent stages: initial production followed by decay. The first corresponds to the *Hermitian conjugate* of the process $R\to e^+e^-$ while the second evidently describes the decay of the resonance R into the state $a\bar{a}$. On taking the square of the modulus, we are thus equivalently calculating the product of the processes $e^+e^-\to R$ and $R\to a\bar{a}$. The spin factors are kept explicitly separate and so are still correct.

From this description we can also appreciate a little better the nature of the approximation: the resonance R is effectively considered as a real, on-shell particle, which is evidently not true, except precisely at the peak. To some extent, the inclusion of momentum factors, as in Eq. (A.5.26), can corrects for this. However, we are also implicitly treating the couplings at the vertices as point-like whereas an exact formulation would require form factors. Since we do not have a complete

theory of the bound state in the strong interaction, we can only make models and parametrise these.

Inserting the spins of the initial-state electron–positron pair and the intermediate photon, we have

$$\sigma_R \simeq \frac{3\pi\hbar^2}{4p^2} \frac{\Gamma_i \Gamma_f}{(E-E_0)^2 + (\Gamma_{\text{tot}}/2)^2}. \tag{4.2.5}$$

We use the photon spin since we are considering a process initiated by electrons, which do not interact strongly and which cannot therefore directly produce strongly interacting particles. Thus, the first vertex in the diagram must be of the $e^+e^-\gamma$ type. However, the photon may convert into any hadronic resonance having the same quantum numbers, in particular having the same spin.

An example case is the production of the $\rho^0(770)$ meson (with $I, J^{PC} = 1, 1^{--}$). As already noted, it has the (strong) decay channel $\pi^+\pi^-$ and will thus be seen as a BW resonance according to Fig. 4.8 with $R = \rho^0$ and $a\bar{a} = \pi^+\pi^-$.

However, in the case of e^+e^- annihilation it is easier to simply require that the final state contain only hadronic states (all those allowed). We thus now examine the *spectrum* obtained in e^+e^- annihilation when the final states are restricted to those containing only hadrons. See Fig. 4.9, in which the total *hadronic* cross-section is plotted as a function of $\sqrt{s} = E_{CoM}$. A large number of peaks are evident, from very low energies right up to the maximum available.

The first peak corresponds to resonant production of the $\rho^0(770)$ meson together with the $\omega^0(782)$, the former having a width $\Gamma \simeq 146\,\text{MeV}$ and $I, J^{PC} = 1, 1^{--}$ while the latter is much narrower with $\Gamma \simeq 8.5\,\text{MeV}$ and $I, J^{PC} = 0, 1^{--}$; these two certainly overlap owing to the broadness of the former. The reason for the difference is that while $\rho^0(770)$ is part of an isospin triplet, the $\omega^0(782)$ is a singlet; in quark terms it is almost purely $(u\bar{u} + d\bar{d})/\sqrt{2}$ and its principal decay mode is into three pions.

The successive peak is the $\phi^0(1020)$, also $I, J^{PC} = 0, 1^{--}$ and width $\Gamma \simeq 4.3\,\text{MeV}$. This is the SU(3) partner to the $\omega^0(782)$ and is almost purely $s\bar{s}$. The fact that these two states are arranged in this manner, *i.e. neither* being the natural eigenstate superpositions of $u\bar{u}$, $d\bar{d}$ and $s\bar{s}$ states,

$$\phi_0 = \tfrac{1}{\sqrt{3}}(u\bar{u} + d\bar{d} + s\bar{s}) \tag{4.2.6a}$$

and

$$\phi_8 = \tfrac{1}{\sqrt{6}}(u\bar{u} + d\bar{d} - 2s\bar{s}), \tag{4.2.6b}$$

is known as *ideal mixing*. Quite why the mixing should be so nearly ideal is not clear and must have to do with the (poorly understood) bound-state dynamics. The determination of these combinations is obtained through comparison with the

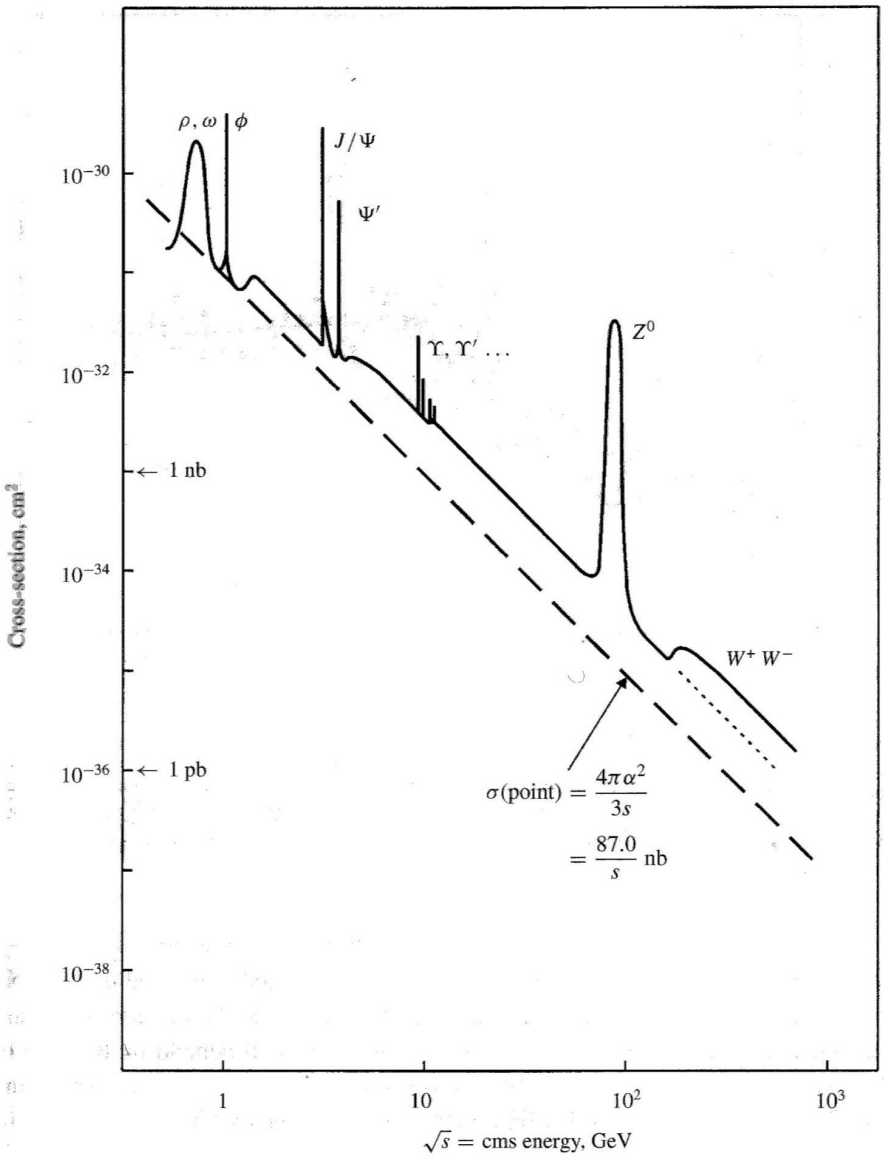


Figure 4.9: The e^+e^- cross-section into hadrons as a function of CM energy. The dashed line indicates the baseline $\mu^+\mu^-$ final-state cross-section.

mass formulæ in SU(3). In any case, these three states belong to the nonet of pseudoscalar mesons.

The OZI or Zweig rule

It is interesting to examine and compare the principal decay modes of the two SU(2)-singlet states $\omega^0(782)$ and $\phi^0(1020)$, displayed in Table 4.1. Naïvely, the

Table 4.1: The principal decay modes of the two lowest-lying, spin-zero, SU(2)-singlet states $\omega^0(782)$ and $\phi^0(1020)$.

$$\phi^0(1020) \to \begin{cases} K^+K^- & 49\% \\ K^0\overline{K}^0 & 34\% \\ \pi^+\pi^-\pi^0 & 15\% \end{cases} \qquad \omega^0(782) \to \begin{cases} \pi^+\pi^-\pi^0 & 89\% \\ \pi^0\gamma & 9\% \\ \pi^+\pi^- & 2\% \end{cases}$$

phase-space factor should favour the three-pion decay mode of the ϕ since the Q-value is 600 MeV, as compared to just 24 MeV for the two-kaon decay. Nevertheless, the three-pion mode is relatively suppressed by a factor of nearly 6. This and other similar observations led to the formulation of the so-called OZI or Zweig rule (Okubo, 1963; Zweig, 1964; Iizuka *et al.*, 1966).

Let us examine the quark diagrams for these decays (see Fig. 4.10). We see

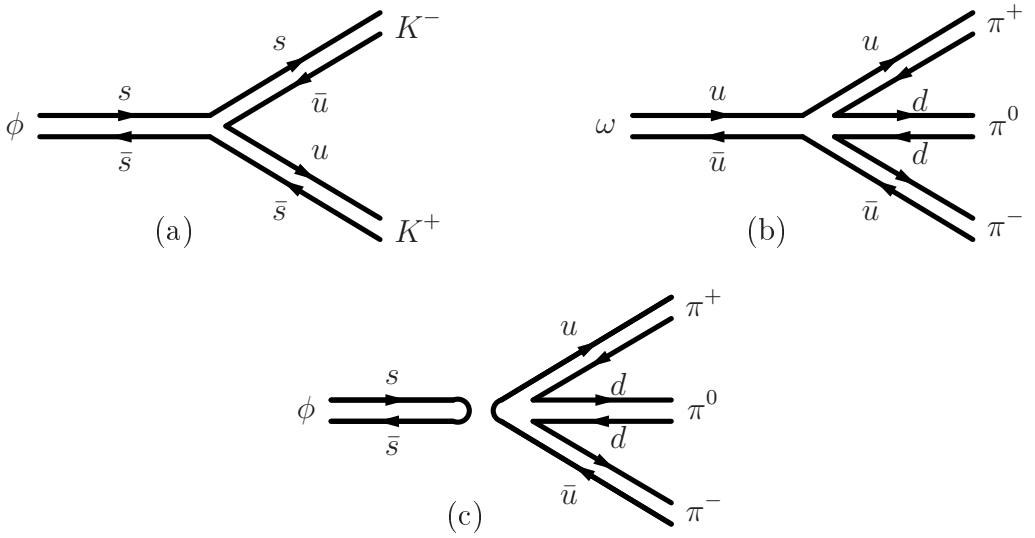


Figure 4.10: Comparison of the diagrams for OZI suppressed and unsuppressed decays of quark–antiquark states: (a) $\phi \to K^+K^-$, (b) $\omega \to \pi^+\pi^-\pi^0$, (c) $\phi \to \pi^+\pi^-\pi^0$.

that while at least some of the quark lines are connected between the initial and final state for the three-pion decay of the ω and for the two-kaon decay of the ϕ, the case of $\phi \to 3\pi$ is distinctive in that the initial and final states are completely disconnected with respect to quark lines. The OZI rule then simply states that such decays are suppressed.

We can understand this suppression in terms of gluon exchange. To connect the initial annihilating quark–antiquark pair to the final state, one evidently needs a number of gluons. Unlike the photon, gluons are carriers of the relevant (colour) charge; a single gluon is therefore not allowed since the intermediate state would then be (colour) charged while the initial and final states are (colour) neutral. A two-gluon state would permit a colour-singlet exchange but would not have the correct J^{PC} quantum numbers; recall that, just as the photon, the gluon has $J^{PC}=1^{--}$. The minimum number of gluons is therefore three. Since each gluon is associated with an extra factor α_s at the amplitude level, the price is high and such a process is evidently suppressed with respect to decays proceeding via connected diagrams. Compare this with the relative decays rates of para- and ortho-positronium, which proceed via two- and three-photon final-state channels respectively. There are many examples of decays in which the OZI rule is at work and it is now considered both well confirmed and well understood.

Discovery of the *c* quark

The next set of visible peaks corresponds to production of the so-called J/ψ and one of its excited states, ψ'. The observation of this resonance marks the discovery of the c or *charm* quark* (at Brookhaven by Aubert *et al.*, 1974 and at SLAC by Augustin *et al.*, 1974).† The principal object, the J/ψ, has the following mass and width:

$$m_{J/\psi} \;=\; 3097 \text{ MeV} \quad \text{and} \quad \Gamma_{J/\psi} \;\simeq\; 93 \text{ keV}, \tag{4.2.7}$$

with spin–parity quantum numbers $J^{PC}=1^{--}$. The second peak is due to the ψ' or $\psi(2S)$ with

$$m_{\psi'} \;=\; 3686 \text{ MeV} \quad \text{and} \quad \Gamma_{\psi'} \;\simeq\; 286 \text{ keV} \tag{4.2.8}$$

and it too obviously has $J^{PC}=1^{--}$. These two states are then evidently $c\bar{c}$ in an s-wave and with spins aligned, the ψ' being a radial excitation. Various other radial-excitation states are now known.

We should remark that the experiment of Aubert *et al.* (1974), performed at the Brookhaven National Lab., did not use colliding electron and positron beams. Instead, the AGS provided 28-GeV protons, which were made to collide with a beryllium target. The experiment consisted in the measurement of the invariant-mass spectrum of e^+e^- and $\mu^+\mu^-$ pairs produced in these collisions. The reaction

* The 1976 Nobel Prize for physics was awarded equally to Burton Richter and Samuel C.C. Ting for "their pioneering work in the discovery of a heavy elementary particle of a new kind."

† Almost immediately after the announcement of the discovery similar evidence was found at the ADONE machine in Frascati (Bacci *et al.*, 1974).

studied was then

$$\begin{aligned} p + \mathrm{Be} \;&\rightarrow\; J/\psi + \mathrm{anything} \\ &\hookrightarrow\; e^+e^-,\;\; \mu^+\mu^-, \end{aligned} \tag{4.2.9}$$

so that standard BW resonance peaks were clearly visible here too.

It is interesting to examine a little more closely the data of the experiment performed at SLAC. The process here is classic e^+e^- annihilation in the Stanford Positron–Electron Accelerating Ring (SPEAR) collider. The possible channels are

$$e^+e^- \;\rightarrow\; J/\psi \;\rightarrow\; \mathrm{hadrons},\, e^+e^- \;\mathrm{and}\; \mu^+\mu^-. \tag{4.2.10}$$

The three related cross-sections are shown in Fig. 4.11. The particularly skewed shapes of some of the curves is typical of interference between competing channels. Here we have both the direct channel (as shown in Fig. 4.7) and the channel with an intermediate photon (Fig. 4.8). It can be shown that the interference pattern corresponds well to interference between these two channels.

This is also an example of how too selective a production mechanism can actually hide states. There exists a so-called η_c (for its similarity to the much lower-mass η), also an s-wave $c\bar{c}$ but with spins antiparallel:

$$m_{\eta_c} \;=\; 2981 \;\mathrm{MeV} \quad \mathrm{and} \quad \Gamma_{\eta_c} \;\simeq\; 30 \;\mathrm{MeV}, \tag{4.2.11}$$

but $I^G, J^{PC} = 0^+, 0^{-+}$. Being spin-zero and also of positive C-parity (and thus the natural $c\bar{c}$ partner to the original pseudoscalar η), it is not accessible in e^+e^-.

Discovery of the *b* quark

The situation then repeats itself a little higher in energies: at around 10 GeV in the centre–of–mass the threshold for upsilon (Υ) production is reached and we find several peaks near together. The first of these is the $\Upsilon(1S)$. Historically, the discovery of the b (beauty or bottom) quark (Herb *et al.*, 1977 and Innes *et al.*, 1977) was actually made at the Fermilab proton machine, where 400-GeV protons were made to collide with copper and platinum targets. The process studied was

$$\begin{aligned} p + \mathrm{Cu, Pt} \;&\rightarrow\; \Upsilon + \mathrm{anything} \\ &\hookrightarrow\; \mu^+\mu^-, \end{aligned} \tag{4.2.12}$$

where again the resonance is evident via a final-state spectrum. The various radial excitations of the fundamental Υ state were observed later in the two most energetic e^+e^- machines of the time: the Doppel-Ring-Speicher (DORIS) at the

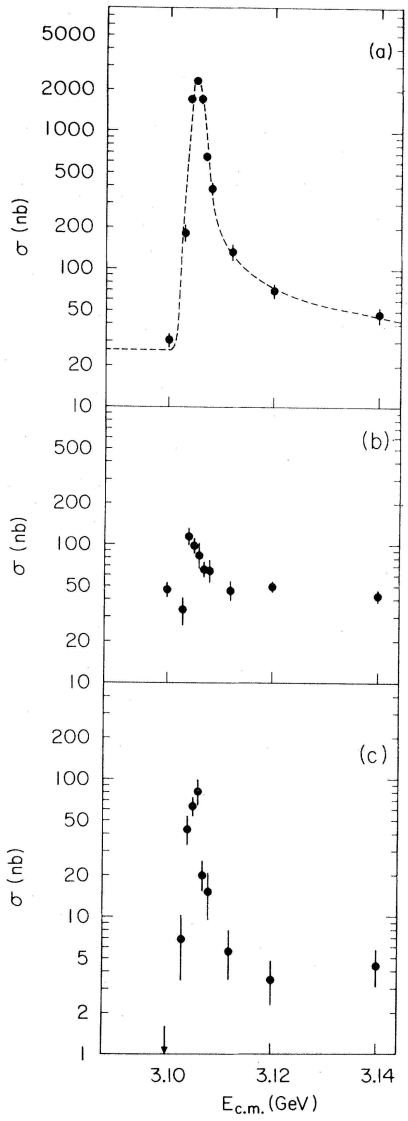


Figure 4.11: The e^+e^- cross-section *vs.* energy for various final states: (a) hadrons, (b) e^+e^- and (c) $\mu^+\mu^-$, $\pi^+\pi^-$ and K^+K^-. The dashed curve in the top figure is the theoretical prediction including a Gaussian energy spread of the beams. The figure is taken from Augustin *et al.* (1974).

Deutches Elektronische Syncrotron (DESY) in Hamburg and the Cornell Electron Storage Ring (CESR) at Cornell University.

The first Υ state has the following mass and width:

$$m_\Upsilon = 9.46 \text{ GeV} \quad \text{and} \quad \Gamma_\Upsilon \simeq 53 \text{ keV}, \tag{4.2.13}$$

with spin–parity quantum numbers $J^{PC} = 1^{--}$ (just as the J/ψ). On a logarithmic energy scale there are a number of states quite close to each other. In fact, in the first Fermilab experiment the resolution was rather poor (around 0.5 GeV) and the set of peaks thus appeared as one or possibly two broad humps. Nevertheless, since the total width of the distribution was around 1.2 GeV, it was immediately deduced that more than one resonance was present (see Fig. 4.12). On further

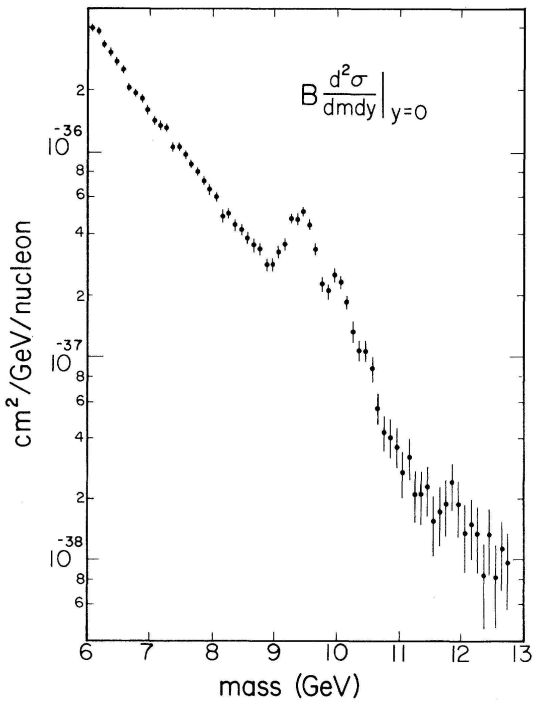


Figure 4.12: The broad overlapping peaks of the lowest-mass Υ states in the dimuon spectrum, as reported by Innes *et al.* (1977).

investigation, others were indeed found; the first and most prominent are those shown in Table 4.2, although there are many others (see too Fig. 4.13). The evident reduction in the widths with growing radial excitation will be discussed shortly.

Table 4.2: The three lowest-mass Υ ($b\bar{b}$, $J^{PC} = 1^{--}$, radial excitation) states.

state	mass/GeV	width/keV
$\Upsilon(2S)$	10.02	43
$\Upsilon(3S)$	10.36	26
$\Upsilon(4S)$	10.58	20

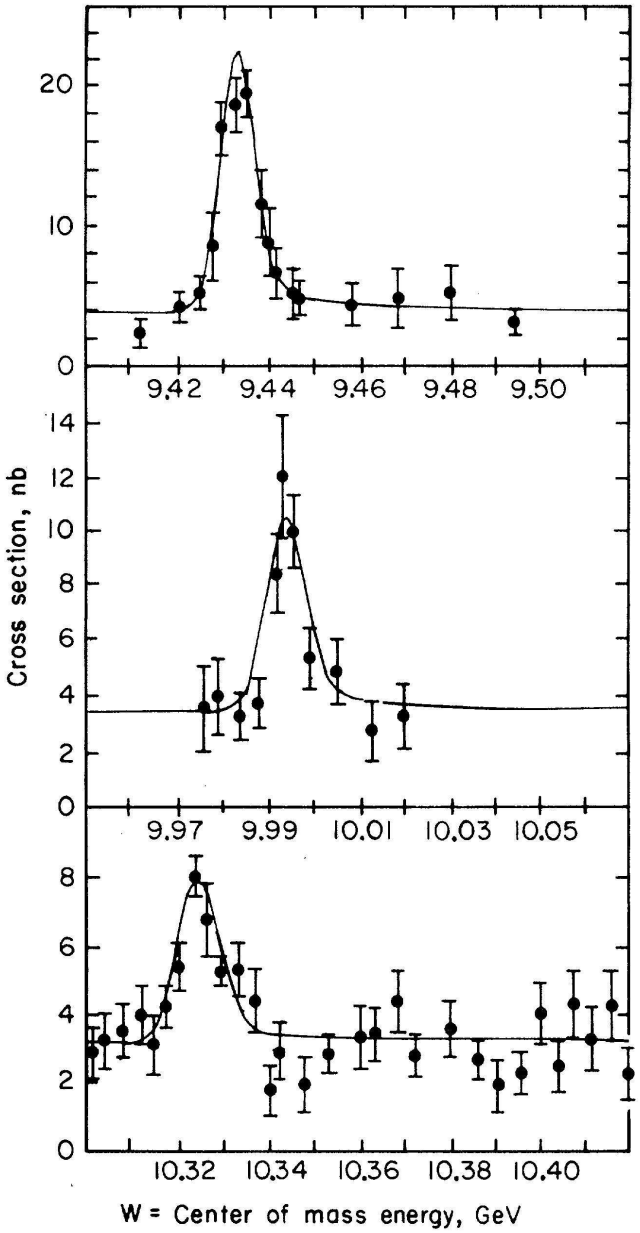


Figure 4.13: The narrow $\Upsilon(1S)$, $\Upsilon(2S)$ and $\Upsilon(3S)$ states observed with the CLEO detector at CESR. The figure is taken from Andrews *et al.* (1980).

Again, these states are very narrow owing to OZI-type suppression. Indeed, they are seen to be even narrower than the typical J/ψ widths. One might have *naïvely* expected just the opposite on the basis of arguments of both phase-space and number of channels available. However, since the OZI rule is evidently oper-

ative here and the decays must proceed via three-gluon intermediate states, there will be a very strong dependence on α_s. Each gluon implies a factor α_s in the total decay rate and thus, if perturbative calculations can be trusted, the widths are expected to be proportional to α_s^3. The reason this is not $(\alpha_s^3)^2$ is that the final coupling to quarks is not included as this part of the calculation is considered as a sort of decay of the gluons produced, which must occur with unit probability. Thus, in complete analogy with positronium decay, we expect one power of α_s for each gluon "emitted".*

Now, as already remarked in the section on the strong interaction, in quantum field theory the couplings are *not* constant, in particular, in QCD the coupling is a decreasing function of energy scale. It can be shown (in quantum field theory) that the correct scale to consider here is of the order of the quark masses involved. For a change of scale from that of the c-quark mass to that of the b-quark, the change in α_s is roughly a factor two. The resulting suppression is, of course, offset by various other effects: phase-space, number of channels and, last but not least, the size of the objects involved: the larger mass of the b quark in a non-relativistic approximation leads to a smaller Bohr radius and thus the wave-function is more concentrated at the origin, thus favouring the $q\bar{q}$ annihilation channel.

It is, of course, not possible to perform complete *ab initio* calculations since we are unable to reliably calculate the bound-state wave-function. However, for a heavy system such as $c\bar{c}$ or $b\bar{b}$ one might imagine non-relativistic quantum mechanics to be sufficient. Moreover, as discussed in the previous chapter, a reasonable assumption as to the effective form of the potential might be

$$V(r) \;=\; -\frac{\alpha(r)}{r} + b\,r, \tag{4.2.14}$$

where the first term represents a typical Coulomb-like potential, including the known scale dependence of the coupling ($Q^2 \sim r^{-2}$), and the second term is a string-like potential, which is believed to be dominant at large distances. The coefficient b is just a parameter of the model (the so-called *string tension*), to be fixed by fitting to the data. Using such a potential, we may solve the Schrödinger equation for a heavy quark–antiquark pair and the resulting "energy levels" will provide the masses of the various excitations (radial and rotational). The model can be refined by the inclusion of spin–orbit and spin–spin effects, with the necessary additional parameters. Indeed, such models already have modest success even with lighter systems and, for example, the string tension turns out to be approximately universal, with a value of the order of $1\,\mathrm{GeV\,fm^{-1}}$.

It is then possible to reliably estimate decay rates. An obvious prediction of

* Note, however, that for exclusive decay-channel rates, where a specific final state must also couple to the gluon, one would still expect a behaviour as $(\alpha_s^3)^2$.

such a model would be precisely the observed decrease in the decay widths of the heavy $q\bar{q}$ states with growing radial excitation. The decay annihilation requires the pair to be found at some instant at the origin and thus the rate is proportional to the wave-function at the origin. The radial-excitation states, having greater spatial extension naturally have smaller wave-functions at the origin. Note also that reliability can be improved by taking ratios: for example, we may compare the two-photon and two-gluon decays of the spin-zero quarkonia states, or the decays via a single photon (to two charged leptons) and three gluons in the case of the $J^{PC} = 1^{--}$ states. In this way, to some extent, the uncertainties in the bound-state parameters cancel in the ratio. The measured values for the decay rates agree well with the values of α_s extracted from other sources.

Toponium

A few words are in order here on the question of toponium ($t\bar{t}$), First of all, as we shall see, the t quark is much more massive than the others, having a mass $173.5 \pm 0.6 \pm 0.8\,\text{GeV}$ (see PDG-2016 – Patrignani *et al.*, 2016). This already naturally precludes its discovery in any e^+e^- collider built to-date, or indeed ever proposed. Moreover, the t quark itself is very unstable and is likely to have a much greater intrinsic decay rate than any bound state. Note that at energies corresponding to the top mass the weak interaction is no longer particularly suppressed by the $W^\pm$ and Z^0 boson masses. This would make the $t\bar{t}$ meson a very short-lived and weakly bound object, rendering it highly unlikely that any such resonances will ever be observed.

4.2.3 Discovery of the Z^0 boson

We now turn to the very last of the pronounced peaks in the cross-section plot for $e^+e^- \to \text{hadrons}$: namely, the Z^0 boson. This is evidently a very different situation to those described in the case of $q\bar{q}$ resonances. The Z^0 boson is an elementary field, with a point-like fundamental coupling to an e^+e^- initial (and/or final) state. However, its experimental appearance is very similar: having a finite (large) mass and a finite (but not small) width, a standard BW shape is observed in the cross-section energy dependence for a variety of final states. It was first produced on-shell in large numbers in LEP at the Centre Europée de Rechèrche Nucleaire (CERN—Geneva) although incontrovertible evidence for its existence had already been obtained in various other experiments. In particular the TRISTAN e^+e^- collider at KEK in Japan had already achieved a maximum $32+32\,\text{GeV}$. At such energies the presence of the Z^0 BW tail is very evident.

At intermediate energies (*i.e.* below the peak), where the Z^0 boson is not completely dominant and where interference with the photon intermediate state is

thus non-negligible, we can check (via the presence of interference) certain of the quantum numbers of this resonance. This is important: the spin and parity could, in principle, be different to that of the photon, since the coupling to fermions could be different. In an e^+e^- collision all possible spin configurations of the initial pair are possible and indeed present:

$$|-\hat{\boldsymbol{z}}\rangle|-\hat{\boldsymbol{z}}\rangle, \quad |-\hat{\boldsymbol{z}}\rangle|+\hat{\boldsymbol{z}}\rangle, \quad |+\hat{\boldsymbol{z}}\rangle|-\hat{\boldsymbol{z}}\rangle, \quad |+\hat{\boldsymbol{z}}\rangle|+\hat{\boldsymbol{z}}\rangle, \tag{4.2.15}$$

where the $\hat{\boldsymbol{z}}$ indicates that these are *not* helicities, but spin projections along the z-axis. They therefore represent the helicities of the particle moving in the positive z direction and *minus* the helicities for the other. These states should then be rearranged into multiplets according to *total spin*:

$$\text{spin-1:} \qquad |-\hat{\boldsymbol{z}}\rangle|-\hat{\boldsymbol{z}}\rangle, \quad \tfrac{1}{\sqrt{2}}\left(|-\hat{\boldsymbol{z}}\rangle|+\hat{\boldsymbol{z}}\rangle + |+\hat{\boldsymbol{z}}\rangle|-\hat{\boldsymbol{z}}\rangle\right), \quad |+\hat{\boldsymbol{z}}\rangle|+\hat{\boldsymbol{z}}\rangle, \tag{4.2.16a}$$

$$\text{spin-0:} \qquad \tfrac{1}{\sqrt{2}}\left(|-\hat{\boldsymbol{z}}\rangle|+\hat{\boldsymbol{z}}\rangle - |+\hat{\boldsymbol{z}}\rangle|-\hat{\boldsymbol{z}}\rangle\right). \tag{4.2.16b}$$

The coupling to a spin-one boson merely selects a particular multiplet; three out of four are acceptable, reflecting the following factor already shown in the BW formulæ:

$$\frac{(2J_R+1)}{(2s_1+1)(2s_2+1)} = \frac{3}{4}. \tag{4.2.17}$$

Note that, for the Z^0 to interfere with the photon, it must have the same J^{PC}. As we shall see in the next chapter, the Z^0 is to be considered a close relative of the photon, as too are the $W^\pm$.

As deduced from the BW-resonance line-shape, the mass and width parameters of the Z^0 are (see PDG-2016 – Patrignani *et al.*, 2016):

$$m_Z = 91.1876 \pm 0.0021 \text{ GeV} \quad \text{and} \quad \Gamma_Z = 2.4952 \pm 0.0023 \text{ GeV}. \tag{4.2.18}$$

It is a spin-one boson, just as the photon, but its parity and charge-conjugation properties are *not* defined: the weak coupling of the Z^0 violates both parity and charge-conjugation (just as that of the $W^\pm$) and it cannot therefore be an eigenstate of either parity or charge conjugation. This last observation does not preclude its interfering with the photon since it should thus be considered as a superposition of parity- and charge-conjugation eigenstates (just as, *e.g.* the K^0 and $\overline{K}^0$ are superpositions of CP eigenstates). Note finally that such precise measurements* on such a broad object require the inclusion of the phase-space corrections to the

* Indeed, such is the precision achieved at LEP, that tidal effects on the length of the ring (variations of order 1 mm), which can contribute up to 40 MeV to the beam energy, must be taken into account (Arnaudon *et al.*, 1995). Even gravitational effects, owing to the nearby Jura mountains and seasonal changes in the level of water in Lake Geneva, have to be considered.

width itself as s various across the BW peak, see Eq. (A.5.26).

Now, recall that the width Γ_Z, as deduced from the line-shape, is the *total* width—that is to say it includes all available channels without discrimination. Remembering that the width is none other than the decay rate, we see that if there are a number of *different* (*i.e.* *non*-interfering) final states, then the total width (or rate) is simply the sum of all the partial widths (or rates). This leads to some very important cross checks.

Z^0 partial decay widths

First of all, universality of the weak coupling may be checked in the neutral-current channel by comparing the measured decay rates or branching ratios into electrons, muons and tau leptons:

$$\Gamma_Z^{e^+e^-}/\Gamma_Z^{\text{tot}} = (3.363 \pm 0.004)\ \%, \tag{4.2.19a}$$

$$\Gamma_Z^{\mu^+\mu^-}/\Gamma_Z^{\text{tot}} = (3.366 \pm 0.007)\ \%, \tag{4.2.19b}$$

$$\Gamma_Z^{\tau^+\tau^-}/\Gamma_Z^{\text{tot}} = (3.370 \pm 0.008)\ \%. \tag{4.2.19c}$$

Note that phase-space differences due to lepton masses are essentially negligible here. One can, of course, also *calculate* the partial decay rates; the generic rate for a real (*i.e.* on mass-shell) Z^0 to decay into a fermion–antifermion pair is

$$\Gamma_Z^{f\bar{f}} = \frac{C\,G_{\text{F}} M_Z^3}{6\sqrt{2}\,\pi}\left[g_{Vf}^2 + g_{Af}^2\right], \tag{4.2.20}$$

where G_{F} is just the usual Fermi weak coupling constant, *e.g.* as measured in μ-decay. The coefficient C is defined as follows:

$$C = \begin{cases} 1 & \text{for lepton pairs,} \\ N_{\text{c}}\left[1 + a + 1.409a^2 - 12.77a^3 - 80.0a^4\right] & \text{for quark pairs.} \end{cases} \tag{4.2.21}$$

The second line here includes the corrections due to the strong interaction (or QCD), where $a = \alpha_s(M_Z)/\pi$ (*i.e.* α_s as evaluated at the energy scale of the Z^0, for which the present world average is 0.1184 ± 0.0007). Finally, the vector- and axial-vector weak coupling constants (or weak charges) are

$$g_{Vf} = I_{3f} - 2Q_f \sin^2\theta_{\text{W}}, \tag{4.2.22a}$$

$$g_{Af} = I_{3f}, \tag{4.2.22b}$$

where I_{3f} is just the third component of *weak* isospin ($+\frac{1}{2}$ for up-type quarks and neutrinos, $-\frac{1}{2}$ for down-type quarks and charged leptons), Q_f is the electric charge of the fermion in units of $|e|$ and $\theta_{\rm W}$ (with $\sin^2\theta_{\rm W} \simeq 0.223$) is the weak mixing angle, which we shall discuss later. Inserting the values of the various parameters (all of which may, in principle, be measured independently in other processes), one obtains

$$\Gamma_Z^{f\bar{f}} \;=\; \begin{cases} 300.26 \pm 0.05 \;\text{MeV} & (u\bar{u},\, c\bar{c}), \\ 383.04 \pm 0.05 \;\text{MeV} & (d\bar{d},\, s\bar{s}), \\ 375.98 \pm 0.03 \;\text{MeV} & (b\bar{b}), \\ 167.22 \pm 0.01 \;\text{MeV} & (\nu\bar{\nu}), \\ 84.00 \pm 0.01 \;\text{MeV} & (\ell\bar{\ell}). \end{cases} \tag{4.2.23}$$

The predicted total width is then

$$\Gamma_Z^{\rm tot} \;=\; 2.4968 \pm 0.0011 \;\text{GeV}, \tag{4.2.24}$$

in excellent agreement with the measurements.

The number of light neutrinos

Hidden inside the numbers and formulæ just presented is an interesting and perhaps surprising measurement: the number of *light* neutrinos. First of all, we say light neutrinos since one would obviously not be sensitive to a neutrino with a mass greater than half that of the Z^0. Now, of course, although the statistics gathered at LEP is considerable (many tens of millions of Z^0 bosons have been produced and detected), there is little chance of actually detecting either the neutrino or anti-neutrino produced in a Z^0 decay. Note that to be certain, one would need to detect *both* and measure their energies, in order to fully reconstruct the mass of the decaying object. However, since the line-shape provides the total width and the other partial widths may all be measured directly, we can deduce the "invisible" width from the difference. The following combination is then usually calculated:

$$N_\nu \;=\; \left(\frac{\Gamma_{\rm inv}}{\Gamma_\ell}\right)_{\rm expt} \left(\frac{\Gamma_\ell}{\Gamma_\nu}\right)_{\rm th} \;=\; \left(\frac{\Gamma_{\rm tot} - \Gamma_{\rm vis}}{\Gamma_\ell}\right)_{\rm expt} \left(\frac{\Gamma_\ell}{\Gamma_\nu}\right)_{\rm th}. \tag{4.2.25}$$

The double ratios in these equalities are so chosen as to improve the reliability of the calculation by allowing cancellations of various systematic effects, both theoretical and experimental. Combining the results from all four LEP experiments,

leads to

$$N_\nu = 2.92 \pm 0.06. \tag{4.2.26}$$

That is, the three known neutrinos (the partners to the electron, muon and tau lepton) are confirmed as the only *light* neutrinos. Incidentally, this number also places very stringent limits on the existence of other light particles coupling to the Z^0 and which might hitherto have gone undetected for some reason.

Many other tests of the SM may be performed at the Z^0 peak. For example, there are various parity-violating angular asymmetries that may be measured and that are directly related to the parameters of the theory (in particular, to $\sin\theta_{\rm W}$). All measurements so far performed provide absolutely *no* evidence of any flaw or shortcoming in the model. We shall discuss these questions more in detail in the following chapter.

4.2.4 Discovery of the τ lepton

The remaining major discovery made in e^+e^- collisions is that of the τ lepton, again found using SPEAR at SLAC (Perl *et al.*, 1975). The experiment obtained just 64 events of the form $e^+e^- \to e^\pm + \mu^\mp +$missing energy, in which no other charged particles or photons were detected. Most of these events were detected at or just above a centre–of–mass energy of 4 GeV (see Fig. 4.14). The missing-energy and

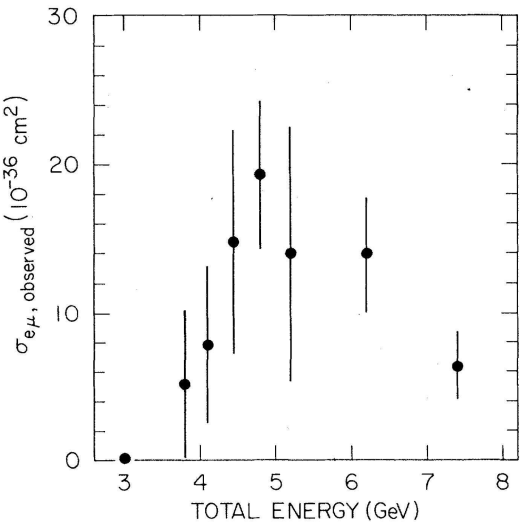


Figure 4.14: The observed cross-section for the $e^\pm + \mu^\mp +$missing-energy events as a function of e^+e^- centre–of–mass energy. The figure is taken from Perl *et al.* (1975).

missing-momentum spectra indicated that at least two additional particles had

been produced in each event. There being no conventional explanation for such events, they were attributed to the production of a new charged-lepton pair $\tau^+\tau^-$.

The processes being observed here are then presumed to be

$$\begin{array}{rcl} e^+e^- & \to & \tau^+\tau^- \\ & & \quad \big\lfloor \hookrightarrow \mu^-\bar{\nu}_\mu\nu_\tau \\ & & \quad \hookrightarrow e^+\nu_e\bar{\nu}_\tau \end{array} \qquad (4.2.27a)$$

and

$$\begin{array}{rcl} e^+e^- & \to & \tau^+\tau^- \\ & & \quad \big\lfloor \hookrightarrow e^-\bar{\nu}_e\nu_\tau \\ & & \quad \hookrightarrow \mu^+\nu_\mu\bar{\nu}_\tau. \end{array} \qquad (4.2.27b)$$

Under normal circumstances, neither of these two (charged) final states is possible at the level measured since they would represent *individual* (electron and muon) lepton-number violation. However, the undetected neutrinos (missing energy) balance all relevant quantum numbers.

4.2.5 Open-flavour and other particle production

Shortly above the thresholds for each new quark, which are first seen as $q\bar{q}$ resonances, the thresholds for *open* production are reached. That is, the centre–of–mass energy is sufficient to produce a meson–antimeson pair, each containing the new quark or antiquark paired up with a light antiquark or quark respectively.

D mesons

Immediately above the J/ψ peak we thus see a fall in the cross-section, back to the previous level, owing to the fact that we are now off resonance but have not yet reached the threshold for open charm. As soon as there is sufficient energy available in the centre–of–mass to produce a D-meson pair, we see a rise to the new level. The lowest-lying D mesons have $J^P = 0^-$; the masses and decay widths are shown in Table 4.3. An extra charged state is now possible by replacing the d

Table 4.3: The lowest-lying D-meson masses and decay widths.

state	quark content	mass/MeV	$\tau/10^{-12}$s
$D^\pm$	$(c\bar{d},d\bar{c})$	1870	1.04
D^0, $\overline{D}^0$	$(c\bar{u},u\bar{c})$	1865	0.41
$D_s^\pm$	$(c\bar{s},s\bar{c})$	1968	0.49

quark with an s—it is naturally a little heavier than the others.

Note that here we do not talk of decay widths, but of lifetimes. These mesons do not have access to strong annihilation channels and only decay via the weak decay of the individual c and $\bar{c}$ quarks. The lifetimes are therefore considerably longer and the resonance widths too narrow to be determined as such. By using modern silicon-strip tracking detectors with vertex resolution capabilities of the order of a few microns, the lifetimes are thus typically determined via precise measurement of track lengths.

B mesons

Slightly above the Υ peak we see a rise in cross-section to a new level corresponding to the threshold for open-b production, which initiates with the production of B mesons. This is in precise analogy with the previous case of D mesons. Again, the lowest-lying B mesons have $J^P = 0^-$; the masses and decay widths are shown in Table 4.4. Here we see yet another new possibility: the doubly heavy, charged,

Table 4.4: The lowest-lying B-meson masses and decay widths.

state	quark content	mass/MeV	$\tau/10^{-12}$s
$B^{\pm}$	$(u\bar{b},b\bar{u})$	5279	1.64
B^0, $\bar{B}^0$	$(d\bar{b},b\bar{d})$	5279	1.53
B_s^0, $\bar{B}_s^0$	$(s\bar{b},b\bar{s})$	5368	1.47
$B_c^{\pm}$	$(c\bar{b},b\bar{c})$	6286	0.46

meson states $c\bar{b}$ and $b\bar{c}$.

W^+W^- pair production

The final structure visible in the plot of the hadronic cross-section in e^+e^- annihilation is the opening of the threshold for the production of W^+W^- pairs. So far such a study has only been possible in one machine: LEP at CERN. After having thoroughly studied the Z^0 resonance with centre–of–mass energies around 91 GeV, the machine was slowly pushed to its design limit of about 100 GeV beam energy, thus providing a total centre–of–mass energy of 200 GeV. This permitted the study of the highest-energy process accessible in this machine W^+W^- production:

$$e^+e^- \rightarrow \gamma, Z^0 \rightarrow W^+W^-. \tag{4.2.28}$$

Note that while one might assume the photon channel to be rather natural, since, after all, the $W^{\pm}$ is charged, the triple weak-boson coupling $Z^0W^+W^-$ is a precise

prediction of the electroweak theory. As such, it must be tested; again, the measurements performed revealed absolutely *no* indication of deviations with respect to theoretical predictions.

4.2.6 Jets in e^+e^- annihilation

Before leaving the topic of e^+e^- collisions, there is one final type of process that deserves detailed examination. We have discussed a great deal the production of hadrons in such collisions. Before the advent of QCD and the quark model, the description of the process $e^+e^- \to$ hadrons took the form of the intermediate creation of what was typically called a *fireball*. In such a picture the centre–of–mass energy was more-or-less uniformly distributed among final-state particles emerging over the entire 4π solid angle. However, once the idea of point-like objects that could be produced in particle–antiparticle pairs was established, this picture changed radically. We have already seen the relevant quark diagrams many times; the main point to appreciate is that, at leading order, the dominant process sees the production of a pair of quarks, with a large amount of kinetic energy and which are therefore projected out of the interaction region *back-to-back*. Thus, if the *hadronisation* process does little to change the direction of motion, we might expect to see something like two back-to-back *jets* of particles (see Fig. 4.15).

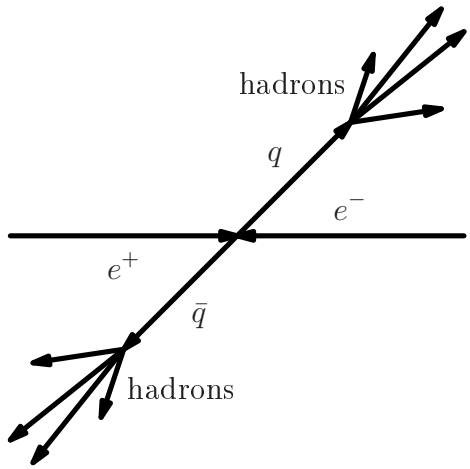


Figure 4.15: A schematic view of the jet-like structure one might expect, based on the assumption of a process initiated by quark–antiquark pair production.

The Lund model

The final detected hadronic state is born then of just two objects, separating at high velocity. Unfortunately, we do not know how these two fundamental fields

convert or *fragment* into the hadrons that eventually materialise in the laboratory and propagate some finite distance until either arriving directly in a detector or decaying into other lighter, more stable hadrons. At any rate, one present conventional picture of the *hadronisation* process is as follows (see Fig. 4.16). As the

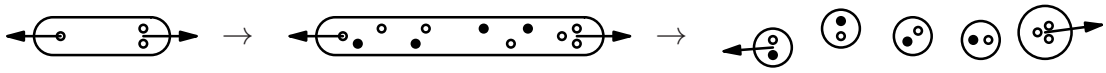


Figure 4.16: The Lund string-model picture of quark fragmentation. The hadrons created at the ends of the string will be the fastest (or so-called *leading*) particles and each will be followed by a jet of slower hadrons.

$q\bar{q}$ separate the field lines stretch out between them to form a string-like object, in which energy is stored according to an approximately linear potential. As this stored potential energy grows, the threshold for real pair production is approached and at some point, the string may break producing a quark–antiquark pair, which will then form the new ends of the two strings so produced. Although a small amount of energy may be converted into transverse momentum, the principal direction of the quarks thus created will be parallel to the initial motion. This process will iterate, with the smaller strings breaking again and again until the energy has been fully dissipated.

The end picture then (the Lund model; Andersson *et al.*, 1983) has a number of final-state particles, over which the energy of the initial quark–antiquark system is shared (in relation to position along the string) and which are moving more-or-less collinearly—this is called a *jet*. At low centre–of–mass energy and consequent low final-state multiplicity, it is hard to distinguish such behaviour. However, at higher energy the two-jet nature of the majority of hard-scattering events clearly emerges. Models based on these ideas, with a simple stochastic choice of the string breaking points provide a good description of the jet-like events observed; an example experimentally observed two-jet event is shown in Fig. 4.17.

A number of variables may be defined to describe the *shape* of a jet-like event: examples are *sphericity* and *thrust*. A value of sphericity is defined for each event as

$$S := \frac{3\left(\sum_i p_{\mathrm{T}i}^2\right)_{\min}}{2\sum_i p_i^2}, \tag{4.2.29}$$

where $p_{\mathrm{T}i}$ is the momentum of the i-th. particle perpendicular to the sphericity axis; the sphericity axis is defined as that minimising S. A perfect two-jet event, with outgoing particles aligned precisely along the axis, has $S=0$ whereas for a perfectly isotropic event $S=1$. The graph in Fig. 4.18 shows measured mean sphericity as a function of centre–of–mass energy in comparison with jet-model (solid line) and isotropic phase-space model (dashed curve) predictions.

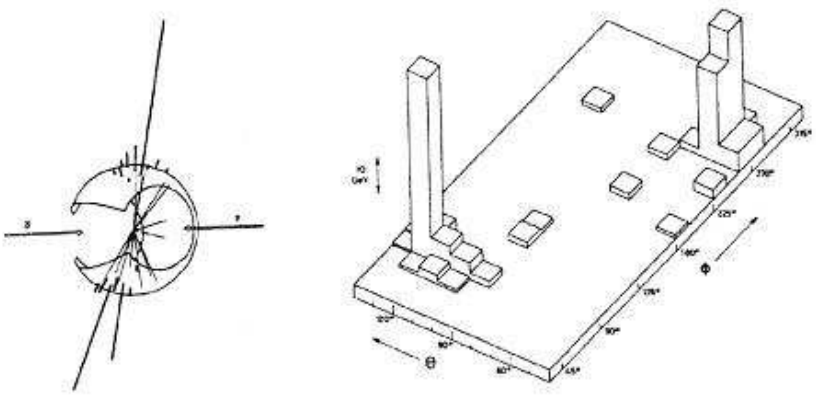


Figure 4.17: An artistic rendering of a two-jet event as experimentally observed in the UA2 detector, together with the corresponding *Lego plot*.

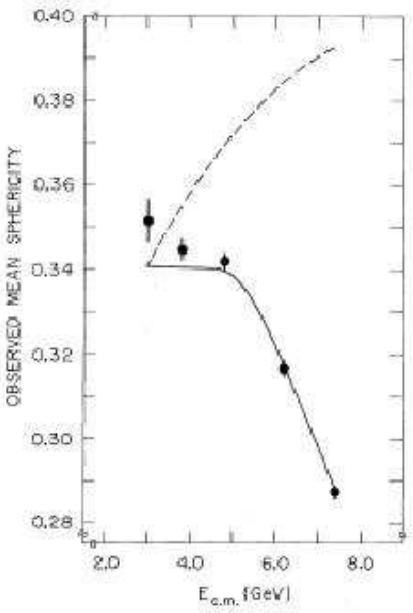


Figure 4.18: The mean sphericity as a function of centre–of–mass energy in comparison with a jet-model (solid line) and an isotropic phase-space model (dashed curve).

So far we have appealed to non-perturbative aspects of QCD and its string-like long-distance potential. There is also a perturbative aspect to the interaction, in which gluons may be emitted, essentially via *Bremsstrahlung*, see Fig. 4.19a. Given the size of α_s, such a process has a non-negligible probability. Moreover, it is quite likely to produce a gluon with large energy and at a large angle with respect to the direction of the emitting quark. There are then three rapidly moving particles,

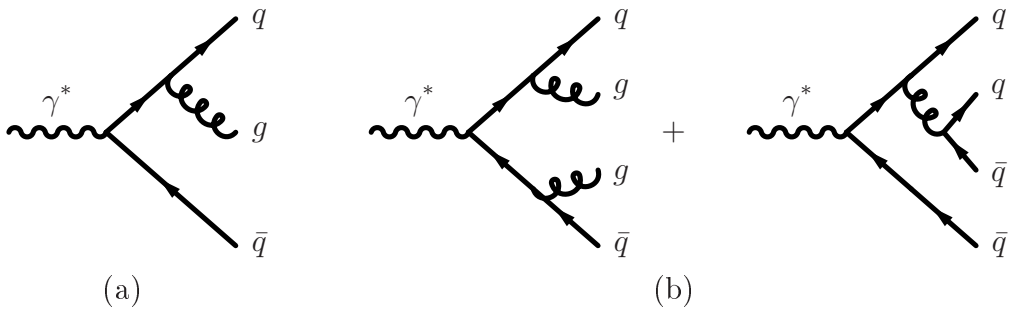


Figure 4.19: Hard gluon *Bremsstrahlung* and $q\bar{q}$ pair-production processes leading to (a) three- and (b) four-jet events.

which, according to the string model, should give rise to three jets. Again, with sufficient centre–of–mass energy, such events are clearly visible. Indeed, since the probability of gluon emission is directly proportional to α_s, the ratio of the numbers of events with three and two jets is a direct measure of α_s. This method of extracting the strong coupling constant agrees well with other results.

One can naturally go further and consider events with four or more jets, see Fig. 4.19b. However, there is evidently a limit to how many jets one can sensibly hope to identify experimentally. The available energy is a severe limitation but there is also the question of adjacent jets overlapping.

Going back to the string model, we see that in the case of a three-jet event the string will stretch from the quark to the gluon and on to the antiquark (see Fig. 4.20). Note how the string is stretched between the quark and gluon and also

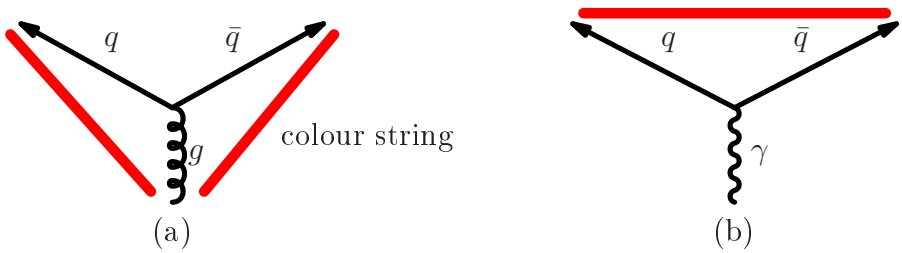


Figure 4.20: The evolution of a three-jet event in the Lund string picture. Of the two possibilities shown, that on the left apparently better describes the data. In particular, it leads to a depletion of particle and energy flow on the opposite side to the gluon jet. The right-hand figure would be as expected for *Bremsstrahlung* of, say, a hard photon.

between the antiquark and gluon, thus generating final-state hadrons mainly in the angular regions between those pairs of partons. However, there is no string directly between the quark and antiquark. Therefore, one would expect this angular region to be depleted with respect to the other two. Now, since the gluon should on average carry less energy, there should be a correlation between the number of

particles in intermediate regions and the energy of the away-side jets. In other words, on average, we expect to find less particles opposite to the lowest-energy jet as compared with the other two, or equivalently less particles between the two most energetic jets. This prediction of the model is also experimentally well verified (see Fig. 4.21).

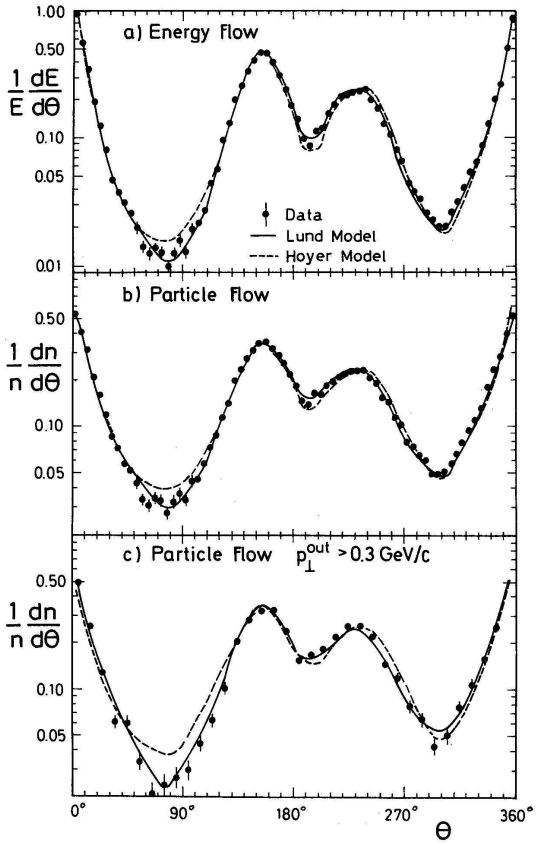


Figure 4.21: Correlations between the energy/particle flow in the angular region between the two jets opposite the lowest-energy jet. Comparisons are presented with a QCD-based model (Hoyer *et al.*, 1979) and with the Lund string model (Andersson *et al.*, 1983); the figure is taken from Bartel *et al.* (1983)

The Marchesini–Webber model

In a different approach, the idea of parton emission (*i.e.* gluon *Bremsstrahlung* and quark–antiquark pair production) led to the development of a perturbative model of hadronisation (Marchesini and Webber, 1984), in which each emission is treated more-or-less independently and thus iterated (see Fig. 4.22). At the end of such a *parton shower* we are again left with a large number of quarks distributed over a

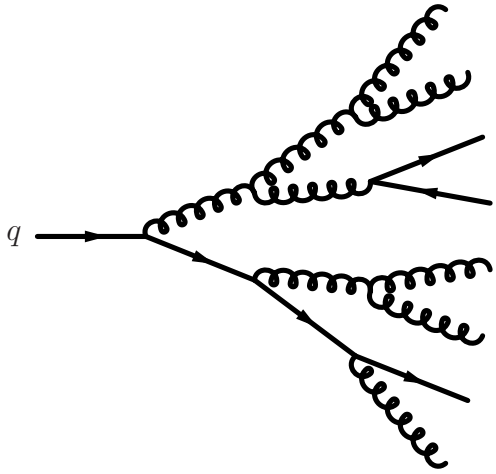


Figure 4.22: The perturbative QCD-based parton-shower model for quark fragmentation.

range of momenta. The hadronisation *ansatz* adopted consists of combining each nearby colourless quark–antiquark pair and assuming it will form the meson state nearest in mass and quantum numbers. Again, a good description of the observed phenomenology is obtained. The depletion phenomenon on the opposite side to the gluon jet is seen here as the result of a so-called colour-coherence effect.

In conclusion, let us remark that in all approaches to the description of a final multi-hadron state there is a great deal of modelling, with necessarily a large number of parameters to be fitted or *tuned* to the data. That these models work as well as they do is a clear indication that our heuristic picture cannot be far from the truth. That said, it must always be borne in mind that we have as yet no first-principles approach to the problem, and then again, with enough parameters, even an incorrect model may appear satisfactory.*

4.3 Proton–antiproton colliders

In this final section of the chapter on new particles we now turn briefly to the case of proton–antiproton (and proton–proton) colliders. All else being equal, one would not normally choose to perform new-particle searches in a hadron collider owing its inherently dirty nature. Conceptually, the problems are essentially two-fold, but with a common origin: the actual collisions that take place in such a collider at the elementary level are not between the hadrons themselves but between the

* One need only recall the Ptolemaic or geocentric model of the universe with its necessary epicycles, to understand how complex but *wrong* models can nevertheless produce apparently or rather approximately correct answers.

partons they contain. The two resulting difficulties can be classified as initial- and final-state.

The difficulty with the initial state is that *a priori* we can have no direct knowledge (*i.e.* event-by-event) as to the details: parton type, energy–momentum, spin *etc.* What we do have is a statistical description: we have parton distributions, measured in other processes, that give us the probability of finding a given parton type with a given momentum fraction inside a given hadron. Indeed, event-by-event we can only hope to reconstruct the initial kinematics from measurement of an exclusive final state, which thus limits what can measured.* Even considering only statistical analyses, we do nevertheless require accurate prior information on the parton densities and, for example, in the case of gluons, or in general for very small values of x_{B}, this is not yet available.

As far as the final state is concerned, the difficulty here is that a large part of the hadron remnants disappears along the beam pipe and cannot therefore be detected. Moreover, it is possible for remnants of the spectator systems (the unstruck quarks) to fall within the detector acceptance. This all makes complete reconstruction impossible and therefore precludes certain types of analyses.

Naturally though, there are situations where there is no choice; typically this has to do with the available energy. The highest centre–of–mass energy achieved to date in an e^+e^- machine is 200 GeV of LEP II whereas the Tevatron at Fermilab ran for many years at a centre–of–mass energy of slightly less than 2 TeV, while the Large Hadron Collider (LHC) at CERN is designed to achieve 14 TeV centre–of–mass energy—although it should be noted that this is a proton–proton machine.† Moreover, we have already seen cases (such as the J/ψ discovery) in which detection is equally as favourable in hadron–hadron as in e^+e^- machines. Indeed, the η_c, for instance, cannot actually be produced directly in an e^+e^- machine.

4.3.1 Discovery of the $W^\pm$ boson

A clear example of the need for a hadron–hadron machine is the case of the production of the $W^\pm$ bosons.‡ Given that the $W^\pm$ couple to a pair of *different* flavour fermions (*e.g.*, $u\bar{d}$, $e^+\nu_e$ *etc.*), a simple e^+e^- collision is no longer useful. Since high-energy collisions between electron or muon and neutrino beams of sufficient intensity are unattainable, the only choice is to collide proton and, preferably, an-

* Incidentally, for somewhat technical theoretical reasons, the use of exclusive final states can seriously impair the reliability of perturbative QCD predictions.

† In its heavy-ion mode LHC is also intended to provide collisions between, *e.g.* lead or gold ions with approximately 2.6 TeV per *nucleon*.

‡ The 1984 Nobel Prize for physics was awarded equally to Carlo Rubbia and Simon van der Meer for "their decisive contributions to the large project, which led to the discovery of the field particles W and Z, communicators of weak interaction."

tiproton beams. For detection of $W^{\pm}$ in $p\bar{p}$ collisions, the following then are the main channels:

$$u + \bar{d} \rightarrow W^+ \rightarrow e^+ + \nu_e, \ \mu^+ + \nu_\mu, \tag{4.3.1a}$$

and

$$\bar{u} + d \rightarrow W^- \rightarrow e^- + \bar{\nu}_e, \ \mu^- + \bar{\nu}_\mu. \tag{4.3.1b}$$

As a by-product, in the same collisions production of the Z^0 is also accessible, although a little more difficult, via

$$u + \bar{u}, \ d + \bar{d} \rightarrow Z^0 \rightarrow e^+e^-, \ \mu^+\mu^-. \tag{4.3.1c}$$

If we recall that valence quarks might be expected to carry on average roughly one third of the total hadronic energy, then we might expect to need of order at least three times the $W^{\pm}$ mass, or $\gtrsim$240 GeV in the centre–of–mass. In fact, the situation is rather worse, for two reasons. First of all, interactions, via which the gluons and sea quarks are generated inside the proton, significantly reduce the energy share of the valence quarks: as mentioned earlier, we find experimentally that the gluons carry slightly more than half the total energy of the proton. Secondly, this effect increases with energy scale. A higher energy scale is equivalent to finer spatial resolution and, as we look at the proton in ever finer detail, we resolve more and more gluons and sea quarks. The resulting mean effective valence and sea-quark fractions relevant to such an experiment are then

$$\langle x_{\rm val} \rangle \approx 0.12 \quad \text{and} \quad \langle x_{\rm sea} \rangle \approx 0.04. \tag{4.3.2}$$

To produce a $W^{\pm}$ in a $p\bar{p}$ machine, one would therefore expect to require nearer $80/0.12 \sim 670$ GeV and for pp around $80/\sqrt{0.12 \times 0.04} \sim 1200$ GeV total centre–of–mass energies. In practice, one can use slightly lower energies and work in the tails of the parton distributions; *i.e.* accepting lower event rates.

The cross-section for $W^{\pm}$ production in a *parton–parton* collision may be calculated using the standard BW form, see Eq. (A.5.15):

$$\sigma(u\bar{d} \rightarrow W^+ \rightarrow e^+\nu_e) \simeq \frac{1}{N_{\rm c}} \frac{1}{3} \frac{(2J+1)}{(2s_u+1)(2s_d+1)} \times \frac{4\pi\hbar^2}{p^2} \frac{(\frac{1}{2}\Gamma_{u\bar{d}})(\frac{1}{2}\Gamma_{e\nu})}{(E_{\rm CM} - M_W)^2 + (\frac{1}{2}\Gamma_{\rm tot})^2}, \tag{4.3.3}$$

where the factor $N_{\rm c}$ in the denominator accounts for the requirement that for a quark of a given colour, the corresponding antiquark must have precisely that (anti-) colour out of the three possible; the further factor 3, associated with spin

degeneracy, arises owing to the requirement that the quark and antiquark be left and right handed respectively; *i.e.*, the $W^{\pm}$ are only produced in one of the three possible helicity states.

At the peak energy ($E_{\rm CM} = M_W$) the maximum total cross-section (in natural units) is therefore

$$\sigma_{\max}(u\bar{d} \to W^+ \to e^+\nu_e) \simeq \frac{4\pi}{3M_W^2} B_{u\bar{d}} B_{e\nu} = \frac{4\pi}{81M_W^2} \simeq 9.2 \text{ nb}. \qquad (4.3.4)$$

The branching ratios used are then based on the observation that each single decay channel (leptonic or coloured quark) has the same weight (neglecting the tiny phase-space variations. There are three leptonic ($e\nu_e$, $\mu\nu_\mu$ and $\tau\nu_\tau$) and two hadronic channels ($u\bar{d}_{\rm W}$ and $c\bar{s}_{\rm W}$) of three colours, giving a total of nine equal-weight channels. We therefore have

$$B_{u\bar{d}} = N_{\rm c}/9 = 1/3 \quad \text{and} \quad B_{e\nu} = 1/9. \qquad (4.3.5)$$

To calculate the cross-section for hadronic collisions, we need to integrate (or average) over the width of the resonance and the relevant partonic distributions. For the actual experiments (UA1: Arnison *et al.*, 1983a and UA2: Banner *et al.*, 1983) performed in the Sp$\bar{\rm p}$S collider at CERN, with very high-intensity 270-GeV* proton and antiproton beams, the cross-sections are

$$\sigma(p\bar{p} \to W^+ \to e^+\nu_e) \sim 1 \text{ nb} \qquad (4.3.6)$$

and

$$\sigma(p\bar{p} \to Z^0 \to e^+e^-) \sim 0.1 \text{ nb}. \qquad (4.3.7)$$

These should be compared to the total $p\bar{p}$ cross-section at such energies, which is measured to be about 40 mb. In other words, the two signals were at a level of 10^{-8} and 10^{-9} of the total, respectively. Such a weak signal requires very distinctive events.

Now, since almost all of the available energy must be used to provide the $W^{\pm}$ mass and since the quark and antiquark will also tend to have similar velocities, the boson so-produced will be almost at rest in the laboratory. When it decays there is then a high probability that the final-state lepton–neutrino pair will come out more-or-less back-to-back (in the laboratory) and at large angles with respect to the beam direction, *i.e.* with high transverse-momentum or $p_{\rm T}$. Moreover, the detected lepton will have large energy $\approx \frac{1}{2}M_W$ while, of course, the neutrino goes undetected. The signature is thus:

- a single, isolated, high-$p_{\rm T}$ electron track in the central tracking detector;

* In later, so-called *ramped*, runs the CERN machine achieved beam energies of 318 GeV.

- a very localised shower in the electromagnetic calorimeter;
- large missing p_{T} when all transverse momenta are added vectorially.

Schematically then, the principal components of the detector setup are as follows (see Fig. 4.23)

- a central tracking detector with a high magnetic field to observe charged particles and measure their momenta;
- electromagnetic shower counters, which detect both electrons and photons;
- hadron calorimeters;
- muon detectors.

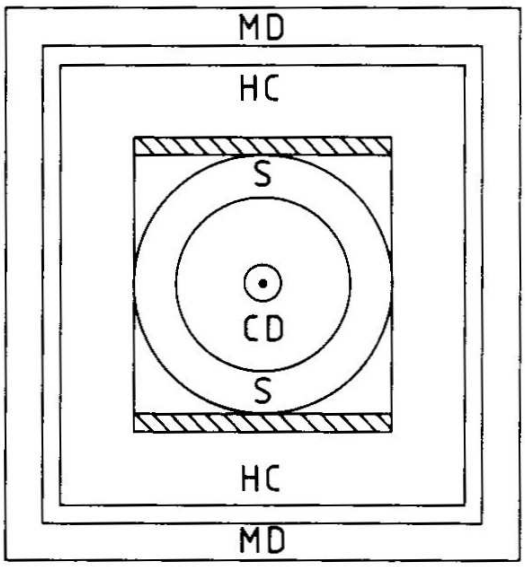


Figure 4.23: A schematic view of a transverse section of the UA1 detector at CERN. The beams travel along the centre (in and out of the page). CD is the central detector, S the shower counters, HC the hadron calorimeters and finally MD are the muon detectors. The shaded area represents the magnetic coils.

Figure 4.24 shows an example of the electromagnetic-calorimeter energy deposition due to a $W^{\pm}$ event. Note that the individual longitudinal momenta of the incident partons are not known and therefore we know nothing of the possible Lorentz boost of the decaying system. It is thus impossible to fully reconstruct the event and so nothing can be said about the $W^{\pm}$ mass or width on an event-by-event basis However, the p_{T} distribution of the detected electrons may be measured. For the purposes of this explanation, we shall assume that the $W^{\pm}$ is produced approximately at rest in the laboratory system—a full (numerical) analysis should

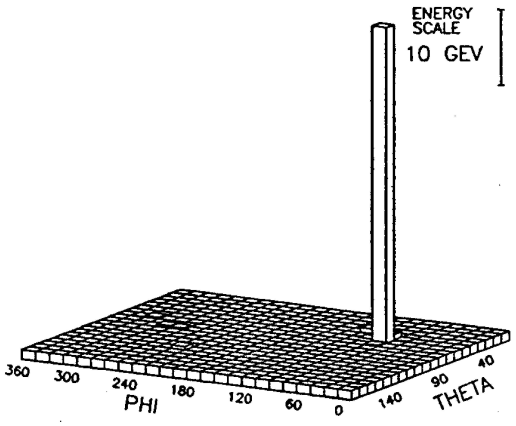


Figure 4.24: The so-called *Lego plot* from the electromagnetic calorimeter due to a $W^{\pm}$ event. The figure is taken from UA2 (Banner *et al.*, 1983).

also use the parton distributions to account for non-zero centre–of–mass motion. The transverse momentum of the outgoing electron is then

$$p_{\mathrm{T}} \approx \tfrac{1}{2} M_W \sin\theta, \tag{4.3.8}$$

where θ is the electron angle with respect to the beam axis. Therefore,

$$\frac{\mathrm{d}\sigma}{\mathrm{d}p_{\mathrm{T}}} = \frac{\mathrm{d}\sigma}{\mathrm{d}\cos\theta}\frac{\mathrm{d}\cos\theta}{\mathrm{d}p_{\mathrm{T}}} \approx \frac{\mathrm{d}\sigma}{\mathrm{d}\cos\theta}\left[1 - 4p_{\mathrm{T}}^2/M_W^2\right]^{-\frac{1}{2}}. \tag{4.3.9}$$

Another often used and related variable is the so-called *transverse mass*:

$$m_{\mathrm{T}} = \sqrt{2p_{\mathrm{T}}^e p_{\mathrm{T}}^{\nu}(1-\cos\phi_{e\nu})}, \tag{4.3.10}$$

where p_{T}^{ν} is the transverse momentum of the undetected neutrino, reconstructed as the missing transverse momentum. The so-called *Jacobian* peak (see Fig. 4.25) thus induced in the p_{T} or m_{T} distribution allows for a fairly precise determination of the mass and width of the $W^{\pm}$ (via a full parton-model analysis); the present-day values are (see PDG-2016 – Patrignani *et al.*, 2016)

$$m_W = 80.399 \pm 0.023 \text{ GeV} \quad \text{and} \quad \Gamma_W = 2.085 \pm 0.042 \text{ GeV}. \tag{4.3.11}$$

We see that the precision is roughly twenty times poorer with respect to the LEP measurements of the Z^0 parameters. Note that the CERN experiments also published data on observation of the Z^0 (UA1: Arnison *et al.*, 1983b and UA2: Bagnaia *et al.*, 1983).

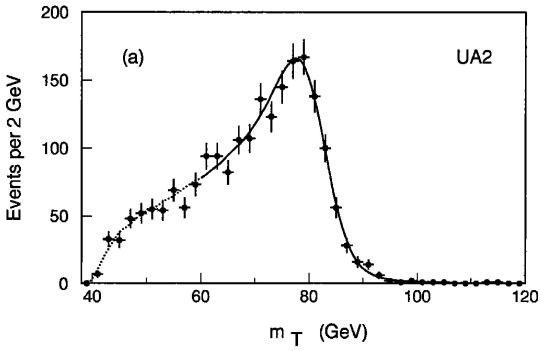


Figure 4.25: The Jacobian peak in the transverse-mass distribution of the electron or positron emitted in the $W^{\pm}$ decay, as seen in the UA2 experiment at CERN. The figure is taken from Alitti *et al.* (1992).

4.3.2 Discovery of the *t* quark

The Higgs boson aside (which we shall discuss later), the last major discovery made in a high-energy hadron collider, indeed in any collider, is that of the *top* quark. This was achieved at Fermilab in 1995 by two experiments: CDF (Abe *et al.*, 1995) and DØ (Abachi *et al.*, 1995). The basic process observed is

$$\begin{array}{rcl} p + \bar{p} & \to & t + \bar{t} \\ & & \quad \big\vert \quad \big\vert \hookrightarrow W^- \bar{b} \\ & & \quad \big\vert \hookrightarrow W^+ b. \end{array} \tag{4.3.12}$$

Of course, both the W's and b's also decay (either leptonically or hadronically) rather rapidly—the cleaner leptonic channels are used here for the $W^{\pm}$ while vertex tracking allows full reconstruction of the semi-leptonic b decays. In such a way, for the initial analyses, CDF collected 19 events and DØ 17 events. Combining all possible production modes (*i.e.* quark–antiquark annihilation and gluon–gluon fusion), the cross-section is estimated to be of order 5 pb at Tevatron energies and for a top-quark mass around 175 GeV, as determined from the reconstructed top-quark mass spectrum (see Fig. 4.26).

These two experiments continued to collect data over a long period and also performed analyses aimed at detecting the weak bosons, in particular the $W^{\pm}$. The current values are (see PDG-2016 – Patrignani *et al.*, 2016):

$$m_t = 173.5 \pm 0.6 \pm 0.8 \text{ GeV} \qquad \text{(direct measurement by CDF and DØ)}, \tag{4.3.13a}$$

$$= 178.1^{+10.4}_{-\ 8.3} \text{ GeV} \qquad \text{(indirect extraction from global SM fits)}, \tag{4.3.13b}$$

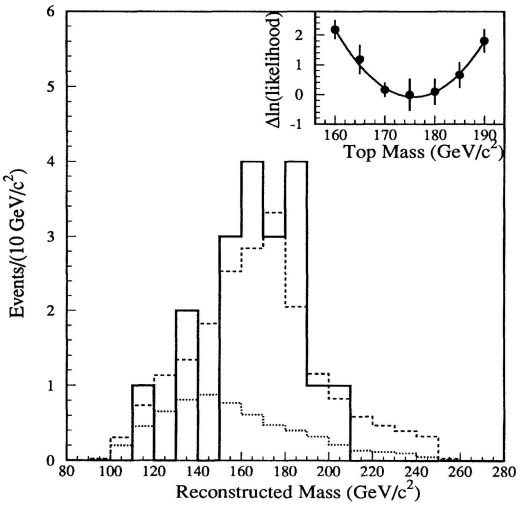


Figure 4.26: The reconstructed top-quark mass spectrum as seen in the CDF experiment at Fermilab. The figure is taken from Abe *et al.* (1995).

$$= 160.0^{+4.0}_{-4.3}\ \text{GeV}. \qquad \text{(deduced from the observed cross-section).} \tag{4.3.13c}$$

The second number is obtained from the precision fits to high-statistics data of various standard-model parameters (*e.g.* $\sin\theta_{\rm W}$), in which the top quark plays a role via mass-dependent, higher-order corrections.

Now, as already commented, the top quark decays so rapidly that there is no chance of forming top mesons (Bigi *et al.*, 1986). The width is measured to be

$$\Gamma_t = 1.99^{+0.69}_{-0.55}\ \text{GeV}, \tag{4.3.14a}$$

while theoretical calculation in the SM gives

$$\Gamma_t \simeq \frac{G_{\rm F} m_t^3}{8\sqrt{2}\,\pi}\left(1-\frac{M_W^2}{m_t^2}\right)^2\left(1+2\frac{M_W^2}{m_t^2}\right)\left[1-\frac{2\alpha_s}{3\pi}\left(\frac{2\pi^2}{3}-\frac{5}{2}\right)\right]. \tag{4.3.14b}$$

The theoretical value, obtained using the measured masses and strong coupling constant, is approximately 1 GeV, in reasonable agreement with experiment. Note that the first-order QCD correction is about 10 %; the second-order correction is also known. Such widths are equivalent to a lifetime of order 10^{-24} s.

4.3.3 The search for quark–gluon plasma

We have seen that phenomenologically the strong interaction confines quarks and gluons, presumably by means of an effective string-like long-distance potential. However, at short distances (or high energies) the quarks and gluons behave as though they were free and not subject to a confining potential. This picture has led to the idea that at very high energy *and* parton density QCD should undergo a transition to a sort of *plasma* phase, of which the principal consequence or feature would be *deconfinement*. That is, there should be no identifiable hadronic states, but rather effectively free quarks and gluons.

Such a phase presumably existed in the very early moments after the big bang and maybe even occurs inside some very dense neutron stars. One might hope to recreate it in the laboratory by generating very high energy and particle densities in high-energy heavy nucleus–nucleus collisions. To achieve this, the Relativistic Heavy-Ion Collider (RHIC) has been built at Brookhaven (USA), in which heavy ions (primarily gold) at around 100 GeV per nucleon collide. The LHC programme at CERN also includes a large fraction of heavy-ion physics.

In this new phase of QCD matter, owing to the effective deconfinement, it is believed that, for example, the strange quark should no longer be suppressed and should exist in roughly equal quantities with the other two light quarks. The signal that is sought then is a sudden and marked enhancement of strange-particle production. To some extent, this is indeed seen at RHIC, but there is still some debate as to the significance of the results.

4.4 Bibliography

Abachi, S. *et al.*, DØ Collab. (1995), *Phys. Rev. Lett.* **74**, 2632.

Abe, F. *et al.*, CDF Collab. (1995), *Phys. Rev. Lett.* **74**, 2626.

Alitti, J. *et al.*, UA2 Collab. (1992), *Phys. Lett.* **B276**, 354.

Anderson, C.D. (1933a), *Phys. Rev.* **43**, 491.

Anderson, C.D. (1933b), *Science* **77**, 432; *Phys. Rev.* 491.

Anderson, C. and Neddermeyer, S.H. (1936), *Phys. Rev.* **50**, 263; Neddermeyer, S.H. and Anderson, C.D., *Phys. Rev.* **51**, 884.

Andersson, B., Gustafson, G., Ingelman, G. and Sjöstrand, T. (1983), *Phys. Rep.* **97**, 31.

Andrews, D. *et al.* (1980), *Phys. Rev. Lett.* **44**, 1108.

Arnaudon, L. *et al.* (1995), *Nucl. Instrum. and Meth.* **A357**, 249.

Arnison, G. *et al.*, UA1 Collab. (1983a), *Phys. Lett.* **B122**, 103.

Arnison, G. *et al.*, UA1 Collab. (1983b), *Phys. Lett.* **B126**, 398.

Aubert, J.J. *et al.* (1974), *Phys. Rev. Lett.* **33**, 1404.

Augustin, J.-E. *et al.* (1974), *Phys. Rev. Lett.* **33**, 1406.

Bacci, C. *et al.* (1974), *Phys. Rev. Lett.* **33**, 1408; *erratum*, *ibid.* 1649.

Bagnaia, P. *et al.*, UA2 Collab. (1983), *Phys. Lett.* **B129**, 130.

Banner, M. *et al.*, UA2 Collab. (1983), *Phys. Lett.* **B122**, 476.

Bartel, W. *et al.*, JADE Collab. (1983), *Z. Phys.* **C21**, 37.

Bigi, I.I.Y., Dokshitzer, Yu.L., Khoze, V.A., Kühn, J.H. and Zerwas, P.M. (1986), *Phys. Lett.* **B181**, 157.

Blackett, P.M.S. and Occhialini, G.P.S. (1933), *Proc. Royal Soc. (London)* **A139**, 699.

Breit, G. and Wigner, E.P. (1936), *Phys. Rev.* **49**, 519.

Conversi, M., Pancini, E. and Piccioni, O. (1947), *Phys. Rev.* **71**, 209; *see also*, Sigurgeirsson, T. and Yamakawa, K.A., *ibid.* 319.

Dirac, P.A.M. (1928), *Proc. Royal Soc. (London)* **A117**, 610.

Dirac, P.A.M. (1930), *Proc. Royal Soc. (London)* **A126**, 360.

Herb, S.W. *et al.* (1977), *Phys. Rev. Lett.* **39**, 252.

Hoyer, P., Osland, P., Sander, H.G., Walsh, T.F. and Zerwas, P.M. (1979), *Nucl. Phys.* **B161**, 349.

Iizuka, J., Okada, K. and Shito, O. (1966), *Prog. Theor. Phys.* **35**, 1061.

Innes, W.R. *et al.* (1977), *Phys. Rev. Lett.* **39**, 1240; *erratum*, *ibid.* 1640.

Marchesini, G. and Webber, B.R. (1984), *Nucl. Phys.* **B238**, 1.

Okubo, S. (1963), *Phys. Lett.* **5**, 165.

Patrignani, C. *et al.*, Particle Data Group (2016), *Chin. Phys.* **C40** [10], 100001.

Perl, M.L. *et al.* (1975), *Phys. Rev. Lett.* **35**, 1489.

Rochester, G.D. and Butler, C.C. (1947), *Nature* **160**, 855.

Street, J.C. and Stevenson, E.C. (1937), *Phys. Rev.* **51**, 1005; **52**, 1003.

Yukawa, H. (1935), *Proc. Phys. Math. Soc. Jap.* **17**, 48.

Zweig, G. (1964), CERN preprint 8419/TH.412.

Chapter 5

The Standard Model and Beyond

In this final chapter we shall deal with the so-called *standard model* (SM) of particle physics. In particular, we shall turn our attention to more dynamical questions and examine the role of interactions. A central issue here will be the generation of mass. Various theoretical considerations lead us to prefer theories in which no masses are *explicitly* present. We find the gauge principle (requiring massless gauge bosons) to be immensely valuable and, indeed, all known phenomena may be described by theories taking the basic form of a QED-like interaction. However, most, if not all, particles have mass and in particular, unlike the photon, the $W^{\pm}$ and Z^0 are very heavy. As we shall see, the solution is to *induce* effective masses via interaction energies and we shall find, moreover, that symmetries and their breaking play an important role here too.

5.1 Fundamental forces and particles

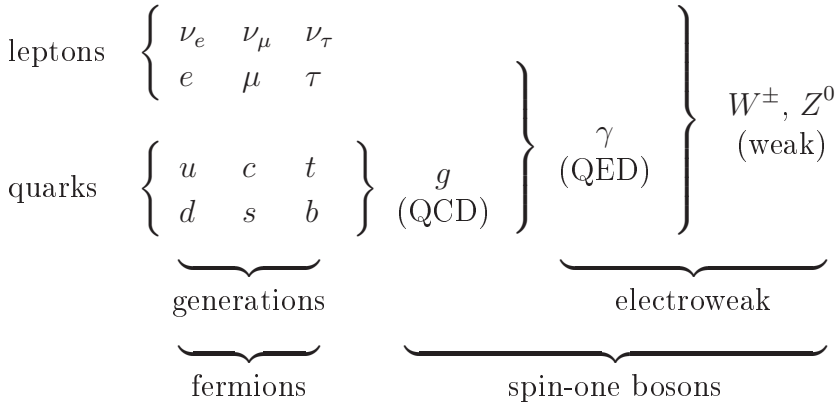
5.1.1 The table of forces and particles

The building blocks we now have in our hands comprise a number of quarks and leptons, interacting via the exchange of various types of spin-one bosons. Leaving aside gravity and not yet wishing to comment on the special case of the Higgs boson, these are grouped according to type and family (or generation) in Table 5.1.

5.1.2 The need for weak-boson masses

A major problem in theoretical physics is precisely how the $W^{\pm}$ and Z^0 might acquire a non-zero mass while remaining true gauge bosons. Let us first of all examine the reasons for requiring them to have a large mass.

Table 5.1: The elementary matter and force fields of the standard model; the only fields missing from this table are the spin-zero Higgs boson and the spin-two graviton.



Unitarity violation in Fermi theory

Fermi theory describes all known low-energy weak-interaction phenomena very well and to high precision. However, once we have the universal four-point interaction introduced by Fermi, we can imagine many new processes that have yet to be observed or studied experimentally. For example, the following process (improbable as it may appear and not easy to study experimentally) becomes possible:

$$\nu_\mu + e^- \;\rightarrow\; \mu^- + \nu_e. \tag{5.1.1}$$

The theory also determines precisely how to calculate the cross-section. If we do this for a sufficiently high energy so that in comparison we may neglect all particle masses, we obtain (in natural units,* $\hbar = 1 = c$)

$$\sigma_{\rm tot}(\nu_\mu e^- \;\rightarrow\; \mu^- \nu_e) \;\approx\; \frac{G_{\rm F}^2 E_{\rm CM}^2}{\pi}; \tag{5.1.2}$$

that is, a cross-section growing with energy. In fact, such behaviour might have been anticipated on dimensional grounds simply by using the knowledge that the Fermi coupling constant $G_{\rm F} \simeq 1.166 \times 10^{-6}\,{\rm GeV}^{-2}$ has dimensions E^{-2} (again in natural units). A cross-section has dimensions of an area, which in natural units is also E^{-2}. Therefore, to compensate the two powers of $G_{\rm F}$, we need two powers of E. It is obvious that such a behaviour must sooner or later violate unitarity; that is, above some critical energy the probability of such an interaction will exceed unity.

* Recall that, since $(\hbar c)^2 = 0.39\,{\rm mb\,GeV}^2$, the conversion is performed by multiplying an expression for a cross-section in ${\rm GeV}^{-2}$ by 0.39 to obtain the answer in mb.

The statement may be made more quantitative and even more stringent by considering the partial-wave expansion for the total cross-section:

$$\sigma_{\rm tot} \;=\; \frac{4\pi}{k^2}\sum_{\ell}(2\ell+1)\sin^2\delta_\ell, \tag{5.1.3}$$

where k is the projectile momentum, ℓ is the angular momentum of the individual partial wave and δ_ℓ is the associated phase-shift, which depends on the precise details of the theory. This may be applied to the above case by exploiting one simple property of the Fermi interaction: it is point-like, which implies vanishing impact parameter and therefore zero orbital angular momentum. One thus only needs to consider $\ell=0$, or s-wave, which leads to the following limiting behaviour:

$$\sigma_{\rm tot} \;=\; \sigma_0 \;\leq\; \frac{4\pi}{k^2} \;\to\; \frac{4\pi}{E_{\rm CM}^2}. \tag{5.1.4}$$

Equality of the two expressions (5.1.2) and (5.1.4) for $\sigma_{\rm tot}$ provides an upper limit to $E_{\rm CM}$ for which Fermi theory can be valid—beyond this energy partial-wave unitarity is violated. Inserting the relevant numbers leads to

$$E_{\rm CM}^{\rm max} \;\approx\; 300\ \text{GeV}. \tag{5.1.5}$$

In other words, before this energy is reached, some new physics must take over that is evidently not visible at lower energies but that tames the unbounded growth of the cross-section.

Intermediate vector bosons

Comparing such behaviour with that of QED, one sees a hint of how it might be cured. The equivalent interaction in QED is not a four-fermion contact interaction but rather a pair of fermion–photon vertices connected via a photon propagator. An indicator of the correct behaviour is the fact that the QED coupling constant is dimensionless. The dimensionality corresponding to $G_{\rm F}$ is provided by the propagator, which introduces the sought-after high-energy suppression. Diagrammatically then, one might be tempted to make the substitution depicted in Fig. 5.1.

In terms of couplings and propagators, this translates into

$$G_{\rm F} \;\to\; \frac{g^2}{|q^2-M^2|}, \tag{5.1.6}$$

where g is some new *dimensionless* coupling and we have allowed for a mass M

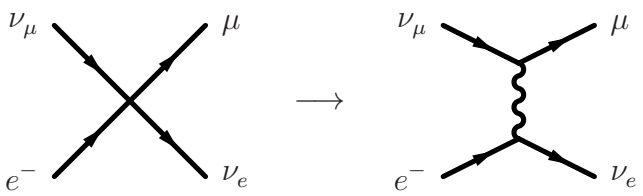


Figure 5.1: A possible cure to the divergent behaviour of Fermi theory via substitution of the four-point coupling with a (charged) boson propagator *à la* QED.

in the new propagator. At very high energies ($|q^2| \gg M^2$), where help is needed, M^2 may be neglected and thus an asymptotic behaviour is attained such as that in QED, *i.e.* acceptable. For low-energy interactions, where $|q^2| \ll M^2$, q^2 may be neglected in the propagator and thus the standard behaviour of Fermi theory is recovered. Moreover, inverting the relation at low energies, one finds

$$g \sim M\sqrt{G}_{\mathrm{F}}. \tag{5.1.7}$$

Now, we have already seen that M cannot be more than about 300 GeV if it is to save unitarity. On the other hand, it cannot be significantly smaller, otherwise the propagator energy dependence would spoil the accurate predictions of Fermi theory. This suggests a value for g that is not so very different from the value of e, the QED coupling. It is reasonable then to speculate that the new theory has indeed a similar coupling to QED and thus M should be around, say, 100 GeV, but we shall justify this a little better shortly.

Charge conservation at the vertices requires that, together with a non-zero mass, the new boson be charged and we should thus introduce something like $W^\pm$. Moreover, if we are to reproduce the $V-A$ interaction associated with Fermi theory, then it must be a spin-one particle, just as the photon. Now, while photon vertices only transfer energy and momentum, interactions involving $W^\pm$ must also change the nature of the fermion. For example, a u quark is converted into a d quark and *vice versa*. We are thus faced with the choice of either continuing to speak of isospin violation or to admit that the new bosons also carry isospin. This latter is, of course, much more attractive. Indeed, it is more natural to include the leptons in such a picture and thus introduce *weak isospin*. To couple objects of isospin one-half we are forced to introduce bosons of isospin one. This implies a multiplicity three $(2I+1)$, that is, a triplet. Thus, just as in the pion case, we should expect a W^0 (a so-called *weak neutral current*) together with the two $W^\pm$.

Weak neutral currents

Now, the W^0 cannot be the photon since it must both have a large mass and interact with neutrinos. We should remark immediately, moreover, that the particle so introduced is not exactly the Z^0 already mentioned, which is in fact a mixture of W^0 and a new field B^0 (while γ is the corresponding orthogonal combination) due to the same mechanism designed to provide the new (gauge) bosons with a non-zero mass. We shall discuss this in more detail later.

The introduction of a new intermediate neutral boson suggests further possible new phenomenology: that of the so-called *weak neutral currents*. An immediate question arises: why had no experimental indication been observed earlier? The answer lies in *both* the similarity *and* the difference with respect to the photon. The neutrality of a W^0 (or Z^0) implies that almost everywhere it may be exchanged, a photon may be exchanged too. However, at low energies photon exchange is not suppressed by a large mass. Indeed, we can estimate the cross-section ratio for W^0 and photon exchange in any given low-energy process as follows:

$$\frac{\sigma(W^0)}{\sigma(\gamma)} \sim \left(\frac{1/(q^2-M^2)}{1/q^2}\right)^2 \approx \left(\frac{q^2}{M^2}\right)^2 \quad \text{for } |q^2| \ll M^2. \tag{5.1.8}$$

Since we expect $M \sim 100\,\mathrm{GeV}$, for energies of the order of a GeV the suppression is of order 10^{-8}. While for interference effects this may improve to the square-root and therefore 10^{-4}, to separate the two contributions, one would still require absolute cross-section measurements of unprecedented precision.

If normal W^0-exchange processes are swamped by photon exchange, then we should look for something not possible in QED. As noted earlier, the photon does not change the nature of matter particles while the $W^\pm$ evidently do. One might then hope that some process exists in which the W^0 also provokes a change of flavour. We are thus led back to the old question of eigenstates of different interactions. The electron and muon *etc.* are evidently eigenstates of QED, which is experimentally seen to be perfectly diagonal in these states. Note that the basis states of QED are precisely those we call mass eigenstates and which propagate in the laboratory. However, the natural eigenstates of the weak interaction are not necessarily the same, that is, they may be superpositions of these states. In other words, if we write a three-component vector describing the charged leptons e^-, μ^- and τ^- then there should be a unitary transformation U that takes us to the weak basis:

$$\psi_i^{\mathrm{W}} = U_{ij}\psi_j. \tag{5.1.9}$$

A similar consideration could naturally be made for the neutrinos, but it is rather useless in this connection since they do not interact with the photon.

Now, we have always described interactions in terms of currents; thus, the QED

(neutral) current is

$$J^{\mu}_{\rm QED} \;=\; e\bar{\psi}\,\gamma^{\mu}\,\mathbb{1}_{\rm flav}\,\psi, \tag{5.1.10}$$

where the unit matrix $\mathbb{1}_{\rm flav}$ just expresses the fact that the interaction is diagonal in this basis. The equivalent *weak* neutral current (*i.e.* mediated by the W^0 or rather Z^0 boson) would then be

$$J^{0\mu}_{\rm W} \;=\; g\bar{\psi}_{\rm W}(c_{\rm V}\gamma^{\mu} - c_{\rm A}\gamma^{\mu}\gamma_5)\,\mathbb{1}_{\rm flav}\,\psi_{\rm W}, \tag{5.1.11}$$

where we are now in the weak basis, the zero index indicates a neutral current and we also remind the reader that this interaction *probably* violates parity conservation, but for the purposes of this discussion the actual values of $c_{\rm V}$ and $c_{\rm A}$ are quite irrelevant. The fact that the components of $\bar{\psi}^{\rm W}$ are mixtures of the physical charged-lepton states could, in principle, then allow for some flavour-changing effect. However, let us rewrite the above current in the physical basis:

$$\begin{aligned} J^{0\mu}_{\rm W} \;&=\; g\overline{U\psi}\,(c_{\rm V}\gamma^{\mu} - c_{\rm A}\gamma^{\mu}\gamma_5)\,\mathbb{1}_{\rm flav}\,U\psi \\ &=\; g\bar{\psi}\,U^{\dagger}(c_{\rm V}\gamma^{\mu} - c_{\rm A}\gamma^{\mu}\gamma_5)\,\mathbb{1}_{\rm flav}\,U\,\psi \\ &=\; g\bar{\psi}\,(c_{\rm V}\gamma^{\mu} - c_{\rm A}\gamma^{\mu}\gamma_5)\,U^{\dagger}\,\mathbb{1}_{\rm flav}\,U\,\psi \\ &=\; g\bar{\psi}\,(c_{\rm V}\gamma^{\mu} - c_{\rm A}\gamma^{\mu}\gamma_5)\,\mathbb{1}_{\rm flav}\,\psi. \end{aligned} \tag{5.1.12}$$

We thus see that it remains perfectly diagonal; in effect, a GIM-like mechanism has protected this situation from the possible consequences of *flavour-changing neutral currents*. Indeed, no evidence of such phenomena is found.*

The experimental discovery of weak neutral currents

We must then find some process in which the photon cannot participate and yet which has a clear signal. This may be provided by the following interactions:

$$\nu_{\mu} + e^{-} \;\rightarrow\; \nu_{\mu} + e^{-} \quad \text{and} \quad \bar{\nu}_{\mu} + e^{-} \;\rightarrow\; \bar{\nu}_{\mu} + e^{-}. \tag{5.1.13}$$

The corresponding experiment was first performed at CERN in the PS and SPS rings by the Gargamelle collaboration (Hasert *et al.*, 1973, 1974). Beginning in the late sixties, the production of high-energy ($200-300\,$GeV) and high-intensity proton beams allowed the generation of high-energy and high-intensity secondary ν_{μ} and $\bar{\nu}_{\mu}$ beams.

The so-called narrow-band neutrino beam was produced via the system shown

* We note for completeness, however, that higher-order quantum corrections, so-called *penguin diagrams*, can and do lead to non-vanishing effectively flavour-changing neutral currents, but the effects are very small.

schematically in Fig. 5.2. A 400-GeV proton beam impinged on a beryllium target

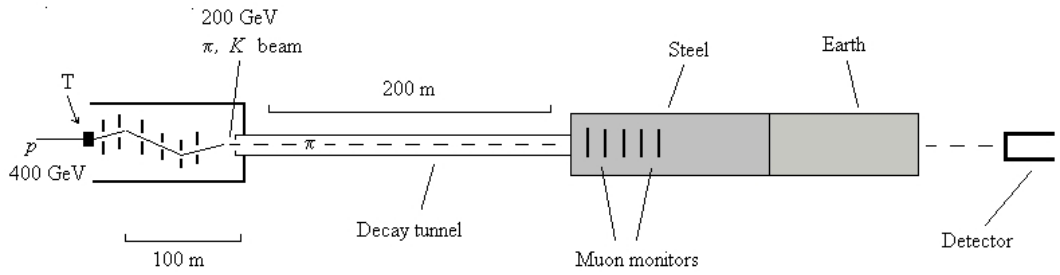


Figure 5.2: A schematic view of the narrow-band neutrino-beam experiment at CERN, in which the neutral current was first observed.

producing a large number of high-energy particles. By judicious choice of electric and magnetic fields, all but charged pions (although the beam did also contain some charged kaons) were filtered out over a length of around 100 m, selecting an energy of around 200 GeV. Inside a vacuum pipe of about 200 m these then decayed in-flight mainly into muons and muon neutrinos. A long shielding block of steel and the earth cleaned up the remaining muons and other stray particles, leaving a pure muon neutrino or antineutrino beam with a rather narrow spectrum. (There is, moreover, a strong correlation between the angle of emission and the energy of the neutrinos.) At the end of all this was Gargamelle, a giant bubble-chamber detector, designed principally for neutrino detection. With a diameter of nearly 2 m and length 4.8 m, it held nearly 12 cubic metres of freon (CF_3Br).

Exercise 5.1.1. *For the given kinematics, that is, a decaying charged-pion beam of energy* 200 GeV*, calculate the energy spread of the resulting neutrino beam.*

The importance of using muon neutrinos is simply that charged currents are excluded in their interactions with electrons (see Fig. 5.3). By requiring the final state to contain an *electron*, while the initial states contains a *muon* neutrino or antineutrino, guarantees the process to be neutral current. Indeed, the process depicted in Fig. 5.3a is impossible without a weak neutral current. The observation then of electrons scattered by the muon neutrino beam constituted a clear indication of the existence of weak neutral currents. Fig. 5.4 displays the first such event recorded by Gargamelle.

One can also look for deeply inelastic type processes where all that is seen to recoil is a purely hadronic system (Hasert *et al.*, 1974). The two possible reactions are then

$$\nu_\mu + N \to \nu_\mu + X \quad \text{and} \quad \bar{\nu}_\mu + N \to \bar{\nu}_\mu + X, \tag{5.1.14}$$

where N is a struck nucleon and X the recoiling hadronic system (see Fig. 5.5).

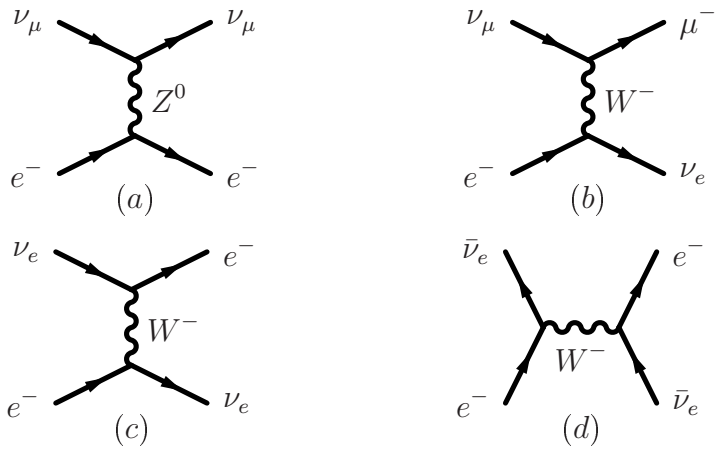


Figure 5.3: A comparison of the possible neutral- and charged-current interactions between muon and electron neutrinos and electrons. (*a*) A neutral-current event with an electron in the final state; (*b*) a charged-current event with instead a muon in the final state; (*c*) and (*d*) electron-neutrino initiated charged-current events with a final-state electron.

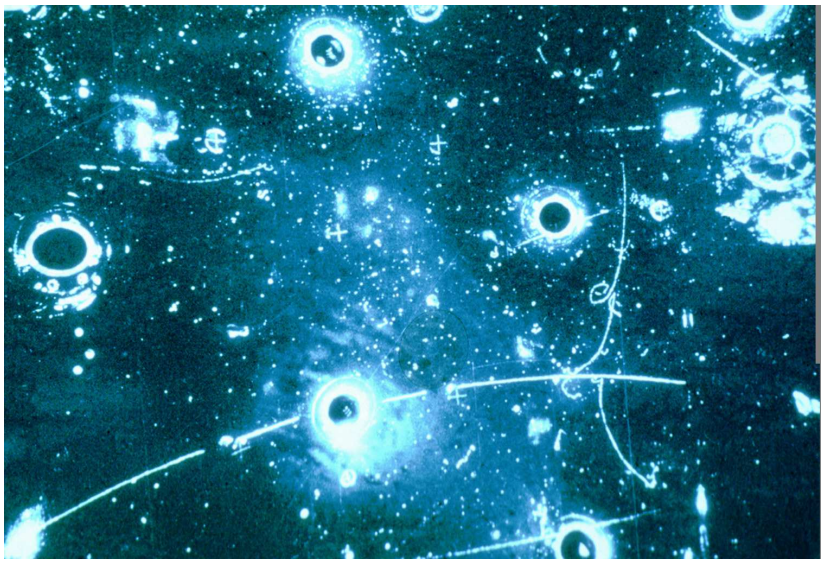


Figure 5.4: The first example of the neutral-current process $\bar{\nu}_\mu + e^- \to \bar{\nu}_\mu + e^-$. The electron is projected forward with an energy of 400 MeV at an angle of $1.5 \pm 1.5^\circ$ to the beam, which enters from the top.

Other evidence of weak neutral currents

Nuclear β-decay proceeds via charged-current exchange and is easily studied. On the other hand, since the neutral currents do not change flavour they cannot give

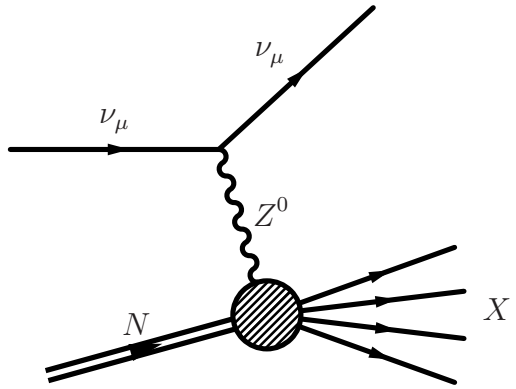


Figure 5.5: Deeply inelastic neutrino–nucleon scattering via neutral-current exchange.

rise to decays. Thus, it is not possible to detect their *direct* role in weak nuclear transitions. However, the presence of a parity-violating interaction in competition with the usual Coulomb potential leads to bound-state nuclear wave-functions that are no longer pure eigenstates of parity. This in turn allows the observation of *indirect* parity violation in, for example, electromagnetic or strong-interaction nuclear transitions.

Three such examples are the following:

$$\begin{aligned} {}^{16}\mathrm{O}^* &\rightarrow {}^{12}\mathrm{C} + \alpha \\ J^P &: \quad 2^- \qquad\quad 0^+ \quad 0^+. \end{aligned} \tag{5.1.15}$$

This process violates parity but proceeds via the strong interaction and should therefore be absolutely prohibited. However, neither the initial nor the final nuclear state is a pure parity eigenstate and the decay is experimentally observed with a width $\Gamma \sim 10^{-10}\,\mathrm{eV}$.

$$\begin{aligned} {}^{19}\mathrm{F}^* &\rightarrow {}^{19}\mathrm{F} + \gamma \\ J^P &: \quad \tfrac{1}{2}^- \qquad\quad \tfrac{1}{2}^- \quad 1^-. \end{aligned} \tag{5.1.16}$$

In this case, we have an electromagnetic transition, which again cannot directly violate P. Nevertheless, given the spin of the decaying ${}^{19}\mathrm{F}^*$, we may polarise the initial state and examine the up–down photon-momentum asymmetry. An asymmetry of $A_\gamma \sim 10^{-4}$ is found.

$${}^{181}\mathrm{Ta}^* \rightarrow {}^{181}\mathrm{Ta} + \gamma. \tag{5.1.17}$$

Here the initial state is not polarised, but the photon is experimentally found to emerge with a slight preference for being left handed: $P_\gamma \sim -4 \times 10^{-6}$. Finally,

we might remark that there are analogous, smaller but still measurable, effects in atomic physics.

5.1.3 Further unitarity problems

The problem of growing cross-sections with respect to unitarity limits is not, however, completely solved: the $\nu_\mu e^- \to \mu^- \nu_e$ process previously examined still grows logarithmically but would exceed the bounds only at extremely high energies. More problematically, the new states invoked to tame the Fermi-theory divergences naturally induce new processes, which although often rather exotic are nevertheless theoretically possible and thus too should not violate unitarity.

A particular example is $\nu_e \bar{\nu}_e$ annihilation into a W^+W^- pair (but there are many others). The first obvious contribution comes from the diagram in Fig. 5.6a. This diagram leads to a total cross-section that again grows as s. One solution

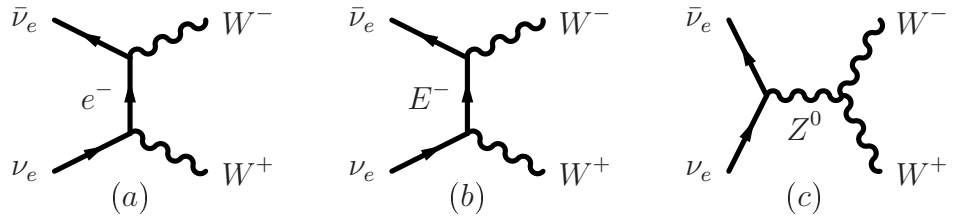


Figure 5.6: Various possible contributions to the annihilation process $\nu_e \bar{\nu}_e \to W^+W^-$: ($a$) the standard t-channel electron exchange graph, (b) a t-channel "heavy-electron" exchange graph, (c) the s-channel neutral-current annihilation graph.

might be to invoke a GIM-like cancellation with a new (heavy) lepton, as shown in Fig. 5.6b, having suitably chosen *ad hoc* couplings. No such heavy lepton has ever been detected though, whereas there are stringent upper limits on the mass it may have for the cancellation to still be effective. Of course, there are now also other possible, new, but more natural contributions that might be included, such as the neutral-current annihilation graph of Fig. 5.6c. This is evidently a different process, however, and has no specific relationship to the first, *unless there is some larger symmetry linking them.*

It turns out that any such attempt to patch up the cross-section growth-rate problem still runs foul of the requirement of renormalisability: when quantum corrections are evaluated, this type of theory is not generally renormalisable. Moreover, if we examine the $\nu_e \bar{\nu}_e \to W^+W^-$ cross-section in detail, we discover that it is precisely the "extra", longitudinal, component of the $W^\pm$ bosons that is responsible for the unbounded growth.

Now, had the gauge principle somehow been enforced, such contributions would have been absent. Indeed, QED suffers *none* of these problems. That is, QED is

renormalisable and has no longitudinal component. However, such highly desirable properties are a direct consequence of the masslessness of the photon; local gauge symmetry is broken by any *explicit* mass term. We recall briefly that gauge symmetry requires that the Lagrangian (and all *physical quantities*) be invariant under the following (local gauge) transformation:

$$A^\mu(x) \;\rightarrow\; A'^\mu(x) \;=\; A^\mu(x) + \partial^\mu \Lambda(x) \tag{5.1.18}$$

and

$$\psi(x) \;\rightarrow\; \psi'(x) \;=\; \mathrm{e}^{-\mathrm{i}e\Lambda(x)}\,\psi(x), \tag{5.1.19}$$

where $\Lambda(x)$ is any scalar function of x; $A^\mu(x)$ is the gauge field (for instance, the photon) while $\psi(x)$ is a spinor field representing, say, the electron.

The field equations governing the motion and interaction of such fields are derived from the following Lagrangian:

$$\mathcal{L} \;=\; \bar{\psi}\left(\mathrm{i}\gamma^\mu D_\mu - m\right)\psi - \tfrac{1}{4}F^{\mu\nu}F_{\mu\nu}, \tag{5.1.20}$$

where the covariant derivative is

$$D^\mu \;:=\; \partial^\mu + \mathrm{i}eA^\mu(x) \tag{5.1.21}$$

and the field-strength tensor $F^{\mu\nu}$ is

$$F^{\mu\nu} \;:=\; \partial^\mu A^\nu - \partial^\nu A^\mu. \tag{5.1.22}$$

It is straightforward to verify explicitly that the purely gauge term in the above Lagrangian leads to Maxwell's equations for the photon field. The term $-eA_\mu\bar{\psi}\gamma^\mu\psi$ leads to the standard fermion–photon interaction. A mass term for the gauge field would take the form $-\frac{1}{2}m^2A^\mu A_\mu$, which would plainly violate gauge invariance. We thus see that, to render the propagation of the vector fields $W^\pm$ massive, some other way of providing a mass must be found, via an interaction, for example.

5.2 Spontaneous symmetry breaking: the Higgs mechanism

In this section we shall briefly describe the mechanism by which it is possible to retain gauge symmetry (with all the consequent benefits) while allowing the gauge bosons to acquire significant *effective* masses. The concept of a broken symmetry, with the consequent generation of *massless* states was first discussed by Goldstone (1961). That a spontaneously broken symmetry (*i.e.* a symmetry of the Lagrangian broken by the vacuum) could circumvent the Goldstone theorem and thus

avoid the existence of massless states was proposed by Anderson (1963)* within the context of superconductivity and successively adapted to particle theory by Nambu and Jona-Lasinio (1961)—see footnote on p. 30. Spontaneous symmetry breaking in a gauge theory and the resulting mass generation, known as the Higgs (or Higgs–Kibble) mechanism, was first applied to quantum field theory by Higgs (1964), Englert and Brout (1964) and Guralnik *et al.* (1964).† Quigg (2007) has published an interesting review article on the general problem of mass generation.

5.2.1 A real scalar field

The simplest case is a purely scalar theory with a mass term of the "wrong sign"‡ and a φ^4 self-interaction (which guarantees stability):

$$\mathcal{L} \;=\; \tfrac{1}{2}(\partial^\mu\varphi)(\partial_\mu\varphi) + \tfrac{1}{2}m^2\varphi^2 - \tfrac{1}{4!}\lambda\varphi^4. \tag{5.2.1}$$

The corresponding potential

$$V(\varphi) \;=\; -\tfrac{1}{2}m^2\varphi^2 + \tfrac{1}{4!}\lambda\varphi^4 \tag{5.2.2}$$

has the form shown in Fig. 5.7, which is evidently symmetric under the discrete transformation $\varphi \to -\varphi$. However, the point $\varphi = 0$ is no longer the minimum while

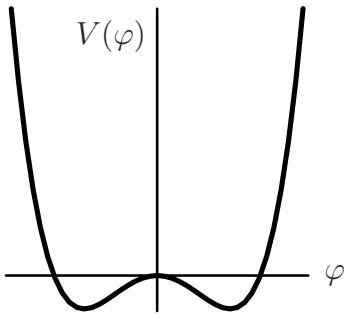


Figure 5.7: The shape of the potential for a single, real, scalar field in the case of a quadratic term with a negative sign.

* The 1977 Nobel Prize for Physics was awarded jointly to Philip Warren Anderson, Sir Nevill Francis Mott and John Hasbrouck van Vleck "for their fundamental theoretical investigations of the electronic structure of magnetic and disordered systems."

† The 2013 Nobel Prize for Physics was awarded jointly to François Englert and Peter W. Higgs "for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider"

‡ Note that it is the term that has the "wrong" sign and not the mass itself.

there are now two equivalent minima at $\varphi = \pm\sqrt{6m^2/\lambda}$. The vacuum or ground state of the system may be one or other, but not both at the same time, and will therefore break the original symmetry.

Since perturbation theory is to be performed around the minimum, we should shift to $\varphi' := \varphi - v$, where for the purposes of example and without loss of generality we shall take $v = +(6m^2/\lambda)^{1/2}$. The Lagrangian then takes the form

$$\mathcal{L} \;=\; \tfrac{1}{2}(\partial^\mu\varphi')(\partial_\mu\varphi') - m^2\varphi'^2 - \tfrac{1}{3!}\lambda\varphi'^3 - \tfrac{1}{4!}\lambda\varphi'^4, \tag{5.2.3}$$

where an irrelevant constant term has been eliminated. Thus, the true physical state of the theory has a mass $\sqrt{2}m$ and a cubic self-interaction has appeared.

5.2.2 A two-component scalar field

The case of a continuous symmetry is rather more interesting. Consider now a *complex* scalar field φ with the following Lagrangian

$$\mathcal{L} \;=\; (\partial^\mu\varphi^*)(\partial_\mu\varphi) + m^2\varphi^*\varphi - \tfrac{1}{2}\lambda(\varphi^*\varphi)^2. \tag{5.2.4}$$

The corresponding potential

$$V(\varphi,\varphi^*) \;=\; -m^2\varphi^*\varphi + \tfrac{1}{2}\lambda(\varphi^*\varphi)^2 \tag{5.2.5}$$

has the so-called Mexican-hat form shown in Fig. 5.8 and is symmetric under the (global gauge) transformation $\varphi \to \mathrm{e}^{\mathrm{i}\phi}\varphi$. The possible vacuum states now belong to

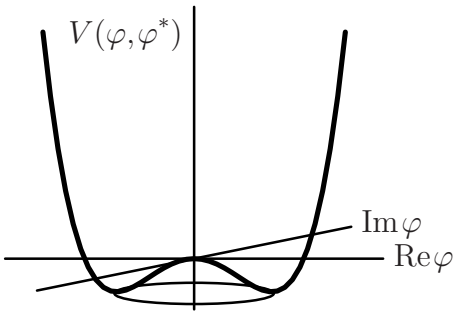


Figure 5.8: The so-called Mexican-hat form of the potential in the case of a complex scalar field with a negative quadratic term; it possesses a cylindrical symmetry about the vertical axis.

a continuum, corresponding to the variable $\phi \in [0,\pi]$ in the above transformation. Let us choose, without loss of generality, the ground state as $\varphi = v = (m^2/\lambda)^{1/2}$ and again make the shift to $\varphi' := \varphi - v$, with v real. It is now convenient to re-express

the fields in terms of two real fields $\varphi_{1,2}$:

$$\varphi' \;=\; \tfrac{1}{\sqrt{2}}(\varphi_1 + \mathrm{i}\varphi_2). \tag{5.2.6}$$

The Lagrangian then takes the form

$$\begin{aligned}\mathcal{L} \;=\; \tfrac{1}{2}(\partial^\mu\varphi_1)(\partial_\mu\varphi_1) + \tfrac{1}{2}(\partial^\mu\varphi_2)(\partial_\mu\varphi_2) - m^2\varphi_1^2 \\ - \lambda\upsilon\varphi_1(\varphi_1^2+\varphi_2^2) - \tfrac{1}{4}\lambda(\varphi_1^2+\varphi_2^2)^2.\end{aligned} \tag{5.2.7}$$

In this case we see that, together with interaction terms of various degrees, there is just a single non-zero mass; that is, one state is left massless. It is easy to see why: the original continuous symmetry implies that there is always a direction (that of the symmetry) in which the derivative of the field is zero. This is the essence of the Goldstone theorem: for each broken continuous symmetry there is a massless field, known as the *Goldstone boson*. We should stress here that the theorem only applies to continuous symmetries and is thus only relevant to multidimensional symmetry subspaces.

5.2.3 The Higgs–Kibble mechanism

We now examine the case of spontaneous symmetry breaking in the presence of a *local* gauge symmetry. We shall find that this provides an exception to the Goldstone theorem, in that the massless bosons do not appear as true individual states of the theory, but are absorbed (or *eaten*) by the gauge bosons to provide the third components for the gauge fields, which in turn become massive. The important point to realise though is that the underlying gauge symmetry survives; the theory thus remains renormalisable and satisfies unitarity in the usual way.

Consider the simplest example of a single charged scalar field, minimally coupled to a gauge field. The Lagrangian then takes on the standard form

$$\mathcal{L} \;=\; (D^\mu\varphi^*)(D_\mu\varphi) + \mu^2\varphi^*\varphi - \lambda(\varphi^*\varphi)^2 - \tfrac{1}{4}F^{\mu\nu}F_{\mu\nu}, \tag{5.2.8}$$

where the covariant derivative D^μ is defined as

$$D^\mu\varphi(x) \;:=\; \big(\partial^\mu - \mathrm{i}eA^\mu(x)\big)\,\varphi(x) \tag{5.2.9a}$$

and

$$D^\mu\varphi^*(x) \;:=\; \big(\partial^\mu + \mathrm{i}eA^\mu(x)\big)\,\varphi^*(x). \tag{5.2.9b}$$

It is convenient to reparametrise the field φ exponentially in terms of two real fields η and ξ (the "radial" and "angular" components respectively):

$$\varphi(x) \;=\; \tfrac{1}{\sqrt{2}}\big(\upsilon + \eta(x)\big)\exp\big(\mathrm{i}\xi(x)/\upsilon\big), \tag{5.2.10}$$

with

$$\upsilon \;=\; \sqrt{\mu^2/\lambda}. \tag{5.2.11}$$

In the absence of the gauge field and the resultant coupling, the field ξ *would be* the massless Goldstone boson associated with the spontaneous breaking of a global U(1) symmetry. However, the presence of a gauge symmetry and corresponding fields induces mixing of the ξ and A^μ fields. To see this, let us rewrite the Lagrangian in terms of ξ and η:

$$\begin{aligned}\mathcal{L}(\xi,\eta,A^\mu) \;=\; &-\tfrac{1}{4}F^{\mu\nu}F_{\mu\nu} + \tfrac{1}{2}\partial^\mu\xi\partial_\mu\xi + \tfrac{1}{2}\partial^\mu\eta\,\partial_\mu\eta \\ &+ \tfrac{1}{2}e^2\upsilon^2A^\mu A_\mu - e\upsilon A_\mu\partial^\mu\xi - \mu^2\eta^2 \\ &+ \text{terms higher than quadratic order.}\end{aligned} \tag{5.2.12}$$

The higher-order terms describe various interactions and are of interest for the full phenomenology, but here we are only interested in the question of mass. At first sight, we appear to have the fields of the previous example: η a boson of mass $\sqrt{2}\mu$ and ξ a massless Goldstone boson, together now with a massive gauge field (since a term in $A^\mu A_\mu$ has been generated). However, the presence of the term in $A_\mu\partial^\mu\xi$ complicates matters: it directly mixes the ξ and A^μ fields, which cannot then be the true asymptotic or physical states of the theory, and we must therefore be a little more careful.

We may exploit the gauge invariance of the Lagrangian to apply the following gauge transformation, which effectively diagonalises the mass terms:

$$\varphi \;\to\; \varphi' \;=\; \exp[-\mathrm{i}\xi(x)/\upsilon]\,\varphi \;=\; \tfrac{1}{\sqrt{2}}(\upsilon+\eta) \tag{5.2.13a}$$

and

$$A_\mu \;\to\; A'_\mu \;=\; A_\mu - \frac{1}{e\upsilon}\,\partial_\mu\xi. \tag{5.2.13b}$$

With this, the Lagrangian becomes

$$\begin{aligned}\mathcal{L}(\xi,\eta,A^\mu) \;=\; &-\tfrac{1}{4}F'^{\mu\nu}F'_{\mu\nu} + \tfrac{1}{2}e^2\upsilon^2A'^\mu A'_\mu + \tfrac{1}{2}\partial^\mu\eta\,\partial_\mu\eta - \mu^2\eta^2 \\ &+ \tfrac{1}{2}e^2A'^\mu A'_\mu\eta(2\upsilon+\eta) - \lambda\upsilon\eta^3 - \tfrac{1}{4}\lambda\eta^4.\end{aligned} \tag{5.2.14}$$

The gauge field has thus acquired an effective mass $e\upsilon$ and the scalar field η has a mass $m=\sqrt{2}\mu$ while the field ξ has simply disappeared from the theory.

The physical interpretation should be rather obvious: a massive vector field necessarily has three degrees of freedom whereas the original massless gauge field only had two; the third is provided by the *would-be Goldstone boson* ξ, which, we say figuratively, has thus been 'eaten' by the gauge field.

As a by-product, the presence of a scalar field with non-vanishing vacuum

expectation value allows the generation of effective mass terms for the matter fields, which may then be initially defined as massless too. It is simply necessary to add a Yukawa-type coupling to each fermion for which a mass is desired:

$$\mathcal{L}_{\text{int}} \;=\; -g\,\varphi\,\bar{\psi}\psi. \tag{5.2.15}$$

Shifting the scalar field and rewriting it as above, we obtain

$$\mathcal{L}_{\text{int}} \;=\; -g(\upsilon+\varphi')\bar{\psi}\psi. \tag{5.2.16}$$

The first term in brackets is evidently none other than a mass term for the field ψ and we thus have

$$m_{\psi} \;=\; g\,\upsilon. \tag{5.2.17}$$

Since the coupling g is arbitrary, the mass is not determined. However, given the measured mass, the relation may be inverted to provide g in terms of the Higgs vacuum expectation value. Indeed, this relation tells us that the heavier fermions will have the strongest couplings.

5.2.4 The Glashow–Salam–Weinberg model

To describe the weak interaction correctly it is found necessary to include the theory of electromagnetism at the same time—although, of course, the photon remains rigorously massless. That is, we are led to the construction of a (*quasi*) *unified* model of the electromagnetic and weak interactions, or *electroweak theory* (Glashow, 1961; Salam, 1968; Weinberg, 1967).*

We start then with a weak-isospin triplet of massless spin-one bosons $W_{\mu}^{(1)}$, $W_{\mu}^{(2)}$ and $W_{\mu}^{(3)}$, where now the associated SU(2) symmetry is taken to be a *local gauge* symmetry, thus guaranteeing unitarity and renormalisability but requiring masslessness. In addition, we include a single (isoscalar) neutral gauge boson B_{μ}^{0}. It turns out that this is *not* to be associated directly with the photon and electric charge, but with the weak *hypercharge*. The W triplet may be more suggestively rewritten as

$$W_{\mu}^{\pm} \;=\; W_{\mu}^{(1)} \pm \mathrm{i}W_{\mu}^{(2)} \tag{5.2.18a}$$

and

$$W_{\mu}^{0} \;=\; W_{\mu}^{(3)}. \tag{5.2.18b}$$

* The 1979 Nobel Prize for Physics was awarded equally to Sheldon Glashow, Abdus Salam and Steven Weinberg for "their contributions to the theory of the unified weak and electromagnetic interaction between elementary particles, including, *inter alia*, the prediction of the weak neutral current."

The scalar system necessary for spontaneous symmetry breaking here consists of two doublets, which we may express as (*cf.* the $K^{\pm}$, K^0, $\overline{K}^0$ system)

$$\begin{pmatrix}\varphi_1 \pm \mathrm{i}\varphi_2\\ \varphi_3 \pm \mathrm{i}\varphi_4\end{pmatrix} \quad \text{or} \quad \begin{pmatrix}\varphi^+\\ \varphi^0\end{pmatrix} \text{ and } \begin{pmatrix}\bar{\varphi}^0\\ \varphi^-\end{pmatrix}. \tag{5.2.19}$$

The construction of the standard interaction part of the Lagrangian, via generalisation of the electromagnetic case, then leads to the following two terms:

$$\mathcal{L}_{\text{int}} \;=\; g\boldsymbol{J}^{\mu}_{\text{W}}\cdot\boldsymbol{W}_{\mu} + g' J^{\mu}_{\text{Y}} B_{\mu} + \text{Hermitian conjugate}, \tag{5.2.20}$$

to which we must add terms coupling the scalar and gauge fields. The two currents introduced here are $\boldsymbol{J}^{\mu}_{\text{W}}$ (weak) and J^{μ}_{Y} (hypercharge). The SU(2) symmetry of the Higgs potential is spontaneously broken in such a way that the two charged $W^{\pm}$ acquire a mass just as described in the preceding section while the case of the two neutral vector bosons is a little more complicated. The neutral fields that finally emerge as the physical degrees of freedom of the theory are the mutually orthogonal combinations

$$A_{\mu} \;=\; \frac{g' W^0_{\mu} + g B_{\mu}}{\sqrt{g^2 + g'^2}} \quad \text{(the photon)} \tag{5.2.21a}$$

and

$$Z_{\mu} \;=\; \frac{g W^0_{\mu} - g' B_{\mu}}{\sqrt{g^2 + g'^2}} \quad \text{(the weak neutral boson).} \tag{5.2.21b}$$

The two independent coupling constants g and g' thus play the role of a mixing angle here and it is therefore more convenient to introduce

$$\sin\theta_{\text{W}} \;:=\; \frac{g'}{\sqrt{g^2 + g'^2}} \quad \text{and} \quad \cos\theta_{\text{W}} \;:=\; \frac{g}{\sqrt{g^2 + g'^2}}. \tag{5.2.22}$$

In other words,

$$\tan\theta_{\text{W}} \;:=\; \frac{g'}{g}. \tag{5.2.23}$$

As to the scalar fields, the two charged scalars $\varphi^{\pm}$ are absorbed into the $W^{\pm}_{\mu}$ respectively to provide the longitudinal components, as before, while the combination $\frac{1}{\sqrt{2}}(\varphi^0 - \bar{\varphi}^0)$ is absorbed by the combination of W^0_{μ} and B_{μ} corresponding to Z_{μ}. This all leaves just one physical, massive, scalar field:

$$H^0 \;:=\; \tfrac{1}{\sqrt{2}}(\varphi^0 + \bar{\varphi}^0), \tag{5.2.24}$$

which is precisely the object known as the *Higgs boson*.

Using the fact that hypercharge $Y = Q - I_3$ (a factor 2 has been introduced to avoid spurious factors of $\frac{1}{2}$), so that

$$J^\mu_{\rm Y} \;=\; J^\mu_{\rm EM} - J^\mu_{(3)}, \tag{5.2.25}$$

we may now rewrite the above in terms of the asymptotic or physical fields. The original interaction Lagrangian

$$\begin{aligned}\mathcal{L}_{\rm int} \;&=\; g\left[J^\mu_{(1)}W^{(1)}_\mu + J^\mu_{(2)}W^{(2)}_\mu\right] + J^\mu_{(3)}\left[gW^{(3)}_\mu - g'B_\mu\right] + g'J^\mu_{\rm EM}B_\mu \\ &\qquad + \text{Hermitian conjugate}\end{aligned}$$

then becomes

$$\begin{aligned}&=\; \frac{g}{\sqrt{2}}\left[J^\mu_- W^+_\mu + J^\mu_+ W^-_\mu\right] + \frac{g}{\cos\theta_{\rm W}}\left[J^\mu_{(3)} - \sin^2\theta_{\rm W} J^\mu_{\rm EM}\right]Z_\mu + g\sin\theta_{\rm W}\, J^\mu_{\rm EM}A_\mu \\ &\qquad + \text{Hermitian conjugate.}\end{aligned} \tag{5.2.26}$$

The first bracketed term represents the charged-current interaction, the second the weak neutral current, which we see mixes the parity-conserving and maximally parity-violating couplings (we shall expand on this later), while the third is identified with the standard electromagnetic interaction and thus immediately leads to the relation

$$e \;=\; g\sin\theta_{\rm W}. \tag{5.2.27}$$

Leaving aside the fermion mass parameters, we see that the theory is determined by a very small number of physical constants. For example, if we take $G_{\rm F}$ and the electromagnetic coupling constant α as known, then we can predict both of the heavy-boson masses in terms of the same mixing parameter $\sin\theta_{\rm W}$:

$$M_W \;=\; \left(\frac{\sqrt{2}g^2}{8G_{\rm F}}\right)^{1/2} \;=\; \left(\frac{\sqrt{2}e^2}{8G_{\rm F}\sin^2\theta_{\rm W}}\right)^{1/2} \;=\; \frac{37.4}{\sin\theta_{\rm W}}\ \text{GeV} \tag{5.2.28}$$

and

$$M_Z \;=\; \frac{M_W}{\cos\theta_{\rm W}} \;=\; \frac{74.8}{\sin 2\theta_{\rm W}}\ \text{GeV}. \tag{5.2.29}$$

As we shall see, $\sin\theta_{\rm W}$ is also determined by various other independent physical quantities; the present world average value is (see PDG-2016 – Patrignani *et al.*, 2016)

$$\sin^2\theta_{\rm W} \;\simeq\; 0.231, \tag{5.2.30}$$

although care must be taken in comparison with, *e.g.*, the boson masses, as there are important quantum corrections to be taken into account, which are different for the various physical quantities.

Unfortunately, one important physical quantity is left entirely undetermined: the Higgs-boson mass m_{H}. This is because the vacuum expectation value of the Higgs field is undetermined. However, the decay width of the Higgs boson is only a function of its mass and may thus be calculated; one finds $\Gamma_{\mathrm{H}} \sim G_{\mathrm{F}} m_{\mathrm{H}}^3$. Since to have any chance of "seeing" the Higgs particle (*i.e.* as a Breit–Wigner resonance peak) its width should be less than its mass, this means that the mass should therefore be less than about $1/\sqrt{G_{\mathrm{F}}}$. A more accurate analysis in terms of partial-wave unitarity in WW scattering places an upper limit of about 1 TeV. Moreover, since the quantum corrections to various physical quantities and processes contain contributions depending on the mass, global fits to SM data can actually provide a window of acceptable masses, which at the 90 % CL was (see PDG-2016 – Patrignani *et al.*, 2016)

$$m_{\mathrm{H}} = 99^{+28}_{-23}\ \mathrm{GeV}. \tag{5.2.31}$$

Despite the high-precision data available, the limits are not very stringent owing to the weak (logarithmic) dependence. In Fig. 5.9 the constraints on the Higgs-

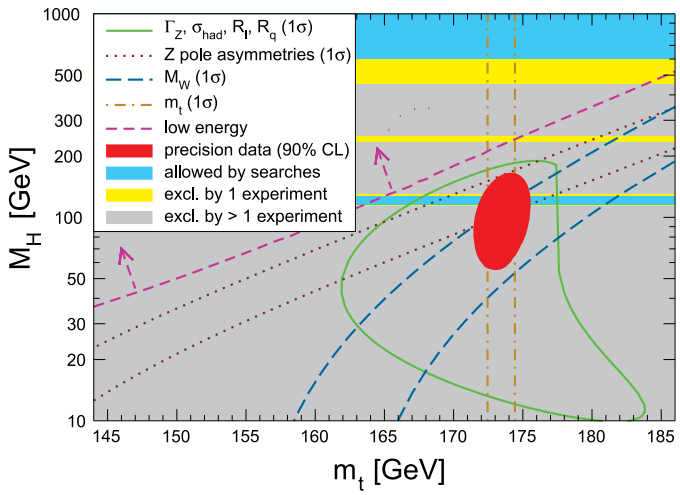


Figure 5.9: The one-standard-deviation limits for the Higgs-boson mass m_{H} as a function of the top mass m_t as deduced from various experimental measurements. The red ellipse marks the 90 % CL allowed region combining all the data. The figure is taken from PDG-2016 (Patrignani *et al.*, 2016).

boson mass m_{H} as a function of the top mass m_t are displayed as 90 % CL allowed regions. The central ellipse marks the 90 % CL allowed region combining all data.

EW couplings of the fermions

The charged $W^\pm$ bosons maximally violate P, inasmuch as they only interact with left-handed fermions, and so too does the extra $W^{(3)}$ or W^0, as it belongs to the same multiplet. On the other hand, the photon has equal left- and right-handed couplings and therefore the Z^0, being a mixture of the two, only partially violates P in its interactions with the charged fermions (both quarks and leptons), though still maximally with neutrinos.

The leptonic sector may be characterised as follows ($\ell = e,\, \mu,\, \tau$):

$$\psi_L = \begin{pmatrix} \nu_\ell \\ \ell \end{pmatrix}_L \qquad I = \tfrac{1}{2} \qquad I_3 = \begin{cases} +\frac{1}{2} \\ -\frac{1}{2} \end{cases} \qquad Q = \begin{cases} 0 \\ -1 \end{cases} \qquad Y = -\tfrac{1}{2}, \quad (5.2.32a)$$

$$\psi_R = \quad \ell_R \qquad I = 0 \qquad I_3 = 0 \qquad\qquad Q = \quad -1 \qquad Y = -1. \quad (5.2.32b)$$

Recall that here we use the definition $Y = Q - I_3$. Now, the Z^0 couples to the current $J^\mu_{(3)} - \sin^2\theta_\mathrm{W} J^\mu_\mathrm{EM}$ and we can therefore define left- and right-handed coupling constants:

$$g_L = I_3 - Q\sin^2\theta_\mathrm{W} \quad \text{and} \quad g_R = -Q\sin^2\theta_\mathrm{W}, \qquad (5.2.33)$$

where Q is the fermion charge in units of $|e|$. We may also define the vector and axial-vector couplings c_V and c_A, $g_{L/R} = \frac{1}{2}(c_\mathrm{V} \pm c_\mathrm{A})$; they are shown in Table 5.2 as functions of $\sin\theta_\mathrm{W}$ for the different fermion species. Note that, since $\sin^2\theta_\mathrm{W} \sim 1/4$,

Table 5.2: The Z^0 vector and axial-vector couplings c_V and c_A, $g_{L/R} = \frac{1}{2}(c_\mathrm{V} \pm c_\mathrm{A})$ as functions of $\sin\theta_\mathrm{W}$ for the various fermion species; ℓ are the charged leptons while U and D are up- and down-type quarks.

	c_V	c_A	g_L	g_R
ν_ℓ	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0
ℓ	$-\frac{1}{2}+2\sin^2\theta_\mathrm{W}$	$-\frac{1}{2}$	$-\frac{1}{2}+\sin^2\theta_\mathrm{W}$	$\sin^2\theta_\mathrm{W}$
U	$\frac{1}{2}-\frac{4}{3}\sin^2\theta_\mathrm{W}$	$\frac{1}{2}$	$\frac{1}{2}-\frac{2}{3}\sin^2\theta_\mathrm{W}$	$-\frac{2}{3}\sin^2\theta_\mathrm{W}$
D	$-\frac{1}{2}+\frac{2}{3}\sin^2\theta_\mathrm{W}$	$-\frac{1}{2}$	$-\frac{1}{2}+\frac{1}{3}\sin^2\theta_\mathrm{W}$	$\frac{1}{3}\sin^2\theta_\mathrm{W}$

the vector coupling almost vanishes for the charged leptons and the up-type quarks and these channels (coupling to the Z^0) thus have almost no parity violation while the neutrino channel remains maximally parity violating.

Neutrino scattering via neutral currents

We first examine the various possible neutrino scattering cross-sections in which the *charged* current intervenes (see Figs. 5.3c and d): at *low energies* the purely charged-current contributions are ($y := E'_e/E_\nu$ in the laboratory frame—initial electron at rest, $E'_e =$ final electron energy):

$$\frac{\mathrm{d}\sigma_{CC}(\nu_e e \,\to\, \nu_e e)}{\mathrm{d}y} \;=\; \frac{G_{\mathrm{F}}^2 s}{\pi} \qquad (LL \to LL,\; \Rightarrow\; J=0) \qquad (5.2.34\mathrm{a})$$

and

$$\frac{\mathrm{d}\sigma_{CC}(\bar\nu_e e \,\to\, \bar\nu_e e)}{\mathrm{d}y} \;=\; \frac{G_{\mathrm{F}}^2 s}{\pi}(1-y)^2 \qquad (RL \to RL,\; \Rightarrow\; J=1). \qquad (5.2.34\mathrm{b})$$

Exercise 5.2.1. *Neglecting the electron mass, show that* $y = \frac{1}{2}(1+\cos\theta_e)$*, where* θ_e *is the centre–of–mass electron scattering angle with respect to the neutrino beam direction. Thus, show that the* $J=1$ *cross-section above behaves as* $(1-\cos\theta_e)^2$*.*

Recall that for an intermediate vector particle one normally expects an angular dependence of the form $1+\cos^2\theta_e$. The difference here is that only one helicity state of the intermediate vector boson is available, corresponding to the amplitude $1-\cos\theta_e$. The other helicity *would* have provided an amplitude $1+\cos\theta_e$ and then the combination *would* have given

$$\tfrac{1}{2}(1-\cos\theta_e)^2 + \tfrac{1}{2}(1+\cos\theta_e)^2 \;=\; 1+\cos^2\theta_e, \qquad (5.2.35)$$

as expected.

These are to be compared with the corresponding *neutral*-current cross-sections (see Fig. 5.10). For $W^\pm$ we have $g_L = 1$ and $g_R = 0$ whereas for Z^0 exchange they

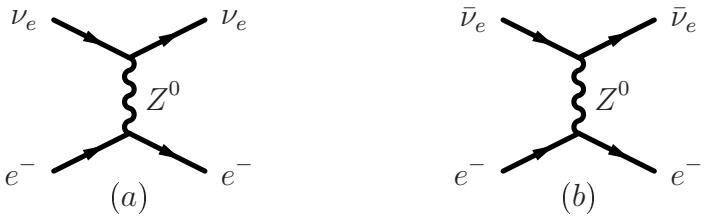


Figure 5.10: The two possible diagrams describing the neutral-current interaction between electrons and either (a) electron neutrinos or (b) electron anti-neutrinos.

are as listed above. Thus, for purely a neutral-current contribution, we have

$$\frac{\mathrm{d}\sigma_{NC}(\nu_e e \to \nu_e e)}{\mathrm{d}y} = \frac{G_{\mathrm{F}}^2 s}{\pi}\Big[g_L^2 + g_R^2(1-y)^2\Big], \tag{5.2.36a}$$

$$\frac{\mathrm{d}\sigma_{NC}(\bar{\nu}_e e \to \bar{\nu}_e e)}{\mathrm{d}y} = \frac{G_{\mathrm{F}}^2 s}{\pi}\Big[g_R^2 + g_L^2(1-y)^2\Big]. \tag{5.2.36b}$$

Note that while the electron may be either L or R here, the ν_e is only L and the $\bar{\nu}_e$ only R.

Now, for the scattering of muon neutrinos off electrons, only the neutral-current contributions survive and so

$$g_L = -\tfrac{1}{2} + \sin^2\theta_{\mathrm{W}} \quad \text{and} \quad g_R = \sin^2\theta_{\mathrm{W}}. \tag{5.2.37}$$

Adding the charged-current diagrams for the electron-neutrino case, we have

$$\nu_e : \begin{cases} g_L &= -\tfrac{1}{2} + \sin^2\theta_{\mathrm{W}} + 1 &=& \tfrac{1}{2} + \sin^2\theta_{\mathrm{W}}, \\ g_R &= \sin^2\theta_{\mathrm{W}} + 0 &=& \sin^2\theta_{\mathrm{W}}, \end{cases} \tag{5.2.38a}$$

$$\bar{\nu}_e : \quad \text{as above with } g_L \leftrightarrow g_R. \tag{5.2.38b}$$

The value deduced from the experimental comparison of these cross-sections in the case of muon neutrinos is $\sin^2\theta_{\mathrm{W}} = 0.2324 \pm 0.0083$, in good agreement with the measurement of the M_W/M_Z ratio. Note, as always, that due account must be made for important quantum corrections.

Polarisation asymmetries

A further independent method for extracting $\sin\theta_{\mathrm{W}}$ is provided by polarisation asymmetries measured in electron–nucleon scattering. The first such experiments were performed at SLAC in 1978 using a polarised electron beam of $16-18\,\mathrm{GeV}$, provided by the linear accelerator. The process studied was electron–deuteron DIS, with an unpolarised deuteron:

$$e^-_{L,R} + d \to e^- + X. \tag{5.2.39}$$

Together with the dominant QED photon-exchange diagram, there is also a weak neutral-current contribution coming from Z^0 exchange. Interference between the two allows a measurable parity-violating asymmetry

$$\mathcal{A} := \frac{\sigma_R - \sigma_L}{\sigma_R + \sigma_L}. \tag{5.2.40}$$

The two contributing amplitudes are

$$\mathcal{M}_{\rm EM} \sim \frac{e^2}{q^2} \quad \text{and} \quad \mathcal{M}_{\rm W} \sim G_{\rm F}. \tag{5.2.41}$$

In the denominator of the asymmetry the electromagnetic contribution dominates while in the numerator it cancels between σ_R and σ_L, leaving the interference term to dominate. The resulting asymmetry may thus be estimated as

$$\begin{aligned} \mathcal{A} \sim \frac{2\mathcal{M}_{\rm W}\mathcal{M}_{\rm EM}}{\mathcal{M}_{\rm EM}^2} &\sim \frac{2\,G_{\rm F}\,q^2}{e^2} \sim \frac{2\left(10^{-5}/m_p^2\right)q^2}{4\pi/137} \\ &\sim 10^{-4}\,q^2 \qquad (\text{for } q^2 \text{ in GeV}^2). \end{aligned} \tag{5.2.42}$$

More precisely, as a function of $y = E'_e/E_e$ (laboratory energies) we have

$$\mathcal{A} \;=\; -\frac{9G_{\rm F}q^2}{20\sqrt{2}\pi\alpha}\left[c_1 + c_2\,\frac{1-(1-y)^2}{1+(1-y)^2}\right], \tag{5.2.43}$$

where

$$c_1 \;=\; 1 - \tfrac{20}{9}\sin^2\theta_{\rm W} \quad \text{and} \quad c_1 \;=\; 1 - 4\sin^2\theta_{\rm W}. \tag{5.2.44}$$

The value obtained was $\sin^2\theta_{\rm W} = 0.22 \pm 0.02$, again, in good agreement with other determinations.

The Higgs boson

The final *prediction* of the Glashow–Salam–Weinberg model that we shall examine is the existence of the Higgs boson. The Higgs mechanism, as applied to the construction of the Glashow–Salam–Weinberg electroweak model, unavoidably leads to a neutral, massive, scalar boson H^0. The model also determines very precisely the form of the Higgs-particle interactions with the other fields in the theory, generating a number of trilinear and quadrilinear couplings, such as those in Fig. 5.11. It therefore couples to *all* other fields in the theory and to itself—it will even couple

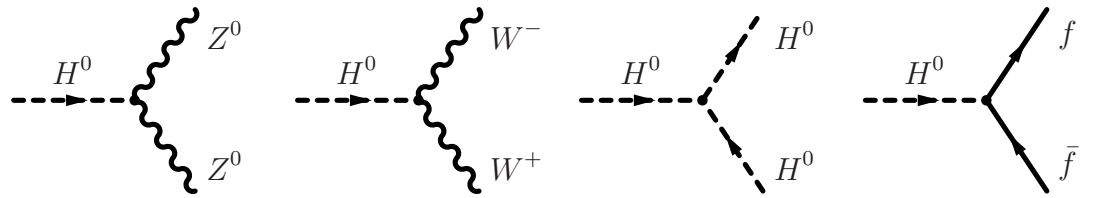


Figure 5.11: Examples of the Higgs trilinear couplings to gauge and matter fields in the standard electroweak model, f stands for any *massive* fermion.

(albeit very weakly) to any massive neutrino. At the first writing of these notes it was the one undetected particle in the SM and a major quest was thus to demonstrate its existence. The search may be approached in two mutually independent ways:

- its direct production and detection through the final decay-product system configuration for specific decay channels,
- its indirect contribution via quantum corrections to the various accurately measured electroweak processes and parameters, using combined fits.

Non-detection of the process $e^+e^- \to Z^0H^0$ at LEP placed a lower limit (see PDG-2016 – Patrignani *et al.*, 2016) on the mass:

$$m_{\rm H} > 114.4 \text{ GeV} \qquad (95 \text{ \% CL}). \tag{5.2.45}$$

On the other hand, indirect evidence from consideration of electroweak quantum corrections can actually place both upper and lower limits:

$$54 \text{ GeV} < m_{\rm H} < 219 \text{ GeV} \qquad (95 \text{ \% CL}), \tag{5.2.46}$$

with a central value of around 100 GeV, which is thus already effectively excluded by direct searches.

Direct searches at LEP were performed by checking the following two possible production processes:

$$\begin{aligned} e^+e^- \;\to\; Z^0 \;&\to\; H^0 + \ell^+ + \ell^-, \\ &\to\; H^0 + \nu_\ell + \bar{\nu}_\ell, \end{aligned} \tag{5.2.47}$$

where $\ell = e$, μ or τ, as shown in Fig. 5.12. We may calculate the branching ratio

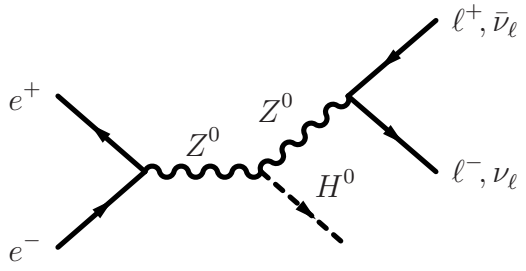


Figure 5.12: The simplest and most sensitive process for Higgs-particle production and detection at LEP ($\ell = e$, μ or τ).

for such a channel in Z^0 decay:

$$10^{-4} \gtrsim \frac{\Gamma(Z^0 \to H^0 \ell^+ \ell^-)}{\Gamma^{Z^0}_{\rm tot}} \gtrsim 3 \times 10^{-6} \quad \text{for } 10 \text{ GeV} \lesssim m_{\rm H} \lesssim 50 \text{ GeV} \tag{5.2.48}$$

and $B_\nu \simeq 2B_\ell$. The LEP I data thus led to a lower limit of about 60 GeV. LEP II raised the centre–of–mass energy to 200 GeV and was therefore able to search for the direct channel $e^+e^- \to Z^0 \to Z^0 H^0$. In this way a lower limit of about 114 GeV was obtained.

The LHC is, however, a proton–proton machine and the colliding quarks will evidently not have all of the laboratory 14 TeV available. Nevertheless, it should be possible to produce the Higgs particle directly ($pp \to H^0 + X$) via, for example, the process depicted in Fig. 5.13, W fusion. This is a good channel for the LHC,

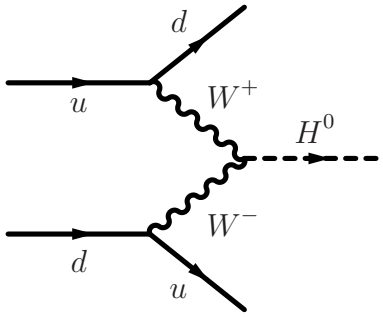


Figure 5.13: Production of the Higgs boson in proton–proton collisions via the so-called W^+W^- fusion process.

in which both beams are protons, as both the u and d are then valence quarks. Naturally, the Higgs boson must be detected through its decay products. If, as appeared evermore likely, $m_{\rm H} \geq 2m_{Z^0}$ then the two following decay channels are possible:

$$H^0 \to Z^0 + Z^0 \quad \text{and} \quad W^+ + W^-. \tag{5.2.49}$$

The two weak bosons must be detected via their decay products; in the case of W^+W^- two of the final-state particles will be neutrinos, which will escape detection. The simplest and cleanest signal for Higgs production (in this mass range) thus involves a final state of four charged leptons:

$$\begin{aligned} p + p \to\ & H^0 + X \\ & \hookrightarrow Z^0 + Z^0 \to \ell^+ + \ell^- + \ell^+ + \ell^-. \end{aligned} \tag{5.2.50}$$

The gold-plated channel is the four-muon final state. Such a process allows for Higgs searches in the range 200 GeV $\lesssim m_{\rm H} \lesssim$ 500 GeV. Of course, the mass may be

measured by reconstructing the invariant mass of the four final-state muons, for which we expect to see a classic BW distribution. The branching ratio here is

$$\frac{\Gamma(H^0 \rightarrow 4\ell)}{\Gamma(H^0 \rightarrow 2Z^0)} \simeq 4\ \%. \tag{5.2.51}$$

Note that for $m_{\rm H}$ very large, the width becomes comparable to the mass and the BW resonance shape is lost with the signal merging invisibly into the standard continuum background.

Should the Higgs be lighter than the threshold for double weak-boson production, the search becomes, perhaps surprisingly, rather more difficult. This is because we shall need to extract the signal from processes with hadronic final states, for example $H^0 \rightarrow b\bar{b}$. In this case the b quarks will give rise to a pair of jets and such signals risk being swamped by a dominating standard QCD jet-production background. A possible alternative in this case is the rare but very distinctive $H^0 \rightarrow \gamma\gamma$ channel depicted in Fig. 5.14. In the SM the branching ratio for this decay is about 10^{-3} and it is dominated by the two processes shown in Fig. 5.14. The two diagrams represent virtual-particle loops, in which the only

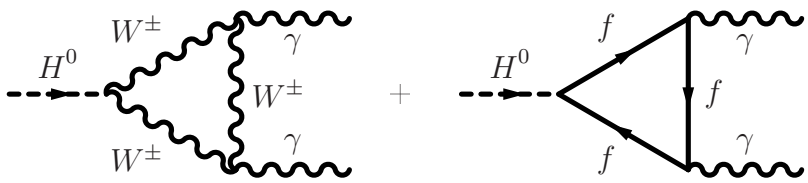


Figure 5.14: The probably dominant contributions to Higgs-boson decay into a photon pair; f may be any charged fermion (since it must couple to the photon).

requirement is a charge for the circulating field. Thus, both quarks and leptons may contribute, together with the charged weak bosons $W^\pm$.

Early in 2012 both ATLAS (Aad *et al.*, 2012a,b) and CMS (Chatrchyan *et al.*, 2012a,b) published papers on the search for the Higgs boson at the LHC. The results indicate an excess of events (with respect to background) in the region of $m_H \simeq 125\,\mathrm{GeV}^2$, at a combined level of $4.3\,\sigma$. At around the same time the two Tevatron experiments CDF and DØ (Aaltonen *et al.*, 2012) also announced combined results in agreement with the CERN data although statistically less significant.

5.2.5 The CKM matrix

Perhaps one of the richest areas to emerge in hadronic physics in the 90's is that of the CKM matrix. Recall that this matrix describes the unitary transforma-

tion between the asymptotic quark basis states and those of the weak interaction. Since we have, as yet, no accepted grand unified theory (GUT), the CKM matrix parametrises our ignorance of any possible link between the electroweak theory and the physics of the strong interaction (or QCD). As such, it is presently seen by many as an important window onto possible physics beyond the SM. First of all, it is the best candidate we have for the origin of CP violation, which should then be described by just a single imaginary phase appearing in various matrix elements. Secondly, and perhaps more importantly at this point, the property (or requirement) of unitarity can be tested experimentally.

One should note that the existence of a non-trivial (*i.e.* non-diagonal) mixing matrix also requires that the all various quarks have different masses (at least separately within the two classes of up and down types). While this is already experimentally verified, it also indicates that any measured CP-violating effects will be proportional to the mass(-squared) differences.

Before proceeding let us briefly examine the present status of the experimental determination of the CKM matrix. We should first define its elements:

$$V_{\mathrm{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} . \tag{5.2.52}$$

As we have shown, in its most general form, this matrix may be parametrised by three Euler angles and just one imaginary phase. The standard choice of representation is the following (Chau *et al.*, 1984):

$$V_{\mathrm{CKM}} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}\,\mathrm{e}^{-\mathrm{i}\delta_{CP}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}\,\mathrm{e}^{\mathrm{i}\delta_{CP}} & c_{12}c_{23} - s_{12}s_{23}s_{13}\,\mathrm{e}^{\mathrm{i}\delta_{CP}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}\,\mathrm{e}^{\mathrm{i}\delta_{CP}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}\,\mathrm{e}^{\mathrm{i}\delta_{CP}} & c_{23}c_{13} \end{pmatrix} , \tag{5.2.53}$$

where $s_{ij} = \sin\theta_{ij}$ and $c_{ij} = \cos\theta_{ij}$, while δ_{CP} is the single allowed phase, which may then be responsible for CP violation. Note also that, by suitable global rotation of both quark bases, all the angles may be taken to lie in the range $[0, \pi/2]$ so that all sines and cosines are non-negative.

If we now restrict consideration to the first two families (*i.e.* the first two rows and columns), then we just have the Cabibbo matrix, with θ_{12} being the Cabibbo mixing angle (we should set $\theta_{13} = 0 = \theta_{23}$ and thus the phase terms disappear). Experimentally, we have already noted that this angle is small; that is, the diagonal elements are near to unity and are much larger than those off-diagonal. Extending the discussion to the 3×3 case, this hierarchy continues and one finds that the

elements furthest from the diagonal are smallest; in other words,

$$\sin\theta_{12} \gg \sin\theta_{23} \gg \sin\theta_{13}. \tag{5.2.54}$$

Indeed, if we measure the scale of smallness of the near off-diagonal terms via a parameter λ, then unitarity implies that those further off-diagonal are order λ^3. This observation leads to an alternative parametrisation due to Wolfenstein (1983):

$$V_{\rm CKM} = \begin{pmatrix} 1-\lambda^2/2 & \lambda & A\lambda^3(\rho-\mathrm{i}\eta) \\ -\lambda & 1-\lambda^2/2 & A\lambda^2 \\ A\lambda^3(1-\rho-\mathrm{i}\eta) & -A\lambda^2 & 1 \end{pmatrix} + O(\lambda^4). \tag{5.2.55}$$

The parameter λ is then essentially $\sin\theta_{\rm C}$. The current best fits to the world experimental data for the moduli of the elements (see PDG-2016 – Patrignani *et al.*, 2016) give

$$V_{\rm CKM} = \begin{pmatrix} 0.97425 \pm 0.00022 & 0.2252 \pm 0.0009 & 0.00415 \pm 0.00049 \\ 0.230 \quad \pm 0.011 & 1.006 \quad \pm 0.023 & 0.0409 \quad \pm 0.0011 \\ 0.0084 \quad \pm 0.0006 & 0.0429 \pm 0.0026 & 0.89 \quad \pm 0.07 \end{pmatrix}. \tag{5.2.56}$$

There is clearly too little information as yet to provide a precise value for the last element; by imposing unitarity one finds that the magnitude is constrained to be

$$V_{tb} = 0.999146^{+0.000034}_{-0.000004}, \tag{5.2.57}$$

which is thus likely to be very near to unity.

The unitarity triangles

Let us begin with the question of unitarity; the equation is simple:

$$V^\dagger_{\rm CKM} V_{\rm CKM} = \mathbb{1}. \tag{5.2.58}$$

Since in the SM this is a 3×3 matrix equation, it actually represents nine equations or constraints. The three "diagonal" equations each have a left-hand side involving the square of one large component, which dominates the sum and the comparison with the large right-hand side. They are therefore rather difficult to test experimentally. In contrast, the six "off-diagonal" equations dilute the dominance of the large diagonal components and are therefore less critical. It can thus even be hoped, for example, that possible physics *beyond* the SM might be made manifest via the *non*-vanishing of these sums of products.

Now, the off-diagonal equations (each containing just three terms) may be expressed as triangles in the complex plane. The most commonly used triangle (being the most sensitive to CP-violation) is generated by the "d–b" product:

$$V_{ud}\,V_{ub}^* + V_{cd}\,V_{cb}^* + V_{td}\,V_{tb}^* \;=\; 0. \tag{5.2.59}$$

First of all, we take the best determined term $V_{cb}V_{cd}^*$ as a reference length and use it to rescale all three terms thus:

$$\frac{V_{ud}\,V_{ub}^*}{V_{cd}\,V_{cb}^*} + 1 + \frac{V_{td}\,V_{tb}^*}{V_{cd}\,V_{cb}^*} \;=\; 0. \tag{5.2.60}$$

We may then choose to place the middle term (now unity) along the positive real axis, running from the origin to the point (1,0). Considering the other two as complex numbers, the sum above then represents a triangle in the complex plane. We thus arrive at the geometric representation shown in Fig. 5.15. The

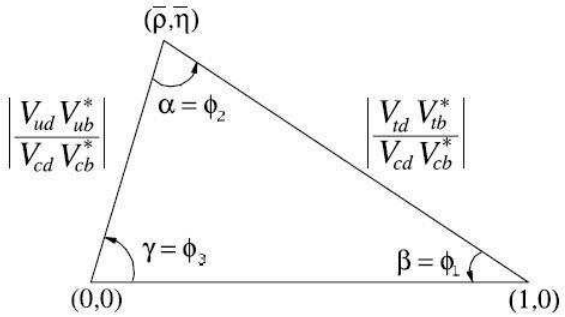


Figure 5.15: A schematic representation of the CKM-matrix unitarity triangle.

new variables $\bar{\rho}$ and $\bar{\eta}$ are equivalent to ρ and η up to corrections of order λ^2, which are thus order λ^4 corrections to the matrix elements themselves. The angles of the triangle are just the phases of the various ratios of elements:

$$\phi_1 \;=\; \beta = \arg\!\left(\frac{V_{cd}\,V_{cb}^*}{V_{td}\,V_{tb}^*}\right), \tag{5.2.61a}$$

$$\phi_2 \;=\; \alpha = \arg\!\left(\frac{V_{td}\,V_{tb}^*}{V_{ud}\,V_{ub}^*}\right), \tag{5.2.61b}$$

$$\phi_3 \;=\; \gamma = \arg\!\left(\frac{V_{ud}\,V_{ub}^*}{V_{cd}\,V_{cb}^*}\right). \tag{5.2.61c}$$

First and foremost, one should experimentally verify whether the three terms actually do form a non-trivial triangle (*i.e.* they do not collapse to a single line). This would then demonstrate that the CP-violating phase is indeed non-zero.

Secondly, one should note that the three angles are given by different combinations of matrix elements and are therefore experimentally independent. Unitarity of the matrix evidently requires that the sum of the angles be exactly 180° and that the two upper sides end at the same point; if this were found not to hold, then we would have a signal for physics beyond the SM. The most obvious case would be a fourth generation of fermions.

A further observation is that the six possible triangles so formed all have the same area. An obvious necessary condition for the CKM matrix to generate CP violation is that the area should be non-zero. Now, it is a fairly simple exercise in geometry to show that it is given by half the Jarlskog (1985) invariant J:

$$J \;=\; \mathrm{Im}[V_{ij}\,V_{kl}\,V_{il}^{*}\,V_{kj}^{*}]/\sum_{mn}\varepsilon_{ikm}\varepsilon_{jln}. \tag{5.2.62}$$

Note that this definition is entirely phase-convention independent and J is thus indeed an invariant. In terms of the general parametrisation of the CKM matrix given earlier, we have

$$J \;=\; c_{12}\,c_{23}\,c_{13}^{2}\,s_{12}\,s_{23}\,s_{13}\sin\delta_{CP}. \tag{5.2.63}$$

The present value (from global fits) for this parameter is (see PDG-2016 – Patrignani *et al.*, 2016)

$$J \;=\; 3.04^{+0.21}_{-0.20}\times 10^{-5}. \tag{5.2.64}$$

The fact that the Jarlskog invariant involves elements from *all three* generations is intimately related to the observation that for less than three generations there can be no CP violation at the level of the CKM matrix. In other words, if an experimental (physical) quantity depends on elements involving less than three generations then, by means of a suitable, unitary transformation of the matrix itself, any CP-violating phase could be rotated away. Conversely, this also implies that for an experimentally measurable quantity to be sensitive to CP violation all three generations must contribute to the process. For example, in the case of the K^0–$\overline{K}^0$ system a three-family GIM-like mechanism would actually cause all effects to cancel and it is the large mass of the t quark that partially deactivates the cancellation.

Measurement of the CKM-matrix elements

Let us briefly outline how the magnitudes of at least some of the CKM-matrix elements are determined experimentally. For up-to-date experimental values see PDG-2016 – Patrignani *et al.* (2016).

$|V_{ud}|$: The most precise determinations of $|V_{ud}|$ are provided by the so-called superallowed $0^+ \to 0^+$ nuclear beta decays. Recall that while (vector) Fermi transitions involve an $e\nu_e$ pair in a spin-zero state, (axial-vector) Gamow–Teller transitions produce a spin-one pair and cannot therefore contribute to the superallowed decays considered, which are thus purely vector transitions. The nine most precise determinations combine to lead to (PDG-2016 – Patrignani *et al.*, 2016)

$$|V_{ud}| \;=\; 0.97417 \pm 0.00021. \tag{5.2.65}$$

The error is dominated by theoretical uncertainties in the nuclear corrections.

The neutron lifetime measurement also affords a precise determination of $|V_{ud}|$. The theoretical uncertainties are very small here too but the extraction of the CKM element requires precise knowledge of the ratio between the axial-vector and vector couplings (g_A/g_V), which is measured to comparable precision via the decay angular distributions. Finally, the theoretically very clean charged-pion decay $\pi^+ \to \pi^0 e^+ \nu$ may also be used; however, present experimental precision is not yet competitive.

$|V_{us}|$: The magnitude of V_{us} is typically extracted either from semileptonic kaon decays or from the strangeness-changing semileptonic hyperon decays. Considerable experimental effort has been made in recent years with regard to the former. High-statistics measurement of $B(K^+ \to \pi^0 e^+ \nu)$ and a number of measurements of neutral-kaon branching ratios, form factors, and lifetime have been performed. Form-factor input is also required: the theoretical value $f^+(0) = 0.961 \pm 0.008$ is generally adopted. The kaon semileptonic decay rates then lead to

$$|V_{us}| \;=\; 0.2248 \pm 0.0006. \tag{5.2.66}$$

However, it must be said that the theoretical calculations of $f^+(0)$ differ by as much as 2 %, with quoted uncertainties around 1 %.

The determination from hyperon decays has long lacked comparable theoretical understanding although in recent years it has received new input from both experiment and theory. In analogy with the strangeness-conserving decays, the vector form factor is protected against first-order SU(3)-breaking effects by the Ademollo–Gatto theorem (Ademollo and Gatto, 1964). Therefore, one may again use the ratio between the axial-vector and vector form factors (often denoted in this context as g_1/f_1) as experimental input, thereby circumventing the problem of accounting for SU(3)-breaking effects in the axial-vector contribution. The best present extraction is

$$|V_{us}| \;=\; 0.2250 \pm 0.0027, \tag{5.2.67}$$

which does not though include estimates of the theoretical uncertainty due to

second-order SU(3) breaking.

Other determinations of $|V_{us}|$ are provided by leptonic kaon decays and also τ decays. Lattice-QCD calculation of the ratio between the kaon and pion decay constants allows extraction of $|V_{us}/V_{ud}|$ from $K \to \mu\nu$ and $\pi \to \mu\nu$. The KLOE measurement of the $K^+ \to \mu\nu$ branching ratio, combined with the theoretical

$$f_K/f_\pi \;=\; 1.198 \pm 0.003^{+0.016}_{-0.005}, \tag{5.2.68}$$

leads to

$$|V_{us}| \;=\; 0.2245^{+0.0012}_{-0.0031}, \tag{5.2.69}$$

where the accuracy is limited by the lack of knowledge of the ratio of the decay constants.

$|V_{cd}|$: The most precise determination is based on neutrino and antineutrino interactions. The difference in the ratio of double- to single-muon production by neutrino and antineutrino beams is proportional to the charm production cross-section off valence d quarks, and therefore to $|V_{cd}|^2$ times the average semileptonic branching ratio of charm mesons, B_μ (see Fig. 5.16 for the quark-model interpretation). In the muon-neutrino case, the most probable process involves the

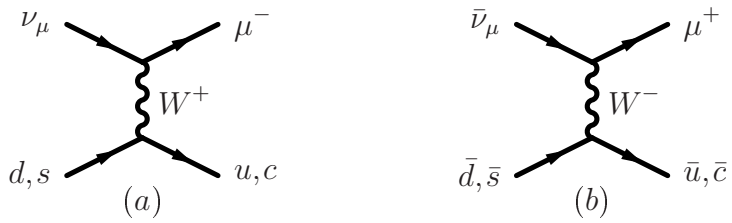


Figure 5.16: The neutrino and antineutrino interactions with a nucleon leading to the production of single- (u and $\bar{u}$ final states) and double-muon (c and $\bar{c}$) events.

transitions $d \to u$ since it is both Cabibbo favoured (*i.e.* proportional to $|V_{ud}|$) and proportional to the relatively large valence d-quark density—this leads to a single muon in the final state. However, there is also the Cabibbo-suppressed transition (*i.e.* proportional to $|V_{cd}|$) $d \to c$; the final charm decay then leads to a second muon. Note that uncertainties in the d-quark density cancel in the ratio. There is, of course, some background due to the presence of s quarks in the proton, but these are sea quarks and therefore much less probable.

This method was used by CDHS, followed by CCFR and CHARM II. The PDG average of their results is

$$B_\mu |V_{cd}|^2 \;=\; (0.463 \pm 0.034) \times 10^{-2}. \tag{5.2.70}$$

Moreover, data from the CHORUS experiment are now sufficiently precise to extract B_μ directly and combining their results with those of the other experiments leads to $B_\mu = 0.0873 \pm 0.0052$, finally giving

$$|V_{cd}| \;=\; 0.220 \pm 0.005. \tag{5.2.71}$$

The magnitude of V_{cd} may be also extracted from semileptonic charm decays using theoretical input from form-factor calculations: *e.g.* lattice-QCD calculations for $D \to \pi\ell\nu$ and $D \to K\ell\nu$. Using these estimates and the isospin-averaged charm semileptonic width measured by CLEO-c, one obtains a compatible but less precise $|V_{cd}| = 0.213 \pm 0.008 \pm 0.021$.

$|V_{cs}|$: An analogous determination of $|V_{cs}|$ from neutrino and antineutrino scattering is much less precise than in the previous case since it requires knowledge of the s-quark density, which is relatively small and not well measured. Other approaches must therefore be adopted.

The direct determination of $|V_{cs}|$ is possible from semileptonic D or leptonic D_s decays, but again theoretical input from hadronic matrix-element calculations is required. The use of $D_s^+ \to \ell^+\nu$ requires lattice-QCD calculation of the decay constant f_{D_s} while for semileptonic D decays form factors are required, which depend on the invariant mass of the lepton pair. Lattice-QCD calculations can predict the normalisation and the shape of the form factor in $D \to K\ell\nu$ and $D \to \pi\ell\nu$. These theoretical results and the isospin-averaged semileptonic widths provide

$$|V_{cs}| \;=\; 0.995 \pm 0.016. \tag{5.2.72}$$

Real $W^\pm$ decays are also sensitive to $|V_{cs}|$ and such measurements were made at LEP-II. The $W^\pm$ branching ratios depend on all six CKM matrix elements involving quarks lighter than M_W. For each lepton flavour we have

$$\frac{1}{B(W \to \ell\bar{\nu}_\ell)} \;=\; 3\left[1 + \left(1 + \frac{\alpha_s(m_W)}{\pi}\right) \sum_{u,c,d,s,b} |V_{ij}|^2\right]. \tag{5.2.73}$$

Assuming lepton universality, the result

$$B(W \;\to\; \ell\bar{\nu}_\ell) \;=\; (10.83 \pm 0.07 \pm 0.07)\;\% \tag{5.2.74}$$

then implies

$$\sum_{u,c,d;s,b} |V_{ij}|^2 \;=\; 2.002 \pm 0.027, \tag{5.2.75}$$

which is a precise test of unitarity. However, $|V_{cs}|$ can only be extracted from

flavour-tagged measurements. The LEP experiment DELPHI measured tagged $W^+ \to c\bar{s}$ decays and obtained

$$|V_{cs}| \;=\; 0.94^{+0.32}_{-0.26} \pm 0.13. \tag{5.2.76}$$

$|V_{cb}|$: Naturally, this matrix element can be extracted from semileptonic decays of B mesons to charm states. Inclusive determinations use the semileptonic decay-rate measurement combined with the leptonic-energy and the hadronic invariant-mass spectra. The basis of the calculation is the so-called heavy-quark expansion, via which the total rate and moments of differential energy and invariant-mass spectra are expressed as expansions in inverse powers of the heavy-quark masses. Since the dependence on m_b, m_c and other parameters occurring at subleading order is different for different moments, the large number of measured moments overconstrains the parameters and tests the consistency of the approach. Inclusive measurements have been performed using B mesons from Z^0 decays at LEP and in e^+e^- machines operated at the $\Upsilon(4S)$ energy. At LEP the large boost of B mesons from Z^0 decays allows determination of the moments throughout phase space, which is not otherwise possible, but the large statistics of the so-called B-factories leads to more precise determinations. An average of the measurements leads to

$$|V_{cb}| \;=\; (42.2 \pm 0.7) \times 10^{-3}. \tag{5.2.77}$$

Exclusive determinations are based on semileptonic B decays into D and D^*. In the limit $m_{b,c} \gg \Lambda_{\rm QCD}$ all form factors are provided by a single so-called Isgur–Wise function, which is a function of the scalar product of the four-velocities, $w = v \cdot v'$, of the initial- and final-state hadrons. Heavy-quark symmetry determines the normalisation of the rate at $w = 1$, the maximum momentum transfer to the leptons, and $|V_{cb}|$ is obtained from an extrapolation to $w = 1$. The exclusive determination,

$$|V_{cb}| \;=\; (39.5 \pm 0.8) \times 10^{-3}, \tag{5.2.78}$$

is less precise than the inclusive method since the uncertainties, both theoretical in the form factor and experimental in the rate near $w = 1$, are around 3 %. The PDG quotes the following average:

$$|V_{cb}| \;=\; (40.5 \pm 1.5) \times 10^{-3}. \tag{5.2.79}$$

$|V_{ub}|$: The natural determination of $|V_{ub}|$ from inclusive $B \to X_u \ell \bar{\nu}$ decays (where X_u represents a meson containing a u-quark) suffers large $B \to X_c \ell \bar{\nu}$ backgrounds. In most phase-space regions where the charm background may be excluded for kinematic reasons there are unknown non-perturbative contributions: the so-called shape functions. In contrast, the non-perturbative physics for $|V_{cb}|$ is encoded in

a few parameters. At leading order in $\Lambda_{\rm QCD}/m_b$ there is only one shape function, which may be extracted from the photon energy spectrum in $B \to X_s\gamma$ and applied to several spectra in $B \to X_u\ell\bar{\nu}$.

Alternatively, one can extend measurements into the $B \to X_c\ell\bar{\nu}$ region to reduce the theoretical uncertainties. Analyses of the electron-energy endpoint from CLEO, BABAR and Belle quote $B \to X_u e\bar{\nu}$ partial rates for $|\boldsymbol{p}_e| \geq 2.0\,\text{GeV}$ and 1.9 GeV, which are well below the charm endpoint. The large and pure B–$\bar{B}$ samples produced at B-factories permit the selection of $B \to X_u\ell\bar{\nu}$ decays in events with the recoiling $\bar{B}$ is fully reconstructed. Using such a full-reconstruction tag method, the four-momenta of both the leptonic and hadronic systems can be extracted.

Exclusive channels may also be used, but then form factors are needed. The better experimental signal-to-background ratios are offset by smaller yields. The $B \to \pi\ell\bar{\nu}$ branching ratio is now known to 8 % and lattice-QCD calculations of the $B \to \pi\ell\bar{\nu}$ form factor for $q^2 > 16\,\text{GeV}^2$ have been performed. So-called light-cone QCD sum rules are applicable for $q^2 < 14\,\text{GeV}^2$ and yield somewhat smaller values for $|V_{ub}|$, $(3.3^{+0.6}_{-0.4}) \times 10^{-3}$. The theoretical uncertainties in extracting $|V_{ub}|$ from inclusive and exclusive decays are different. The PDG quotes the following average:

$$|V_{ub}| \;=\; (4.09 \pm 0.39) \times 10^{-3}, \tag{5.2.80}$$

which is dominated by the inclusive measurement.

$|V_{td}|$, $|V_{ts}|$ and $|V_{tb}|$: The CKM matrix elements involving the t quark are rather more difficult to access, first and foremost owing to the exceedingly limited number of top quarks so-far produced (and detected) in laboratory experiments. However, the top quark plays an important role in the intermediate states in B–$\bar{B}$ oscillation phenomena and also in higher-order corrections (coming from so-called penguin diagrams. Unfortunately, present experimental precision does not yet allow significant measurements to be performed.

On the other hand, the Tevatron experiments CDF and DØ have found evidence for single top-quark production. The basic process studied is the annihilation of a quark and antiquark of different flavours via production of a virtual $W^{\pm}$, which subsequently decays into $\bar{b}t$ or $b\bar{t}$. Special techniques are necessary to extract the single top-quark signal from a large background. The cross-section measurements can be used to provide a direct measurement of the CKM matrix element $|V_{tb}|$.

The CP-violating phase and unitarity-triangle angles: The angles of the unitarity-triangle are evidently non-trivial, *i.e.* the triangle is not flat, *if and only if* CP-violation has its origins in the CKM matrix itself. It is therefore evident that their measurement requires the study of CP-violating effects. Different processes

provide more-or-less direct access to different angles and thus it is, in principle, possible to verify that the sum of the three angles is indeed 180°. The field is in continual evolution and here we shall limit ourselves to a presentation of the current picture. In Fig. 5.17 the combined world constraints on the unitarity

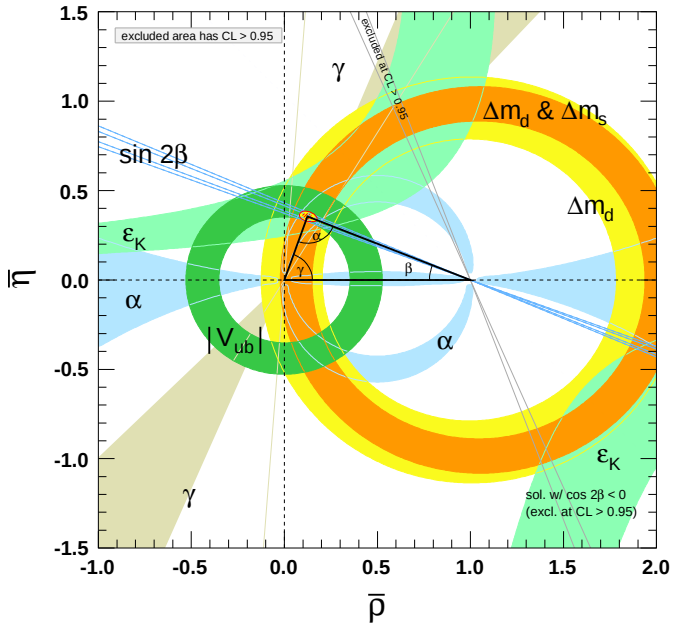


Figure 5.17: World constraints on the unitarity triangle represented in $\bar{\rho}$ and $\bar{\eta}$ plane. The shaded areas are the 95 % CL intervals for the various measurements and the small hatched ellipse indicates the overall constraint on the upper vertex. The figure is taken from PDG-2016 (Patrignani *et al.*, 2016).

triangle are displayed as 95 % CL allowed regions in $\bar{\rho}$ and $\bar{\eta}$ plane. The global agreement is excellent; the consistency worsens noticeably however if $B \to \tau\nu$ data is included in the fit.

5.3 Beyond the standard model

5.3.1 Neutrino masses, oscillation and mixing

Neutrino masses and limits

In the SM the masses of all three neutrino species are rigorously zero. This is related to the observation that only left-handed neutrinos appear to exist in Nature—or rather, only left-handed neutrinos are involved in the observed interactions. A standard Dirac mass term in the Lagrangian describing the propagation of a fer-

mion implicitly connects left- and right-handed states and would therefore violate such observations.*

However, recent data provide strong evidence for oscillations between the different neutrino states.† Such a phenomenon is possible if and only if at least one of the neutrino species has a non-zero mass. In fact, as we saw in the K^0–$\overline{K}^0$ system, the oscillation process requires a mass *difference* and so we are only able to conclude that there is a difference between the masses of at least two neutrino states. Over the next few years more precise and detailed measurements should clarify the issue.

Oscillation phenomena aside, other experimental data are only able to place upper limits on neutrino masses. By far the most stringent limits are those on the electron-neutrino mass and come from nuclear β-decay studies. In particular the end-point (in the Kurie plot) of the β-decay spectrum is lowered if the emitted neutrino is massive (by just the energy equivalent to its mass). The experimental data on tritium decay provide an upper limit of around 3 eV.

It is interesting to note that a similar, though less stringent, limit is provided by the finite spread in arrival times of the few neutrinos collected after the supernova SN 1987A in the Large Magellanic Cloud, approximately 51.4 kpc ($\sim$170 thousand light years) from Earth. A burst of neutrinos was observed at three separate neutrino observatories (Kamiokande II, IMB and Baksan). In total, 24 neutrinos were detected, a significant deviation from the observed background level; 11 were detected by Kamiokande II, 8 by IMB and 5 by Baksan, over a time interval of less than 13 seconds. Analysis of the energy and arrival-time spread provides a limit of about 20 eV (Roos, 1987).

Neutrino oscillation and mixing

If, as then appears to be the case, at least one neutrino is massive, the possibility of oscillation becomes real. Moreover, the possibility of a mixing matrix *à la* CKM naturally arises.‡ Indeed, this is a prerequisite for oscillation. Let us consider the simple case of two neutrino states with different masses (not necessarily both non-zero). Call the two mass eigenstates states $\nu_{1,2}$. The two states we call, for example, $\nu_{e,\mu}$ are then weak-interaction eigenstates, which may then be expressed

* We should note that it is possible to introduce a so-called Majorana mass term for the neutrino, which then becomes its own antiparticle, without altering the observed phenomenology.

† The 2015 Nobel Prize for Physics was awarded jointly to Takaaki Kajita and Arthur B. McDonald "for the discovery of neutrino oscillations, which shows that neutrinos have mass."

‡ Purely leptonic CP violation then also becomes a natural possibility.

as superpositions of the mass eigenstates, thus:

$$\nu_e \;=\; \cos\theta\,\nu_1 + \sin\theta\,\nu_2 \tag{5.3.1a}$$

and

$$\nu_\mu \;=\; -\sin\theta\,\nu_1 + \cos\theta\,\nu_2. \tag{5.3.1b}$$

This may be compared with the case of the K^0–$\overline{K}^0$ system. An important difference here is that typically $E_\nu \gg m_\nu$ and thus we should use *momentum* eigenstates.

Consider an electron neutrino produced in a weak interaction at instant $t=0$:

$$|\nu_e,\boldsymbol{p}\rangle \;=\; \cos\theta|\nu_1,\boldsymbol{p}\rangle + \sin\theta|\nu_2,\boldsymbol{p}\rangle. \tag{5.3.2}$$

We shall now suppress the momentum variable and write the corresponding state evolved to the instant t as

$$|\nu_e,t\rangle \;=\; a_1(t)\,\cos\theta|\nu_1\rangle + a_2(t)\,\sin\theta|\nu_2\rangle, \tag{5.3.3}$$

where the time-dependent coefficients are

$$a_i(t) \;=\; \mathrm{e}^{\mathrm{i}E_i t} \qquad \text{for } i \;=\; 1,2. \tag{5.3.4}$$

Since $m_1 \neq m_2$, for a given well-defined momentum $\boldsymbol{p}$, $E_1 \neq E_2$.

Exercise 5.3.1. *Following similar steps as for the case of K^0–$\overline{K}^0$ oscillations, show that the probability that a state initially produced at $t=0$ as say an electron neutrino will be a muon neutrino a time t is*

$$P(\nu_e \;\to\; \nu_\mu;t) \;=\; \sin^2 2\theta\,\sin^2 \tfrac{1}{2}(E_2 - E_1)t. \tag{5.3.5}$$

We may rewrite the energy difference as follows (using $E^2 = p^2 + m^2$):

$$E_2 - E_1 \;=\; \frac{E_2^2 - E_1^2}{E_2 + E_1} \;=\; \frac{\Delta m^2}{2E_\nu}, \tag{5.3.6}$$

where $\Delta m^2 = m_2^2 - m_1^2$ and E_ν is the average of the two energies. We see that the oscillation depends on two parameters: the mixing angle $\sin\theta$ and mass-squared difference Δm^2, which must both be non-zero. Indeed, in the limit $E_\nu \gg m$ (taking $v_\nu \simeq c$) we may rewrite the time-dependent transition probability as

$$P(\nu_e \;\to\; \nu_\mu;t) \;\simeq\; \sin^2 2\theta\,\sin^2\left(\frac{1.27\Delta m^2 L(t)}{E_\nu}\right), \tag{5.3.7}$$

where Δm^2 is in eV^2, $L(t)$ is the distance travelled in metres and E_ν is the neutrino energy in MeV.

Since this obviously leads to a loss of electron-neutrino flux, it was immediately suggested as a possible explanation of the so-called solar neutrino problem; that is, the result of the Homestake gold-mine chlorine experiment (Bahcall, 1964; Davis, 1964; Davis *et al.*, 1968):*†

$$\Phi_{\nu_e}^{\text{expt}} \sim \tfrac{1}{3}\Phi_{\nu_e}^{\text{standard solar model}}. \tag{5.3.8}$$

Such oscillations have now been confirmed by the Super-Kamiokande experiment in Japan (see, for example, Fukuda *et al.*, 1998; Ashie *et al.*, 2004, 2005). The data on atmospheric neutrinos, produced immediately above the apparatus and also on the far side of the Earth, are consistent with two-flavour $\nu_\mu \leftrightarrow \nu_\tau$ oscillations, $\sin^2 2\theta > 0.92$ and

$$1.5 \times 10^{-3} < \Delta m^2 < 3.4 \times 10^{-3} \text{ eV}^2 \quad \text{at the 90 \% CL.} \tag{5.3.9}$$

However, the data cannot indicate which neutrino superposition is the more massive and various mass hierarchies are presently allowed.‡

5.3.2 Grand unified theories

In earlier sections we have seen how the attempt to correctly describe the weak interaction and, in particular, the massive intermediate bosons it requires led to the development of a (quasi) *unified* theory of the electromagnetic and weak interactions. The concept of a spontaneously broken symmetry allows the bosons to acquire a mass *without* violating the local gauge invariance of the theory and leads to a final remnant $\mathrm{U(1)_{EM}}$ symmetry from the initial larger $\mathrm{U(1)_Y \otimes SU(2)_W}$. Note that since there are two distinct coupling constants, g and g' in the previous sections, one cannot speak of a complete unification.

Moreover, this does not yet include QCD, the theory of the strong interaction. The successes of the unified electroweak theory immediately sparked attempts to include the SU(3) gauge symmetry of QCD. The idea is a little more challenging than a mere extension of $\mathrm{U(1)_Y \otimes SU(2)_W}$ to the larger $\mathrm{U(1)_Y \otimes SU(2)_W \otimes SU(3)_{QCD}}$.

* The ^{37}Cl solar neutrino detector in the Homestake Gold Mine consisted of 615 t of tetrachloroethylene (tetracloroetene), 4200 m of water equivalent underground. It used radiochemical techniques to determine the ^{37}Ar production rate. The detector was built at Brookhaven National Lab. (BNL) in 1965–67 and operated by Brookhaven until 1984.

† One quarter of the 2002 Nobel Prize for Physics was awarded each to Raymond Davis Jr. and Masatoshi Koshiba for "pioneering contributions to astrophysics, in particular for the detection of cosmic neutrinos"; the other half was awarded to Riccardo Giacconi for "pioneering contributions to astrophysics, which have led to the discovery of cosmic X-ray sources."

‡ The 2015 Nobel Prize in Physics was awarded jointly to Takaaki Kajita and Arthur B. McDonald "for the discovery of neutrino oscillations, which shows that neutrinos have mass."

It would, in reality, be preferable to have a *single* gauge group with thus a *single* coupling constant, which is what is known as a GUT. Sensible examples of groups containing the above product are SU(5) or SO(10). The problem of the single coupling is almost automatically solved when one realises that the three known constants actually vary with energy scale: the largest $\alpha_{\rm QCD}$ decreases, as does that associated with the $\mathrm{SU(2)_W}$, while the weakest $\alpha_{\rm QED}$ increases. Indeed, examination of the theoretical variation suggests that they should all have a similar value for $E \sim 10^{15}\,\mathrm{GeV}$.

Now, the enlargement of the group to SU(5), SO(10) or similar involves the addition of further gauge degrees of freedom: new gauge bosons (usually indicated X) must be introduced. These new fields inevitably couple quarks to leptons and thus allow for baryon- and lepton-number changing transitions. In particular, the proton is no longer stable since a u quark may be converted into an electron plus other leptons. In all such models the proton lifetime is estimated to be no greater than about 10^{30} years while present lower limits coming from *non*-observation of proton decay in the various dedicated experiments around the world are of order 10^{34} years.

There are further problems with the *naïve* extensions to obtain a GUT. First of all, even in the simpler electroweak theory, problems to do with renormalisation still remain. The presence of a scalar field (necessary for the Higgs mechanism) upsets the usual renormalisation programme: the contributions coming from virtual scalar loops tend to shift the masses of the particle spectrum up to the upper momentum cut-off, which here should be of order $E \sim 10^{15}\,\mathrm{GeV}$. In order to arrive at the masses of the known quarks and leptons, one therefore has to invoke very delicate cancellations between different contributions. That is, the unphysical (bare) masses need to be fine-tuned to many decimal places in order that the difference of two large numbers be the small number required; this is known as the hierarchy or fine-tuning problem. Moreover, close examination of the running of the three coupling constants reveals that they do *not* all meet at a single energy and thus such a *naïve* grand unification is not strictly speaking possible (see Fig. 5.18).

5.3.3 Supersymmetry

The difficulties encountered in constructing GUT's, as discussed above, suggest some missing ingredient. There are many possibilities, here we shall just mention the presently most favoured: namely, *supersymmetry*. The idea of supersymmetry, in a nutshell, is to enlarge the spectrum of particles and the symmetries to include a symmetry operation that transforms fermions into bosons and *vice versa*. One might hope in such a way to relate the photon to, say, the electron; however, this proves to be impossible. It is then necessary to introduce an entire family of supersymmetric partners to the existing particles. Thus, to every quark there

corresponds a pair of scalar *squarks* (matching the two quark helicity degrees of freedom) while for every gauge boson there is a fermionic *gaugino*. Since no such particle states have ever been observed, they must evidently be very heavy.

With a suitable spectrum (typical masses should be of the order of a TeV or more), both the fine-tuning and GUT-point problems may be solved. The fine-tuning is now automatic: the "opposite" spin nature means that the contributions in quantum corrections enter with the opposite sign* and thus cancellations are guaranteed. Such cancellations will not be exact, precisely because of the large supersymmetric particle masses. Indeed, this observation helps to place an upper limit on the likely masses, in order that the cancellations should not be lost altogether.

Secondly, the extra contributions alter the running of the coupling constants, which can now be arranged to all coincide at one particular energy. The new unification point is typically of order 10^{16} GeV (see Fig. 5.18). Both solutions are achieved if and only if the *sparticle* masses are less than around 10 TeV. Such a limit suggests that they may well be within reach of the LHC.

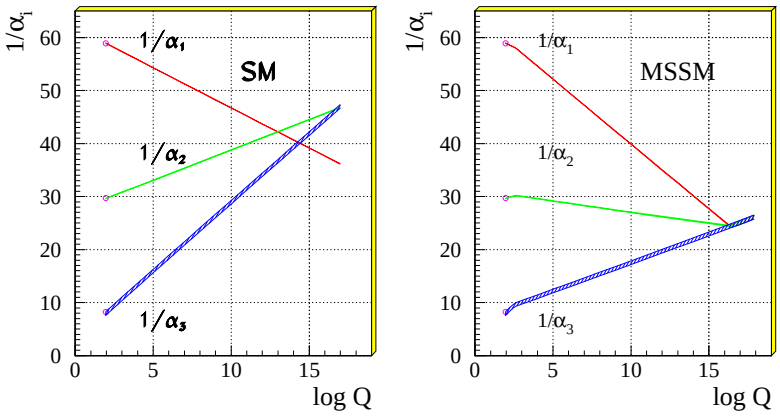


Figure 5.18: The running of the three coupling constants for the SM on the left and the so-called *minimal supersymmetric* SM on the right. From top to bottom the curves are respectively for the QED, weak and strong coupling constants. The figure is taken from Kazakov (2001).

* This is to do with Fermi–Dirac statistics: loop diagrams with a fermion circulating acquire an extra minus sign with respect to boson loops.

5.4 Bibliography

Aad, G. *et al.*, ATLAS Collab. (2012a), *Phys. Lett.* **B710**, 49.

Aad, G. *et al.*, ATLAS Collab. (2012b), *Phys. Lett.* **B716**, 1.

Aaltonen, T. *et al.*, CDF, DØ Collabs. (2012), *Phys. Rev. Lett.* **109**, 071804.

Ademollo, M. and Gatto, R. (1964), *Phys. Rev. Lett.* **13**, 264.

Anderson, P.W. (1963), *Phys. Rev.* **130**, 439.

Ashie, Y. *et al.*, Super-Kamiokande Collab. (2004), *Phys. Rev. Lett.* **93**, 101801.

Ashie, Y. *et al.*, Super-Kamiokande Collab. (2005), *Phys. Rev.* **D71**, 112005.

Bahcall, J.N. (1964), *Phys. Rev. Lett.* **12**, 300.

Chatrchyan, S. *et al.*, CMS Collab. (2012a), *Phys. Lett.* **B710**, 26.

Chatrchyan, S. *et al.*, CMS Collab. (2012b), *Phys. Lett.* **B716**, 30.

Chau, L.-L. and Keung, W.-Y. (1984), *Phys. Rev. Lett.* **53**, 1802.

Davis, J., Raymond (1964), *Phys. Rev. Lett.* **12**, 303.

Davis, J., Raymond, Harmer, D.S. and Hoffman, K.C. (1968), *Phys. Rev. Lett.* **20**, 1205.

Englert, F. and Brout, R. (1964), *Phys. Rev. Lett.* **13**, 321.

Fukuda, Y. *et al.*, Super-Kamiokande Collab. (1998), *Phys. Rev. Lett.* **81**, 1562.

Glashow, S.L. (1961), *Nucl. Phys.* **22**, 579.

Goldstone, J. (1961), *Nuovo Cim.* **19**, 154.

Guralnik, G.S., Hagen, C.R. and Kibble, T.W.B. (1964), *Phys. Rev. Lett.* **13**, 585.

Hasert, F.J. *et al.*, Gargamelle Neutrino Collab. (1974), *Nucl. Phys.* **B73**, 1.

Hasert, F.J. *et al.*, Gargamelle Neutrino Collab. (1973), *Phys. Lett.* **B46**, 138.

Higgs, P.W. (1964), *Phys. Lett.* **12**, 132; Higgs, P.W., *Phys. Rev. Lett.* **13**, 508.

Jarlskog, C. (1985), *Phys. Rev. Lett.* **55**, 1039.

Kazakov, D.I. (2001), in proc. of the *European School of High-Energy Physics—ESHEP 2000* (Caramulo, Aug.–Sept. 2000), eds. N. Ellis and J.D. March-Russell; *CERN Yellow Rep.* **2001-003**, 125.

Nambu, Y. and Jona-Lasinio, G. (1961), *Phys. Rev.* **122**, 345; **124**, 246.

Patrignani, C. *et al.*, Particle Data Group (2016), *Chin. Phys.* **C40** [10], 100001.

Quigg, C. (2007), *Rep. Prog. Phys.* **70**, 1019.

Roos, M. (1987), *Europhys. Lett.* **4**, 953.

Salam, A. (1968), in proc. of *Elementary Particle Theory: Relativistic Groups and Analyticity* (Aspenäsgården, May 1968), eds. N. Svartholm and J. Nilsson (Almqvist and Wiksell), no. 8 in The Nobel Symposium Series, p. 367.

Weinberg, S. (1967), *Phys. Rev. Lett.* **19**, 1264.

Wolfenstein, L. (1983), *Phys. Rev. Lett.* **51**, 1945.

Appendix A

Quantum Mechanics

A.1 Relativistic QM and the Dirac equation

Historically, the first attempts at a relativistic formulation of quantum mechanics are due independently to Klein (1927) and Gordon (1926) although earlier both Fock and Schrödinger had considered such a possibility. Starting from the Einstein energy–momentum relation

$$E^2 \;=\; \boldsymbol{p}^2c^2 + m^2c^4, \tag{A.1.1}$$

they simply applied the canonical approach of transforming the variables E and $\boldsymbol{p}$ into operators:

$$E \;\rightarrow\; \mathrm{i}\hbar\frac{\partial}{\partial t} \quad \text{and} \quad \boldsymbol{p} \;\rightarrow\; -\mathrm{i}\hbar\boldsymbol{\nabla}. \tag{A.1.2}$$

This is represented more conveniently and compactly in four-vector notation as follows:

$$p^2 \;=\; p^\mu p_\mu \;=\; m^2c^2, \tag{A.1.3}$$

where now

$$p^\mu \;\equiv\; (E/c, \boldsymbol{p}) \qquad \text{and } e.g. \qquad x^\mu \;\equiv\; (ct, \boldsymbol{x}). \tag{A.1.4}$$

The operator substitution then becomes

$$p^\mu \;\equiv\; (E/c, \boldsymbol{p}) \;\rightarrow\; \mathrm{i}\hbar\,\partial^\mu \;\equiv\; \mathrm{i}\hbar\frac{\partial}{\partial x_\mu} \;\equiv\; \mathrm{i}\hbar\left(\frac{1}{c}\frac{\partial}{\partial t}, -\boldsymbol{\nabla}\right). \tag{A.1.5}$$

Note the negative sign in front of the spatial components here.

This leads to the following Lorentz-covariant wave equation

$$\left(\partial_\mu\partial^\mu + \frac{m^2c^2}{\hbar^2}\right)\phi \;=\; 0. \tag{A.1.6}$$

For an even more compact notation, it is customary to define the d'Alembertian or wave operator:

$$\Box \equiv \partial_\mu \partial^\mu \tag{A.1.7}$$

and thus write (setting now $\hbar = 1 = c$)

$$\left(\Box + m^2\right) \phi \;=\; 0, \tag{A.1.8}$$

this is the Klein–Gordon equation. Historically, the well-known problems with negative particle probability densities, associated with the possibility of negative-energy solutions (energy enters *squared* in the Einstein's equation), led to its being abandoned and to Dirac's famous alternative.

In a nutshell, Dirac's idea was an attempt to avoid negative energies by effectively taking the square-root of the Klein–Gordon equation and writing

$$\gamma^\mu p_\mu \, \varphi(x) \;=\; m \, \varphi(x), \tag{A.1.9}$$

where γ^μ is some new and unknown vector object (to be determined), necessary to render the left-hand side a scalar quantity, as is the right-hand side. Then, in order that the operator version $\gamma^\mu p_\mu \widehat{=} m$ should agree with Einstein's relation $p^2 = m^2$ for any p^μ, we need a Clifford (or Dirac) algebra: $\{\gamma^\mu, \gamma^\nu\} = g^{\mu\nu}$. Indeed, it is easy to see that with such an algebra we have

$$\gamma^\mu p_\mu \gamma^\nu p_\nu \;=\; \tfrac{1}{2}\{\gamma^\mu, \gamma^\nu\} p_\mu p_\nu \;=\; g^{\mu\nu} p_\mu p_\nu \;=\; p^2. \tag{A.1.10}$$

The simplest way to represent such anticommuting γ^μ is via matrices; the minimal representation has rank four and may be constructed block-wise with the aid of the Pauli matrices. An explicit form (due to Dirac) is

$$\gamma^0 \;=\; \begin{pmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\gamma} \;=\; \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ -\boldsymbol{\sigma} & 0 \end{pmatrix}, \tag{A.1.11}$$

where the sub-matrices are $2{\times}2$. The first immediate consequence is that the wave-functions are represented by *four-component spinors*.* The indices on such a spinor, as too those (implicit) on the matrices γ^μ, are often referred to as Dirac indices and the space over which they run, Dirac space. Note that it cannot be thought of as any sort of vector since, for example, we shall find that a spatial rotation through 2π reverses its sign.

We now simply list some of the basic properties of the γ-matrices:

$$\gamma^{0\dagger} \;=\; \gamma^0, \quad \boldsymbol{\gamma}^\dagger \;=\; -\boldsymbol{\gamma}, \quad \text{or} \quad \gamma_\mu \;=\; \gamma^{\mu\dagger} \;=\; \gamma^0 \gamma^\mu \gamma^0. \tag{A.1.12}$$

* These are not to be confused with Lorentz four-vectors!

Finally then, the relativistic wave equation or Dirac equation is ($\hbar = 1 = c$)

$$\left[\mathrm{i}\gamma^{\mu}\partial_{\mu} - m\mathbb{1}\right]\varphi(x) \;=\; 0, \tag{A.1.13}$$

where $\mathbb{1}$ is just the rank-four unit matrix.

As one might imagine from the appearance of the Pauli matrices, the different spinor components have to do with the spin states of the electron. Indeed, if we consider the coupling to a magnetic field and take the non-relativistic limit we find that the two-by-two block form of the Dirac equation simply reduces to the Schrödinger equation augmented with the Pauli construction to describe the coupling of the *intrinsic* electron magnetic moment. And the gyromagnetic ratio for the electron g_e is *predetermined* to be precisely two, as found experimentally.* In the next section we shall also see how the concept of antiparticle arises naturally from the Dirac equation.

A.2 The discrete symmetries $\mathcal{C}$, $\mathcal{P}$ and $\mathcal{T}$

As stressed at various points during the lectures, the concept of symmetry plays a central role in the development of physics in general and particularly in elementary particle physics. While many of the symmetries encountered are continuous (*e.g.* spatial and temporal translations, rotations *etc.*), there are three fundamental discrete symmetries, which we shall discuss here within the context of a quantum-mechanical description of particle interactions. These symmetry transformations are the operations of:

- $\mathcal{C}$: transforming particle into antiparticle and *vice versa*;
- $\mathcal{P}$: spatial inversion, *i.e.* $\boldsymbol{x} \to -\boldsymbol{x}$;
- $\mathcal{T}$: time reversal, *i.e.* $t \to -t$.

Recall that the first two are *linear* transformations while the last is *antilinear*; that is, together with the obvious coordinate transformation, one must apply complex conjugation to all $\mathbb{C}$-number parameters (*e.g.* masses, coupling constants *etc.*) involved.

Since we shall need to deal with current–current interactions, it will be useful to know in advance the transformation properties of the possible currents under the action of the above operations. Generalised currents may be constructed from bilinears of spin-half fields. We shall thus now examine each separately within the context of the Dirac equation.

* To be precise, owing to quantum mechanical corrections, the value is not exactly two. Nevertheless, the measured and calculated values coincide perfectly to a very high precision.

A.2.1 Charge-conjugation invariance ($\mathcal{C}$)

The obvious way in which to detect theoretically (and indeed experimentally) the presence of an antiparticle is through the relative sign of its coupling to an external classical electromagnetic field $A^\mu = (\Phi, \boldsymbol{A})$. Indeed, the Dirac equation for an electron in the presence of a real (for simplicity) electromagnetic field is

$$\left[\mathrm{i}\gamma_\mu \partial^\mu + e\gamma_\mu A^\mu - m\right] \psi(x) \;=\; 0. \tag{A.2.1}$$

The sign adopted for the electromagnetic coupling is, of course, purely conventional. It does determine, though, the corresponding sign for the coupling of a positron. Indeed, consider the Coulomb part of the potential $A^0(x)$: it couples via γ^0, which, as we have seen has positive entries for the upper two diagonal elements and negative below. The lower two spinor components thus couple to the electric field with opposite sign and hence we have the anti-electron or positron.

If the physics of the positron is to be the same as that of the electron, we must seek a transformation ($\mathcal{C}$) that results in the following equation:

$$\left[\mathrm{i}\gamma_\mu \partial^\mu - e\gamma_\mu A^\mu - m\right] \psi^{\mathcal{C}}(x) \;=\; 0. \tag{A.2.2}$$

Now, the action of complex conjugation evidently only affects the sign of the first term (for a real electromagnetic field). Recall too the following relation (which is a direct result of the Clifford algebra and does *not* depend on the particular representation adopted):

$$\gamma^0 \gamma_\mu \gamma^0 \;=\; \gamma_\mu^\dagger, \tag{A.2.3}$$

while the following representation-dependent relation holds in the standard Dirac representation:

$$\gamma^2 \gamma_\mu \gamma^2 \;=\; -\gamma_\mu^{\mathsf{T}}, \tag{A.2.4}$$

where T simply indicates matrix transposition. Thus, by applying complex conjugation to the Dirac equation (A.2.1) and then multiplying from the left by $\gamma^0\gamma^2$, one can easily verify that the precise form shown in Eq. (A.2.2) is obtained with the identification

$$\mathcal{C} \;:\; \psi(x) \;\rightarrow\; \psi^{\mathcal{C}}(x) \;\equiv\; \mathrm{i}\gamma^0\gamma^2\, \bar{\psi}^{\mathsf{T}}(x), \tag{A.2.5}$$

where the conjugate spinor $\bar{\psi} := \psi^\dagger \gamma^0$. Note that the presence of the factor "i" corresponds to an arbitrary, but conventional, phase choice.

A.2.2 Spatial-inversion invariance ($\mathcal{P}$)

Following a similar procedure to that of the previous section we shall now derive the form of the operator generating spatial inversion. The starting point will again

be the Dirac equation (A.2.1), although the coupling to an electromagnetic field is now superfluous. The transformation of the equation under $\boldsymbol{x}$ to $-\boldsymbol{x}$ may be represented by simply replacing ∂^μ with ∂_μ (since the lowering of the index implies a sign change in the spatial components). Now, the effect of γ^0, already noted above, may be equally expressed as

$$\gamma^0\gamma_\mu\gamma^0 \,=\, \gamma^\mu. \tag{A.2.6}$$

Thus, we recover the original equation via the identification

$$\mathcal{P} \,:\, \psi(x) \,\to\, \psi^{\mathcal{P}}(x') \,\equiv\, \gamma^0\,\psi(x'), \qquad \text{with} \qquad x' \,=\, (t,-\boldsymbol{x}), \tag{A.2.7}$$

where once again the implicit phase choice is conventional.

An immediate consequence of the above form for the parity transformation is that, since the matrix γ^0 is block-diagonal $\pm\mathbb{1}$, the parities of the upper and lower components of ψ are opposite. That is, fermion and antifermion have opposite parities. By convention, the parity of fermions is chosen positive and antifermions negative. This particular choice has, of course, no physical consequence as fermions are always produced in fermion–antifermion pairs (for which the overall intrinsic parity is predetermined to be -1).

A.2.3 Time-reversal invariance ($\mathcal{T}$)

Finally, we turn to the case of time reversal. Note first that the transformation $t\to -t$ also implies exchange of initial and final states. Since, as remarked above, this transformation has the peculiar property of being antilinear, let us start with the simpler case of the Schrödinger equation for a free particle:

$$\mathrm{i}\,\frac{\partial}{\partial t}\,\psi(t,\boldsymbol{x}) \,=\, -\frac{1}{2m}\,\boldsymbol{\nabla}^2\psi(t,\boldsymbol{x}). \tag{A.2.8}$$

The eigen-solutions are plane-waves and may be written as

$$\psi(t,\boldsymbol{x}) \,=\, u(\boldsymbol{p})\,\mathrm{e}^{-\mathrm{i}(Et-\boldsymbol{p}\cdot\boldsymbol{x})}, \tag{A.2.9}$$

where, of course, the energy and momentum satisfy $E=p^2/2m$. It should be immediately obvious that the first choice of simply changing t to $-t$ in the above does not satisfy the original equation, nor indeed does it even correspond to a particle with momentum $-\boldsymbol{p}$, as it should for time reversal. However, the choice

$$\psi^*(-t,\boldsymbol{x}) \,=\, \psi_0^*\,\mathrm{e}^{-\mathrm{i}(Et+\boldsymbol{p}\cdot\boldsymbol{x})} \tag{A.2.10}$$

respects all requirements. We thus see the necessity for an antilinear (complex conjugation) operator. That is, the transformation $t \to -t$ is accompanied by complex conjugation applied to all $\mathbb{C}$-number quantities.

Now, complex conjugation applied to the Dirac equation (A.2.1) leads to

$$\left[-\mathrm{i}\gamma_\mu^{\mathsf{T}}\partial^\mu - m\right]\bar{\psi}^{\mathsf{T}}(t,\boldsymbol{x}) \;=\; 0, \tag{A.2.11}$$

where, as always, T simply stands for matrix transposition and once again the commutation properties of the γ matrices with γ^0 have been exploited. Multiplication from the left by γ^0 together with the simultaneous transformation $x^\mu \to x'^\mu \equiv (-t,\boldsymbol{x})$ then leads to

$$\left[\mathrm{i}\gamma_\mu^{\mathsf{T}}\partial'^\mu - m\right]\gamma^0\,\bar{\psi}^{\mathsf{T}}(-t',\boldsymbol{x}') \;=\; 0. \tag{A.2.12}$$

All that remains is to find the unitary transformation (which must exist) between the Dirac bases for γ_μ and γ_μ^{T}: it is simply $\mathrm{i}\gamma^1\gamma^3$, where the factor "i" is again conventional. Thus, the time-reversal operation is given by

$$\mathcal{T}\,:\,\psi(x) \;\to\; \psi^{\mathcal{T}}(x') \;\equiv\; \mathrm{i}\gamma^1\gamma^3\,\bar{\psi}^{\mathsf{T}}(x'), \qquad \text{with} \qquad x' \;=\; (-t,\boldsymbol{x}). \tag{A.2.13}$$

$\boldsymbol{T}$ and Complex Potentials

As a closing remark to this section, let us illustrate the role of a complex contribution to the potential describing particle interactions. While there is evidently no counterpart in classical mechanics, in quantum mechanics all quantities are potentially complex. Typically, when particles (or radiation) may be emitted or absorbed (created or destroyed), an imaginary contribution to the scattering matrix elements is found. Consider schematically the temporal evolution of a state of definite energy E,

$$\phi(t,\boldsymbol{x}) \;=\; a(\boldsymbol{x})\,\mathrm{e}^{-\mathrm{i}Et}\,. \tag{A.2.14}$$

The probability density is just $\rho(t,\boldsymbol{x}) := \phi^*\phi = |a(\boldsymbol{x})|^2$, which in this case is time independent. If, however, we introduce an imaginary contribution, $-\frac{1}{2}\mathrm{i}\Gamma$ say, to the energy, that is $E \to E - \frac{1}{2}\mathrm{i}\Gamma$, then something interesting occurs:

$$\rho(t,\boldsymbol{x}) \;=\; |\phi(t,\boldsymbol{x})|^2 \;=\; |a(\boldsymbol{x})|^2\,\mathrm{e}^{-\Gamma t}\,. \tag{A.2.15}$$

The probability density thus follows the usual decay law, with rate Γ.

Note that the resulting time dependence evidently violates time-reversal invariance. That is, an imaginary phase in the interaction Hamiltonian automatically provokes a violation of $\mathcal{T}$. This may be immediately understood as a direct consequence of the *anti*linearity property of the temporal inversion operator.

A.2.4 Dirac-spinor bilinears and $\mathcal{CPT}$

Armed with the previously derived transformation operators, it is now easy to determine the transformation property of any spinor-field bilinear and thus any current we may wish to employ in the description of particle interactions. The natural electric four-current associated with the Dirac equation, is shown by standard methods to be $j^\mu := \bar{\psi}\gamma^\mu\psi$, the temporal component of which is $\rho = \psi^\dagger\gamma^0\gamma^0\psi = \psi^\dagger\psi$, a natural and positive-definite probability density.

However, we also wish to describe all possible interactions (include the weak and strong nuclear forces) and thus we must consider all possible currents. The complete basis of generalised currents is $S = \bar{\psi}\psi$, $P = \bar{\psi}\gamma_5\psi$, $V = \bar{\psi}\gamma^\mu\psi$, $A = \bar{\psi}\gamma^\mu\gamma_5\psi$ and $T = \bar{\psi}\sigma^{\mu\nu}\psi$. Their transformation properties under the discrete transformations $\mathcal{C}$, $\mathcal{P}$ and $\mathcal{T}$ are summarised in Table A.1.

Table A.1: The properties of the five spinor-bilinear currents (S, P, A, V and T) under the discrete transformations $\mathcal{C}$, $\mathcal{P}$ and $\mathcal{T}$.

	S	P	V	A	T
$\mathcal{C}$	+	+	−	+	−
$\mathcal{P}$	+	−	+	−	+
$\mathcal{T}$	+	−	+	+	−
$\mathcal{CPT}$	+	+	−	−	+

In order to construct a current–current interaction, it is necessary to combine two (or possibly more) currents by *completely* contracting or saturating the indices. For the currents listed here, it is immediately noticeable that the CPT signature is $(-1)^{n_i}$, where n_i is the number of indices. Since complete saturation implies n_i even, it follows that it is *impossible* to obtain a product that is overall CPT odd. In other words, within our present knowledge and method of constructing (current–current) interactions in field theory, CPT cannot be violated. That is not to say that it is absolutely impossible, but simply that we do not know how.

For completeness, let us mention that a *possible* consequence of CPT violation is a difference between particle and antiparticle masses. At present the most stringent limits come from the study of the K^0–$\overline{K}^0$ system (which is examined in some detail in Sec. 2.5.3):

$$\left|\frac{m_{K^0} - m_{\overline{K}^0}}{m_{K^0}}\right| \leq 0.9 \times 10^{-18} \quad (90\ \%\ \text{CL}). \tag{A.2.16}$$

Note, in contrast, that any of C, P or T may be violated individually (or in pair products) by a suitable choice of interfering currents; *e.g.* the product $V{\cdot}A$

violates both C and P but not T. What is not included above is the possibility of a complex coupling (as in the elements of the $V_{\rm CKM}$ matrix, see Sec. 2.5.2). Such a contribution would naturally induce a violation of time-reversal invariance, which is rather difficult (though not impossible) to detect experimentally. For this reason and since the conservation of CPT requires a simultaneous compensating violation of the product CP, one normally talks of CP violation and not T violation, although the two are entirely equivalent in this context.

A.2.5 C and P of simple composite systems

Many simple composite systems, such as positronium (an e^+e^- bound state) and $q\bar{q}$ pairs, but also two- or multi-pion final states, may possess well-defined symmetry properties under the operations of $\mathcal{C}$ and $\mathcal{P}$. In theories in which these symmetries are respected such properties naturally lead to the idea of associated conserved quantum numbers C and P, even for composite objects (the same is true too for T). Such discrete quantum numbers are *multiplicative* in nature and thus in the case of a composite system all the relevant quantum numbers of the parts must simply be multiplied together. We shall now present a few instructive examples.

From the foregoing discussion on CPT, we see that, by complementarity, once the properties under C and P are understood an explicit discussion of T is superfluous. That said, in condensed matter physics, foe example, a consequence of T for systems of fermions is the so-called Kramers degeneracy, whereby even certain highly disordered systems must have degenerate energy eigenstates.

Charge conjugation in composite systems

First of all, note that C can clearly only be associated with neutral systems, such as positronium or neutral $q\bar{q}$ states.

$C_{\pi^+\pi^-}$: Let us start by considering the two-pion state $\pi^+\pi^-$. The action of $\mathcal{C}$ is to interchange the two pions and this will introduce a factor $(-1)^L$, where L is the orbital angular momentum quantum number, owing to the parity of the spatial part of the wave-function. Since there is no other effect, we have $C_{\pi^+\pi^-} = (-1)^L$.

$C_{e^+e^-}$: Next we examine the case of positronium. This is a little more complicated owing to the spin effects. For the spatial exchange of the electron and positron there is the same factor $(-1)^L$ above. However, the spin part of the wave-function must also be considered: for the spin-0 singlet state this is antisymmetric and for the spin-1 triplet, symmetric. This leads to another factor conveniently expressed as $-(-1)^S$. Finally, the full relativistic theory of electrons generates a further -1

for every interchange of two identical fermion or antifermion states. Putting all this together, we obtain a charge-conjugation quantum number $C_{e^+e^-} = (-1)^{L+S}$.

Parity in composite systems

Here we should recall that all particles (elementary or not) may be ascribed an intrinsic parity: for fermions this is not determined absolutely, but Dirac theory predicts it to be opposite for fermion and antifermion. A fermion–antifermion pair thus have overall negative intrinsic parity, which must then be multiplied by the parity of the relative spatial wave-function. In the case of bosons a scalar particle has, by definition $P = +1$ while a pseudoscalar (such as a pion) has $P = -1$.

$P_{\pi^+\pi^-}$: Given an even number of pions, it is not necessary to know the intrinsic parity of the pion since in this case we have $P^2 = +1$, whatever the value of $P = \pm 1$. We therefore need only consider the parity of the spatial wave-function, which leads to $P_{\pi^+\pi^-} = (-1)^L$.

$P_{e^+e^-}$: As noted above the product of intrinsic parities for the e^+e^- pair is -1, which again must be multiplied by the spatial contribution. Here then we have $P_{e^+e^-} = -(-1)^L$.

The J^{PC} classification of mesons

We can now study the classification of mesons ($q\bar{q}$ states) in terms of the three quantum numbers J, P and C. Let us examine the angular momentum first. The purely spin part may be either spin-0 (antisymmetric, singlet) or spin-1 (symmetric, triplet). The orbital angular momentum part is naturally any integer from zero up. The J^{PC} assignments may then be deduced easily from the previous analysis of the positronium case.

The lowest-lying mesons (π^0, K^0, η and η') are $L = 0$, $S = 0$ states, which must then have $J^{PC} = 0^{-+}$.* The slightly heavier s-wave $S = 1$ states (ρ, K^*, ϕ and ω) have $J^{PC} = 1^{--}$. For increasing mass the J^{PC} quantum numbers then follow a natural sequence. A complete list of all the known mesonic states and their J^{PC} assignments may be found in the PDG Review of Particle Physics (see PDG-2014 – Olive *et al.*, 2014).

Exercise A.2.1. *From the foregoing classification, show that for a standard $q\bar{q}$ state the assignments $J^{PC} = 0^{+-}$, 0^{--}, 1^{-+}, 2^{+-} etc. are not admissible while 0^{++}, 1^{++} etc. are allowed.*

* While C is clearly not defined for a charged meson, the J^P assignments given still hold.

Finally, for completeness, we might add that the photon has $J^{PC}=1^{--}$, the gluons (being colour charged with therefore C undefined) have $J^{P}=1^{-}$ while the spin-1 $W^{\pm}$ and Z^{0} weak bosons have neither P nor C well-defined. The neutral Higgs boson in the SM has $J^{PC}=0^{++}$.

A.3 The double well and quantum oscillation

The double potential well in quantum mechanics nicely demonstrates one of the more surprising phenomena associated with quantisation: namely, *quantum oscillation*. Consider the situation in which there are two identical square wells (in one dimension for simplicity) sufficiently separated so that the form of the solution to the Schrödinger equation locally in the neighbourhood of either well is not appreciably affected by the presence of the other. In other words, the solution within the regions of the wells is very similar to that of a single isolated well. The form of the well and the x dependence of the corresponding two lowest-energy eigenstates are represented in Fig. A.1.

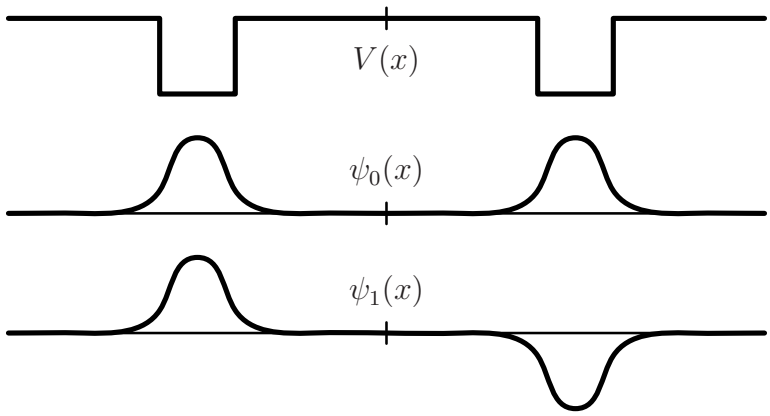


Figure A.1: The double potential well and the x dependence of the corresponding two lowest-energy eigenstates, $E_1 > E_0$.

The solutions can be represented generically as

$$\psi_{1,2}(x,t) \;=\; u_{1,2}(x)\,\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_{1,2}t}, \tag{A.3.1}$$

where the precise form of $u_{1,2}(x)$ is entirely irrelevant for the present purposes.

Now, these represent the "unperturbed" eigenstates of the system and are those with independent temporal evolution. However, if now some "interaction" with the system introduces a particle into *one* of the wells (the left, say) the state induced does not correspond to any single pure eigenstate. In other words, a different basis

is necessary to describe this external interaction:

$$u_{\mathrm{L,R}}(x) \;=\; \tfrac{1}{\sqrt{2}}\left[u_1(x) \pm u_2(x)\right]. \tag{A.3.2}$$

Such a state describes the system at the instant $t=0$, when the particle is introduced into the left-hand well. At later times the evolution is given by (we shall now take the initially left-hand case for definiteness)

$$\psi_{\mathrm{L}}(x,t) \;=\; \tfrac{1}{\sqrt{2}}\left[u_1(x)\,\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_1 t} + u_2(x)\,\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_2 t}\right]. \tag{A.3.3}$$

However, if we wish to know the probability of finding the particle in one or other well, we should decompose over the L/R basis:

$$\begin{aligned}\psi_{\mathrm{L}}(x,t) \;&=\; \tfrac{1}{2}\left[\Big(u_{\mathrm{L}}(x)+u_{\mathrm{R}}(x)\Big)\,\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_1 t} + \Big(u_{\mathrm{L}}(x)-u_{\mathrm{R}}(x)\Big)\,\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_2 t}\right] \\ &=\; \tfrac{1}{2}\left[u_{\mathrm{L}}(x)\left(\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_1 t}+\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_2 t}\right) + u_{\mathrm{R}}(x)\left(\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_1 t}-\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_2 t}\right)\right]. \end{aligned} \tag{A.3.4}$$

Introducing now the average energy $\bar{E} \equiv \frac{1}{2}(E_1+E_2)$ and the energy difference $\Delta \equiv E_2 - E_1$, we finally obtain

$$\psi_{\mathrm{L}}(x,t) \;=\; \left[u_{\mathrm{L}}(x)\cos\!\left(\tfrac{\Delta}{2\hbar}\,t\right) - \mathrm{i}\,u_{\mathrm{R}}(x)\sin\!\left(\tfrac{\Delta}{2\hbar}\,t\right)\right]\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}\bar{E}t}. \tag{A.3.5}$$

One sees that the coefficient of, for example, $u_{\mathrm{L}}(x)$, which determines the probability of finding the particle in the left-hand well, oscillates in time:

$$\mathcal{P}_{\mathrm{L}}(t) \;=\; \cos^2\!\left(\tfrac{\Delta}{2\hbar}\,t\right) \;=\; \tfrac{1}{2}\left[1-\cos\!\left(\tfrac{\Delta}{\hbar}\,t\right)\right]. \tag{A.3.6}$$

Note that the frequency is thus $\Delta/\hbar$ (and not half that). In other words, the particle effectively oscillates between the two potential wells with a frequency determined by the difference of the natural frequencies associated with the two states involved. Note too the close parallel with the case of coupled oscillators in classical mechanics.

To conclude this section, let us underline a common aspect of such phenomena: the original physical system possesses a symmetry (in this case under parity or spatial inversion), which the state created externally does *not* respect (*i.e.* it is not an eigenstate of the basic Lagrangian). This is a sufficient (and in fact necessary) condition for the induced mixing, which lies at the heart of the quantum oscillation phenomenon.

A.4 The partial-wave expansion

A.4.1 Scattering in perturbation theory

In quantum mechanics scattering by weak finite-range potentials may be treated in a perturbative manner. That is, we assume to have solved the free-field equations and that the scattering potential maybe treated as a small perturbation. In this section we shall present a simplified derivation of the cross-section for the case of a spherically symmetric potential. The basis is the knowledge we have of solutions to the Schrödinger equation in such a case—the angular part is provided by just the spherical harmonic functions already encountered in the solution of the hydrogen atom. Moreover, for a finite-range potential the asymptotic form of the radial wave-function is also predetermined.

We consider then the simple case of a particle of momentum $\boldsymbol{k}$ incident on a spherically symmetric potential $V(r)$, centred at the origin. The initial state, in the absence of $V(x)$, may be represented as a plane-wave

$$u_0(\boldsymbol{x}) \;=\; \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\,. \tag{A.4.1}$$

We suppress the normalisation since our aim is to calculate a cross-section, which is a ratio of fluxes. In the presence of the scattering potential, but *outside* its range, the full solution to the Schrödinger equation will be of the form

$$u(\boldsymbol{x}) \;\xrightarrow{r\to\infty}\; \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} + f(\theta,\phi)\frac{1}{r}\,\mathrm{e}^{\mathrm{i}kr} \tag{A.4.2}$$

where θ and ϕ are the polar scattering angles in the laboratory frame with respect to $\boldsymbol{k}$, which we then take to be along the z-axis, while $r=|\boldsymbol{x}|$. That is, we have the incident plane-wave plus an outgoing spherical wave centred on the origin.

In perturbation theory one can always calculate the form of $f(\theta,\phi)$ from first principles if the potential is known. We shall just quote the result here: in the Born (or leading-order) approximation we have

$$\begin{aligned} f(\theta,\phi) \;&=\; -\frac{2m}{\hbar^2}\,\frac{(2\pi)^3}{4\pi}\,\langle\boldsymbol{k}'|V|\boldsymbol{k}\rangle \\ &=\; -\frac{2m}{\hbar^2}\,\frac{(2\pi)^3}{4\pi}\int \mathrm{d}^3\boldsymbol{x}'\,\frac{\mathrm{e}^{-\mathrm{i}\boldsymbol{k}'\cdot\boldsymbol{x}'}}{(2\pi)^{3/2}}\,V(\boldsymbol{x}')\,\frac{\mathrm{e}^{+\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}'}}{(2\pi)^{3/2}}, \end{aligned} \tag{A.4.3}$$

where the final-state momentum $\boldsymbol{k}'$ is direct along (θ,ϕ) and energy-momentum conservation requires that $k=k'$. Gathering together exponentials and constant

factors, this becomes

$$f(\theta,\phi) \;=\; -\frac{2m}{4\pi\hbar^2}\int \mathrm{d}^3\boldsymbol{x}'\; \mathrm{e}^{\mathrm{i}\left(\boldsymbol{k}-\boldsymbol{k}'\right)\cdot\boldsymbol{x}'}\; V(\boldsymbol{x}'). \tag{A.4.4}$$

That is, as remarked elsewhere in these notes, the Born-approximation to the scattering amplitude is just a Fourier transform of the potential, in the variable $\boldsymbol{q}=\boldsymbol{k}-\boldsymbol{k}'$, the momentum transfer. In practice then, we may model a potential and calculate the corresponding phase shifts; alternatively, at least in principle, measured phase shifts may be used to reconstruct (or fit) a potential model.

Exercise A.4.1. *Using the Yukawa potential,*

$$V_{Yuk.}(r) \;\equiv\; \frac{V_0\,\mathrm{e}^{-\mu r}}{r}, \tag{A.4.5}$$

and taking the limit $\mu\to 0$, show that the quantum expression for Rutherford scattering takes on the same form as the classical result.

Often, however, we do not possess *a priori* a realistic model of the scattering potential, *e.g.* for pion scattering off a nucleon. In such cases we require a suitably parametrised description of the cross-section. In what follows we shall exploit our knowledge of the general solution to the Schrödinger equation in the case of a spherically symmetric potential to construct a faithful and simple parametrisation, which may then be fit to experimental data.

A.4.2 The partial-wave formula

Now, the cross-section is defined in terms of the ratio of the scattered flux divided by the incident flux. Recall that in quantum mechanics the flux corresponding to a wave-function solution to the Schrödinger equation has the form

$$\boldsymbol{j}(\boldsymbol{x}) \;=\; \frac{\hbar}{2m\mathrm{i}}\left[u^*(\boldsymbol{x})\boldsymbol{\nabla}u(\boldsymbol{x}) - u(\boldsymbol{x})\boldsymbol{\nabla}u^*(\boldsymbol{x})\right]. \tag{A.4.6}$$

In Eq. (A.4.2) for $u(\boldsymbol{x})$, we should consider the two parts as spatially separated: the incident wave will, in practice, be collimated along a narrow region around the z-axis, where the outgoing spherical wave will have little weight. The incident flux (j_{inc}) therefore simply corresponds to the plane-wave piece while the scattered flux (j_{scatt}) corresponds to the other, with no interference terms (which are negligible as far as the present discussion is concerned, but see later). The partial cross-section

for scattering into a given solid angle $\mathrm{d}\Omega(\theta,\phi)$ is defined as

$$\begin{aligned} \mathrm{d}\sigma &= \frac{\text{scattered flux in } \mathrm{d}\Omega}{\text{incident flux}} \\ &= \frac{j_{\text{scatt}}\, r^2 \mathrm{d}\Omega}{j_{\text{inc}}} = \left|f(\theta,\phi)\right|^2 \mathrm{d}\Omega\,. \end{aligned} \tag{A.4.7}$$

The differential cross-section is therefore quite simply

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \left|f(\theta,\phi)\right|^2. \tag{A.4.8}$$

We thus see that the object containing all the necessary information is just the angular modulation $f(\theta,\phi)$.

The spherical symmetry of the problem suggests transformation to a basis of solutions in terms of spherical harmonic functions. In the case of a plane-wave, we have the following decomposition

$$\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} = \mathrm{e}^{\mathrm{i}kr\cos\theta} = \sum_{\ell=0}(2\ell+1)\,\mathrm{i}^{\ell}\, j_\ell(kr)\, P_\ell(\cos\theta), \tag{A.4.9}$$

where $P_\ell(\cos\theta)$ are Legendre polynomials and $j_\ell(kr)$ are spherical Bessel functions, the solutions to the reduced radial equation. Note that the imaginary factor may also be rewritten as

$$\mathrm{i}^{\ell} \equiv \mathrm{e}^{\frac{1}{2}\mathrm{i}\ell\pi} \tag{A.4.10}$$

and asymptotically the Bessel functions take the form

$$j_\ell(kr) \xrightarrow{r\to\infty} \frac{\sin(kr-\frac{1}{2}\ell\pi)}{kr} = \frac{1}{2\mathrm{i}kr}\left[\mathrm{e}^{+\mathrm{i}(kr-\frac{1}{2}\ell\pi)} - \mathrm{e}^{-\mathrm{i}(kr-\frac{1}{2}\ell\pi)}\right]. \tag{A.4.11}$$

Now, considering the full wave-function in the presence of the scattering potential, we may make a similar expansion and write for r large

$$u(\boldsymbol{x}) = \sum_{\ell=0}^{\infty}(2\ell+1)\,\mathrm{i}^{\ell}\frac{1}{2\mathrm{i}kr}\left[\mathrm{e}^{+\mathrm{i}(kr-\frac{1}{2}\ell\pi+2\delta_\ell)} - \mathrm{e}^{-\mathrm{i}(kr-\frac{1}{2}\ell\pi)}\right] P_\ell(\cos\theta). \tag{A.4.12}$$

In writing this expression we have taken into account certain simple general properties of the solutions to the Schrödinger equation. First of all, each term in the sum corresponds to a component of well-defined orbital angular momentum ($\ell\hbar$) and, since angular momentum is conserved, term-by-term the partial amplitudes cannot change in magnitude but only in phase. Moreover, the two terms in square brackets on the right-hand side represent outgoing and incoming waves

respectively. Only the outgoing wave has had the opportunity to interact with the potential $V(\boldsymbol{x})$ and thus only these components may have a phase shift, $2\delta_\ell(E_k)$, the factor 2 is for later convenience. Note that the phase shift can (and therefore will) only be a function of the incident beam energy.

Comparing now (A.4.2) and (A.4.12), taking into account (A.4.9), we find

$$f(\theta,\phi) \;=\; \sum_{\ell=0}^{\infty}(2\ell+1)\,\mathrm{i}^{\,\ell}\frac{1}{2\mathrm{i}k}\left[\mathrm{e}^{\mathrm{i}(-\frac{1}{2}\ell\pi+2\delta_\ell)}-\mathrm{e}^{\mathrm{i}(-\frac{1}{2}\ell\pi)}\right]P_\ell(\cos\theta) \tag{A.4.13}$$

and using the expression for $\mathrm{i}^{\,\ell}$ given earlier, this becomes

$$= \sum_{\ell=0}^{\infty}(2\ell+1)\,\frac{1}{2\mathrm{i}k}\left[\mathrm{e}^{2\mathrm{i}\delta_\ell}-1\right]P_\ell(\cos\theta) \tag{A.4.14}$$

$$= \frac{1}{k}\sum_{\ell=0}^{\infty}(2\ell+1)\,\mathrm{e}^{\mathrm{i}\delta_\ell}\sin\delta_\ell\,P_\ell(\cos\theta). \tag{A.4.15}$$

Let us make a few comments: firstly, we see that there is no ϕ dependence, as would be expected for a spherically symmetric potential—since $\boldsymbol{k}$ is directed along the z-axis, there can be no z-component of orbital angular momentum. Secondly, we see that all the scattering information is contained in the phase-shifts $\delta_\ell(E_k)$; they determine not only the phase of each partial amplitude (through the factor $\mathrm{e}^{\mathrm{i}\delta_\ell}$), but also the magnitude (through the factor $\sin\delta_\ell$). Finally, let us remark on the validity of the above formula: it may be used for two-body scattering, provided we use the centre–of–mass system and reduced mass *etc.*, and even carries over to the relativistic case.

The importance of the partial-wave expansion is that, as just stated, it provides a useful (and, in particular, model-independent) parametrisation of the scattering amplitude, even when nothing is known about the potential. This permits unbiased analysis of experimental scattering data, which may then be compared with theoretical predictions. As it stands though, the expansion contains an infinite number of parameters (the phase shifts δ_ℓ for $\ell=0,1,\ldots,\infty$), which would render any experimental fit totally impracticable. However, for a finite-range potential (say $r<a$) and finite incident momentum, the maximum orbital angular momentum that can be generated is

$$\ell_{\max}\,\hbar \;\approx\; a\,|\boldsymbol{k}|. \tag{A.4.16}$$

Therefore, recalling that $\hbar c\sim 200\,\mathrm{MeV\,fm}$, we see that, for example, a typical nuclear-potential of range $O(1\,\mathrm{fm})$ and a beam of momentum 200 MeV could generate up to $\ell=O(1)$. In general then, low-energy scattering off a finite-range

potential involves only a very limited number of partial waves. Incidentally, this also explains why the expansion only has limited use in high-energy hadronic physics: a beam of momentum 20 GeV, say, would involve up to $\ell = O(100)$ terms in the expansion. Note that with a small number of partial waves contributing (each with its own angular dependence), the phase shifts may be extracted directly from the measured angular distributions.

A.4.3 The optical theorem

We can now provide a simplified proof of a very important theorem in scattering. Let us first calculate the total cross-section:

$$\begin{aligned} \sigma \;=\; \int \mathrm{d}\Omega \, \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \;&=\; \int \mathrm{d}\Omega \, \left| f(\theta,\phi) \right|^2 \\ &=\; \int \mathrm{d}\Omega \, \left| \frac{1}{k} \sum_{\ell=0}^{\infty} (2\ell+1)\, \mathrm{e}^{\mathrm{i}\delta_\ell} \sin\delta_\ell \, P_\ell(\cos\theta) \right|^2 , \end{aligned}$$

which, using the orthogonality of the Legendre polynomials, reduces to

$$= \; \frac{4\pi}{k^2} \sum_{\ell=0}^{\infty} (2\ell+1) \sin^2 \delta_\ell . \tag{A.4.17}$$

Consider now the forward amplitude, *i.e.* for $\theta = 0$ (or $\cos\theta = 1$):

$$f(0) \;=\; \frac{1}{k} \sum_{\ell=0}^{\infty} (2\ell+1)\, \mathrm{e}^{\mathrm{i}\delta_\ell} \sin\delta_\ell , \tag{A.4.18}$$

where we have used the fact that $P_\ell(1) = 1$ for all ℓ. The imaginary part of this last expression is just the total cross-section, up to a factor $4\pi/k$:

$$\sigma \;=\; \frac{4\pi}{k} \, \mathrm{Im}\, f(0). \tag{A.4.19}$$

This is precisely the optical theorem: the *total cross-section* is proportional to the *forward scattering amplitude*. Its validity actually extends beyond the simple proof provided here.

The apparent contradiction in a left-hand side that is, by definition, proportional to an amplitude *squared* and a right-hand side *linear* in the amplitude may be reconciled by carefully considering the origin of the scattering cross-section. In Eq. (A.4.2) we see that the general form of the wave-function is a sum of two

terms:

$$\psi(\boldsymbol{x}) \propto \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} + \frac{\mathrm{e}^{\mathrm{i}kr}}{r} f(\theta,\phi). \tag{A.4.20}$$

Now, the flux *loss* in the forward direction (which is clearly proportional to the total cross-section) must be due to the *interference* between these two terms (to see this, consider the difference in flux along the z-axis between points before and after the scattering centre) and thus is indeed linear in $f(\theta,\phi)$. In fact, if we perform the calculation in this way, it becomes clear that the theorem is very general and holds even in the presence of *inelastic* scattering (or absorption). As one might imagine, the expression *optical theorem* is borrowed from classical optics, where the phenomenon is well known: a bright central spot that appears behind a black disc diffracting a light source of suitable wavelength.

A.5 Resonances and the Breit–Wigner form

A.5.1 Resonances in classical mechanics

In classical mechanics the equation of motion for a forced oscillator subject to friction is

$$m\ddot{x} + \gamma\dot{x} + kx \;=\; F\cos\omega t, \tag{A.5.1}$$

Defining the natural frequency of the oscillator as $\omega_0 = \sqrt{k/m}$, the solution is

$$x \;=\; x_{\max}\cos(\omega t + \phi), \tag{A.5.2}$$

where the phase difference ϕ is given by

$$\tan\phi \;=\; \frac{-\gamma\omega}{m(\omega^2 - \omega_0^2)} \tag{A.5.3}$$

and the oscillation *amplitude* is

$$x_{\max} \;=\; \frac{F}{[m^2(\omega^2-\omega_0^2)^2 + \gamma^2\omega^2]^{1/2}}. \tag{A.5.4}$$

The total energy $(E_{\mathrm{kin}} + E_{\mathrm{pot}})$ of the oscillator is thus

$$E \;=\; \frac{kF^2}{[m^2(\omega^2-\omega_0^2)^2 + \gamma^2\omega^2]}. \tag{A.5.5}$$

Notice that the presence of dissipation (in the form of friction, $\gamma \neq 0$) tames the otherwise divergent behaviour for $\omega = \omega_0$.

A.5.2 BW resonances in quantum mechanics

A similar behaviour occurs in quantum mechanics for the production of intermediate so-called resonant (virtual) states when the natural energy of the virtual state is near to that of the real energy of the system. A formal description may be provided by considering the variation of phase shifts in the partial-wave decomposition of scattering amplitudes. We shall start from the standard form of the elastic scattering amplitude in quantum mechanics, taken for some particular partial wave ℓ (recall ℓ is the total orbital angular momentum quantum number):

$$f_\ell \;=\; \frac{\hbar}{2\mathrm{i}p}\,(2\ell+1)\left(a_\ell\,\mathrm{e}^{2\mathrm{i}\delta_\ell}-1\right)P_\ell(\cos\theta), \tag{A.5.6}$$

where p is the centre–of–mass initial-state momentum, a_ℓ is the amplitude of the of the ℓ-th partial scattered wave ($0\leq a_\ell\leq 1$, $a_\ell<1$ implies absorption) and δ_ℓ is the so-called phase-shift, a function of energy, which contains all relevant information on the scattering potential.

In the purely elastic case, *i.e.* with *zero* absorption (*i.e.* $a_\ell=1$), the corresponding partial cross-section is then

$$\sigma_\ell^{\mathrm{el}} \;=\; \frac{\pi\hbar^2}{p^2}\,(2\ell+1)\left|\mathrm{e}^{2\mathrm{i}\delta_\ell}-1\right|^2. \tag{A.5.7}$$

This expression has a maximum whenever $\delta_\ell=(n+\frac{1}{2})\pi$ with n integer. The maximum value is

$$\sigma_\ell^{\max} \;=\; (2\ell+1)\,\frac{4\pi\hbar^2}{p^2}. \tag{A.5.8}$$

Note that for the case of *total* absorption (*i.e.* $a_\ell=0$), the cross-section is just a quarter of this.

Experimentally, cross-sections are often observed with a well-pronounced peak at some particular centre–of–mass energy. The question then is how one might describe such effects in the absence of a complete theory of the interaction involved. From the above formula, we immediately deduce that if the cross-section attains a maximum at some energy for some partial wave, then the corresponding phase-shift evidently passes through a value $(n+\frac{1}{2})\pi$, so that $\cot\delta_\ell$ passes through a zero. Now, the only independent variable for each individual partial wave is the centre–of–mass energy (the angular dependence is already coded into the spherical harmonics) so the phase-shift δ_ℓ is only a function of E. Thus, those *energies* for which $\delta_\ell=(n+\frac{1}{2})\pi$ and for which the cross-section is therefore maximal, correspond to *resonances*. We shall assume that such a resonance is sufficiently well separated from any others so that it dominates the cross-section for $E\sim E_0$, the

resonant energy. It is instructive to study the behaviour of the amplitude in the neighbourhood of such a point.

The identity

$$\frac{\mathrm{e}^{2\mathrm{i}\delta}-1}{2\mathrm{i}} \equiv \frac{1}{\cot\delta - \mathrm{i}} \tag{A.5.9}$$

may be used to rewrite the elastic partial-wave amplitude as

$$f_\ell = \frac{\hbar/p}{(\cot\delta_\ell - \mathrm{i})}. \tag{A.5.10}$$

At the resonance $\cot\delta_\ell = 0$; therefore, performing a Taylor expansion about this point in the energy E and retaining only the leading term, linear in $E - E_0$, we may write

$$\cot\delta_\ell(E) \simeq -\frac{2}{\Gamma}(E - E_0). \tag{A.5.11}$$

The sign choice is conventional but physically motivated: indeed, suppose the phase-shift *grows* with energy in the neighbourhood of the resonance. Therefore, $\cot\delta_\ell(E)$ *decreases* and the parameter Γ is *positive*. Inserting this into Eq. (A.5.10) leads to the standard BW form for the amplitude:

$$f_\ell(E) = -\frac{\hbar}{p}\,\frac{\Gamma/2}{(E - E_0) - \mathrm{i}\Gamma/2}\,(2\ell+1)\,P_\ell(\cos\theta). \tag{A.5.12}$$

That is, the elastic cross-section is described by the form

$$\sigma_\ell \simeq \frac{4\pi\hbar^2}{p^2}\,(2\ell+1)\,\frac{(\Gamma/2)^2}{(E - E_0)^2 + (\Gamma/2)^2}. \tag{A.5.13}$$

More simply, we may rewrite this as

$$\sigma_\ell \simeq \sigma_\ell^{\max}\,\frac{(\Gamma/2)^2}{(E - E_0)^2 + (\Gamma/2)^2}. \tag{A.5.14}$$

For any production process that passes through a (well-isolated) resonant channel, one thus finds that the spectrum or cross-section assumes the BW form shown in Fig. A.2 (Breit and Wigner, 1936). Note that, in practice, the peak behaviour is superimposed over the energy dependence of $\sigma_\ell^{\max}$, which from Eq. (A.5.8), for example is typically a rapidly falling function of energy.

Finally, including the multiplicity factors associated with spin states in the case of initial particles of arbitrary spin, the BW approximation to the total cross-

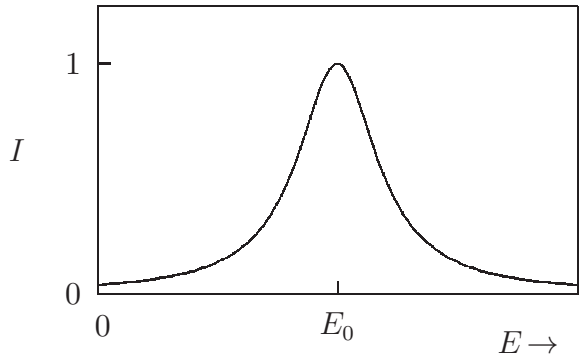


Figure A.2: The standard Breit–Wigner resonance form; the underlying $1/p^2$ behaviour of $\sigma_\ell^{\max}$ has been factored out.

section for particles 1 and 2 scattering via a resonance R may be expressed as

$$\sigma_R \simeq \frac{4\pi\hbar^2}{p^2} \frac{(2J+1)}{(2s_1+1)(2s_2+1)} \frac{(\Gamma_{12}/2)(\Gamma/2)}{(E-E_R)^2+(\Gamma/2)^2}, \tag{A.5.15}$$

where the denominators $(2s_{1,2}+1)$ provide the usual *average* over the initial-state spins $s_{1,2}$ and J is the spin of the resonance R. The numerator factor Γ_{12} represents the *partial* width for resonance two-body decay into particles 1 and 2 (this may be better understood considering the Feynman diag4rams introduced in the following subsection). The possible multiplicity factors for any final-state spins are subsumed (indeed, summed not averaged) in the numerator Γ.

A.5.3 BW resonances in quantum field theory

We shall now try to motivate such a form in particle physics without recourse to detailed calculation. First, recall that for a metastable state the decay rate is the inverse of the mean life (up to factors of $\hbar$ and c): $\Gamma = \tau^{-1}$. Note that here we are not necessarily limited to discussing purely elastic processes, the main aspect is the resonant nature of the intermediate state. The probability density for the decaying state then takes the following form:

$$\mathcal{P}(t) \propto \mathrm{e}^{-\Gamma t}. \tag{A.5.16}$$

In quantum mechanics this should simply be the squared modulus of the wave-function describing the state. Thus, one is led to the following form for the time

dependence of the wave-function:

$$\psi(t) \propto \mathrm{e}^{-\mathrm{i}Et}\mathrm{e}^{-\frac{1}{2}\Gamma t} = \mathrm{e}^{-\mathrm{i}\left(E-\mathrm{i}\frac{1}{2}\Gamma\right)t}. \tag{A.5.17}$$

Considering a particle of physical mass m_0 in its rest frame, the total energy E may be replaced by m_0. The wave-function is then seen to represent a state of *complex* mass $m \equiv m_0 - \mathrm{i}\Gamma/2$.

If we now make the *plausible* step of using this mass in the propagators appearing in any Feynman diagram where such an unstable particle might propagate internally, we are led to the following substitution (assuming $\Gamma \ll m_0$):

$$\frac{1}{p^2 - m^2} \to \frac{1}{p^2 - m_0^2 + \mathrm{i}m_0\Gamma}. \tag{A.5.18}$$

One can show, by explicit calculation, that the effect of the self-interaction induced by the decay channels (*i.e.* the possibility of temporary spontaneous fluctuations into the decay final states) is precisely this. However, the full armoury of quantum field theory is needed to attack such a problem.

As an example, let us finally examine the effect of such a substitution on the interaction probability of the process $e^+e^- \to Z^0 \to \mu^+\mu^-$ (see Fig. A.3). At lowest

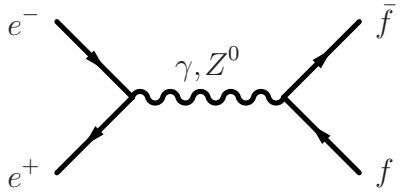


Figure A.3: An example of the Feynman diagram for a typical particle-physics process; in this case, e^+e^- annihilation leading to an intermediate virtual photon or Z^0 boson, which then finally "decays" into a fermion–antifermion pair.

order in perturbation theory the propagator associated with the intermediate Z^0 state is normally:

$$\frac{1}{q^2 - m_Z^2}, \tag{A.5.19}$$

where q^μ is the Z^0 four-momentum and m_Z its mass. According to the above discussion, one should thus adopt the following form:

$$\frac{1}{s - m_Z^2 + \mathrm{i}m_Z\Gamma_Z}, \tag{A.5.20}$$

where use has been made of the Mandelstam variable

$$s \equiv (p_1 + p_2)^2 \equiv q^2 \equiv E_{\mathrm{CM}}^2. \tag{A.5.21}$$

The interaction probability is proportional to the modulus squared of the amplitude and thus we should really examine

$$\left|\frac{1}{s-m_Z^2+\mathrm{i}m_Z\Gamma_Z}\right|^2 = \frac{1}{(s-m_Z^2)^2+m_Z^2\Gamma_Z^2}. \tag{A.5.22}$$

Note once again that the presence of dispersion (in this case particle decay) tames a potential divergence for $E_{\rm CM}=m_Z$. This also demonstrates that higher-order corrections are not merely a theoretical luxury to be easily foregone. The form shown in Eq. (A.5.22) is relevant for relativistic field-theory calculations; noting that for energies near the *pole* mass ($E_{\rm CM}\sim m_Z$)

$$\begin{aligned} s-m_Z^2 &= E_{\rm CM}^2-m_Z^2 \\ &= (E_{\rm CM}+m_Z)(E_{\rm CM}-m_Z) \simeq 2m_Z(E_{\rm CM}-m_Z), \end{aligned} \tag{A.5.23}$$

one readily obtains the standard BW form:

$$\propto \frac{1}{(E_{\rm CM}-m_Z)^2+\frac{1}{4}\Gamma_Z^2}. \tag{A.5.24}$$

The final complete form for $e^+e^-\to Z\to ab$ is then

$$\sigma_{e^+e^-\to Z\to ab} \simeq \frac{4\pi\hbar^2}{p^2}\,\frac{(2s_Z+1)}{(2s_1+1)(2s_2+1)}\,\frac{(\Gamma_{12}/2)(\Gamma_{ab}/2)}{(E-m_Z)^2+(\Gamma_{\rm tot}/2)^2}, \tag{A.5.25}$$

where s_Z and $s_{1,2}$ are the spins of the Z^0, electron and positron respectively. The partial widths in the numerator, Γ_{12} and Γ_{ab} are those for Z^0 decay into e^+e^- and ab final states while for the denominator the total width $\Gamma_{\rm tot}$ must be used in *all* cases.

One further refinement is necessary for greater precision in those cases where the momenta of the outgoing particles varies appreciably over the width Γ of the resonance. Since Γ may also be correctly interpreted as the decay rate, it will depend on the final-state momentum $\boldsymbol{p}$ (evaluated in the rest-frame). One can then show, by explicit calculation (see Flatté, 1976), that for a resonance of mass M the generally correct form is

$$\Gamma(s) = \left|\frac{\boldsymbol{p}}{\boldsymbol{p}_0}\right|^{2\ell+1}\Gamma(M^2), \tag{A.5.26}$$

where $\Gamma(M^2)$ is the on-shell decay rate, *i.e.* for $s=M^2$; $\boldsymbol{p}$ ($\boldsymbol{p}_0$) is the off-shell (on-shell) final-state momentum and ℓ is the intrinsic spin of the resonance. The effect

of this is seen in a certain skew of the otherwise symmetric BW form. If not taken into account, it would induce a shift in the extracted mass of the resonance.

A few final observations are in order before concluding. Here we have only considered the simplified case of a single resonance contributing to a given channel. Now, while it is true that only one intermediate state is likely to be *resonant* at any one energy, if two or more resonances are near in mass (*i.e.* with respect to their widths), then interference effects can become important. In such cases care must be taken to sum over all possible contributing *amplitudes*, after which the cross-section (or decay rate) may be calculated from the square of the total amplitude so obtained.

Now, it is indeed quite likely that more than one process contributes to the overall width or rate Γ for the decay of any given resonance while we may only be interested experimentally in a particular channel. In such a case the procedure is quite simple: the width Γ appearing in the denominator, being effectively the imaginary part of the physical mass, must be taken as the *total* decay width. However, the width appearing in the numerator should be that corresponding to the particular channel under study.

Finally, there is evidently an implicit approximation in the derivation of the BW form. Quite simply, the intermediate objects should not be too broad—one often speaks of the "narrow-resonance approximation." Now, while the above correction for the intrinsic energy dependence of the width goes some way to allowing even relatively broad states to be accurately included, this can in no way take into account the non-elementary nature of many of the particles involved. The BW form has its base in a treatment of all processes as involving only elementary particle and although something can be done to include form-factor like effects there is no well-defined way in which the substructure of the resonance (nor indeed of the initial and/or final states) may be reliably accounted for.

A.6 Bibliography

Breit, G. and Wigner, E.P. (1936), *Phys. Rev.* **49**, 519.

Flatté, S.M. (1976), *Phys. Lett.* **B63**, 224.

Gordon, W. (1926), *Z. Phys.* **40**, 117.

Klein, O. (1927), *Z. Phys.* **41**, 407.

Olive, K.A. *et al.*, Particle Data Group (2014), *Chin. Phys.* **C38**, 090001.

Rayleigh (1871), *Phil. Mag.* **41**, 107; 274.

Sellmeier, W. (1871), *Pogg. Ann. Physik. Chem.* **143**, 271; *ibid.* **145** (1872) 399; *ibid.* **145** (1872) 520; *ibid.* **147** (1873) 386; *ibid.* **147** (1873) 525.

Appendix B

Background Notes

B.1 The muon

As a forerunner to the problem of describing the strong interaction, as we see it today, let us examine the case of the muon. Discovered independently by Anderson and Neddermeyer (1936) and Street and Stevenson (1937) in cosmic-ray experiments, the muon was considered a prime candidate as the particle (then known as the *mesotron*) suggested by Yukawa (1935)* as the exchange field responsible for the strong interaction. With a mass of 106 MeV, it appeared more similar to the baryons than to the other known charged lepton at that time, the electron. According to Yukawa's theory, such a mass would lead to a range of action around 1 fm or so, which corresponded well to the observed finite range of the strong nuclear force.

The question then arises as to how to ascertain whether or not such an interpretation is correct. Apart from the experimental evidence, which we shall shortly discuss, there are theoretical reasons (not, however, available at that time) for *not* accepting such a role for the muon. Conservation of angular momentum requires that the exchange particle have integer spin—we now know that the muon is a fermion. Moreover, the flavour or isospin symmetry of the strong interactions requires that the exchange particle have integer isospin too (the proton and neutron belong to an isospin one-half doublet). A singlet state would not interact (or at best its interactions would be suppressed) and therefore it should have at least one unit of isospin. Finally, the multiplicity of an isospin-one system is three while there exist only two states for the muon: $\mu^{\pm}$, there being no neutral state.

Evidently though, it was necessary to examine the strength of its interaction to really understand the nature of the muon. Its decay is seen to be weak ($\tau_{\mu} \sim 2\,\mu$s),

* The 1949 Nobel Prize for Physics was awarded to Hideki Yukawa for "his prediction of the existence of mesons on the basis of theoretical work on nuclear forces."

but this alone cannot be interpreted as excluding its strong interaction; the type of interaction through which a particle may decay is also determined by the various conservation laws. In this case conservation of energy is sufficient to exclude a strong decay: the muon is lighter than all known strongly interacting particles and, in fact, decays primarily to $e\bar{\nu}_e\nu_\mu$, none of which are strongly interacting. We must then study the behaviour of the muon in a strongly interacting environment, *e.g.* inside the nucleus. The strong interaction has a time scale of the order of 10^{-23} s (a typical decay time for the heavier hadrons such as Δ^{++} *etc.*), so we might expect a muon to be absorbed on such a time scale by a nucleus (inside which the energy balance may easily be redressed).

In 1947 Conversi, Pancini and Piccioni set out to measure the lifetime of what had then been dubbed the *mesotron* (the present-day muon) by studying its stopping behaviour in nuclear matter; they had already measured the free lifetime. The experiments they performed turned out to be a disproof of the strong-interaction hypothesis.

In matter (negatively charged) muons lose energy via electromagnetic interactions until they are eventually captured by an atom and become bound, just as an electron. Since the muon is evidently distinguishable from the electrons, it does not suffer Pauli exclusion and, via photon emission, may cascade down to the ground state. At this point, owing to its relatively large mass, it is much nearer to the nucleus than the corresponding K-shell electron would be. Indeed, for a charged particle of mass m, the Bohr radius in an atom with atomic number Z is

$$R^B = \frac{Z}{m\,\alpha}, \tag{B.1.1}$$

which leads to $R_e^B \simeq Z{\times}0.6{\times}10^{-10}$ m for an electron. And since then

$$R_\mu^B = \frac{m_e}{m_\mu}\,R_e^B, \tag{B.1.2}$$

for the given mass ratio of approximately 200, this leads to $R_\mu^B \simeq Z{\times}3{\times}10^{-13}$ m for a muon.

The strong interaction is evidently negligible at such distances; however, the smaller radius implies that the wave-function for the muon will have a higher density inside the nucleus than would the corresponding electron, by roughly a factor 200^3. As we shall now show, this represents a sufficiently long time spent *inside* the nucleus to test the strong-interaction hypothesis. Conversi *et al.* measured a decay lifetime (*i.e.* for the disappearance or so-called K-capture of muons) of 0.88 µs (to be compared to the free decay time of 2 µs). Thus, some form of interaction evidently occurs. In order to evaluate the strength of this interaction it is necessary to estimate the mean free path of muons in nuclear matter.

A simple (back-of-the-envelope) estimate may be performed by considering the volume of the nucleus itself as a fraction f of the total volume occupied by a K-shell muon. This is just the ratio $(R_{\rm nucl}/R_\mu^B)^3$. Recall that empirically $R_{\rm nucl}=R_0A^{\frac{1}{3}}$, where A is just the atomic mass and $R_0\simeq 1.2\,{\rm fm}$. Using this and Eq. (B.1.2) we obtain

$$f \;=\; \left(\frac{R_{\rm nucl}}{R_\mu^B}\right)^3 \;=\; 0.27\,A\left(\frac{Z}{137}\right)^3. \tag{B.1.3}$$

For aluminium one has $Z=13$ and $A=27$, giving

$$f \;\simeq\; 6\times 10^{-3}. \tag{B.1.4}$$

This will be roughly the fraction of its lifetime that a muon spends inside an aluminium nucleus. Already, one might anticipate that, as a fraction of a microsecond, this still leads to a survival time inside the nucleon many orders of magnitude larger that the 10^{-23} s one might have expected. However, let us first estimate the mean free path (for a strongly interacting particle it should not be much larger than about 1 fm). The mean velocity of the muon may be estimated from the Heisenberg uncertainty principle by setting $p_\mu\sim\hbar/R_\mu^B$. In a non-relativistic approximation, this leads to an estimated velocity $v_\mu\sim Z\alpha$, which, given that typically $Z\ll\alpha^{-1}$, justifies the approximation *a posteriori*. Finally, the mean free path is

$$\Lambda \;=\; vf\tau, \tag{B.1.5}$$

where τ is the lifetime of such a K-shell state.

Now, since decay rates are additive and inversely proportional to lifetimes (*i.e.* $\Gamma=\Gamma_d+\Gamma_c$ and $\Gamma\propto 1/\tau$), the rule for combining lifetimes is

$$\frac{1}{\tau} \;=\; \frac{1}{\tau_d}+\frac{1}{\tau_c}, \tag{B.1.6}$$

where τ_d and τ_c stand for the free-decay and capture lifetimes respectively. The measured values are $\tau_d=2.16\,\mu{\rm s}$ and $\tau=0.88\,\mu{\rm s}$ in aluminium. We thus obtain

$$\tau_c \;\sim\; 1.5\;\mu{\rm s}. \tag{B.1.7}$$

Inserting this into the formula for the mean free path, Eq. (B.1.5), leads to

$$\Lambda \;\sim\; 20-30\;{\rm cm}. \tag{B.1.8}$$

In other words, muon survival inside the nucleus far exceeds the expectations for a strongly interacting particle (Fermi *et al.*, 1947). Indeed, the K-capture time τ_c is more suggestive of a *weak* interaction; this came very much as a surprise:

> "This result was completely unexpected, and we believed at first that there might be some malfunction in our apparatus."
>
> Marcello Conversi

Indeed, the revelation that apparently the muon therefore had no particular role in the general scheme of particle physics prompted Isidor Rabi to comment publicly, in astonishment:

> "Who ordered that?"

B.2 Isospin and SU(2)

In nuclear and particle physics a number of symmetries are apparent. One of the simplest is the existence of a large number of so-called *mirror nuclei*: that is, pairs of nuclei that differ only by interchange of the number of protons and neutrons. An example is

$$^{11}_{5}\mathrm{B} \leftrightharpoons {}^{11}_{6}\mathrm{C}. \tag{B.2.1}$$

While the chemical properties of ^{11}B and ^{11}C atoms are obviously rather different, the nuclei are very similar indeed. When one takes into account the variations due to the effects of Coulomb repulsion, one might even claim they are identical, as far as the strong interaction is concerned, that is.

Evidently, such a symmetry must have to do with a corresponding symmetry at the nucleon level. That is, we assume it to be just the manifestation of a deeper proton–neutron symmetry. In fact, at the hadronic level in general we see much the same sort of mirror behaviour in various particles:

- The masses of the neutron and the proton are very similar; indeed, although the proton has a positive charge and the neutron is neutral, they are almost identical in all other respects. In fact, inasmuch as electromagnetic effects may be ignored with respect to the strong interaction and taking into account that, as we now know, there is a small up–down quark mass difference, they might be considered as two different states of the same fundamental field.

- The strong interaction between any pair of nucleons is identical, independently of whether they are protons or neutrons. That is, the proton–proton, proton–neutron and neutron–neutron forces are the same. Again, to see this phenomenologically, one must first subtract electromagnetic effects.

- In a similar fashion, the three known pion states π^+, π^0 and π^- are also very similar. Indeed, the two charged pions have exactly the same mass while the neutral pion is just slightly lighter. Moreover, apart from very systematic differences, which are in fact *explained* by the *isospin* picture we shall now

discuss, their strong interactions with matter (protons and neutrons) are also the same.

In 1932 Heisenberg thus introduced the notion of *isotopic spin* (or *isobaric spin*) to explain these observations. The standard contraction of the name is now *isospin*.

We know from quantum mechanics that when the Hamiltonian of a system possesses a discrete symmetry, *e.g.* with respect to spatial inversion, this manifests itself through a degeneracy of the energy states of the system. Consider, for example, the various energy levels of the hydrogen atom.

In particle physics mass is equivalent to energy (since $E=mc^2$) and so the near mass degeneracy of the neutron and proton indicates a symmetry of the Hamiltonian describing the strong interactions. The neutron does have a slightly higher mass and so the degeneracy is not exact. However, here (as the case would be in general for quantum mechanics) the appearance of a symmetry may be imperfect as it can be perturbed by other forces, giving rise to slight differences between otherwise degenerate states. Indeed, the proton is charged while the neutron is not and therefore electromagnetism must play a different role.

Heisenberg noted that the mathematical description of the observed symmetry rendered it similar to the symmetry structure of orbital angular momentum or *spin*, hence the term isotopic spin or isospin. In mathematical terms, the isospin symmetry is due to an invariance of the strong-interaction Hamiltonian under the action of the (Lie) group SU(2). The neutron and the proton are placed in a doublet (a spin-½ or fundamental representation) of SU(2). The pions, being evidently a triplet are assigned a spin-one or adjoint (rank-3) representation of SU(2).

The mathematical structure (or algebra) is then quite simply that of the usual angular momentum. Isospin is described by two quantum numbers: the total isospin I and the spin projection along the quantisation axis I_3. The proton and neutron thus both have $I=\frac{1}{2}$; the proton has $I_3=+\frac{1}{2}$ or 'isospin up' while the neutron has $I_3=-\frac{1}{2}$ or 'isospin down'. The pions naturally belong to the $I=1$ triplet, with π^+, π^0 and π^- having $I_3=+1$, 0 and -1 respectively. In Dirac notation, for the nucleon pair, we write

$$|p\rangle \;=\; |\tfrac{1}{2},+\tfrac{1}{2}\rangle \quad \text{and} \quad |n\rangle \;=\; |\tfrac{1}{2},-\tfrac{1}{2}\rangle \tag{B.2.2}$$

while the pion triplet becomes

$$|\pi^+\rangle \;=\; |1,+1\rangle, \quad |\pi^0\rangle \;=\; |1,\;0\rangle \quad \text{and} \quad |\pi^-\rangle \;=\; |1,-1\rangle. \tag{B.2.3}$$

The pairs of quantum numbers above then have the same *mathematical* significance as the j,m pairs for angular-momentum states.

An important consequence of isospin symmetry and its mathematical structure is the possibility to apply Clebsch–Gordan coefficients to combinations (or composition) of particles. A simple example is the strong decay of the so-called Δ resonances (spin-$\frac{3}{2}$, isospin-$\frac{3}{2}$), which may be generically described as $\Delta \to N\pi$ (N being a nucleon, p or n). For concreteness, let us consider the state Δ^+, whose *isospin* designation is indicated as $|\frac{3}{2},\frac{1}{2}\rangle$. There are two distinct possible final states: $p\pi^0$ and $n\pi^+$ or $|\frac{1}{2},+\frac{1}{2}\rangle|1,0\rangle$ and $|\frac{1}{2},-\frac{1}{2}\rangle|1,1\rangle$ respectively. Now, a glance at a table of Clebsch–Gordan coefficients tells us that a spin-$\frac{3}{2}$ state may be decomposed into the following combination of spin-one and spin-$\frac{1}{2}$ objects:

$$|\tfrac{3}{2},\tfrac{1}{2}\rangle \;=\; \sqrt{\tfrac{2}{3}}\,|\tfrac{1}{2},+\tfrac{1}{2}\rangle|1,0\rangle + \sqrt{\tfrac{1}{3}}\,|\tfrac{1}{2},-\tfrac{1}{2}\rangle|1,1\rangle. \qquad \text{(B.2.4)}$$

The squares of the coefficients provide the branching fractions: namely $\frac{2}{3}$ into $p\pi^0$ and $\frac{1}{3}$ into $n\pi^+$. These fractions are experimentally well verified.

B.3 Bibliography

Anderson, C. and Neddermeyer, S.H. (1936), *Phys. Rev.* **50**, 263; Neddermeyer, S.H. and Anderson, C.D., *Phys. Rev.* **51**, 884.

Conversi, M., Pancini, E. and Piccioni, O. (1947), *Phys. Rev.* **71**, 209; .

Fermi, E., Teller, E. and Weisskopf, V.F. (1947), *Phys. Rev.* **71**, 314; **72**, 399.

Heisenberg, W. (1932), *Z. Phys.* **77**, 1; 156.

Street, J.C. and Stevenson, E.C. (1937), *Phys. Rev.* **51**, 1005; **52**, 1003.

Yukawa, H. (1935), *Proc. Phys. Math. Soc. Jap.* **17**, 48.

Appendix C

Scattering Theory

C.1 Electron scattering

Through the study of α-particle scattering, Rutherford arrived at a new understanding of the atom and its internal structure. However, the α-particle is not a point-like object and can also interact via the strong force. Thus, the details of a small nuclear target obtained via such scattering are clouded by the internal structure of the probe used. On the other hand, the electron has only electromagnetic interactions (the weak interaction may usually be neglected in comparison) and so far it has exhibited a purely point-like behaviour (at least up to the highest energies presently available).

For low-energy scattering ($E \ll m_W$, where the weak force is particular suppressed) the interaction of the electron with a nucleon or nucleus is governed purely by the theory of QED and is therefore completely known. This makes the electron an ideal probe to study the internal structure of the nucleus and, going deeper, of the nucleon.

C.1.1 Non-relativistic point-like elastic scattering

The simplest example of scattering with electrons is the elastic case. Here energy and momentum are transferred from an electron to a nucleus (or nucleon) exchange via (single) photon exchange and the final nucleus remains intact (see Fig. C.1). The four-momenta in the problem are then the initial (final) electron momentum k^μ (k'^μ) and the initial (final) nucleus momentum p^μ (p'^μ). Conservation of energy and momentum requires

$$k^\mu + p^\mu = k'^\mu + p'^\mu. \tag{C.1.1}$$

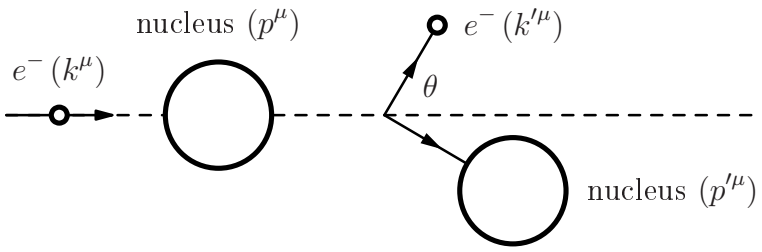


Figure C.1: Elastic electron–nucleus scattering in the laboratory or nucleus rest frame.

Since the momentum of the final nucleus is not usually measured, it is convenient to rewrite this as

$$p'^2 \;=\; M^2 \;=\; (k^\mu - k'^\mu + p^\mu)^2, \tag{C.1.2}$$

where M is the nuclear mass. It is then straight-forward to derive the relation

$$EM \;=\; E'E(1-\cos\theta) + E'M, \tag{C.1.3}$$

where E and E' are the laboratory-frame initial and final electron energies respectively, θ the laboratory-frame electron scattering angle and we have neglected the electron mass, note that in the laboratory frame $p^\mu = (M, \mathbf{0})$. This may then be rearranged to express E' as a function of E and θ:

$$E' \;=\; \frac{E}{1+(1-\cos\theta)E/M}. \tag{C.1.4}$$

which provides the well-known result that, for *elastic* scattering, the final energy is determined by the angle (and *vice versa*).

Let us now simplify to the non-relativistic limit (we shall treat the relativistic case later). The scattering cross-section for electrons is then given by the Rutherford formula (1911), with the obvious substitution $z=-1$ for the electron:

$$\frac{\mathrm{d}\sigma^{e^-}}{\mathrm{d}\Omega} \;=\; \frac{1}{16}\left(\frac{Z\alpha}{E_\infty}\right)^2 \mathrm{cosec}^4\,\tfrac{\theta}{2}. \tag{C.1.5}$$

However, this is the cross-section for a point-like target, whereas we may wish to study the charge *distribution* inside the nucleus or nucleon; we shall discuss later how a distributed charge modifies the formula. Let us first consider though the modifications due to relativistic effects.

C.1.2 Relativistic elastic scattering—the Mott formula

To resolve the internal structure of a nucleus (*i.e.* to be sensitive to energy dependence in the form factors, see later), we require the wavelength of the exchange photon to be small compared to the nuclear size. Since $\hbar c \sim 200\,\mathrm{MeV\,fm}$, we deduce that the energy required is of the order of 100's of MeV. This implies that the electrons will certainly be relativistic, in which case we should really perform a calculation based on the Dirac equation. The relativistic calculation for a point-like object then leads to the Mott cross-section (1929):

$$\frac{\mathrm{d}\tilde{\sigma}}{\mathrm{d}\Omega}^{\mathrm{Mott}} = \left(1 - \beta \sin^2 \tfrac{\theta}{2}\right) \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}^{\mathrm{Ruther.}}, \tag{C.1.6}$$

where the tilde indicates that this is not yet the full (high-energy) Mott formula since we are still neglecting the nuclear recoil—we shall call this the reduced Mott formula. The new factor is due to the conservation of angular momentum and the role played by the spin of the electron. Note that in the ultra-relativistic limit, where $\beta \to 1$, the spin factor becomes simply $\cos^2 \frac{\theta}{2}$. The large-angle Mott cross-section falls off more rapidly than that of Rutherford; indeed, for $\theta = 180^\circ$ it vanishes. Note though that the effect conveniently factorises.

Let us just take a moment to try and understand this behaviour in physical terms. For Dirac theory in the relativistic limit, $\beta \to 1$ (which is evidently equivalent to the limit $m \to 0$), it turns out that the helicity or projection of the particle spin onto the direction of motion $h \equiv \hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{p}}$ is a conserved quantum number if the interactions are of a purely vector or axial-vector type (*e.g.* via photon or weak-boson exchange). Indeed, starting from the Dirac equation, it can be shown that helicity-flip amplitudes are proportional to m/E (where E is some characteristic energy scale of the interaction (*e.g.* the centre–of–mass energy). We thus speak of right- and left-handed fermions as having $h = \pm 1$ respectively and in the massless limit they cannot flip (*i.e.* the two helicities do not communicate). Consider now the extreme cases of forward and backward scattering, in which the incoming electron collides and either continues unaltered or returns in the direction from where it came ($\theta = 0$ or π in Fig. C.2). Now, assuming a spin-zero nucleus, since any orbital angular momentum between the electron–nucleus pair $\boldsymbol{L} = \boldsymbol{r} \wedge \boldsymbol{p}$ must lie in the plane orthogonal to $\boldsymbol{p}$, the spin of the electron must be conserved on its own. For forward scattering this is trivially the case and for backward scattering it is evidently impossible. For intermediate cases, one needs to understand how spin-projection eigenstates are constructed for arbitrary directions. For a state of positive helicity travelling in a direction θ with respect to the chosen quantisation axis ($\hat{z}$ say), we find

$$|+,\theta\rangle = \cos \tfrac{\theta}{2} |+,z\rangle + \sin \tfrac{\theta}{2} |-,z\rangle. \tag{C.1.7}$$

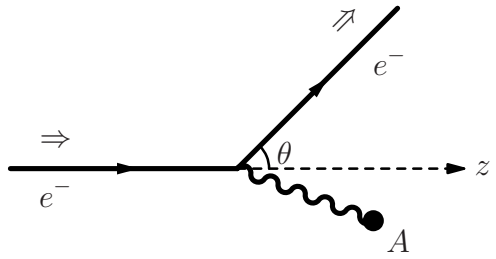


Figure C.2: Helicity conservation in high-energy scattering processes; θ is the electron scattering angle (the blob A represents the target nucleus).

Thus, the amplitude $\langle +,\theta|+,z\rangle = \cos\frac{\theta}{2}$ and, squaring, we have the Mott result. Of course, if the nucleus also possesses an intrinsic angular momentum (due to internal motion or spins of the constituent nucleons), then the situation is a little more complex and one also needs to understand the mechanism by which the nucleus, as a whole, may change or flip its spin projection.

Finally we should take into account the recoil of the struck nucleus and the consequent modification of the final-state phase space. The final, full Mott formula is then (in the target rest frame)

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}^{\text{Mott}} = \frac{E'}{E}\frac{\mathrm{d}\tilde{\sigma}}{\mathrm{d}\Omega}^{\text{Mott}} = \frac{E'}{E}\left(1-\beta\sin^2\tfrac{\theta}{2}\right)\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}^{\text{Ruther.}}. \tag{C.1.8}$$

Again, note the fortunate factorisation of all the new effects.

C.2 Form factors

C.2.1 Elastic scattering off a distributed charge

In describing the case of scattering off a distributed charge, it will be helpful to make two simplifying approximations:

(i) $E_e \ll m_A$—in order to neglect the nuclear recoil,

(ii) $Z\alpha \ll 1$—to permit the Born approximation (*i.e.* single-photon exchange).

We shall also consider the quantum-mechanical treatment, as the concepts of interference and coherence will become relevant here. Our starting point will therefore be Fermi's golden rule:

$$W = \frac{\sigma v_a}{V} = \frac{2\pi}{\hbar}\left|\langle\psi_f|\mathcal{H}_{\text{int}}|\psi_i\rangle\right|^2\frac{\mathrm{d}n}{\mathrm{d}E_f}, \tag{C.2.1}$$

where the third factor in the last line is the density of final states and E_f is the total final energy $K+m_A$ (K is the final kinetic energy of the electron) and, since m_A is constant, $\mathrm{d}E_f = \mathrm{d}E' = \mathrm{d}E$.

We first need the description of the initial and final electron states, which we shall naturally take to be plane-waves:

$$\psi_i \;=\; \frac{\mathrm{e}^{\frac{\mathrm{i}}{\hbar}\boldsymbol{p}\cdot\boldsymbol{x}}}{\sqrt{V}} \quad \text{and} \quad \psi_f \;=\; \frac{\mathrm{e}^{\frac{\mathrm{i}}{\hbar}\boldsymbol{p}'\cdot\boldsymbol{x}}}{\sqrt{V}}, \tag{C.2.2}$$

where the normalisation is one particle in a volume V (of course, V will not appear in the final answer). With this choice, the density of final states is

$$\mathrm{d}n(p) \;=\; V\,\frac{p^2\mathrm{d}p\,\mathrm{d}\Omega}{(2\pi\hbar)^3}, \tag{C.2.3}$$

where $p \equiv |\boldsymbol{p}|$ (recall that we are using a non-relativistic approximation). We therefore have

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \;=\; \frac{V^2E'^2}{(2\pi)^2}\,\big|\mathcal{M}_{fi}\,\big|^2\,\frac{1}{\hbar^4 v}. \tag{C.2.4}$$

We thus need to find the transition matrix element $\mathcal{M}_{fi} = \langle\psi_f|\mathcal{H}_{\mathrm{int}}|\psi_i\rangle$.

For a non-relativistic electron (with charge $-e$) the interaction is given in terms of the electromagnetic scalar potential $\phi(\boldsymbol{x})$: $\mathcal{H}_{\mathrm{int}}(\boldsymbol{x}) = e\phi(\boldsymbol{x})$. We therefore have (suppressing $\hbar$ for clarity)

$$\langle\psi_f|\mathcal{H}_{\mathrm{int}}|\psi_i\rangle \;=\; \frac{e}{V}\int \mathrm{d}^3\boldsymbol{x}\;\mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}}\,\phi(\boldsymbol{x}), \tag{C.2.5}$$

where we have substituted $\boldsymbol{q} = \boldsymbol{p} - \boldsymbol{p}'$, defined to be the *momentum transfer*. It is immediately obvious that this is none other than the Fourier transform of the potential $\phi(\boldsymbol{x})$, which is in turn determined by the charge-density distribution $\rho(\boldsymbol{x})$ that generates it:

$$\boldsymbol{\nabla}^2\phi(\boldsymbol{x}) \;=\; -\rho(\boldsymbol{x}). \tag{C.2.6}$$

The plane-wave form allows us to rewrite the expression for the matrix element directly in terms of $\rho(\boldsymbol{x})$ by noting that

$$\boldsymbol{\nabla}^2\,\mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}} \;=\; -q^2\,\mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}}\,. \tag{C.2.7}$$

If we now apply Green's theorem or integration by parts, we obtain

$$\langle\psi_f|\mathcal{H}_{\mathrm{int}}|\psi_i\rangle \;=\; \frac{e}{Vq^2}\int \mathrm{d}^3\boldsymbol{x}\;\mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}}\,\rho(\boldsymbol{x}), \tag{C.2.8}$$

where $q := |\boldsymbol{q}|$.

It is convenient to define a normalised density,

$$\rho(\boldsymbol{x}) \;=:\; Ze\, f(\boldsymbol{x}), \tag{C.2.9}$$

such that

$$\int \mathrm{d}^3\boldsymbol{x}\, f(\boldsymbol{x}) \;=\; 1. \tag{C.2.10}$$

We thus obtain

$$\langle \psi_f | \mathcal{H}_{\mathrm{int}} | \psi_i \rangle \;=\; \frac{Z\alpha}{Vq^2} \int \mathrm{d}^3\boldsymbol{x}\; \mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}}\, f(\boldsymbol{x}). \tag{C.2.11}$$

The integral on the right-hand side, the Fourier transform of the charge density, is known as the *form factor*:

$$F(\boldsymbol{q}) \;\equiv\; \int \mathrm{d}^3\boldsymbol{x}\; \mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}}\, f(\boldsymbol{x}). \tag{C.2.12}$$

Putting everything together, we have the differential cross-section:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \;=\; \frac{Z^2\alpha^2 E'^2}{q^4} \left|F(\boldsymbol{q})\right|^2. \tag{C.2.13}$$

It is easy to show (neglecting the electron mass) that

$$\boldsymbol{q}^2 \;=\; (\boldsymbol{p} - \boldsymbol{p}')^2 \;=\; 4EE' \sin^2 \tfrac{\theta}{2} \tag{C.2.14}$$

and thus we may finally write

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \;=\; \frac{Z^2\alpha^2}{16E^2 \sin^4 \frac{\theta}{2}} \left|F(\boldsymbol{q})\right|^2, \tag{C.2.15a}$$

or

$$\;=\; \left|F(\boldsymbol{q})\right|^2 \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}^{\mathrm{Ruther.}}. \tag{C.2.15b}$$

In other words, the substructure of the nucleon has the effect of introducing a multiplicative form factor $F(\boldsymbol{q})$, which simply modulates the cross-section. From the definition, we see that the standard Rutherford cross-section is recovered for a point-like distribution, which is just a δ-function (for which the Fourier transform is just unity). Note that this is also the limiting case for low-energy scattering; for $q \ll \hbar/r_{\mathrm{nucl}}$ or $\lambda \gg r_{\mathrm{nucl}}$, $f(\boldsymbol{x})$ does not vary appreciable over the nuclear volume. For q large, however, the q-dependence of $F(\boldsymbol{q})$ makes itself felt and thus changes the energy dependence with respect to that of the point-like formula.

C.2.2 The phenomenology of form factors

We have just seen that the effect of an extended charge distribution is factorisable into a form factor, which only depends on the momentum transfer and which, being a Fourier transform, contains (at least in principle) all necessary information on the charge distribution. That is, if we were able to measure $F(\boldsymbol{q})$, by comparing data with the point-like Mott expression,

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}^{\mathrm{expt}} = \left|F(\boldsymbol{q})\right|^2 \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}^{\mathrm{Mott}}. \tag{C.2.16}$$

over the *entire* range of $\boldsymbol{q}$ from zero to infinity, we could then perform the inverse Fourier transform to obtain $f(\boldsymbol{x})$. Needless to say, this is impossible; the momentum transfer is always limited by the beam (or centre–of–mass) energy. However, the lack of higher frequencies (or shorter wavelengths) simply translates into a lack of resolution. An example of the sort of information one extracts from experiment on $F(\boldsymbol{q})$ is shown in Fig. C.3. Note the pattern of maxima and min-

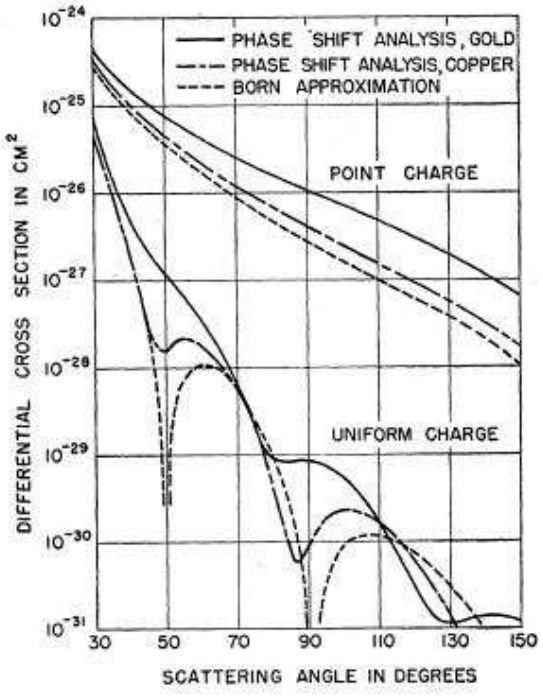


Figure C.3: The theoretical angular dependence of electron–nucleus elastic scattering. The upper curves represent the point-like Mott cross-section while the lower curves correspond to a sphere of uniformly distributed charge (dashed curves) and typical experimental results (full line), where the edge effects of smearing are felt.

ima, reminiscent of diffraction in classical optics. Note also that the rapid fall-off with angle severely limits the maximum q effectively available. The first measurements of this type were made in the fifties at SLAC, with a beam energy of around 500 MeV (for which the effective absolute resolution us $\lambda_{\min} \sim 0.4\,\mathrm{fm}$).

In Fig. C.3 we also see a comparison of the point-like Mott cross-section with that from scattering off a continuous extended charge distribution. At very small angles, for which the momentum transfer is kinematically constrained to be low, the exchange photon has a very long wavelength and thus cannot resolve the internal nuclear structure. In this region there is no difference between a point-like and extended charge distribution. However, if the beam energy is sufficient, with growing angle the photon wavelength may eventually become short enough to resolve the internal structure of point-like charges. In this case the cross-section is much less suppressed, even out to large angles and does not fall off as rapidly as in the unresolved, extended case. This comparison will be important in the discussion of very high-energy electron scattering off a single proton and the question of the proton substructure.

C.2.3 Fitting form factors to trial functions

By appealing to simple parametrisations for plausible charge-density distributions, parametrisations of the form factors may be obtained. Such parametrisations are then compared to the data to extract the parameters. Note that the large-angle dependence provides information on the internal structure while as $\theta \to 0$ (for which eventually $\lambda > R_{\mathrm{nucl}}$) one should see a return to the typical q^{-4} behaviour. By exploiting spherical symmetry, we may integrate out the angular dependence and thus simplify the expression for $F(\boldsymbol{q})$. This leaves

$$F(q^2) \;=\; \frac{4\pi}{q} \int_0^\infty \mathrm{d}r\, r \sin qr\, f(r), \tag{C.2.17}$$

where $q = |\boldsymbol{q}|$. Note that the density $f(r)$ is then normalised as

$$4\pi \int_0^\infty r^2 \mathrm{d}r\, f(r) \;=\; 1. \tag{C.2.18}$$

In Table C.1 we provide a list of typical functional forms used. As an example, consider the case of a uniform sphere; the first minimum lies at $qR \simeq 4.5$. Thus, referring to the graph in Fig. C.3, we find $R \simeq 2.5\,\mathrm{fm}$ for $^{12}\mathrm{C}$. Moreover, the fact that the minima are not as sharp as is predicted for a uniform sphere indicates the existence of an outer "skin" of finite depth.

Table C.1: A collection of possible forms of charge distribution inside the nucleus together with the corresponding form factors. In all cases R represents a measure of the nuclear radius.

form	$f(r)$	$F(q^2)$	behaviour
point	$\delta(r)/4\pi$	1	constant
exponential	$\frac{1}{8\pi R^3}\mathrm{e}^{-r/R}$	$(1+q^2R^2)^{-2}$	"dipole"
Gaussian	$\frac{1}{(2\pi)^{3/2}R^3}\mathrm{e}^{-\frac{1}{2}r^2/R^2}$	$\mathrm{e}^{-\frac{1}{2}q^2R^2}$	Gaussian
uniform sphere	$\frac{3}{4\pi R^3}\quad(r<R)$	$\frac{(\sin\rho-\rho\cos\rho)}{\rho^3}\quad(\rho\equiv qR)$	oscillatory

C.2.4 Physical interpretation

Let us conclude this section by providing a physical interpretation of the form-factor effects. For wavelengths much greater than the size of an extended target object, the latter is not resolved and acts as an effective point-like charge. However, for a given wavelength there is a limited region over which scattering may be coherent; that is, over which the subregions all interfere constructively. A simple comparison of trajectories reveals that the size of this region is of the order of one wavelength. Thus, for objects much larger than the wavelength used, only a small fraction of their total charge actually effectively contributes to the scattering. This leads to a rapid decrease of the cross-section with decreasing wavelength, or correspondingly increasing energy–momentum transfer.

C.3 Quasi-elastic scattering

Let us now consider a case intermediate between pure elastic scattering and the process known as deeply inelastic scattering (DIS, in which the proton is completely broken up into many pieces). The process we wish to consider is thus called *quasi*-elastic scattering. As we have already seen, elastic scattering kinematics imposes a one–to–one relation between the scattering angle and the energy of the outgoing electron. We should thus expect a single spectral line. Let us examine what one observes in practice for scattering off a nucleus at high energy. The example we shall use is that of scattering off a water molecule (see Fig. C.4). There are essentially three prominent features to be found in the spectrum, two pronounced spikes and one broad underlying peak. The larger of the spikes (at around 165 MeV) is evidently due to elastic scattering off one of the hydrogen nuclei (this is determined simply from the position of the spike with respect to the initial

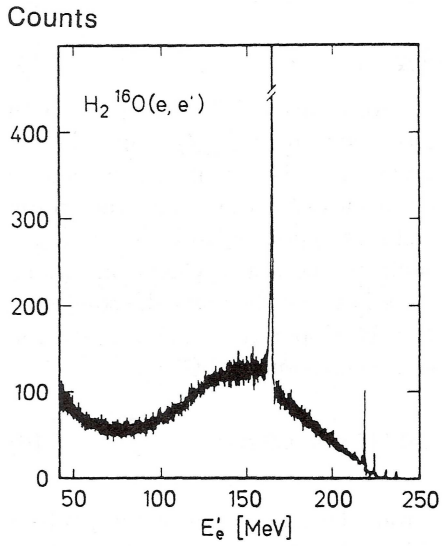


Figure C.4: Quasi-elastic scattering of electrons off water molecules as a function of outgoing-electron energy for a beam energy of 246 MeV at a scattering angle of 148.5°.

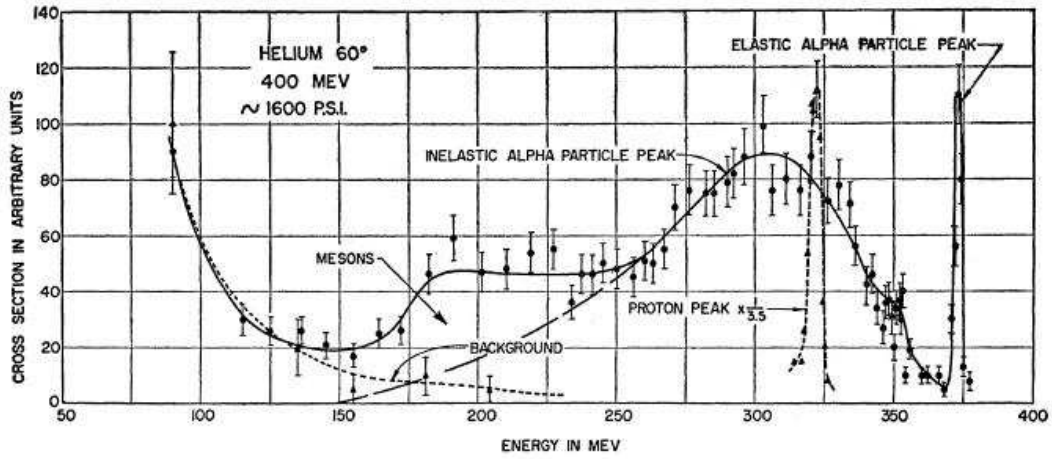


Figure C.5: Quasi-elastic scattering of electrons off helium nuclei as a function of outgoing-electron energy for a beam energy of 400 MeV at a scattering angle of 60°.

energy and fixed scattering angle) while the lower (at around 218 MeV) corresponds to the oxygen nucleus—the shift in position being due to the different mass of the scattering object. The question then remains of the underlying, broader structure (centred roughly around 147 MeV), which nevertheless has the clear form of a peak. We shall now show that this may be attributed to scattering off a proton bound inside an oxygen nucleus; the proton is then ejected from the nucleus and this is what is known as *quasi-elastic* scattering. Another example, scattering off helium,

is shown in Fig. C.5.

We known that, for example, the Fermi-gas model makes rather precise predictions: the nucleons lie in a potential well of approximate depth 40 MeV with a Fermi level corresponding to a momentum of the order of 250 MeV. The depth of the well represents an energy that must be supplied above and beyond the kinematical needs while the Fermi motion will induce smearing of the total centre–of–mass energy, leading to a smearing of the final-state spectrum. Let us now examine in detail how this works. The process we wish to study has a *three*-body final state: the electron, the proton and the recoiling nuclear remnant (see Fig. C.6). We begin

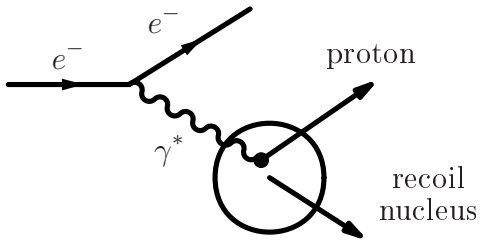


Figure C.6: A schematic view of quasi-elastic electron–nucleus scattering: via exchange of a virtual photon a proton is ejected from the nucleus.

by introducing the necessary kinematic variables in the laboratory frame:

$$\begin{aligned} \boldsymbol{p} &= \text{initial electron momentum,} && \text{(C.3.1a)} \\ \boldsymbol{p}' &= \text{final electron momentum,} && \text{(C.3.1b)} \\ \boldsymbol{P} &= \text{initial struck proton momentum,} && \text{(C.3.1c)} \\ \boldsymbol{P}' &= \text{final struck proton momentum,} && \text{(C.3.1d)} \\ \boldsymbol{q} &= \boldsymbol{p} - \boldsymbol{p}' \\ &= \text{momentum transfer.} && \text{(C.3.1e)} \end{aligned}$$

Finally, a useful variable is $\nu \equiv E_e - E'_e$, the energy transfer in the laboratory frame. It then a simple exercise to show that we have

$$\nu = \frac{q^2}{2M} + V_0 + \frac{2qP\cos\alpha}{2M}, \qquad \text{(C.3.2)}$$

where $q = |\boldsymbol{q}|$, $P = |\boldsymbol{P}|$ and α is the angle between the vectors $\boldsymbol{q}$ and $\boldsymbol{P}$.

Exercise C.3.1. *Derive the above expression for the energy transfer ν.*

Now, assuming the initial momentum $\boldsymbol{P}$ to be distributed uniformly inside the

Fermi sphere, we can easily evaluate both $\bar{\nu}$ and σ_ν:

$$\begin{aligned}\bar{\nu} &= 2\pi \int_0^{p_\mathrm{F}} p^2 \mathrm{d}p \int_{-1}^{1} \mathrm{d}\cos\alpha \left[\frac{q^2}{2M} + V_0 + \frac{2qP\cos\alpha}{2M}\right] \Big/ \left[\frac{4\pi p_\mathrm{F}^3}{3}\right] \\ &= \frac{q^2}{2M} + V_0 \end{aligned} \tag{C.3.3}$$

and

$$\begin{aligned}\sigma_\nu &\equiv \sqrt{\overline{\nu^2} - \overline{\bar{\nu}^2}} = \frac{q}{M}\sqrt{\overline{P^2}\cdot\overline{\cos^2\alpha}} = \frac{q}{M}\sqrt{\tfrac{1}{3}\overline{P^2}} \\ &= \frac{1}{\sqrt{5}}\frac{qp_\mathrm{F}}{M}, \end{aligned} \tag{C.3.4}$$

where we have used the standard result that, for a uniform spherically symmetric momentum distribution, $\overline{P^2} = \frac{3}{5}p_\mathrm{F}^2$.

Exercise C.3.2. *From the peak position and width, as shown in Fig. C.4 and/or Fig. C.5, estimate the corresponding values of p_F and V_0.*

We should now comment on the implicit approximation made here: the fact that the cross-section may be calculated as though the nucleons (or protons in this case) were *free* inside the nuclear volume is not entirely trivial. If the struck particle sits in a potential well with a strong spatial dependence, then it is presumably permanently subject to forces comparable to the scattering potential itself and that should therefore be added to those operating during the scattering process. However, the data lead to a picture in which the nucleons apparently lie in a potential well with an effectively *flat* bottom. That is, deep inside the nucleus they are not subject to any forces except for the brief moments in which they make contact with the boundary or surface of the nucleus and thus we are justified in making what is known as the *impulse approximation*.

It is perhaps also worth mentioning that the same approximation is used in describing the scattering of electrons off single quarks inside the nucleon (a process known as *deeply inelastic scattering*). In this case the struck quark does not actually materialise as a free particle in the laboratory owing to the absolute confining effect of the strong interaction. Nevertheless, calculations (first performed by Feynman) in the impulse approximation describe the data surprisingly well. In this case the explanation is not the triviality of the potential, but a phenomenon known as *asymptotic freedom*. According to this property of QCD, the strength of the interaction decreases with increasing energy scale or, equivalently, at short distances. Thus, provided the energy of the probe is sufficiently large, then the time and distance scales become such that the struck quark can interact only weakly with the parent nucleon.

A more quantitative and rigorous way of defining the impulse approximation is in terms of interaction times. The struck objects, bound inside a more complex structure (be they nucleons inside the nucleus or quarks inside a nucleon) move freely for a time that may be roughly estimated as the diameter of the surrounding structure divided by their average velocity. The interaction has a time scale which may be estimated as the inverse of the energy transfer. The impulse approximation is then justified if the interaction time is less than that of the mean free motion.

Exercise C.3.3. *Estimate the mean free time for a nucleon bound inside an iron nucleus and thus estimate the energy transfer necessary to guarantee applicability of the impulse approximation.*

C.4 Bibliography

Mott, N.F. (1929), *Proc. Royal Soc. (London)* **124**, 425.

Rutherford, E. (1911), *Phil. Mag.* **21**, 669.

Appendix D

A selection of suitable examination topics

In this appendix we provide a short list of suitable topics to prepare for the course (oral) examination:

- Gell-Mann's quarks and SU(3)—motivation, structure and consequences.
- Feynman's parton model—motivation, structure and consequences.
- The Higgs model for the electroweak interaction—motivation, structure and consequences.
- The CKM matrix, with particular reference to CP violation.
- The physics of e^+e^- machines, with particular reference to particle discovery.
- The role of symmetries and their violation in particle physics.

Appendix E

Glossary of Acronyms

ABJ: Adler–Bell–Jackiw

AGS: alternating gradient synchrotron

BW: Breit–Wigner

BNL: Brookhaven National Lab.

CERN: Centre Européee de Rechèrche Nucleaire

CESR: Cornell Electron Storage Ring

CKM: Cabibbo–Kobayashi–Maskawa

DIS: deeply inelastic scattering

DESY: Deutches Elektronische Syncrotron

DORIS: Doppel-Ring-Speicher

GIM: Glashow–Iliopoulos–Maiani

GUT: grand unified theory

LEP: Large Electron–Positron Collider

LHC: Large Hadron Collider

PDG: Particle Data Group

QCD: quantum chromodynamics

QED: quantum electrodynamics

RHIC: Relativistic Heavy-Ion Collider

SLAC: Stanford Linear Accelerator Center

SM: standard model

SPEAR: Stanford Positron–Electron Accelerating Ring

SSC: Superconducting Super-Collider