Introduction to Nuclear and Subnuclear Physics

(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 26th February 2018)

Preface

This is a written version of lecture notes for the course on *Nuclear and Subnuclear Physics*, presented in the third year of the first degree course in Physics during the periods 2002–05 and 2011 to the present in Como. However, they have been augmented and edited with the aim of being self-contained and thereby of more general use. They are thus primarily intended for use by students who already have a basic knowledge of classical electromagnetism, quantum mechanics and special relativity.

An important part of the training of any serious physics student, irrespective of possible intentions as to a future career in physics, is that of *problem solving*. To this end, a large and hopefully comprehensive collection of exercises is provided, together with solutions. The level of integration within the text is also hopefully such that students will find it as natural as possible to treat these notes not merely as a "reader" but as a complete training in the basics of nuclear and particle physics or, more simply, a preparation that will engender the flexibility of approach associated with the mentality typically acquired by physicists in general.

Turning now more specifically to the motivation behind the choice of content and aims of the specific presentation, beyond the obvious intent of providing a basic introduction to nuclear and particle physics from both experimental and theoretical standpoints, the nature of the subject matter treated lends itself to a consolidation of many of the preceding physics courses. In particular, together with the courses on the structure of matter, nuclear physics represents a first playground in which to become more familiar with the physical implications of quantum mechanics. In addition, the natural requirements of particle physics of high-energy experiments leads to the inclusion of a relativistic description. One should also not forget the role of electromagnetism as one of the fundamental interactions.

A more general aspect of this course is the emphasis placed on the use of the concept of symmetry in physics. This, more noble issue, of the search for similarity and unity in the fundamental laws of nature is, perhaps, the most vital and far-reaching lesson of all. It is probably no exaggeration to say that every major break-through in physics has been triggered by the recognition of some common denominator between two or more apparently unrelated phenomena, or of the

equivalence of *a priori* unconnected approaches or viewpoints. Or at the very least the observation that som apparent symmetry of Nature is not fully satisfied.

On a more humanist note, at the time of writing this preface, the first major international conflict of the third millennium was under way in Iraq. Without wishing in any way to trivialise the horror of war, nor to appear too *naïve* in regard of its causes, until humanity as a whole learns to appreciate the importance and beauty, not to say vitality, of uncovering similarities and common traits between the peoples of the world, there will *never* be peace among mankind.

Before turning to the concrete task of studying nuclear and particle physics let me spend a little more time on questions that may be regarded as somewhat philosophical, but which nevertheless have some bearing on the activity that awaits us. One of the difficulties associated with studies at the boundaries of human knowledge is the distinction between real fundamental theory and more mundane model description. It is often necessary to start with a model in order to make progress. Unfortunately, if deeper understanding is too slow in arriving, by dint of familiarity what was once merely conjecture becomes accepted as fact without real proof. While the practised eye of the seasoned researcher should discern the difference, for the inexperienced student this is often impossible. In the text (and more so during my in-class lectures) I attempt to clarify where hard proven theory is being applied and where a merely well-worn model is being propounded. It may also be helpful to review the opinions of two great thinkers of the past.

On experiment and theory

It is, for me, an unfortunate fact of modern physics (not to say modern life) that compartmentalisation or specialisation forces researchers very early in their careers to take on the epithet of "experimentalist" or "theorist".* The problem is then that, having donned the mask of one or other character, identification with the chosen role often becomes so pre-eminent as to limit contact with the other. To my mind, this is a great loss for both areas.

The following is an excerpt from the work of Roger Bacon (Franciscan monk and controversial theologian, considered to be one of the most influential philosophers of the thirteenth century), taken from the Paul Halsall's Medieval Sourcebook:

> "On Experimental Science, 1268
>
> Having laid down the main points of the wisdom of the Latins as regards language, mathematics and optics, I wish now to review the principles of wisdom from the point of view of experimental science, because without experiment it is impossible to know anything thoroughly.

* I prefer to define myself a phenomenologist, thus shunning either label as too exclusive. This means, however, that experimental colleagues usually consider me a theoretician while many theoreticians would also locate me outside their defined sphere of activity.

> There are two ways of acquiring knowledge, one through reason, the other by experiment. Argument reaches a conclusion and compels us to admit it, but it neither makes us certain nor so annihilates doubt that the mind rests calm in the intuition of truth, unless it finds this certitude by way of experience. Thus, many have arguments toward attainable facts, but because they have not experienced them, they overlook them and neither avoid a harmful nor follow a beneficial course. Even if a man that has never seen fire, proves by good reasoning that fire burns, and devours and destroys things, nevertheless the mind of one hearing his arguments would never be convinced, nor would he avoid fire until he puts his hand or some combustible thing into it in order to prove by experiment what the argument taught. But after the fact of combustion is experienced, the mind is satisfied and lies calm in the certainty of truth. Hence argument is not enough, but experience is.
>
> This is evident even in mathematics, where demonstration is the surest. The mind of a man that receives that clearest of demonstrations concerning the equilateral triangle without experiment will never stick to the conclusion nor act upon it till confirmed by experiment by means of the intersection of two circles from either section of which two lines are drawn to the ends of a given line. Then one receives the conclusion without doubt. What Aristotle says of the demonstration by the syllogism being able to give knowledge, can be understood if it is accompanied by experience, but not of the bare demonstration. What he says in the first book of the Metaphysics, that those knowing the reason and cause are wiser than the experienced, he speaks concerning the experienced who know the bare fact only without the cause. But I speak here of the experienced that know the reason and cause through their experience. And such are perfect in their knowledge, as Aristotle wishes to be in the sixth book of the Ethics, whose simple statements are to be believed as if they carried demonstration, as he says in that very place."
>
> Roger Bacon

On simplicity—Occam's Razor

Ockham's Razor ("Occam" is the latinisation) describes the principle, espoused by William of Ockham (yet another Franciscan monk and controversial theologian, considered to be one of the most influential philosophers of the fourteenth century), that: "*Pluralitas non est ponenda sine neccesitate*", which may be translated as: "*Entities should not be multiplied unnecessarily.*" In other words, the simplest theories with least parameters, particles or states should preferred over those in which there is unnecessary proliferation. Put more simply, but still in Latin, "*lex parsimoniae*". Sir Isaac Newton also espoused such an approach: "*Truth is ever to*

be found in simplicity, and not in the multiplicity and confusion of things."

The notion of the "razor" or "simplicity" lies in the observation that for any given set of facts there is an infinite number of theories that could explain them. A data set of ten points may almost always be described by a ten-parameter fit, but such a description almost never conveys any real physical understanding, whereas a successful one- or two-parameter description most probably contains real physics. One should therefore *discard* unnecessary complication in favour of simplicity (often related to a perception of the *beauty* of a theory).

Progress in physics is, of course, often made by renewing or even overthrowing long-standing principles or theories; however, this should not in any way be taken as in contradiction with Ockham's razor. It attempts no predictions as to future directions, but merely advocates, at any given moment, use of a minimal theory.

> "You can recognize truth by its beauty and simplicity. When you get it right, it is obvious that it is right – at least if you have any experience – because usually what happens is that more comes out than goes in."
>
> Richard P. Feynman

And, on the other hand, our true guide can only be the hard facts of the real physical world:

> "It doesn't matter how beautiful your theory is, it doesn't matter how smart you are. If it doesn't agree with experiment, it's wrong."
>
> Richard P. Feynman

Finally, in preparing these notes, I have let my approach be guided by the following observation, which might be viewed as a synthesis of all of the above considerations:

> "It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience."
>
> Albert Einstein

This is often interpreted as *Einstein's razor*, or (paraphrased as) "*everything should be made as simple as possible, but no simpler*"; *i.e.* it is taken as a warning to avoid oversimplification, even in the name of communication.

Contents

Part I

Nuclear Physics

Chapter 1

Introduction to Part I

1.1 Supplementary reading for Part I

A (very short) list of suggested supplementary reading material follows. The book by Povh *et al.* is particularly recommended for its style (very different to that of the present notes), clarity and relevance also for the second course in this series. Beyond this, each chapter presented here contains a more or less complete list of the works cited (both original papers and review articles), which may be consulted for further study.

We should also mention the numerous, continually updated, review articles (not forgetting the mountain of technical information) contained in the Particle Data Group (PDG) biennial publication (current version: Patrignani *et al.*, 2016, also accessible online at `http://pdg.lbl.gov/`).

Reading list

Alonso, M. and Finn, E.J. (1968), *Quantum and Statistical Physics* (Addison–Wesley).

Cottingham, W.N. and Greenwood, D.A. (1986), *An Introduction to Nuclear Physics* (Cambridge U. Press), 1st. edition.

Enge, H.A. (1966), *Introduction to Nuclear Physics* (Addison–Wesley).

Krane, K.S. (1987), *Introductory Nuclear Physics* (John Wiley & Sons), 3rd. edition.

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

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

Wichmann, E.V.H. (1967), *Quantum Physics*, vol. 4 of the *Berkeley Phys. Course* (McGraw–Hill).

Wong, S.S.M. (1998), *Introductory Nuclear Physics* (John Wiley & Sons), 2nd. edition.

1.2 Conventions and notation

We shall indicate defining equations using the notation ":=", where it is understood that the right-hand side defines the left-hand side, while the notation "$\widehat{=}$" will be used to indicate equivalence though not necessarily strict equality (such as between operators).

1.2.1 Factors of 2π and "*bar*" notation

The momentum-space measure naturally conjugate to $\mathrm{d}x$ is $\mathrm{d}p/(2\pi)$ and in most cases the Dirac δ-function for momenta is also accompanied by a 2π factor; thus, $2\pi\,\delta(p)$. Since such 2π factors generally simplify in the final expressions, for clarity of the intermediate steps by avoiding such inessential factors, we shall adopt the following extensions to the usual "bar" notation (in analogy with the standard $\hbar$ and common $\lambda\!\!\!^{-}$):

$$đ := \frac{\mathrm{d}}{2\pi} \text{ and } \delta\!\!\!^{-} := 2\pi\,\delta(p). \tag{1.2.1}$$

In particular then, one has

$$\int đp'\,\delta\!\!\!^{-}(p'-p) \;=\; 1. \tag{1.2.2}$$

1.2.2 Natural units

For clarity of notation, we shall also generally adopt the universal "natural" units of the high-energy physicist, in which $\hbar$ and c are set to unity and therefore disappear from all expressions. However, it will occasionally be useful, in order to fully appreciate the quantum or relativistic nature of a given situation to render the dependence on these two parameters explicit; in such cases $\hbar$ and c will be temporarily reinstated. Indeed, this restoration may always be achieved by noting the dimensions of the object under study: within the natural system all physical

quantities have dimensions of powers of energy (say) and multiplication by suitable powers of $\hbar$ and c will regain the true physical dimensions.

1.2.3 Errors

Measurable physical quantities obtained experimentally are always presented with corresponding errors. These are of two types: statistical and systematic. The first relates to the fact that all measurements are carried out using a number of sample data and then, in some way, averaged. The statistical precision of such an average improves with the number of data values obtained (to be precise, as the inverse square-root). Systematic errors are more difficult to control and can only be estimated on the basis of knowledge of the experimental setup and the theoretical analysis. In any case all experimental (and many theoretical) quantities are presented together with both statistical and systematic errors; for example,

$$\frac{g_\mu - 2}{2} \;=\; (11659208.9 \pm 5.4 \pm 3.3) \times 10^{-10}, \tag{1.2.3}$$

where conventionally the first error is statistical and the second systematic. It is sometimes useful to combine the two and also compactify the notation as follows:

$$\frac{g_\mu - 2}{2} \;=\; 11659208.9(6.3) \times 10^{-10}, \tag{1.2.4}$$

where the statistical and systematic errors have been combined in quadrature* and placed in parenthesis as significant figures.

1.2.4 Indices and vectors

As a final question of notation, to aid with identification of indexed objects, we shall attempt to adhere as strictly as possible to the following convention: letters from the middle of the Greek alphabet μ, ν, ρ, σ, λ will be used to indicate Lorentz four-vector indices (and thus run from 0 to 3); the Latin letters i, j, k, l will usually stand for purely spatial indices (running from 1 to 3); the beginning of the Greek alphabet α, β, γ will be used as Dirac-space indices (*i.e.* in Dirac spinors or gamma-matrix elements); and the beginning of the Latin alphabet a, b, c will typically be used for internal spaces of various types. Unindexed four vectors will be printed in a normal math-italic font while unindexed three vectors will appear in bold face.

* The overall final error, combined in quadrature, is $\epsilon = \sqrt{\epsilon_{\rm stat}^2 + \epsilon_{\rm syst}^2}$.

1.3 Aims and philosophy

This first part of the lecture series is loosely aimed at and tied to the topic of nuclear physics. However, as already hinted in the preface, a more general goal is to be borne in mind. Here the nature of nuclear physics in particular lends itself to the exemplification and consolidation of much of the *assumed* prior study of quantum mechanics.

As noted in the preface nuclear and particle physics (just as the study of condensed matter) provides a natural quantum mechanics playground for the beginner, in which to experiment with early notions and methods. With this in mind, the author has endeavoured to stress the role played by quantum physics in the description of nuclear properties and phenomenon:

> "Nature isn't classical, dammit, and if you want to make a simulation of nature, you'd better make it quantum mechanical, and by golly it's a wonderful problem, because it doesn't look so easy."
>
> Richard P. Feynman in "Simulating Physics with Computers"

Furthermore, we shall attempt to put tha question of rigour into perspective: the study of physics, by the very nature of the subject, has at least as much to do with experimental confirmation as with abstract mathematical constructs.

> "Pure mathematics is just such an abstraction from the real world, and pure mathematics does have a special precise language for dealing with its own special and technical subjects. But this precise language is not precise in any sense if you deal with real objects of the world, and it is only pedantic and quite confusing to use it unless there are some special subtleties which have to be carefully distinguished."
>
> Richard P. Feynman in "New Textbooks for the 'New' Mathematics"

Let us be clear on this issue: the final aim of Physics is to describe and understand an experimentally accessible reality. That said, however, mathematical simplicity has always been considered an important guide in the choice of theoretical path to be followed. Indeed, the art of being a physicist might be summed up as the ability to walk the tightrope between an all-too-human desire or need to "describe" the world and the scientifically necessary constraint of self-consistency or logic.

While it would be easy to provide a list of examples of what has now become known as "pathological science" and thus merely stress the futility of unbridled fantasy, the many examples of the need to be capable of suspending belief in current dogma should not be forgotten either. While the author has no wish here to become embroiled in a deontological discussion, it is probably worthwhile to briefly remind the reader of concrete situations in which even eminent physicists were led astray. One of the most celebrated examples of the tragedy that may ensue from incautious insistence on fantasy-driven lines of research is perhaps the case of

N-rays (Blondlot, 1905), which was consumed in the early years of the twentieth century on the wake of the many discoveries of new forms of radiation.* On the other hand, Feynman's description of deeply inelastic scattering processes in 1972 could reasonably have been considered untenable in the light of the immense forces then assumed to be at work inside the nucleon while Pauli had this to say of Dirac's antimatter:

> "Dirac has tried to identify holes with antielectrons ... We do not believe that this explanation can be seriously considered."
>
> Wolfgang Pauli

Thus, at the end of the day, although Einstein probably had good reason to claim: "*Common sense is the collection of prejudices acquired by age eighteen.*", completely unbridled flights of fantasy rarely lead to real progress, indeed,

> "Philosophy is written in this grand book (I mean the universe) which stands continually open to our gaze, but it cannot be understood unless one first learns to comprehend the language in which it is written. It is written in the language of mathematics, and its characters are triangles, circles, and other geometric figures, without which it is humanly impossible to understand a single word of it; without these, one is wandering about in a dark labyrinth."
>
> Galileo Galilei

With this last (rather famous) quote, I also wish to underline the importance of, alas oft-neglected, basic and solid mathematical training, which is vital for *all* physicists—theoreticians and experimentalists alike.

1.4 A preface on nuclear physics

Let me now make a few remarks that will hopefully place the area of nuclear physics into a perspective from which the student might better understand the problems to be faced in approaching the subject.

Nuclear physics, it must be admitted, is still rather far from having a complete, fundamental and well-defined theory—that is, as compared to the theory we have, say, of the atom, via quantum mechanics and the electromagnetic interaction. What we possess is a series of models based in part on theoretical notions (such as, indeed, quantum mechanics) and in part on phenomenology or observation.

* The exposure by Wood (1904) of the non-existence of N-rays, has been claimed (probably erroneously) to have led to the madness and death of Blondlot, their "discoverer." This episode and others are also documented and discussed within the framework of pathological science by, for example, Langmuir (1904) and Gardner (1957).

Put bluntly, the latter means many formulæ are that way simply because that is what agrees with the data. The vision we have of the nucleus is of a ball (of radius a few femtometres) made up of other little balls (the nucleons—protons and neutrons). The latter are held together by the so-called "*strong* nuclear force" of which, unfortunately we know very little in concrete terms.

That is to say, we believe that at a fundamental or elementary level the nuclear or *strong* interaction is described by a quantum field theory known as quantum chromodynamics (QCD). This theory describes a world of quarks* interacting via the exchange of gauge fields called gluons, photocopying, as it were, quantum electrodynamics (QED) and its photon. Unfortunately, the nature of QCD is such as to render it impervious to perturbative treatment (our main tool for describing quantum dynamics) except in the high-energy regime (above the scale of around 1 GeV). Since the natural scale of the nucleus is that of only a few MeV (*i.e.*, the typical binding energy of a single nucleon), this means we should seek other instruments. Since there are essentially no other methods of a fundamental nature, we are thus left in an unfortunate situation: it is a though we were trying to study the structure of molecules or crystals without first having a working knowledge of hydrogen and other atoms.

1.4.1 A simple model

Let us explore the analogy between nuclei and molecules a little further. After all, as we shall see shortly, the best model we have for the nucleus is based on a picture similar to that of electrons bound in an atom (indeed, we speak of the *atomic nucleus*). Since atoms are electrically neutral, the force between, say two hydrogen atoms in an H_2 molecule may be described in terms of the van der Waals force. This force may be viewed as the result of an electromagnetic force field that somehow "leaks out" of the neutral system. In quantum mechanics such a force may be explained generically (*i.e.*, with little knowledge of the precise nature of the underlying interaction) as the attraction one finds between two potential wells, see Fig. 1.1. In other words, owing to the tunnel effect a particle "trapped" inside one of the wells actually partially "sees" a larger volume: that of the second well—and the occupation of the available volume increases with decreasing separation. According to the Heisenberg uncertainty principle, the minimum momentum and therefore minimum kinetic energy (the so-called zero-point energy) is determined by the dimensions of the confining potential. Putting this all together, we see that the energy of the system is lowered as the wells approach each other—thus

* Note that while, according to Gell-Mann, the spelling of the word "quark" was taken from James Joyce's *Finnegan's Wake* the pronunciation he himself chose was that which rhymes with "talk" and *not* "park" (see Gell-Mann, 1995, p. 180).

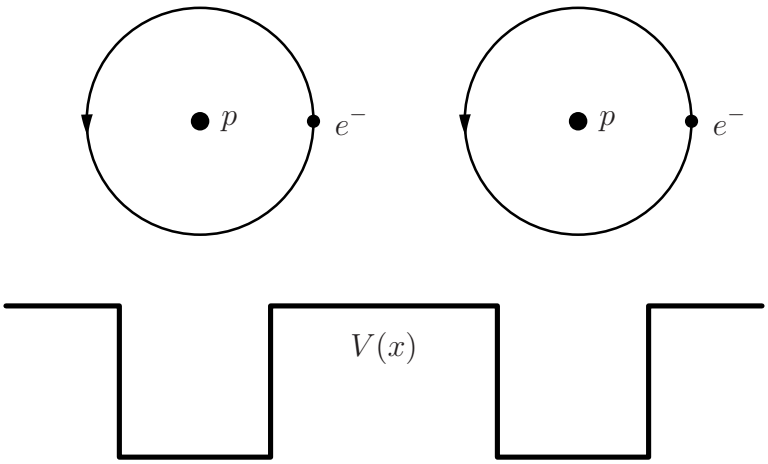


Figure 1.1: The van der Waals force in an H_2 molecule seen as being generated via the interaction between two potential wells in QM.

implying an attractive force. In nuclear physics we imagine a similar mechanism to be at work, due to the underlying QCD theory of quarks and gluons.

1.4.2 The problem with QCD

Considerable difficulty is posed, however, by the experimental observation that quarks (and gluons) are never found outside the confines of a hadron (*i.e.* a qqq system such as a baryon or a $q\bar{q}$ meson), a property known as *confinement*. We should thus in any case talk in terms of objects such as pions (being the lightest) as somehow providing the nuclear glue. Yukawa (1935) hypothesised a potential of the following form:

$$U_{\mathrm{Y}}(r) \propto \frac{\mathrm{e}^{-\mu r}}{r}, \tag{1.4.1a}$$

cf. the standard Coulomb potential

$$U_{\mathrm{C}}(r) \propto \frac{1}{r}, \tag{1.4.1b}$$

where μ, the inverse range of the resulting force, is just the mass of the exchange object (in our case $m_\pi \sim 140\,\mathrm{MeV}$). There are, however, many such particles, only a little more massive, that should be included and so a complete theory along these lines rapidly becomes unmanageable.

Moreover, if one does accept this view, given our knowledge of molecular physics, there is an immediate question: why are nuclei always more or less spherical

and never chain-like? This is an example of how phenomenological input is, on the one hand, vital but, on the other, also clouds the issue of fundamental predictions: the theory does not predict spherical nuclei, this property must be put in by hand. Similarly, there are many other aspects of nuclear physics that are necessarily inserted *ad hoc* and that cannot therefore be explained. Moreover, in parametrising the structures and effects observed, natural biases are inserted, which can then only reappear in the "predictions" that might be made.

1.5 Bibliography

Blondlot, R.P. (1905), *The N-Rays* (Longmans, Green and Co.).

Dirac, P.A.M. (1930), *Proc. Royal Soc. (London)* **A126**, 360.

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

Gardner, M. (1957), *Fads and Fallacies in the Name of Science* (Dover).

Gell-Mann, M. (1995), *The Quark and the Jaguar: Adventures in the Simple and the Complex* (Henry Holt and Co.).

Langmuir, I. (1904), *Phys. Today*, 36.

Wood, R.W. (1904), *Nature* **70**, 530; *Z. Phys.* (1904) 789.

Yukawa, H. (1935), *Proc. Phys. Math. Soc. Jap.* **17**, 48.

Chapter 2

A Brief Review of Modern Physics

In this chapter we shall review very briefly a selection of fundamental aspects of quantum mechanics and special relativity, with a particular eye to their relevance in the understanding of nuclear and particle physics. Note that the appendices also contain further topics that are of relevance to this course, but with which many readers will already be sufficiently familiar. In the next few sections it is hoped to provide the student, not only with more and useful tools, but also with an insight that will help and complement the later studies presented in these chapters.

As quantum mechanics has to do with the laws of physics at very small scales (we shall discuss the definition of small shortly) while the nucleus and its components are indeed very small, it is inevitable that quantum mechanics will play a very important role in our understanding of many aspects of the dynamics involved. Moreover, since the study of the nucleus, its substructure and the various elementary particles found in the universe requires ever-increasing energies, we shall often have to deal with the problem of objects moving relativistically (*i.e.*, with velocities close to that of light). In the following sections then we shall try to present physically intuitive pictures for the relevant aspects of these two branches of modern physics and also lay down some of the foundations for their application to the subject matter at hand.

2.1 Quantum mechanics

While we cannot hope and do not wish to present here a detailed derivation of any of the principal results of quantum theory*, we shall endeavour to provide an intuitive framework that will allow the student to appreciate some of the subtleties involved and, more importantly, to understand its relevance to the dynamics of

* Indeed, it is, of course, assumed that the student has at least a basic prior knowledge of quantum mechanics, its foundations and mathematical development.

real physical objects. It will, moreover, be helpful to remember that the theory of quantum behaviour is not merely a mathematical construct, but a deduction from experimental observations that do not find explanation within the classical laws of physics. Thus, the spirit of these lectures is embodied in the saying: "*The proof of the pudding is in the eating.*" We shall normally sacrifice mathematical rigour in exchange for a clearer physical picture providing it is, at any rate, known to be correct.

Let us begin this review by examining two fundamental characteristics of the quantum-mechanical description of physics:

- wave-corpuscle duality,
- quantisation.

Both emerge naturally in a wave-mechanical description of the motion not only of electromagnetic radiation, but also of matter (*i.e.*, the particles of which the universe is composed). Indeed, it might be argued that the term *quantum* is something of a misnomer and that one ought simply to talk of *wave* mechanics; all else follows.*

2.1.1 Wave-corpuscle duality

The dual nature, often presented as some sort of contraposition, is quite simply a question of scale. On the scale of a football pitch the ball itself may be considered as point-like while on the scale of a proton it would appear as an immense universe.† A localised wave packet may thus appear as a point-like particle (*i.e.* as having corpuscular dynamics) on some large scale; however, if examined on a sufficiently small scale, it will have all the behaviour (interference, diffraction *etc.*) of classical radiation, see Fig. 2.1. In other words, a photon in the visible region behaves as a point-like object when it is registered on a photographic plate, but if it encounters an obstacle with dimensions the order of a fraction of a micron, its wave nature emerges. Specific examples are:

corpuscle: the photoelectric effect—absorption and emission times are longer than the temporal duration of a wave packet.

wave: Young's double-slit experiment—the slit spacing is of the order of the wavelength used;

* After all, one does not usually speak of the quantisation of musical notes produced by a piano, a flute, violin or guitar strings.

† The title "The Fly in the Cathedral" of a book by Cathcart on early nuclear discoveries is meant to emphasis precisely this point.

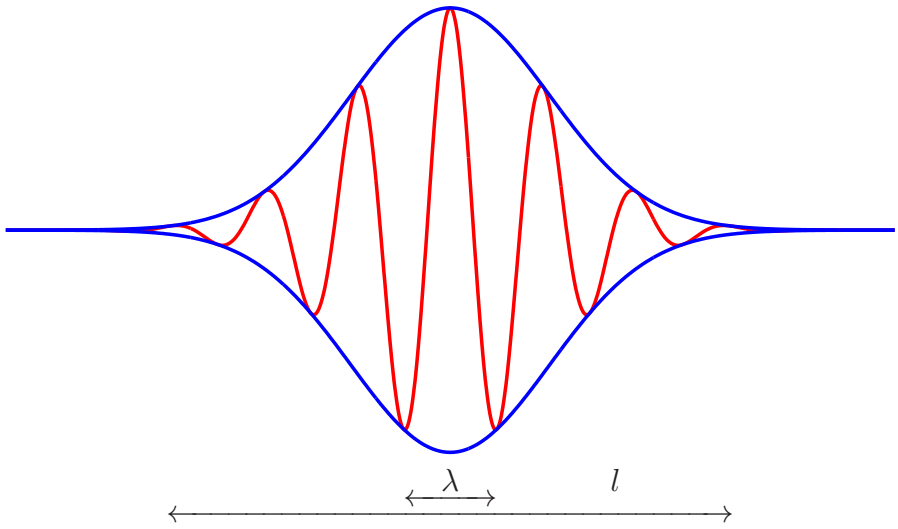


Figure 2.1: A typical wave packet; the overall length l of the packet determines the scale at which it will start to appear point-like while the wavelength λ determines the scale at which wave mechanics will come into play. Note that l and λ may be very different.

Now, all of the above may be repeated in consideration of the behaviour of neutrons of suitable energy in appropriate circumstances.

Indeed, the important point is that all this is found in almost perfect carbon copy for the case of matter fields, such as electrons, protons, neutrons *etc.* In this case one has the Einstein–de Broglie relations:

$$\begin{aligned} E &= h\nu \quad \text{and} \quad p = h/\lambda \\ &= \hbar\omega, \qquad\qquad = \hbar k, \end{aligned} \tag{2.1.1}$$

where $\omega \equiv 2\pi\nu$ is the angular frequency and $k = 2\pi/\lambda$ is the wave-number. In units more congenial to the nuclear or particle physicist, Planck's constant is

$$\hbar \simeq 6.58211928(15) \times 10^{-16}\ \text{eV} \cdot \text{s} \qquad (1\ \text{eV} \simeq 1.6 \times 10^{-19}\ \text{J}). \tag{2.1.2}$$

This again brings us to the concept of scale: the Planck constant effectively fixes a scale for quantum mechanics. As an example, if the energy involved is of order 1 eV (*e.g.*, the electron binding energy in an atom), then quantum mechanics will come into action when the time scales are of the order of or less than 10^{-16} s.

2.1.2 The fundamental constants of Nature

There are three fundamental constants of Nature having dimensions: $\hbar$ (Planck's constant), c (the speed of light in the vacuum) and G (Newton's gravitational

constant); each therefore sets the scale for different types of dynamics. They can also be combined in various ways, for example,*

$$\hbar c \simeq 197 \text{ MeV} \cdot \text{fm} \simeq 1.97 \text{ keV} \cdot \text{Å}. \tag{2.1.3}$$

Moreover, since quantum mechanics is a wave mechanics, we may talk of *resolution*. Hence, according to the above product, to resolve the structure of a nucleus (*i.e.* at the femtometre scale) we need a beam energy of order at least about 200 MeV. In other words, this is the energy for which (photon) wavelengths are of order 1 fm or less. The speed of light, according to the special theory of relativity, may be used to convert mass into the equivalent rest energy, $E = mc^2$, and thus it becomes natural to describe particle masses in terms of electronvolts. In this connection it may be useful to recall a few fundamental energy and length scales in physics, see Table 2.1; the standard prefixes are as shown in Table 2.2.

Table 2.1: A collection of energy and length scales relevant to atomic, nuclear and particle physics.

$1\text{eV} = 1.6 \times 10^{-19}\text{J}$	$\cdots$	atomic E_{binding}
$1\text{keV} = 10^{3}\text{eV}$	$\cdots$	soft X-rays
$1\text{MeV} = 10^{6}\text{eV}$	$\cdots$	nuclear E_{binding}, hard X-rays
$1\text{GeV} = 10^{9}\text{eV}$	$\cdots$	nucleon rest mass (energy equivalent)
$1\text{TeV} = 10^{12}\text{eV}$	$\cdots$	accelerators (*e.g.*, LHC, Tevatron & RHIC)
$1\text{Å} = 10^{-10}\text{m}$	$\cdots$	atomic radius
$1\text{fm} = 10^{-15}\text{m}$	$\cdots$	proton or nuclear radius

Now, an interesting exercise can be performed using all three fundamental constants: since there are only three fundamental dimensions, length, time and mass, by choosing suitable multiplicative combinations of the three, we may define a "fundamental" length, time, mass and also energy. These scales take on the names Planck length, time, mass and energy (*e.g.* m_{P}), the last being the most commonly quoted (as if it were a mass though):

$$m_{\text{P}} \sim 10^{19} \text{ GeV}. \tag{2.1.4}$$

Exercise 2.1.1. *Derive the combination of $\hbar$, c and G required to form two parameters with dimensions of length and time.*

* We note in passing that $\hbar c$ also corresponds to precisely one unit of angular momentum.

Table 2.2: Standard SI prefixes commonly used in atomic, nuclear and particle physics.

prefix	symbol	factor	prefix	symbol	factor
milli	m	10^{-3}	kilo	k	10^{3}
micro	μ	10^{-6}	mega	M	10^{6}
nano	n	10^{-9}	giga	G	10^{9}
pico	p	10^{-12}	tera	T	10^{12}
femto	f	10^{-15}	peta	P	10^{15}

Exercise 2.1.2. *What is the physical significance of, say, m_P?*

2.1.3 Dimensional analysis

The above now naturally brings us to the topic of *dimensional analysis*. The concept should already be familiar from classical mechanics, where one can derive, for example, the form of the expression for the period of a pendulum as a function of its length and g (and also show that the period is independent of its mass). In quantum mechanics, we only have the constants $\hbar$ and c to provide dimension, along with any natural parameters of the system under study (*e.g.*, its size or the centre–of–mass energy), and this often allows us to make rather general statements with little or no prior knowledge of the underlying theory.

A good example is a rather general derivation of the high-energy behaviour of scattering cross-sections. Let us assume that the only dimensional parameters a given theory contains will be particle masses and that the coupling constants (such as $\alpha_{\rm QED} \simeq 1/137$) will all be dimensionless. The form of the cross-section σ (physically an area) for a *general* process can then be easily derived:

$$\begin{aligned} \sigma &= \text{effective area to be ``struck''} \\ &= [L^2] \propto \left(\frac{\hbar c}{E}\right)^2, \end{aligned} \tag{2.1.5}$$

where E should be something like the centre–of–mass energy. Note that we have exploited the fact that $E \gg m$ and that all masses may thus be neglected. In other words, in the absence of any other natural energy parameter, a cross-section at high energies must scale as

$$\sigma \propto E^{-2}. \tag{2.1.6}$$

Since σ gives us the interaction probability of a laboratory collision, we immedi-

ately see one of the negative aspects of the push to ever-higher energies: the number of events produced decreases as the inverse square of the energy. On the other hand, we also have $\lambda \sim \hbar c/E$ so that higher resolution requires higher energy. This is the source of an impending crisis in high-energy physics: at the energies required to resolve the structures of interest we shall eventually not be able to produce a useful number of events in a feasible time span.

2.1.4 Quantisation

We now turn to another fundamental aspect of wave mechanics: *quantisation*. The concept should already be familiar, not only within the framework of quantum mechanics itself, but also from the classical theory of waves. The description of the vibrations of the strings of a musical instrument is an example of quantisation. All that is necessary is a confined system described in terms of waves. Then, since ν and λ have a direct quantum mechanical correspondence with energy and momentum, the boundary conditions (provoked by the confinement) select discrete frequencies and/or wavelengths and thus naturally lead to the quantisation of energy levels, momentum, angular momentum *etc.*

In quantum mechanics one of the first concrete examples is Bohr's description of the hydrogen atom. In an inexact but functional description of the electron as being *in orbit around the nucleus* (or proton), the cyclic boundary conditions on the wave-function (*i.e.* it must be single valued) lead to the following condition:

$$n\lambda = 2\pi r \qquad (n \in \mathbb{N}), \tag{2.1.7}$$

where λ is the de Broglie wavelength associated with the electron motion and r is the (notional) radius of its orbit. Note that this may be rewritten as

$$L = pr = n\hbar \qquad (n \in \mathbb{N}), \tag{2.1.8}$$

where L is just the orbital angular momentum. We shall justify *a posteriori* the use of non-relativistic kinematics and thus write the equation of motion as

$$\frac{\alpha\hbar c}{r^2} = \frac{m_{\mathrm{e}} v^2}{r}. \tag{2.1.9}$$

Eliminating r from the previous two equations gives:

$$E_{\mathrm{B}} = \frac{\alpha^2 m_{\mathrm{e}} c^2}{2n^2} = \frac{13.6 \text{ eV}}{n^2}. \tag{2.1.10}$$

It is now easy to see that the kinetic energy of a bound atomic electron is much

less than its rest mass and that it is thus non-relativistic.

2.1.5 Diffraction

In the foregoing we have pointed out that quantum mechanics implies a wave-like behaviour of particles such as electrons, neutrons and even entire nuclei. Rather spectacularly, one can even perform diffraction experiments using "slow" neutrons incident on crystals.* Recall the Bragg condition for constructive interference:

$$2a\cos\theta \;=\; n\lambda \qquad (n \;\in\; \mathbb{N}), \tag{2.1.11}$$

where a is the separation of the crystal planes, θ is the angle of incidence and λ is the wavelength used. If we are to use such an effect to study the crystal-plane separation, then we clearly require λ to be smaller than a. Translating to units more appropriate to atomic physics, we have recall $\hbar c \sim 2\,\mathrm{keV}\cdot\mathrm{\AA}$. We thus require beams with *momenta* of order 2 keV, which in the case of neutrons implies *kinetic energies* of order 2 meV.

Exercise 2.1.3. *Make a detailed estimate of a suitable neutron energy in order for diffraction effects to be observed in a typical crystal?*

2.1.6 Heisenberg's uncertainty principle

The Heisenberg uncertainty principle is a direct consequence of the wave-like nature of matter and radiation. And, in fact, the answer to the question "Is it possible (at least in principle) to simultaneously measure or know any set of variables to arbitrary precision (limited only by apparatus)?" in quantum mechanics is very definitely (and importantly) "no".

Now, already in classical mechanics all physical quantities exist in conjugate pairs and in quantum mechanics these mutually exclude perfect measurement. Standard examples are x and p_x or ϕ and L_z; the case t and E should always be considered on a slightly different footing owing to the different nature of time (it is a parameter rather than a free variable), nevertheless it does still constitute a limit on precise measurement. The following are examples of the physical consequences of this principle:

- Energy conservation may be violated for very short periods, *i.e.*, provided the time interval Δt and the amount of energy violation ΔE are such that

* The 1994 Nobel Prize in Physics was awarded "for pioneering contributions to the development of neutron scattering techniques for studies of condensed matter" jointly with one half to Bertram N. Brockhouse "for the development of neutron spectroscopy" and one half to Clifford G. Shull "for the development of the neutron diffraction technique."

$\Delta t \Delta E < \hbar$. This is easily understood in physical terms: in such a short time it would be impossible to verify an energy violation of such a small entity. A consequence is the possibility of transitions via classically prohibited (or so-called *virtual*) intermediate states, such as in the tunnel effect.

- Perfectly linear or *collinear* motion is not possible since this would imply simultaneously having $\Delta x_\perp = 0$ and $\Delta p_\perp = 0$. The phenomenon implied by this restriction goes under the name of *Zitterbewegung*.

For the physicist, the relevance of all this goes beyond the mere *observation* of any specific phenomena and touches upon the very *dynamics* to be described. Particles interacting with one another "suffer" the same uncertainties: for example, two particles interacting in a limited region of space need not locally conserve momentum (within the limits outlined above) provided that overall, when observed at large distances, conservation is obeyed. Likewise, for sufficiently short time periods, energy conservation may also be violated and so on.

Let us now quickly examine two specific cases to exemplify the question.

Position measurement

Let us imagine a rather schematic apparatus for measuring position: a simple lens aimed at collecting light scattered by an electron (see Fig. 2.2). Suppose the initial momentum of the electron is known with arbitrary precision. Then, the question is: "Can we measure its position while retaining the knowledge we already possess of its momentum?"

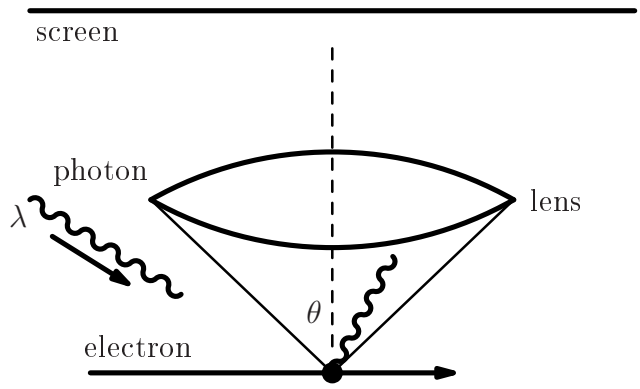


Figure 2.2: Position measurement via a schematised optical method leading to the standard Heisenberg uncertainty.

From the classical theory of optics, we know that the resolution with which the

position of the electron can be determined is

$$\Delta x \gtrsim \frac{\lambda}{\sin\theta}, \tag{2.1.12}$$

where λ is the wavelength of the light used to "illuminate" the electron and θ is the semi-angle subtended by the lens with respect to the electron. The angular aperture of the lens also limits the angle at which the photon is reflected or scattered, although, of course, we have no way of knowing in what precise direction it actually travels. Thus, the diffusion process provides a momentum kick to the electron along the x-axis of maximum magnitude $p_\gamma \sin\theta$, which thus becomes the uncertainty in the final momentum of the electron. From Einstein's relation $p_\gamma = h/\lambda$, we have

$$\Delta p_x \gtrsim \frac{h}{\lambda}\sin\theta. \tag{2.1.13}$$

Multiplying the two uncertainties, we finally obtain

$$\Delta x\,\Delta p_x \gtrsim h. \tag{2.1.14}$$

Notice that by varying the wavelength used and/or the aperture of the lens, we may alter the individual resolutions of position and momentum but there is no way to improve on the Heisenberg limit on the two.

While this may seem an academic exercise, the reader is reminded that many forms of investigation are little more than a variation on the apparatus just described. Indeed, an important example is provided by the experiments performed at the Stanford Linear Accelerator Center (SLAC) in the late sixties. In order to examine the internal structure of the proton (*i.e.*, at length scales of order $\frac{1}{10}$ fm), the scattering of high-energy electrons off hydrogen targets was studied. The process is essentially electromagnetic and involves the exchange of a (virtual) photon between the interacting electron and proton; the latter is thus effectively "observed". The momentum transfer is $\Delta p \sim p_\gamma \lesssim p_e$. Recalling then that $\hbar c \sim 200\,\mathrm{MeV}\cdot\mathrm{fm}$, we immediately conclude that the required beam energy is approximately 2 GeV, which is indeed that used.

Velocity (momentum) measurement

As a second and final example, let us examine a possible method for velocity measurement based on the well-known Doppler effect. Consider an atom in an excited state of energy E^* and in a well-determined position. Suppose we now wish to determine its velocity v by observing the frequency of the emitted photon on de-excitation. If the ground-state energy is E, then the natural emission frequency will be $h\nu_0 = E^* - E$.

The Doppler effect causes the measured frequency to shift according to the approximate (non-relativistic) formula (v is in the direction of the observer)

$$\nu \simeq \left(1+\frac{v}{c}\right)\nu_0, \tag{2.1.15a}$$

or inverting

$$v \simeq \left(\frac{\nu}{\nu_0}-1\right)c, \tag{2.1.15b}$$

where c is just the velocity of light. Now, in order to measure a frequency with precision $\Delta\nu$ it is necessary to carry out observations for an interval of at least $\tau \simeq 1/\Delta\nu$ (*i.e.*, at least one complete period must be observed), which naturally introduces an uncertainty on the instant of emission of the same time lapse. In such an interval the momentum of the atom changes by $\delta p = h/\lambda = h\nu/c$ and hence the velocity changes by an amount $\delta v = h\nu/mc$. Given the time interval, the uncertainty on the final position of the atom is then

$$\Delta x \simeq \delta v\,\tau = \frac{h\nu}{mc}\,\tau. \tag{2.1.16}$$

At the same time the uncertainty in the velocity measurement implies an uncertainty in the final momentum of order

$$\Delta p = m\Delta v \simeq \frac{mc}{\nu_0}\,\Delta\nu \simeq \frac{mc}{\nu_0\tau}. \tag{2.1.17}$$

Once again, up to unimportant numerical factors, the product of the two uncertainties agrees with the Heisenberg inequality.

Mathematical derivation

More formally, the above limits on determination may be obtained by considering a particle described (according to a wave-mechanical formalism) by a Gaussian wave packet (for simplicity, we suppress the irrelevant time variation):

$$\psi(x) = A\,\exp\left[-\frac{1}{2}\left(\frac{x-x_0}{\Delta x}\right)^2\right]\cos\left(\frac{2\pi x}{\lambda_0}\right), \tag{2.1.18}$$

where λ_0 is the principal wavelength. However, we know that to form a packet of finite spatial extension, a range of wavelengths must be present.

Recall that the wave-number $k = 2\pi/\lambda$ and that $p = \hbar k$. The Fourier transform

of the wave-function to wave-number (or momentum) space is then

$$\tilde{\psi}(k) \;=\; \int_{-\infty}^{+\infty} \mathrm{d}x\; \mathrm{e}^{\mathrm{i}kx}\, \psi(x) \;\propto\; \exp\left[-\frac{1}{2}\left(\frac{k-k_0}{\Delta k}\right)^2\right], \tag{2.1.19}$$

where $k_0 = 2\pi/\lambda_0$ and $\Delta k = 1/\Delta x$. This last equality then implies

$$\Delta x\, \Delta k \;=\; 1, \tag{2.1.20}$$

which multiplying by $\hbar$ becomes

$$\Delta x\, \Delta p \;=\; \hbar. \tag{2.1.21}$$

Finally, since it can be shown that the Gaussian is the indeed "best case", the equality becomes the following general inequality:

$$\Delta x\, \Delta p \;\geq\; \hbar. \tag{2.1.22}$$

2.1.7 Classical electromagnetism

Before leaving the topic of quantum mechanics let us briefly examine the classical theory of electromagnetism in the light of the development and interpretation of quantum mechanics. The Lorentz force on a charged particle moving in an electromagnetic field is given by

$$\boldsymbol{F} \;=\; q\left(\boldsymbol{E} + \frac{1}{c}\,\boldsymbol{v}_{\wedge}\boldsymbol{B}\right), \tag{2.1.23}$$

where q is the charge of the particle. Such a force may be derived from the following Hamiltonian:

$$H_{\mathrm{EM}} \;=\; \frac{1}{2m}\left(\boldsymbol{p} - \frac{q}{c}\,\boldsymbol{A}\right)^2 + qV. \tag{2.1.24}$$

The electromagnetic scalar and vector potentials V and $\boldsymbol{A}$ generate the fields $\boldsymbol{E}$ and $\boldsymbol{B}$ thus:

$$\boldsymbol{E} \;=\; -\boldsymbol{\nabla}V - \frac{1}{c}\,\frac{\partial \boldsymbol{A}}{\partial t} \tag{2.1.25a}$$

and

$$\boldsymbol{B} \;=\; \boldsymbol{\nabla}_{\wedge}\boldsymbol{A}. \tag{2.1.25b}$$

Choosing now the so-called radiation gauge, defined by $V = 0$, the general solu-

tion of these equations is simply

$$\boldsymbol{A}(\boldsymbol{x},t) \;=\; 2\boldsymbol{\xi}\, A_0\, \cos(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t + \phi), \tag{2.1.26}$$

with $|\boldsymbol{k}| = 2\pi/\lambda$, the wave-number. The arbitrary phase ϕ has no physical meaning and we may set it to zero here. Using this solution we may calculate the energy density

$$\begin{aligned}\rho_E(\boldsymbol{x},t) \;&=\; \frac{1}{8\pi}\,(\boldsymbol{E}^2 + \boldsymbol{B}^2) \\ &=\; \frac{1}{\pi}\,k^2 A_0^2 \sin^2(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t).\end{aligned} \tag{2.1.27}$$

In other words, we see that the energy density is proportional to $|\boldsymbol{A}|^2$. For a large number of photons, we may divide this expression by $E_\gamma = h\nu$, the single-photon energy, and thus obtain the number density for the photons in the radiation wave so-described. This leads naturally to a probabilistic interpretation when applied to a single photon: we may identify $\boldsymbol{A}(\boldsymbol{x},t)$ with the probability *amplitude* to find the photon at position $\boldsymbol{x}$ at instant t. Thus, in a sense, the *interpretation* of quantum mechanics is really a simple extension of notions already present in the classical wave theory of electromagnetism.

Let us make one last comment: great care should always be exercised when considering the physics of what actually occurs in Nature. All aspects of a problem should be considered before arriving at a conclusion. For example, a common source of confusion is the fact that the wave-function in quantum mechanics may be non-zero *outside* the classically permitted region. This does not, of course, imply that one has a finite probability of directly detecting the particle at a classically prohibited point in space—this would violate the conservation of energy! So the warning is that, in some sense, the interpretation is often none other than a classical analogy, whose sole purpose is to allow human minds to picture and express in words what is happening. However, as always, analogies can be *dangerous* and should never be taken too seriously.

2.2 Special relativity

In 1887 Michelson and Morley published the results of their attempts to measure the velocity of the Earth with respect to the æther (the medium in which light was assumed to propagate and which should then represent a universal and fixed frame of reference). The surprising fact was that the Earth's velocity appeared to be precisely and constantly *zero*.* The interpretation of these findings led Einstein

* The discovery led to the award of the 1907 Nobel Prize for Physics to Albert A. Michelson.

(1905) to formulate the *Principle of Relativity*.* This work was a synthesis of the Lorentz contraction and time dilation in a unifying picture of the so-called Lorentz transformations, in which it is the *velocity of light* that is a fixed point of reference: *i.e.*, the measured speed of light is the same c for all *inertial* observers. Recall that the Michelson–Morley experiment uses a double-arm spectrometer in which interference effects of *light* are studied. Let us therefore first quickly see separately how time and space are affected in moving (inertial) frames by considering the motion of light.

2.2.1 Time dilation

In order to study the effect of motion on time, we need a clock. Since it is the central role played by light that forces all consequences, let us construct a *light clock* by considering a beam reflected between two mirrors separated from one another by a distance l, see Fig. 2.3. The clock is taken to be moving in a direction

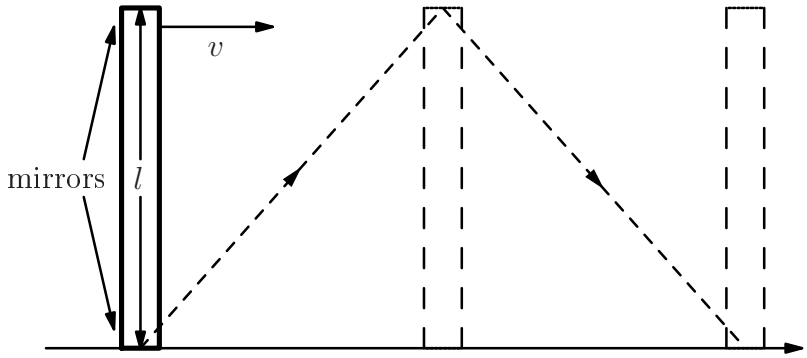


Figure 2.3: The schematised light clock used in considering the effect of time dilation in a moving frame.

perpendicular to the light beam it contains and we shall consider one complete cycle (or "tick") of the clock, in which the light beam travels back and forth between the two mirrors.

Defining τ, the period (or reference "second") of one such cycle in its *rest-frame* (*i.e.*, the frame in which the clock is stationary), we clearly have $c\tau = 2l$. In an inertial frame moving with velocity v with respect to the clock's rest-frame the light beam must follow the longer path indicated by the dotted line in Fig. 2.3. However, according to the principle of relativity, in this frame the measured velocity of light must still be c and therefore, if s is the total path length, we must

* Note that Albert Einstein was awarded the 1921 Nobel Prize for Physics for his 1905 paper on the photoelectric effect.

have $c\tau' = s$, where τ' will be the period in the moving frame. Clearly, $\tau' \neq \tau$ as the paths are necessarily of different lengths (for $v \neq 0$) while, by hypothesis, the perceived velocity must be the same. From the geometry of the setup we have $s = 2\sqrt{(v\tau'/2)^2 + l^2}$ and therefore

$$c\tau' \;=\; 2\sqrt{(v\tau'/2)^2 + l^2}. \tag{2.2.1}$$

Now, substituting for l and rearranging, we obtain

$$\tau' \;=\; \tau/\sqrt{1 - v^2/c^2}. \tag{2.2.2}$$

It is convenient to define*

$$\beta \;:=\; \frac{v}{c} \qquad \text{and} \qquad \gamma \;:=\; \frac{1}{\sqrt{1-\beta^2}}. \tag{2.2.3}$$

We then have $\tau' = \gamma\tau$ and the clock appears to slow down in the moving reference frame as seen from the stationary frame—an effect known as *time dilation*.

2.2.2 Spatial (Lorentz) contraction

Next, consider the same clock but now moving with velocity v in a direction parallel to the internal light beam, see Fig. 2.4. We shall now use the clock as a standard

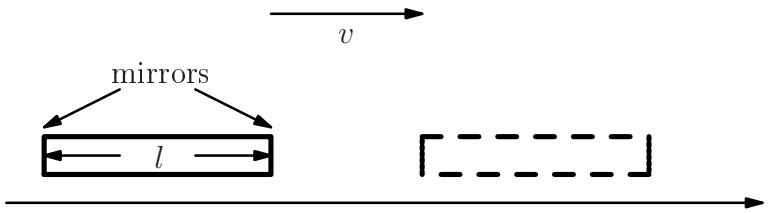


Figure 2.4: The schematised light clock used in considering the effect of Lorentz spatial contraction in a moving frame.

metre of length l in the moving frame, but as seen from the stationary frame we must now allow for a possible change in its length, which we shall thus call l'.

Exercise 2.2.1. *Show that similar considerations to the preceding lead to a spatial contraction since the length of the moving metre is given by:*

$$l' \;=\; l/\gamma. \tag{2.2.4}$$

* Note that we also often use β (and even occasionally γ) to indicate a particle velocity.

That is, objects appear shorter (in the direction of motion) in a moving frame—an effect known as *Lorentz spatial contraction.*

2.2.3 Lorentz transformations

The preceding two results may be combined into a single transformation law for so-called Lorentz *boosts* (*i.e.* velocity shifts), see Table 2.3. Note that the non-

Table 2.3: A comparison of the Lorentz and Galileo transformation laws for inertial frames (the boost is taken along the positive z-axis), the variables refer to space–time separations between events.

		Lorentz	Galileo
ct'	$\rightarrow$	$\gamma(ct-\beta z)$	ct
x'	$\rightarrow$	x	x
y'	$\rightarrow$	y	y
z'	$\rightarrow$	$\gamma(-\beta ct+z)$	$-vt+z$

relativistic limit $\beta \rightarrow 0$ ($\gamma \rightarrow 1$) is equivalent to $c \rightarrow \infty$ while keeping $v = \beta c$ fixed and finite. With a little care, it is then easy to show that the Galilean transformations are just the non-relativistic limit of the Lorentz case.

Exercise 2.2.2. *Show that in the non-relativistic limit the Lorentz transformations do indeed correctly reduce to Galilean transformations.*

2.2.4 Four-vectors

A very convenient notation is the Lorentz covariant four-vector formalism, which we shall now introduce. First, a little exercise:

Exercise 2.2.3. *Show that the quantity $c^2t^2 - \boldsymbol{x}^2$ is invariant under Lorentz transformations (i.e. that it is a scalar quantity).*

Since $c^2t^2 - \boldsymbol{x}^2$ thus behaves as a scalar under Lorentz transformations , it is convenient to define a space–time *four-vector* as follows:

$$x^\mu := (ct, \boldsymbol{x}) \qquad (\mu = 0,1,2,3) \qquad (2.2.5a)$$

and

$$x_\mu := (ct, -\boldsymbol{x}), \qquad (2.2.5b)$$

where, note, $x_0 := x^0 \equiv ct$. The first form is known as a contravariant vector and the second covariant.* The scalar product of such four-vectors may then be defined

* To avoid possible confusion, we have $x = x^1 = -x_1$, $y = x^2 = -x_2$ and $z = x^3 = -x_3$.

as

$$x^2 \;=\; x{\cdot}x \;=\; x^\mu x_\mu \;=\; x_\mu x^\mu \;:=\; (ct)^2 - \boldsymbol{x}^2. \tag{2.2.6}$$

According to the sign of x^2 we say that the four-vector x^μ is time-like (positive) or space-like (negative) and in the special case of $x^2 = 0$ we say it is light-like. For two events to be causally related they must have a time-like separation.

It should be obvious that the distinction between upper and lower indices is vital to this notation. The relation between the two types of vectors is defined in terms of the *metric tensor*

$$g^{\mu\nu} \;=\; g_{\mu\nu} \;:=\; \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \tag{2.2.7}$$

using which we may write

$$x^\mu \;=\; g^{\mu\nu} x_\nu, \quad x_\mu \;=\; g_{\mu\nu} x^\nu \tag{2.2.8}$$

and scalar products may now be expressed in a more symmetric fashion:

$$x{\cdot}y \;=\; x^\mu y_\mu \;=\; x_\mu y^\mu \;=\; g_{\mu\nu} x^\mu y^\nu \;=\; g^{\mu\nu} x_\mu y_\nu. \tag{2.2.9}$$

One can show that a generic such scalar product is invariant. Let is first write the Lorentz transformation in matrix form: define $\Lambda^\mu{}_\nu$ such that

$$x'^\mu \;=\; \Lambda^\mu{}_\nu x^\nu. \tag{2.2.10}$$

For a boost v (velocity) of the reference frame along the positive z-axis say (spatial rotational invariance is manifest), we then clearly have

$$\Lambda^\mu{}_\nu \;=\; \begin{pmatrix} \gamma & 0 & 0 & -\gamma\beta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\gamma\beta & 0 & 0 & \gamma \end{pmatrix} \;=\; \begin{pmatrix} \cosh\chi & 0 & 0 & -\sinh\chi \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\sinh\chi & 0 & 0 & \cosh\chi \end{pmatrix}, \tag{2.2.11}$$

with $\tanh\chi \equiv \beta$. So now

$$\begin{aligned} x'{\cdot}y' \;&=\; g_{\mu\nu} x'^\mu y'^\nu \\ &=\; g_{\mu\nu} \Lambda^\mu{}_\rho x^\rho \Lambda^\nu{}_\sigma y^\sigma \\ &=\; \Lambda^\mu{}_\rho \Lambda_{\mu\sigma} x^\rho y^\sigma \\ &=\; g_{\rho\sigma} x^\rho y^\sigma \;=\; x{\cdot}y, \end{aligned} \tag{2.2.12}$$

where we have used the fact that $\cosh^2\chi - \sinh^2\chi = 1$ and thus $\Lambda^\mu{}_\rho \Lambda_{\mu\sigma} = g_{\rho\sigma}$. Note that from this last identity the invariance of $g_{\mu\nu}$ also follows trivially.

There is one further useful invariant tensor, the Levi–Civita or alternating symbol:

$$\begin{aligned} \varepsilon_{\mu\nu\rho\sigma} &= \begin{cases} +1 & \text{if } (\mu\nu\rho\sigma) \text{ is an } \textit{even} \text{ permutation of (0123)}, \\ -1 & \text{if } (\mu\nu\rho\sigma) \text{ is an } \textit{odd} \text{ permutation of (0123)}, \\ 0 & \text{otherwise.} \end{cases} \\ &= -\varepsilon^{\mu\nu\rho\sigma} \end{aligned} \tag{2.2.13}$$

Note that with this sign convention $\varepsilon_{0123} = +1$ and $\varepsilon^{0123} = -1$.*

Exercise 2.2.4. *Show that the Levi–Civita alternating tensor $\varepsilon^{\mu\nu\rho\sigma}$ is indeed invariant under (proper) Lorentz transformations.*

Other conventions

Let us now simply list two other four-vector conventions that may be found in various texts. The simplest variation is to introduce an overall sign into the metric tensor $g^{\mu\nu}$, thus

$$\tilde{g}^{\mu\nu} = \tilde{g}_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.2.14}$$

Another form, which has the virtue of eliminating the need for upper and lower indices, is to introduce an imaginary time as the fourth (rather than zeroth) component:

$$x^\mu = x_\mu \equiv (\boldsymbol{x}, ict) \equiv (x_1, x_2, x_3, x_4). \tag{2.2.15}$$

The obvious drawback is, of course, that such a convention introduces spurious complexity into quantities that are in fact real.

Other specific four-vectors

We can now define various other specific four-vectors using, *e.g.*, the de Broglie and Einstein relations for length–momentum and time–energy respectively. We thus have

$$p^\mu := (E/c, \boldsymbol{p}). \tag{2.2.16}$$

* Beware that in some textbooks the *opposite* sign convention is adopted.

This definition of the energy–momentum vector immediately returns Einstein's relation for the energy, momentum and mass of a particle:

$$p^2 \;=\; E^2/c^2 - \boldsymbol{p}^2 \;=\; m^2c^2. \tag{2.2.17}$$

Trivially, we have that for $\boldsymbol{p}=0$, $E=mc^2$ while for $\boldsymbol{p}\neq 0$, $E=\gamma mc^2$, moreover $\boldsymbol{p}c=\boldsymbol{\beta}E$ (where $\boldsymbol{\beta}:=\boldsymbol{v}/c$).

Another useful four-vector is the four-gradient:

$$\partial^\mu \;\equiv\; \frac{\partial}{\partial x_\mu} \;:=\; \left(\frac{\partial}{c\partial t}, -\boldsymbol{\nabla}\right). \tag{2.2.18}$$

Note the minus sign in front of the space part. From which we have the Laplace operator of Minkowski space, known as the d'Alembert operator (represented by $\Box$) or more simply the d'Alembertian or wave operator:

$$\begin{aligned}\Box \;&=\; \partial^\mu\partial_\mu \;=\; g^{\mu\nu}\partial_\nu\partial_\mu \;=\; \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2} \\ &=\; \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2 \;=\; \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \Delta.\end{aligned} \tag{2.2.19}$$

Analogously, we may define a four-momentum derivative

$$\frac{\partial}{\partial p_\mu} \;:=\; \left(c\frac{\partial}{\partial E}, -\frac{\partial}{\partial \boldsymbol{p}}\right). \tag{2.2.20}$$

Natural units

At this point it becomes natural to use the two universal constants derived from quantum mechanics and special relativity, $\hbar$ and c respectively, to define a system of natural units—to do this we simply set $\hbar=1=c$. Thus, t and x are now measured in the same units (say seconds, with a multiplicative factor of c to convert times into lengths) as too are E, p and m (in GeV, say). Finally then, one can also measure time, say, in GeV^{-1}. An important conversion factor is the product already introduced

$$\hbar c \;=\; 197\ \mathrm{MeV}\cdot\mathrm{fm}. \tag{2.2.21}$$

Similarly, the conversion between E^2 and cross-section is given by

$$(\hbar c)^2 \;=\; 0.389\ \mathrm{GeV}^2\cdot\mathrm{mb}. \tag{2.2.22}$$

Note that these quantities then set the natural scales for quantum mechanical and/or relativistic effects.

2.2.5 Lorentz-invariant phase space

The final topic of this section is the integration measure in momentum space—used in the description of final-state momenta distributions. In particular, we note that there is a need for a Lorentz-invariant definition of phase space. This is rather obvious: experimental conditions may take any one of three possible and common configurations:

(*i*) fixed target with high-energy beam,

(*ii*) high equal-energy colliding beams,

(*iii*) asymmetric high-energy colliding beams

and one last important case

(*iv*) particle decay in flight.

Only in the second case is the (preferred) centre–of–mass system immediate, in the others a Lorentz boost must be performed. There are countless examples of all four cases in the experimental study of high-energy particle physics.

The obvious problem is the transformation of the sub-volume in momentum space $\mathrm{d}^3\boldsymbol{p} = \mathrm{d}p_x \mathrm{d}p_y \mathrm{d}p_z$: while a boost along the z-axis leaves $\mathrm{d}p_x \mathrm{d}p_y$ unaltered, $\mathrm{d}p_z$ must be transformed correctly. Note that the cross-section itself (being defined in the transverse plane with respect to the beam direction) is automatically invariant. Thus, while we could in principle accept a non-invariant definition of the phase-space element together with a non-invariant differential cross-section, this does involve an unnecessary complication of adding reference to a particular system when defining or measuring *differential* cross-sections. It makes sense then to seek an invariant definition of the phase-space element, with which also to define an invariant differential cross-section.

As noted above, the problem lies with the piece $\mathrm{d}p_z$ and one might be tempted to try a form such as $\mathrm{d}^3\boldsymbol{p}/p_z$. Now, the Lorentz transformation for p_z is

$$p_z \;=\; \gamma(\beta E^* + p_z^*), \tag{2.2.23}$$

where we are using E^*, p_z^* (and therefore also β^* and γ^*) for the particle in the observer's system while β and γ indicate generic boosts of the system (without loss of generality always in the z-direction though). Differentiating the above, we obtain

$$\mathrm{d}p_z \;=\; \gamma(\beta \mathrm{d}E^* + \mathrm{d}p_z^*). \tag{2.2.24}$$

For a real (on-shell) particle $E^{*2} = \boldsymbol{p}^{*2} + m^2$ and thus, for p_x^* and p_y^* fixed, we have

$$2E^* \mathrm{d}E^* \;=\; 2p_z^* \mathrm{d}p_z^*. \tag{2.2.25}$$

Substituting for $\mathrm{d}E^*$ in the previous expression, we have

$$\mathrm{d}p_z \;=\; \gamma(\beta p_z^*/E^* + 1)\,\mathrm{d}p_z^*, \tag{2.2.26}$$

from which we find

$$\frac{\mathrm{d}p_z}{p_z} \;=\; \frac{(\beta p_z^*/E^* + 1)}{(\beta E^*/p_z^* + 1)}\,\frac{\mathrm{d}p_z^*}{p_z^*}. \tag{2.2.27}$$

This is evidently *not* invariant. However, consider now the transformation applied to E^*:

$$E \;=\; \gamma(\beta p_z^*/E^* + 1)\,E^*, \tag{2.2.28}$$

We may thus achieve our objective with the following:

$$\frac{\mathrm{d}p_z}{E} \;=\; \frac{\mathrm{d}p_z^*}{E^*}. \tag{2.2.29}$$

It is instructive to examine another approach to the problem. We start from the observation that the natural four-vector phase space d^4p is automatically invariant.

Exercise 2.2.5. *Show that* d^4p *is invariant under general Lorentz transformations.*

Since we are dealing with *real* final-state particles, we must enforce the (four-vector) mass-shell condition $p^2 = m^2$. Given that this is an invariant equation, the simplest and most general realisation is via a Dirac δ-function: $\delta^+(p^2 - m^2)$, where '+' indicates that only the positie-energy solution is to be considered. Thus, the product $\mathrm{d}^4p\,\delta^+(p^2 - m^2)$ also remains invariant.

Exercise 2.2.6. *Expand* $\mathrm{d}^4p\,\delta^+(p^2 - m^2)$ *in terms of the individual space–time components and thus use the* δ*-function to perform the energy integral. The result should be as above (with an extra factor* 2 *in the denominator).*

The final, generally adopted, expression for the so-called Lorentz-invariant phase space (LIPS) is therefore

$$\begin{aligned} \mathrm{dLIPS} \;&=\; \left.\frac{\mathrm{d}^3\boldsymbol{p}}{(2\pi)^3 2E}\right|_{E^2=\boldsymbol{p}^2+m^2} \\ &=\; \frac{\mathrm{d}^4\boldsymbol{p}}{(2\pi)^4}\,2\pi\delta(p^2 - m^2), \end{aligned} \tag{2.2.30}$$

where the 2π factors are conventional. Note that the chosen form makes no explicit reference to a particular direction and is therefore invariant under boosts in *any* direction. A particularly convenient, though little-used, convention allowing tidier expressions exploits the following definitions:

$$\text{đ} \;:=\; \mathrm{d}/2\pi \qquad \text{and} \qquad \delta \;:=\; 2\pi\delta, \tag{2.2.31}$$

from which we have

$$\begin{aligned} \text{dLIPS} &= \frac{đ^3\boldsymbol{p}}{2E} \\ &= đ^4\boldsymbol{p}\ \delta(p^2 - m^2). \end{aligned} \tag{2.2.32}$$

2.3 Bibliography

Cathcart, B. (2005), *The Fly in the Cathedral: How a Group of Cambridge Scientists Won the International Race to Split the Atom* (Farrar, Straus and Giroux), 1st. American edition.

Einstein, A. (1905), *Annalen Phys.* **18**, 639.

Michelson, A.A. and Morley, E.W. (1887), *Am. J. Sci.* **34**, 333; *Phil. Mag.* 449.

Chapter 3

Nuclear Structure

3.1 Nuclear masses

Before turning to an examination of the various models describing the structure of nuclei, and in particular their masses and related binding energies, let us discuss the general method used for measuring such small masses.

3.1.1 Nuclear reactions

First of all, note that whereas absolute measurements are understandably not easy, the precision obtained on relative masses is rather high; one need only measure the energy released in nuclear transitions. A typical example of a reaction used in this manner is

$$^{26}_{12}\mathrm{Mg} + p \rightarrow {}^{23}_{11}\mathrm{Na} + \alpha, \tag{3.1.1}$$

where, we recall, the notation is such that ${}^{A}_{Z}\mathrm{X}$ describes a nucleus corresponding to the chemical element X, containing A nucleons (protons and neutrons) of which there are Z protons (occasionally, a redundant following lower index N indicates the number of neutrons, where, of course, $N = Z - A$). Formally, one refers to the various nucleus types as *nuclides*, nuclei of the same mass number A as *isobars* and of the same Z (*i.e.*, corresponding to the same chemical element) as *isotopes*. Nuclear physicists often prefer a more compact notation for reactions, in which Eq. (3.1.1) may be represented by

$$^{26}\mathrm{Mg}(p,\alpha)^{23}\mathrm{Na}, \tag{3.1.2}$$

where the first object inside the brackets (p here) is understood to be the projectile and the second (α here) a typically small object effectively ejected.

Now, provided the final nucleus is produced directly in the ground state (which

is checked by observing that, *e.g.*, *no* photons are emitted); then, given that the masses of the proton and the α-particle are known, we can determine the mass difference between the two nuclei directly from the change in kinetic energy, thus

$$Q \;=\; \Delta T \;=\; \Delta m\, c^2, \tag{3.1.3}$$

where Q is the so-called Q-value (or energy released) of the reaction, ΔT is the variation in total kinetic energy and Δm is the overall mass difference between the initial and final states. Since kinetic energies (which here are typically of order MeV) may be measured with a precision of the order of keV, one immediately sees that the precision on the relative mass is of order one part per» million.

Exercise 3.1.1. *Using relativistic kinematics show that in a two-to-two-body reaction the exact formula for the laboratory threshold kinetic energy (the minimum energy required for the process to proceed) for a projectile incident on a fixed target is given by*

$$E_{kin} \;=\; -Q \sum_{i=1}^{4} m_i/2m_2,$$

where the projectile is particle 1, the target 2 and the two final-state particles 3 and 4 while Q is the energy released (it is negative for an endothermic reaction). Show also that in the non-relativistic limit ($|Q| \ll m_1+m_2$) this reduces to the more standard

$$E_{kin} \;=\; -Q(1 + m_1/m_2).$$

To what does the extra energy required ($-Qm_1/m_2$) correspond?
Consider the case $m_1 = 0$ and try to understand why $E_{kin} = -Q$ in that limit.
What is its value in the (relativistic) case $|Q| \approx m_2$?

3.1.2 Nuclear mass units

This is a good point to introduce the standard unit for describing atomic or nuclear masses: the *atomic mass unit* (amu or more simply u). From the preceding discussion, it should be clear that it is better to define nuclear masses (we here shall use nuclear and atomic interchangeably although there is an obvious difference*) relative to some fixed standard mass. The chosen standard is the mass of an atom

* Clearly, at a certain level of precision corrections should be applied to account for the masses of the atomic electrons and their binding energies.

of ^{12}C and we thus define

$$\begin{aligned} 1 \text{ u} &:= \tfrac{1}{12} m_{^{12}\text{C}} \\ &\simeq 1.66 \times 10^{-27} \text{ kg} \\ &\simeq 931.494 \text{ MeV}. \end{aligned} \tag{3.1.4a}$$

This may be compared with the proton and neutron masses:

$$m_p \simeq 938.3 \text{ MeV}, \tag{3.1.4b}$$

$$m_n \simeq 939.6 \text{ MeV}. \tag{3.1.4c}$$

Note that the officially quoted mass values are then usually rather more precise when presented in atomic units (u) rather than as absolute values such as MeV.

Exercise 3.1.2. *Why is the atomic mass unit smaller than both the proton and neutron masses?*

3.1.3 Mass spectroscopy

The standard techniques for nuclear-mass measurement exploit the Lorentz force, $\boldsymbol{F} = Q(\boldsymbol{E} + \boldsymbol{B}_\wedge \boldsymbol{v})$ on a moving charged particle (*e.g.*, an ionised atom) and the fact that the radius of curvature of the particle track in the presence of magnetic or electric fields is precisely related to its mass. For a particle of mass M and electric charge Q (we ignore any possible intrinsic magnetic moment) moving in electric or magnetic fields E and B, the two radii of curvature are

$$r_E = \frac{M}{Q}\frac{v^2}{E}, \tag{3.1.5a}$$

$$r_B = \frac{M}{Q}\frac{v}{B}. \tag{3.1.5b}$$

For fixed, known radius of curvature and field strength, together with a measurement of the particle velocity, one could use either of these equations to determine the mass. However, by using the two in parallel it is possible to eliminate the velocity and thus simplify the measurement:

$$\frac{M}{Q} = \frac{r_E}{r_B^2}\frac{E}{B^2}. \tag{3.1.6}$$

This leads us to an apparatus known as the *double-focus spectrometer*, see Fig. 3.1. The ions of varying velocities first pass through a collimator and then into the

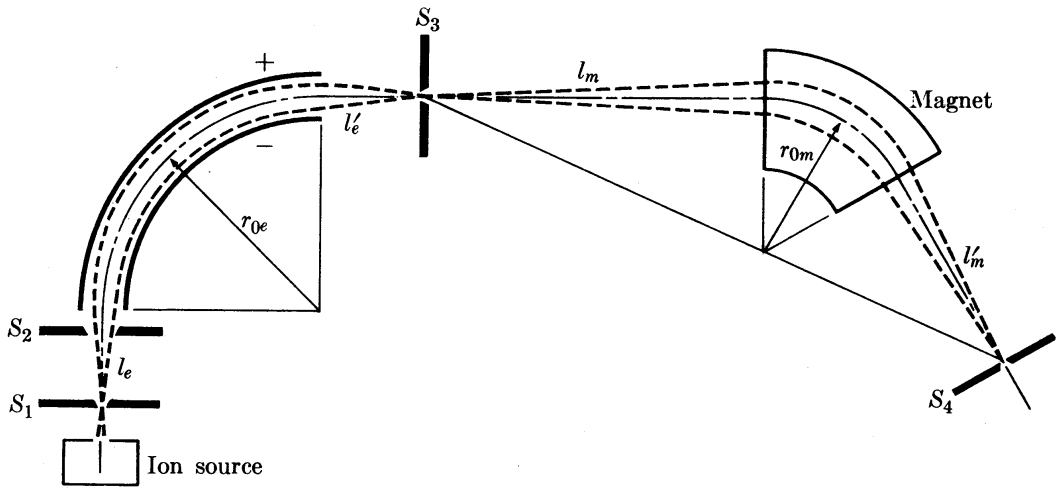


Figure 3.1: A schematic representation of the double-focus spectrometer.

first arm of the spectrometer. Here only those of the velocity corresponding to formula 3.1.5a pass. The second arm, receiving a monochromatic beam, only passes ions if their mass corresponds to formula 3.1.5b. Thus, by varying the field strength of one or other arm until ions are detected at the end of the system, the mass may be determined from formula 3.1.6. A continuously updated experimental database is published by the International Union of Pure and Applied Chemistry (for the latest version, see Meija *et al.*, 2016).

3.1.4 Nuclear masses and binding energies

Armed with a (rather large) table of isotopic masses, we can now examine the behaviour or variation with respect to Z and A. The simplest plot we can imagine is that of the masses of the most stable isotopes as a function of A. Let us first remark that the absolute mass itself is of little theoretical interest; the more relevant quantity is the binding energy, which may easily be extracted from its definition:

$$B(A,Z) := Zm_pc^2 + (A-Z)m_nc^2 - m_{A,Z}c^2, \tag{3.1.7}$$

where $m_{p,n}$ are the *free* proton and neutron masses and $m_{A,Z}$ is the nuclear mass in question. It is also useful to factor out the "size" of the single nucleus and therefore plot B/A, the average biding energy per nucleon, as a function of A, see Fig. 3.2.

Two aspects of the plot are obvious: firstly, it rises fairly linearly from zero for small A and, secondly, reaches an approximate plateau for values of A around

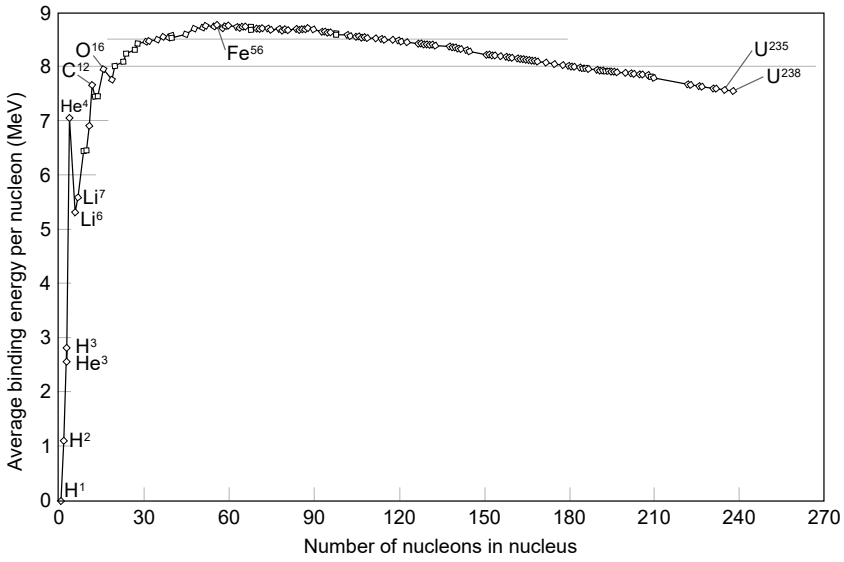


Figure 3.2: The variation of (β) stable nuclear binding energies (per nucleon) with respect to A.

that of iron ($A = 56$).* It then falls very slowly for higher A. The first observation implies a *total* binding energy that grows as A^2 (for small A), which is reminiscent of the case of a growing uniform spherical distribution of electric charge (although there the force would be repulsive) or mass in gravity. We might thus imagine that each nucleon "binds" to all the others with more or less uniform strength. However, the second observation indicates that the binding energy essentially saturates (at nearly 9 MeV per nucleon) for large enough A.

If we now add to this the experimental fact that nuclear density is very constant with respect to atomic number and mass,† then we may interpret this as indicating a finite range for the nuclear force, in such a way that it is limited to essentially the nearest neighbours. In other words, each nucleon in the bulk only interacts with the others that it actually touches and therefore can only bring a constant contribution (*i.e.* independent of A) to the total binding energy. This is already an important observation: unlike the electrostatic (and gravitational) force, with the typical Coulomb behaviour, the nuclear force has a rather sharp cut-off at a distance of the order of 1 or 2 fm.

The modern theory of the nuclear or *strong* interaction explains this in terms of a phenomenon called *confinement*. It is now believed that the interaction underlying the nuclear force is described by a gauge theory similar to electromagnetism but with two important differences: the natural strength is more than an order of

* Note that the actual peak of binding energy is for ^{62}Ni, which has $B/A = 8.8$ MeV.

† We shall discuss the experimental evidence later in the course.

magnitude larger and the carriers of the force (the *gluons*, which are the equivalent of the photon in electromagnetism) are themselves charged—note that here we are speaking not of the electric charge, but of what is known as the *colour* or strong charge. The nature of the theory is then believed to be such that an essentially infinite energy would be required to free any object carrying colour charge and that therefore the only free objects we find in Nature are colour neutral. In particular, the gluon is not allowed to propagate over the same long distances as the photon and the force has therefore an effectively finite range.

3.2 The liquid-drop model

The constant density or incompressibility of nuclear matter combined with the finite range of the interaction naturally leads to analogy with a liquid drop. The liquid-drop model was first proposed by Gamow and later developed by Bohr and Wheeler, who in 1939 established the equivalence between a nucleus and a liquid drop. Water is known to be highly incompressible and the binding via so-called *hydrogen bonds* leads to a short-range interaction, *i.e.* involving only nearest neighbours. A spherical nucleus has the largest binding energy. Moreover, if the surface-tension energy of the nucleus is greater than the electrostatic Coulomb repulsion between the positively charged protons then the nucleus is stable, otherwise it may deform and decay via fission. Thus, von Weizsäcker (1935) was led to propose a semi-empirical mass formula based on the liquid-drop model, which might reasonably be expected to work well for nuclei near to and above the saturation point. For simplicity, we shall present this as a formula for the binding energy (the difference being the overall sign and the baseline).

The first and, *a priori*, presumably dominant term should be that of the bulk binding contribution, which, being proportional to the volume (in a liquid drop), is usually represented as

$$B_{\rm vol} = c_v A, \tag{3.2.1}$$

where c_v is a (positive) constant giving the bulk binding strength. Here and in the following we shall introduce a series of (positive) phenomenological constants that should then be fixed by an overall fit to the data. Note that implicit in this term is the empirical fact that the proton–proton, proton–neutron and neutron–neutron strong interaction forces are essentially identical (*cf.* the gravitational force).*

In order of importance, it turns out that the next contribution is due to the Coulomb repulsion between the protons. The (negative) Coulomb energy of a

* We shall see later that this is related to what is known as the *isotopic-spin* (or *isospin*) symmetry between protons and neutrons.

uniform distribution of Z charges in a sphere of radius R is

$$E_{\text{Coulomb}} = -\frac{3}{5}\frac{Z(Z-1)}{R}\frac{e^2}{4\pi\epsilon_0}. \tag{3.2.2}$$

Let us make a short aside here and convert the complicated fraction involving the charge into something more readily remembered. It is standard practice to define the so-called *fine structure constant* as

$$\alpha := \frac{e^2}{4\pi\epsilon_0}\frac{1}{\hbar c} \simeq \frac{1}{137}. \tag{3.2.3}$$

For typical nuclear Z and R, the above expression does not estimate well the absolute size of the effect—obviously, the assumption, among others, of uniformity is rather dubious—and so the following simple parametrisation is usually adopted

$$B_{\text{Coulomb}} = -c_c\frac{Z^2}{A^{1/3}}, \tag{3.2.4}$$

where the constancy of the density has been exploited to set $A \propto r^3$. Such a term naturally explains a further empirical fact: as Z grows, there is a tendency for the most stable nuclei (for A fixed) to contain more neutrons than protons.

The third and final term, motivated by the liquid-drop picture, is effectively a surface term. Quite simply, the nucleons at or near the surface do not "see" the same number of nearest neighbours as those situated in the bulk. We are thus led to introduce another negative term that should grow as the surface area of the nucleus or $A^{2/3}$:

$$B_{\text{surface}} = -c_s A^{2/3}. \tag{3.2.5}$$

Already with just these three terms, apart from some sharp local fluctuations, the overall curve in Fig. 3.2 is well described.

However, the nuclear spectrum displays richer behaviour than can be seen from the figure: there is a fine structure not shown in the plot and there is also a marked dependence on the relative numbers of protons and neutrons (for A fixed). The most stable of the lighter nuclei always have equal numbers of protons and neutrons and the mass spectrum, interpreted in terms of a binding energy, displays a so-called *valley of stability* in the $Z-N$ plane running along a contour $Z \lesssim N$, with Z becoming progressively less and less with respect to N as A increases. This last variation is clearly due to the Coulomb repulsion of the protons. To explain the valley of stability, we need to exploit a little knowledge of quantum effects. Although we shall present a more detailed model shortly (the Fermi-gas model), let us give a more heuristic derivation immediately.

Without attempting any real justification, let us assume that the nucleus may be described as a collection of nucleons bound inside some mean potential well (which for simplicity we approximate as square). There are two different species of fermions (protons and neutrons), which must separately obey the Pauli exclusion principle. That is, no two protons (or neutrons) may occupy precisely the same quantum-mechanical state. As a first approximation then we shall ignore the effects of the Coulomb repulsion and thus imagine the nucleus as depicted in Fig. 3.3. The idea then is simply to fill the states (separately for protons and

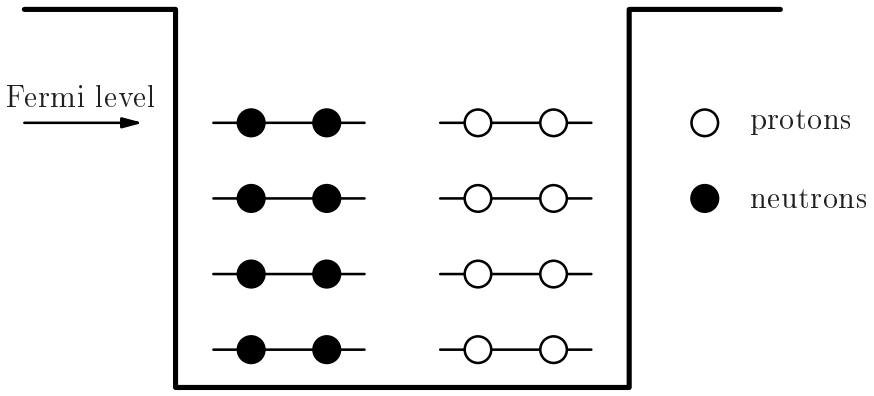


Figure 3.3: A schematic representation of the nuclear potential well and the level occupation of protons and neutrons corresponding to the lowest nuclear energy state for a given fixed A.

neutrons) progressively from the lowest upwards, placing two nucleons in each level (corresponding to spin up and down). The upper-most level defines the so-called Fermi energy (see Fig. 3.3). For a given fixed A, the configuration shown corresponds to the lowest overall energy state for the nucleus in question.*

It is immediately obvious that to replace, say, a neutron with a proton, the new proton must necessarily occupy an energy state higher than that of the current Fermi level. Let us assign ΔE to the energy difference between levels (which we assume roughly constant), then the energy cost will be precisely ΔE. However, no more than two nucleons may be exchanged at this price; the next pair will be removed from a lower energy level and be replaced at an even higher level, thus costing $3\Delta E$ per nucleon. The energy required to exchange n such pairs (either protons or neutrons) is thus given by:

$$E_{\rm asym} \;=\; 2\big[1+3+\cdots+(2n-1)\big]\Delta E \;=\; 2n^2\Delta E. \qquad (3.2.6)$$

* Note that the proton and neutron Fermi levels must more-or-less coincide for a nucleus to be stable against β-decay.

Now, for such an exchange, $|Z-N|=4n$ and therefore we finally have

$$E_{\text{asym}} = \frac{(Z-N)^2}{8}\Delta E. \tag{3.2.7}$$

Experimentally, it is found that this contribution is less important for large nuclei (evidently the level spacing decreases with A) and thus the form used in the semi-empirical mass formula is

$$B_{\text{asym}} = -c_a\frac{(Z-N)^2}{A} = -c_a\frac{(A-2Z)^2}{A}, \tag{3.2.8}$$

where the substitution $N=A-Z$ has been applied.

One last contribution may be included, which takes into account the observation that the masses of nuclei with even numbers of protons and/neutrons are lower with respect to those with odd numbers (again, this clearly cannot be explained in a purely classical liquid-drop picture). There are then three cases to consider: from the most energetically favourable, "even–even", to the least, "odd–odd", with the "even–odd" and "odd–even" cases being equally favoured at a mid-point between the two extremes. We are thus led to a so-called pairing term of the form

$$B_{\text{pairing}} = c_p\frac{\varepsilon(Z,N)}{A^{3/4}}, \tag{3.2.9}$$

with $\varepsilon(Z,N)=+1,0,-1$ for the pair (Z,N) respectively (even, even), (even, odd) or (odd, even) and (odd, odd). Note that the two extreme cases both correspond to A even while the intermediate case always has A odd. A justification for such a term may be given via the *shell* model (to be discussed shortly): an even number of protons or neutrons implies that each orbital will be occupied by precisely two nucleons (one spin up and one down), which will therefore spend a large amount of time in the same physical space. Since the nuclear force is basically attractive, this will enhance the binding-energy contribution for the pair. In contrast, a lone nucleon spends less time in close contact with other nucleons and thus contributes less. As already noted, the nuclear force is also essentially flavour blind and treats protons and neutrons equally; hence, equally spaced splittings. The power of A in the denominator is not well determined and values, ranging from $\frac{1}{2}$ to 1, may be found in the literature.

It is remarkable that with the final five-parameter formula,

$$B(A,Z) = c_vA - c_c\frac{Z^2}{A^{1/3}} - c_sA^{2/3} - c_a\frac{(A-2Z)^2}{A} + c_p\frac{\varepsilon(Z,N)}{A^{3/4}}, \tag{3.2.10}$$

it is possible to obtain a fit to all available data to within an MeV or so over the

entire range (hundreds of nuclides). Note that the contributions of the individual terms are of the order of a few MeV per nucleon while total binding energies reach the order of 10^3 MeV. On the other hand, there are noticeable deviations: ^{4}He has a total binding energy of 28.3 MeV or slightly more than 7 MeV per nucleon whereas nearby in the binding-energy plot the binding energies are much less. Indeed, the most striking deviations are peaks in the binding energies corresponding to certain numbers of protons and/or neutrons—these are known as *magic numbers* (following atomic physics) and will be explained shortly within the framework of the *shell* model.

Typical values for the parameters of the semi-empirical von Weizsäcker formula are listed in Table 3.1. While the size of the Coulomb term might appear rather

Table 3.1: Typical global-fit values (in MeV) of the five parameters appearing in the semi-empirical von Weizsäcker formula.

$c_v = 15.68$
$c_c = \ 0.717$
$c_s = 18.56$
$c_a = 28.1$
$c_p = 34.0$

small (and therefore insignificant) in comparison with the others, being proportional to $Z^2/A^{1/3}$ it is indeed suppressed for small values of Z (as it should be), but grows rapidly for larger nuclei. This is easily seen by substituting the approximation $Z \approx A/2$, whence the term behaves as $A^{5/3}$ and thus manifestly grows with respect to, *e.g.*, the volume term: the relative suppression factor of ~ 20 in the coefficient is recovered entirely for $A \approx 90$. Likewise, the asymmetry term may appear particularly large. However, in the B/A plot shown earlier it plays little or no role, since by definition $Z - N \approx 0$ on average there while it does correctly describe the empirically steep rise away from the bottom of the valley of stability.

Of course, the equivalent mass formula may now be readily written,

$$m_{A,Z} \ = \ Zm_p + Nm_n - c_v A + c_c \frac{Z^2}{A^{1/3}} + c_s A^{2/3} + c_a \frac{(A-2Z)^2}{A} - c_p \frac{\varepsilon(Z,N)}{A^{3/4}}, \quad (3.2.11)$$

Exercise 3.2.1. *Use the semi-empirical mass formula Eq. (3.2.11) to calculate the curve in the Z–N plane (or rather Z as a function of A) of maximum stability (or minimum mass).*

3.3 Nuclear fission

With this simple model for the nuclear binding energy we are now in a position to discuss one of the many forms of nuclear decay: namely, *fission*. To do so, we shall only need the more classical and empirical pieces of the formula constructed in the previous section; we shall, however, need to apply concepts taken from quantum mechanics.

We have seen that binding energies peak around iron and then, on average, decrease slowly with growing A. Therefore, in principle, a nucleus with $Z \gtrsim 40$ may split into two smaller (more or less equal) pieces with an overall gain in energy. For $Z \lesssim 40$, the daughter nuclei lie too far below iron and are therefore also poorly bound, leaving no energy gain. For nuclei in the region $A \approx 120$, $B/A \simeq 8.5\,\mathrm{MeV}$, whereas for $A \approx 240$ (uranium and its neighbours) $B/A \simeq 7.6\,\mathrm{MeV}$. Thus, if a heavy nucleus undergoes fission, the energy released is potentially enormous: in the above case we have

$$\Delta E \;\simeq\; 240\,(8.5 - 7.6) \;\simeq\; 220\ \mathrm{MeV}. \tag{3.3.1}$$

In reality, in any fission process a varying number of neutrons (usually around 2 or 3) are ejected, which reduces the gain from binding energy and therefore the true energy release is a little less, around 200 MeV.

Spontaneous fission does not occur frequently (the half-life for *e.g.* ^{235}U *fission* is $3.5 \times 10^{17}\,\mathrm{yr}$);* so there must be a barrier hindering the decay. We may imagine that the process is essential one of deformation of the parent nucleus into an elongated form (ellipsoid), which eventually splits into the two daughter nuclei. Considering that the surface energy increases in magnitude during this process while the Coulomb energy decreases, we may assume a form of the type shown in Fig. 3.4. If we make the approximation of an ellipsoid for the deforming nucleus, then the two contributions to the potential energy (note the opposite sign with respect to the binding energy) will have the following behaviour as a function of the deformation or eccentricity ε:

$$E_s \;\simeq\; (1 + \tfrac{2}{5}\varepsilon^2 + \cdots)\, c_s A^{2/3}, \tag{3.3.2a}$$

$$E_c \;\simeq\; (1 - \tfrac{1}{5}\varepsilon^2 + \cdots)\, c_c Z^2 A^{-1/3}, \tag{3.3.2b}$$

where the ε is defined through its relation to the major and minor axes (a and b) and the original radius R for a constant volume (since A is constant and nuclear matter is assumed incompressible), $a \simeq R(1+\varepsilon)$ and $b \simeq R(1-\varepsilon/2)$. Thus, the

* The half-life is the time in which half of any initial sample decays. The relation between half-life and lifetime is $\tau = \tau_{1/2}/\ln 2$.

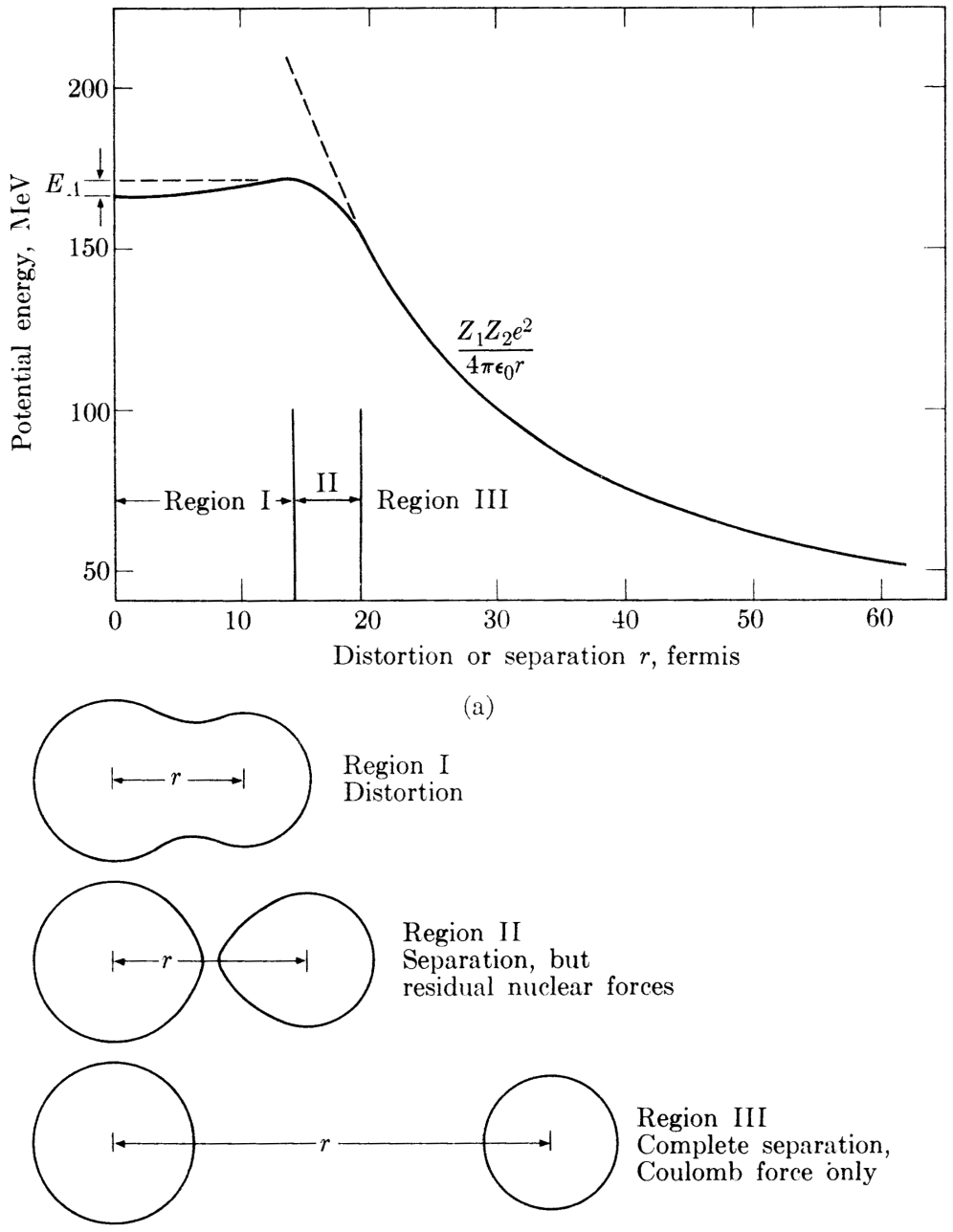


Figure 3.4: A schematic plot of the potential energy of a nucleus in fission as a function of the separation between the two daughter nuclei. E_A is the activation energy.

variation in potential energy is

$$\Delta E \equiv \Delta E_s + \Delta E_c \simeq \tfrac{1}{5}\varepsilon^2 \left(2c_s A - c_c Z^2\right) A^{-1/3}. \tag{3.3.3}$$

This approximation is made for small values of ε and one should not forget that the Coulomb contribution must dominate for large deformations and cause an overall *reduction* in potential energy.

Now, for these small deformations we see that either sign is possible for ΔE. A positive sign indicates the presence of a barrier (the solid curve in Fig. 3.4) and is thus the general case. In the case of negative values, the corresponding nuclei would be absolutely unstable and would indeed never have time to form. When positive, ΔE (the energy required to overcome the barrier) is called the activation energy E_A.

The point $2c_sA = c_cZ^2$ marks the critical values of A and Z, for which the transition from metastable to unstable occurs. In other words, the barrier disappears for

$$\frac{Z^2}{A} \gtrsim \frac{2c_s}{c_c} \approx 47.8. \tag{3.3.4}$$

This is the case for nuclei with $Z \gtrsim 114$ and $A \gtrsim 270$; indeed, no stable nuclei are known for such large values. Bohr and Wheeler defined a *fissionability* parameter:

$$x := \frac{E_c}{2E_s} = \frac{Z^2/A}{2c_s/c_c} = \frac{Z^2/A}{(Z^2/A)_{\text{critical}}}. \tag{3.3.5}$$

Where we have defined $(Z^2/A)_{\text{critical}} := 2c_s/c_c$. The point $x = 1$ marks the transition from metastable to unstable while smaller values indicate lesser predisposition to fission.

So far we have discussed spontaneous fission, such as that of ^{235}U, but we might also imagine inducing such a process with an "injection" of neutrons* and energy in order to provoke a more unstable situation. Indeed, from quantum mechanics, we know that the decay rate depends exponentially on the barrier height (and width) and thus any energy brought into the process will increase the rate considerably. For very heavy nuclei ($Z \gtrsim 92$) the barrier for fission is rather low ($\lesssim 6\,\text{MeV}$) and thus a neutron providing a contribution to the binding energy (and perhaps with some amount of kinetic energy) would be sufficient to tip the balance.

A simple example of such a case is $^{239}_{92}$U, which has a fission threshold of $\sim 5.5\,\text{MeV}$ while the neutron capture by a nucleus of $^{238}_{92}$U releases $\sim 4.9\,\text{MeV}$ of binding energy; that is, the reaction is *exothermic*. This is still a little below the threshold, but a kinetic energy of only $\sim 0.6\,\text{MeV}$ is sufficient to overcome the remaining barrier—this would be termed a "fast" neutron (as opposed to slow or thermal). As an aside, one should also note that the cross-section for neutron capture falls as $1/v$ (where v is the neutron velocity) and thus becomes rapidly less probable for fast neutrons.

* Note that injecting protons is essentially impossible owing to the large Coulomb barrier.

However, the extra neutron also alters the parity of the neutron number N and thus the pairing energy. A nucleus with an initial *odd* number of neutrons therefore gains even more energy from neutron capture—the extra energy thus released can then help to push the system over the barrier. An example is the case of $^{235}_{92}$U, which on capture of a neutron to form $^{236}_{92}$U (with an *even* number of neutrons) releases $\sim$6.4 MeV while the fission barrier of $^{236}_{92}$U is just $\sim$5.5 MeV. Thermal neutrons are thus quite sufficient in this case (and have a higher probability) to induce fission. This is one of the processes commonly used in nuclear reactors (and, unfortunately, weapons); other common *fissile* materials are ^{233}Th and ^{239}Pu. We shall study the more technical details of controlled fission later in the course.

3.4 The Fermi-gas model

As already discussed above, a simple quantum-based model for the nucleus may be constructed assuming a potential well in which two different species of identical fermions are bound independently. Note that such a model makes no attempt to explain the possible origins of the well, which is assumed to arise as some sort of mean-field effect generated by the cooperative action of the nucleons themselves. In such a model the nucleons are taken to behave as a gas of fermions moving freely inside the confining potential. The important restriction is that no two identical nucleons may occupy the same spatial–spin state.

Experimentally, we know that the nuclear density is more or less constant over most of the volume and drops rapidly to zero at the surface. Thus, a first working approximation is to assume a three-dimensional square well—with just two free parameters: the depth and the radius. If we wish to be a little more accurate, we should take into account the so-called *skin-depth* (*i.e.*, that the density falls to zero over some finite, but small, range) and also that the protons see a slightly shallower well due to the Coulomb repulsion effect. As a first approximation we shall thus assume the same picture as is described in Fig. 3.3.

In order to simplify calculations, we shall further assume that the precise *shape* of the well is not important and approximate a sphere with a cube of side L. The quantisation conditions on the single-particle wave-functions then become

$$k_i L \;=\; n_i \pi \qquad (i = 1,2,3;\; n_i \in \mathbb{N}), \tag{3.4.1}$$

where $\boldsymbol{k}$ is the wave-number vector associated with the state in question. Allowing for the fact that standing waves only have positive wave-numbers and that we are therefore only interested in the positive octant (or $\frac{1}{8}$) of the k-sphere, the density

of states is then given by

$$\mathrm{d}^3\boldsymbol{n} \;=\; \frac{1}{8}\left(\frac{L}{\pi}\right)^3 \mathrm{d}^3\boldsymbol{k} \;=\; V\,\frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3}, \tag{3.4.2}$$

where V is the nuclear volume. The de Broglie condition $\boldsymbol{p}=\hbar\boldsymbol{k}$ then leads to the standard expression for the density of states:

$$\mathrm{d}^3\boldsymbol{n} \;=\; V\,\frac{\mathrm{d}^3\boldsymbol{p}}{(2\pi\hbar)^3}, \tag{3.4.3}$$

Therefore, in a shell of thickness $\mathrm{d}n$ we have a total number of states equal to

$$\mathrm{d}N \;\equiv\; 4\pi n^2\mathrm{d}n \;=\; V\,\frac{4\pi p^2\mathrm{d}p}{(2\pi\hbar)^3}. \tag{3.4.4}$$

Before continuing, we should address the question of the occupation of these states; recalling that $kT\sim\frac{1}{40}\,\mathrm{eV}$ at 300 K, it is immediately obvious that we may work in the zero-temperature approximation. The available states will thus be filled strictly from the lowest upwards to some highest-energy state or *Fermi level*, corresponding to a Fermi energy E_F and momentum p_F. Integrating the previous expression then gives

$$N \;=\; 2\times\frac{4\pi V}{(2\pi\hbar)^3}\int_0^{p_\mathrm{F}}\mathrm{d}p\,p^2 \;=\; \frac{Vp_\mathrm{F}^3}{3\pi^2\hbar^3}, \tag{3.4.5}$$

where the factor two in front of the integral takes into account that each level can accommodate two spin states. This then is the number of protons or neutrons contained inside the well. The actual value of the Fermi momentum will, in general, be different for protons and neutrons, reflecting the different depth of the individual wells and therefore the number of each species in a given nuclide.

Going back to the spherical approximation, the formula for V is simply

$$V \;=\; \tfrac{4}{3}\pi R^3 \;=\; \tfrac{4}{3}\pi A R_0^3, \tag{3.4.6}$$

where R_0 is the radius associated with the volume occupied by each nucleon. From electron scattering we know that this volume (and hence R_0) is rather independent of A, having a more or less universal value $R_0\simeq 1.21\,\mathrm{fm}$. We thus have

$$N \;=\; \frac{4AR_0^3p_\mathrm{F}^3}{9\pi\hbar^3}. \tag{3.4.7}$$

Further simplifying to the case $Z \simeq N \simeq \frac{1}{2}A$, we finally obtain a *prediction* for the Fermi momentum (recall that $\hbar c = 197\,\mathrm{MeV}\cdot\mathrm{fm}$):

$$p_\mathrm{F} \approx \frac{\hbar}{R_0}\sqrt[3]{\frac{9\pi}{8}} \simeq 250\ \mathrm{MeV}/c. \tag{3.4.8}$$

The corresponding Fermi energy is then

$$E_\mathrm{F} \equiv \frac{p_\mathrm{F}^2}{2M} \simeq 33\ \mathrm{MeV}, \tag{3.4.9}$$

where M is the nucleon mass. We therefore see *a posteriori* that the non-relativistic approximation is more than justified. Later on we shall discuss a process known as quasi-elastic scattering, from which similar values of p_F and E_F are indeed experimentally inferred, at least for large A; for smaller nuclei the effective value of R_0 is larger and one finds that p_F is smaller.

If we now add one further piece of empirical knowledge: namely, that the nuclear binding energy per nucleon is 7–8 MeV, then we have an estimate for the overall *depth* of the well:

$$V_0 = E_\mathrm{F} + B/A \simeq 40\ \mathrm{MeV}. \tag{3.4.10}$$

An immediate observation is that the typical kinetic energy of the nucleons at the Fermi surface is of the same order as the well depth itself and we thus conclude that the outer nucleons are only relatively *weakly* bound inside a nucleus.

Let us also remark that while the *bottom* of the nuclear potential well sits at a different level for protons and neutrons (owing to the presence of the Coulomb potential), the Fermi level must be the *same* for β-stable nuclides.* Indeed, the Fermi-gas model allows us to predict and even estimate the proton–neutron asymmetry term introduced in the semi-empirical liquid-drop model earlier.

Since the mass of the system is determined by the energies of the components, we shall now calculate the mean kinetic energy per nucleon of a given nucleon species:

$$\langle E_\mathrm{kin}\rangle = \frac{\displaystyle\int_0^{p_\mathrm{F}} p^2\mathrm{d}p\,\frac{p^2}{2M}}{\displaystyle\int_0^{p_\mathrm{F}} p^2\mathrm{d}p} = \frac{3}{5}\,\frac{p_\mathrm{F}^2}{2M}. \tag{3.4.11}$$

Using Eq. (3.4.5) for the number N, we obtain the total kinetic energy of the

* We shall discuss nuclear β decay later; suffice it to say that it involves conversion of a neutron into a proton or *vice versa*.

nucleon species:

$$E_{\rm tot} \;=\; N\,\frac{3}{5}\,\frac{p_{\rm F}^2}{2M} \;=\; \frac{V p_{\rm F}^5}{10M\pi^2\hbar^3}. \tag{3.4.12}$$

Or, inverting Eq. (3.4.7) to express $p_{\rm F}$ in terms of N, we have

$$p_{\rm F} \;=\; \sqrt[3]{\frac{N9\pi\hbar^3}{4AR_0^3}} \tag{3.4.13}$$

and therefore

$$E_{\rm tot} \;=\; \frac{\pi^{2/3}}{5}\left(\frac{3}{2}\right)^{7/3}\frac{\hbar^2}{MR_0^2}\,\frac{N^{5/3}}{A^{2/3}}. \tag{3.4.14}$$

Exercise 3.4.1. *Verify the dimensions of the expression above.*

Adding the kinetic energies for both protons and neutrons, we have

$$E_{\rm tot}^{p+n}(A,Z) \;=\; \frac{\pi^{2/3}}{5}\left(\frac{3}{2}\right)^{7/3}\frac{\hbar^2}{MR_0^2}\,\frac{\left[Z^{5/3}+N^{5/3}\right]}{A^{2/3}}. \tag{3.4.15}$$

Using $N=A-Z$, we may expand the above expression in powers of $Z-N$ for fixed A:

$$E_{\rm tot}^{p+n}(A,Z) \;=\; \frac{\pi^{2/3}}{5}\left(\frac{3}{2}\right)^{7/3}\frac{\hbar^2}{MR_0^2}\left(\frac{1}{2}\right)^{2/3}\left[A+\frac{5}{9}\,\frac{(Z-N)^2}{A}+\cdots\right]. \tag{3.4.16}$$

The first term in brackets is a contribution to the volume term while the second has the precise form already used in the semi-empirical mass formula for the asymmetry term, see Eq. (3.2.8). Note, however, that again the calculated coefficient does not correspond well to a fit of the data; this may be attributed to the lack of account of variations in the potential wells themselves for varying $Z-N$.

3.5 The atomic nucleus or shell model

In this section we turn to the final and most complete model of the nucleus, based, as far as possible, on a correct quantum-mechanical treatment: the so-called *atomic nucleus*. By way of introduction, we shall briefly review the electronic structure of atoms in quantum mechanics.

3.5.1 The hydrogen and hydrogenoid atoms

Energy levels and single-electron states

The simplest case to consider is that of hydrogen and hydrogenoid atoms, in which a single electron is bound to a nucleus containing Z protons via the Coulomb potential

$$V_c(r) = -\frac{Ze^2}{r}. \tag{3.5.1}$$

The solutions for the electron states are well known and are usually expressed as $\psi_{n,l,m,m_s}(r,\theta,\phi)$, where the quantum numbers n, l, m and m_s refer respectively to energy, orbital angular momentum, projection of orbital angular momentum along the z-axis and spin projection along the z-axis (these last two are often referred to as *magnetic* quantum numbers). Note that (neglecting the fine and hyperfine structures) the energy only depends on the principal quantum number n:

$$E_n = -\frac{13.6\,Z^2}{n^2}\ \text{eV}. \tag{3.5.2}$$

Moreover, the other quantum numbers are bounded:

$$\begin{aligned} 0 \leq\ & l\ < n, \\ -l \leq\ & m \leq l, \end{aligned} \tag{3.5.3}$$

while, for a spin-½ electron, $m_s = \pm\frac{1}{2}$.

The standard notation in atomic physics assigns letters to the different orbital angular momentum states:

$$\begin{array}{rcccccc} l = & 0 & 1 & 2 & 3 & 4 & \dots \\ & s & p & d & f & g & \dots \end{array}$$

The classification of states is then as follows:

$$\begin{array}{ccccc} n & \text{levels} & & & \\ 1 & 1s^2 & & & \\ 2 & 2s^2 & 2p^6 & & \\ 3 & 3s^2 & 3p^6 & 3d^{10} & \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{array}$$

The index indicates the total degeneracy: since there are two possible spin alignments for each spatial state, we find that for each l there are $2(2l+1)$ states.

Now, in atomic physics the *fine* and *hyperfine* structures are small splittings of the spectral lines (or equivalently energy levels) due to magnetic (*i.e.* spin–orbit and spin–spin) interactions. For the moment we shall ignore such small effects and concentrate on the variations induced when many electrons are present. Put simply, the presence of extra electrons leads to screening of the nuclear charge. From Gauss' theorem the outer electrons (those with large l) are subject not only to the nuclear Coulomb potential but also to that of the inner electrons. The higher l states thus tend to rise in energy (*i.e.* they are less bound) with respect to those of lower l and the above degeneracy is thereby partially lifted.

It is perhaps worth mentioning here the technique used to attack the rather complex problem when many electrons are present: the Hartree–Fock method. First of all, the problem is that the complexity of interactions between a large number of electrons certainly does not allow analytic solution and does not even permit exact numerical solution. The best one can do is a so-called *mean-field* type of approximation, in which one starts from the single-electron solutions and then arranges the required number of electrons to fill the available states bottom–up. At this point one can evaluate the (time-averaged) mean field generated by such a charge distribution. Adding this extra potential to that of the nucleus, one may then iterate the procedure by finding the single-electron states for the new total potential. In general such a process converges and therefore allows a reasonable approximation to the full problem. Note, however, that it does remain an approximation since it does not take into account the instantaneous field of the electrons, which must naturally vary with time. Such a method may also be applied, with some success, to the (more complex) problem of the nuclear levels of protons and neutrons.

In atomic physics the results lead to appreciable modification of the level structure; the typical sequence is shown in Table 3.2. Elements with complete shells,

Table 3.2: The series of electron energy levels, as calculated in the mean-field or Hartree–Fock approximation.

K-shell		L-shell		M-shell		N-shell		
$1s^2$	|	$2s^2 2p^6$	|	$3s^2 3p^6$	|	$4s^2 3d^{10} 4p^6$	|	$5s^2 4d^{10} 5p^6 \ldots$
	He		Ne		Ar		Kr	
	(2)		(10)		(18)		(36)	

corresponding to the so-called magic numbers (2, 10, 18, 36, 54, 86), are just the *noble gases* (helium, neon, argon, krypton, ...). While the energy differences between levels belonging to the same shell are small, between shells there is a conspicuous energy *gap*. Elements with filled shells are thus harder to ionise

with respect to their neighbours, which instead have either an easily filled empty electron state (hole) or an easily removed extra electron in a higher level.

Angular momentum—orbital and spin

Since the ground state of an atom is formed by filling the available electron states starting from the lowest energy upwards, in a complete shell for every electron with projection $l_z = m$ there will be another with $l_z = -m$; they thus cancel. Likewise, for every spin-up electron there will be another spin-down, thus cancelling the spin contribution. In other words, atoms with filled shells (the noble gases) will have identically vanishing total electron spin–orbit angular momentum and therefore also magnetic moment. Since the parity of the spatial wave-function is $P = (-1)^l$ and there are always pairs of electrons with the same l in a filled-shell situation, then the parity of the noble gases is also fixed as positive.

Moreover, the case of atoms with, *e.g.*, one extra electron or one hole are also rather easy to predict. The total electron angular momentum will simply be that of the extra electron or hole. The only problem in such a case is the spin–orbit alignment. However, the spin–orbit coupling is known (it is a magnetic interaction) and thus even this may be predicted—the agreement is acceptable. More complicated cases of more than one extra electron or hole may also be dealt with in an empirical fashion. One finds that for low-Z (and therefore low-l) atoms it is the spin–spin interaction that dominates while for higher-Z atoms it is again the spin–orbit coupling that determines the alignments.

3.5.2 The single-particle approximation to the nucleus

Early attempts at applying the same ideas to nuclear structure failed. For example, the magic numbers are incorrectly predicted: no matter what form of nuclear potential well is assumed, agreement with known values is unattainable. However, in 1949 Goeppert Mayer and independently Haxel, Jensen and Suess* showed that by assuming a surprisingly large spin–orbit coupling, the ensuing level splittings and rearrangements could be so-engineered as to fit the known magic numbers rather well even up to rather high values.

Before continuing with the more technical discussion of the shell model, let us comment on one obvious difficulty that many textbooks leave untouched: namely, the concept of an *orbit* in such a densely packed situation. It is rather hard to

* The 1963 Nobel Prize in Physics was divided, one half awarded to Eugene Paul Wigner "for his contributions to the theory of the atomic nucleus and the elementary particles, particularly through the discovery and application of fundamental symmetry principles", the other half jointly to Maria Goeppert Mayer and J. Hans D. Jensen "for their discoveries concerning nuclear shell structure."

accept that a nucleon might really be considered in a well-defined orbit when it must come into contact continually with other nucleons and should therefore undergo scattering. The only plausible explanation makes recourse to the Pauli exclusion principle: as we have already noted, the states will all be filled, up to some Fermi level. Therefore, when two nucleons come into contact, there are no available *different* final scattering states into which they may transit. In other words, the phenomenon of *Pauli blocking* saves the orbital description. This is only a plausibility argument and it is rather difficult to prove rigorously. The main point is, of course, that the shell model does actually work rather well!

The nuclear potential well

The starting point will be the infinite square-well potential (see Fig. 3.5), for which the analytic solutions are known and of simple form. The square-well approxim-

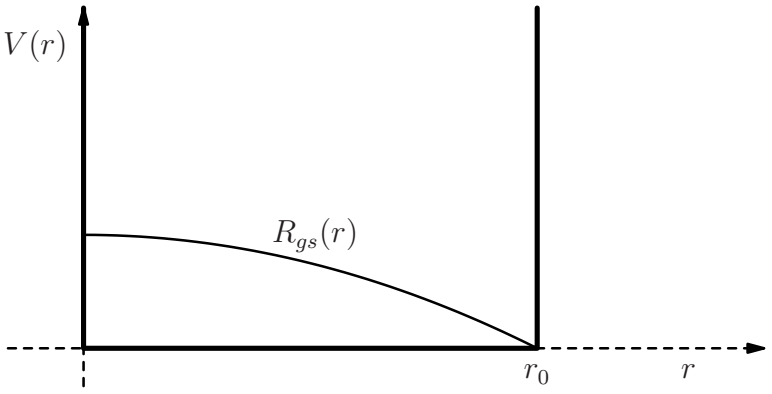


Figure 3.5: The infinite square-well approximation to the nuclear potential. The curve marked $R_{gs}(r)$ represents the lowest-energy or ground-state radial eigenfunction.

ation is justified by the experimental observation that the density is more or less constant over the nuclear volume, suggesting that the nucleons move as if confined inside a spherical container (or bag). As we shall discuss later, the refinement of using different, more reasonable, shapes does not actually alter the description qualitatively, nor does it have significant quantitative effects.

The solution of the problem in quantum mechanics is straightforward. Exploiting the spherical symmetry of the problem, in spherical polar coordinates we may separate the variables:

$$\psi_{lm}(\boldsymbol{r}) \;=\; R_l(r)\, Y_{lm}(\Omega), \tag{3.5.4}$$

where $Y_{lm}(\Omega)$ are the spherical harmonic functions. The problem thus immediately

reduces to that of the radial equation:

$$\left[\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}}{\mathrm{d}r}\right)+\frac{2m}{\hbar^2}\left(E-\frac{l(l+1)\hbar^2}{2mr^2}\right)\right]R_l(r)\;=\;0, \tag{3.5.5}$$

where E is the energy eigenvalue. Note that the mass m here corresponds to the so-called reduced nucleon mass (with respect to the rest of the nucleus),

$$m\;\equiv\;\frac{m_0}{1+m_0/m_{A-1}}\;\approx\;m_0\big[1-\tfrac{1}{A}\big], \tag{3.5.6}$$

which is approximately m_0 for large A, but can be appreciably smaller for lighter nuclei.

The solutions to the above equation are Bessel functions,

$$j_l(\rho)\;=\;(-\rho)^l\left(\frac{1}{\rho}\frac{\mathrm{d}}{\mathrm{d}\rho}\right)^l\left(\frac{\sin\rho}{\rho}\right), \tag{3.5.7}$$

where $\rho=kr$, with $k=\sqrt{2mE}/\hbar$.*

The quantisation condition then arises from the requirement that $R_l(r_0)=0$, where r_0 is the well radius. Indicating the zeroes of j_l by η_{ln}, *i.e.*, $j_l(\eta_{ln})=0$ with n labelling the series of zeroes for each l, we obtain the following expression for the energy eigenvalues:

$$E_{ln}\;=\;\frac{\hbar^2k^2}{2m}\;=\;\frac{\hbar^2\eta_{ln}^2}{2mr_0^2}. \tag{3.5.8}$$

We note now that the energy eigenvalues depend also on angular momentum. Moreover, owing to the infinite depth of the well, there is no restriction on l for any given n. The first few values of η_{ln} are, in order of size:

Wave	$1s$	$1p$	$1d$	$2s$	$1f$	$2p$	$1g$	...
η_{ln}^2	9.9	20.1	33.2	39.5	48.8	...	...	...

(3.5.9)

The level structure is thus already rather different to that of atomic electrons.

The problem at this point is that the magic numbers predicted by such a level series do not correspond (beyond the first three) to those found in Nature. One is immediately tempted to ask whether or not variations in the form of the nuclear potential might not allow other level shifts that could then accommodate the known series. We shall now show a progression of possible well shapes starting with a simple reference point of an harmonic potential (best suited for describing

* There are also the exact solutions known as the Neumann functions, with sine substituted by cosine. These are, however, singular at the origin and are therefore not permitted here.

light nuclei—again, the first three magic numbers are predicted correctly) up to the best available models based on a Woods–Saxon (WS) shape of a smoothed square well, see Fig. 3.6. Also included in the figure is a final series, including an

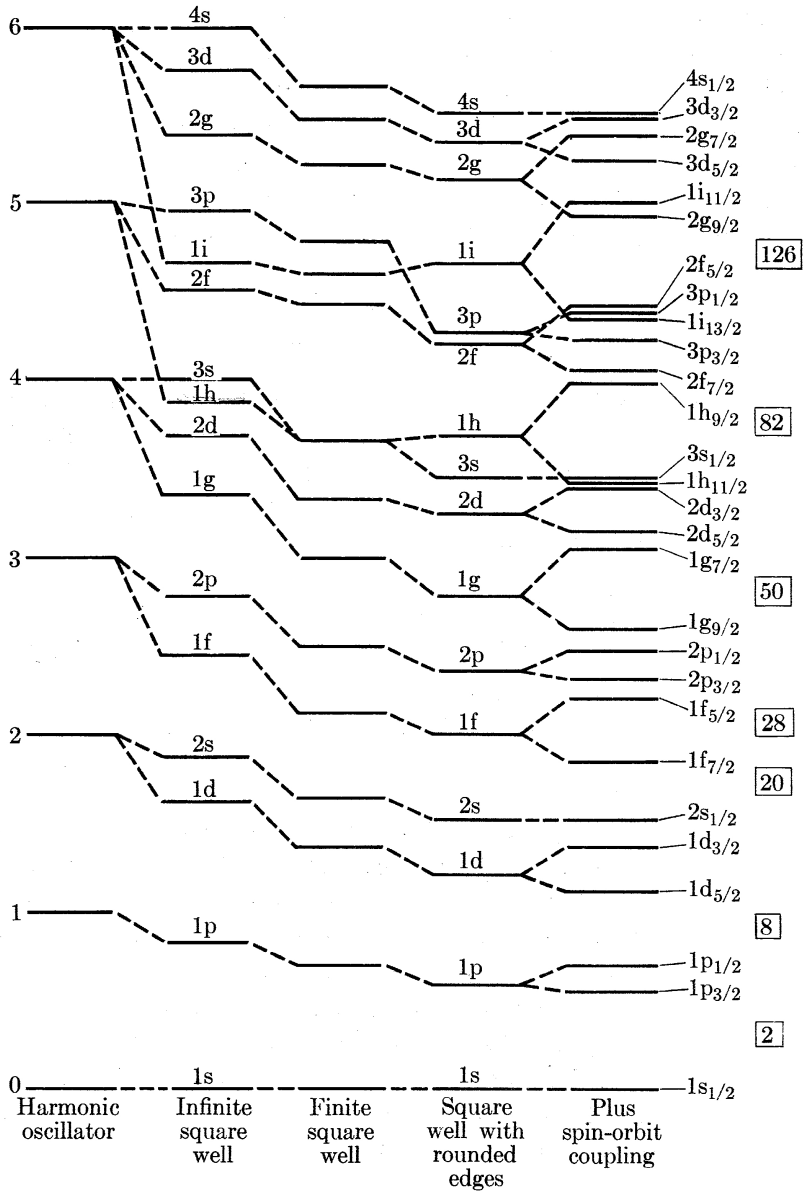


Figure 3.6: Examples of the level series obtained for a variety of potential wells and inclusion of a spin–orbit term. For comparison purposes, the base level is chosen as the lowest energy state and the scale is set by the harmonic oscillator level $n = 4$.

empirical spin–orbit coupling. One immediately sees the futility of simply altering the well shape: the shell grouping of levels is hardly altered in moving from the extreme of an infinite square well to the more physically appealing WS form.

Before continuing with a more detailed discussion of the problems of spin and angular momentum let us briefly discuss the WS potential: in Fig. 3.7 the WS form for the nuclear density distribution is depicted. This form merely tries to

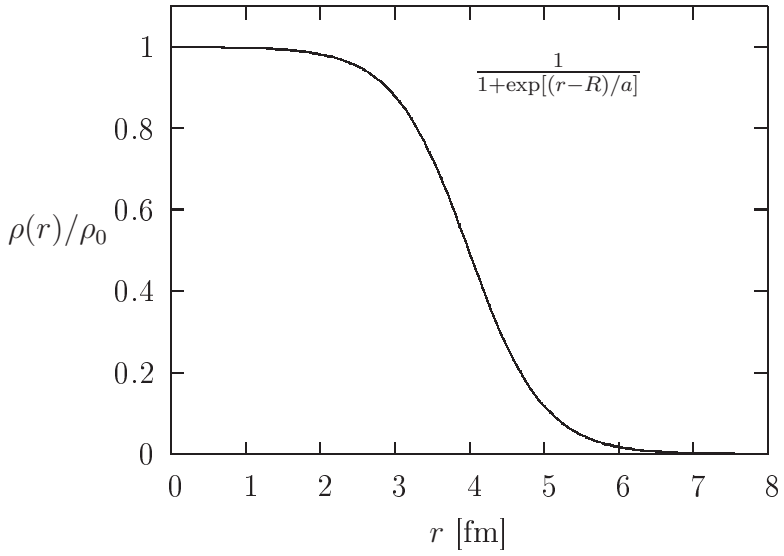


Figure 3.7: The form of the Woods–Saxon potential. The parameters are: ρ_0 (the approximate mean nuclear density), R (the effective radius of the specific nucleus in question) and $a \simeq 0.55\,\mathrm{fm}$ (the so-called *skin depth*).

render the hard sphere a little softer and take into account that realistically the external nucleons, being more weakly bound, will have a lower density (*i.e.*, there will indeed be a skin effect). Now, We can motivate the same choice of form for the potential with the observations that the nuclear *density* is empirically well-described by a WS form (substituting V with ρ above) and that the short range of the nuclear interaction suggests that the potential should follow the spatial dependence of the density rather closely.

The form adopted for the potential is thus

$$V(r) \;=\; -\frac{V_0}{1+\exp[(r-R)/a]}, \tag{3.5.10}$$

where, given the actual values of the other parameters, one finds that $V_0 \simeq 40\,\mathrm{MeV}$ is essentially the potential-well depth; R is the effective nuclear radius (of the specific nucleus under study—for large nuclei it is therefore of the order a few fm);

and $a \approx 0.55\,\mathrm{fm}$ (it only varies experimentally by about $\pm 0.05\,\mathrm{fm}$) is a measure of the radial interval over which the density falls from the internal bulk value of $\sim V_0$ to zero outside (it is thus a sort of *skin depth*).

Now, the magic numbers following from a mean-field approximation using the WS potential are as follows:

2 8 20 34 40 58 92 138

However, for example, marked jumps in the binding energy are found for

2 8 20 28 50 82 126

Thus, while again the first three are correct, there is a rapid and significant departure thereafter.

The spin–orbit interaction

The work of Goeppert Mayer and Haxel, Jensen and Suess showed that the inclusion of a sizable spin–orbit coupling could provide further level shifts and bring the theoretical numbers into line with experiment, see Fig. 3.6. The idea is to introduce a form analogous to that found in atomic theory, namely

$$V_{LS} \equiv f(r)\,\boldsymbol{l}\cdot\boldsymbol{s}, \tag{3.5.11}$$

where $\boldsymbol{l}$ and $\boldsymbol{s}$ are the angular-momentum and spin vectors* associated with a given nucleon and the form of function $f(r)$ may be derived from considerations of the effective *magnetic* field experienced by an electron *moving* in a constant *electric* field:

$$\boldsymbol{B}_{\mathrm{eff}} = -\left(\frac{\boldsymbol{v}}{c}\right)_{\wedge}\boldsymbol{E}, \tag{3.5.12}$$

where the electric field is simply

$$\boldsymbol{E} = -\frac{1}{e}\,\boldsymbol{\nabla} V_c(r). \tag{3.5.13}$$

Taking the magnetic moment of the electron as

$$\boldsymbol{\mu} = \frac{e\boldsymbol{s}}{m_e c}, \tag{3.5.14}$$

* Note that $\boldsymbol{l}$ and $\boldsymbol{s}$ are related to, but not to be confused with the quantum numbers l and s.

we obtain

$$\begin{aligned} V_{LS} &= -\boldsymbol{\mu}\cdot\boldsymbol{B}_{\text{eff}} \\ &= \boldsymbol{\mu}\cdot\left(\frac{\boldsymbol{v}}{c}\right)\wedge\boldsymbol{E} \\ &= \left(\frac{e\boldsymbol{s}}{m_e c}\right)\cdot\left[\frac{\boldsymbol{p}}{m_e c}\wedge\left(\frac{\boldsymbol{x}}{r}\right)\frac{1}{(-e)}\frac{\mathrm{d}V_c(r)}{\mathrm{d}r}\right] \\ &= \frac{1}{(m_e c)^2}\frac{1}{r}\frac{\mathrm{d}V_c(r)}{\mathrm{d}r}\boldsymbol{l}\cdot\boldsymbol{s}. \end{aligned} \tag{3.5.15}$$

Note that this formula is actually incorrect by a factor of two (too large) due to the neglect of an effect known as *Thomas precession*. However, since we only require the general behaviour for our nuclear discussion, we shall not pursue the question further.* For the nuclear case, we shall instead simply *assume* a contribution of the form

$$V_{LS} \equiv a_{LS}\frac{1}{r}\frac{\mathrm{d}V(r)}{\mathrm{d}r}\boldsymbol{l}\cdot\boldsymbol{s}, \tag{3.5.16}$$

where a_{LS} is then a phenomenological parameter to be fitted to the data and $V(r)$ is the nuclear potential adopted.

For a single given nucleon, it is natural to define the *total* angular momentum $\boldsymbol{j}=\boldsymbol{l}+\boldsymbol{s}$. The rules for composing angular momenta tell us that the quantum number j assigned to a nucleon with orbital and spin numbers l and s are restricted to lie in the range

$$|l-s| \le j \le l+s. \tag{3.5.17}$$

Since $s=\frac{1}{2}$, this reduces to $j=l\pm\frac{1}{2}$ for $l\neq 0$ (and trivially $j=\frac{1}{2}$ for $l=0$). Using this fact and a knowledge of the individual expectations values, we shall now derive a relation for $\langle\boldsymbol{l}\cdot\boldsymbol{s}\rangle$. We have

$$\langle\boldsymbol{s}^2\rangle = s(s+1)\hbar^2 = \tfrac{3}{4}\hbar^2, \tag{3.5.18a}$$

$$\langle\boldsymbol{l}^2\rangle = l(l+1)\hbar^2, \tag{3.5.18b}$$

$$\begin{aligned} \langle\boldsymbol{j}^2\rangle &= j(j+1)\hbar^2 \\ &= \langle(\boldsymbol{s}+\boldsymbol{l})^2\rangle \\ &= \langle\boldsymbol{s}^2\rangle + \langle\boldsymbol{l}^2\rangle + 2\langle\boldsymbol{l}\cdot\boldsymbol{s}\rangle. \end{aligned} \tag{3.5.18c}$$

* The interested reader is referred to, *e.g.*, Jackson (1975).

This last equality may be inverted to obtain

$$\begin{aligned}\langle \boldsymbol{l}\cdot\boldsymbol{s}\rangle &= \tfrac{1}{2}\left[\langle \boldsymbol{j}^2\rangle - \langle \boldsymbol{l}^2\rangle - \langle \boldsymbol{s}^2\rangle\right] \\ &= \tfrac{1}{2}\Big[j(j+1) - l(l+1) - s(s+1)\Big]\hbar^2\end{aligned}$$

and substituting $j = l \pm s$ (for $l \neq 0$ and $s = \frac{1}{2}$) gives

$$\begin{aligned}&= \tfrac{1}{2}\Big[(l \pm \tfrac{1}{2})(l \pm \tfrac{1}{2} + 1) - l(l+1) - \tfrac{1}{2}(\tfrac{1}{2}+1)\Big]\hbar^2 \\ &= \tfrac{1}{4}\Big[\pm(2l+1) - 1\Big]\hbar^2. \qquad (3.5.19)\end{aligned}$$

We may thus finally write

$$\langle \boldsymbol{l}\cdot\boldsymbol{s}\rangle = \begin{cases} \tfrac{1}{2}\,l\hbar^2 & \text{for } j = l + \tfrac{1}{2} \text{ (“long” configuration)}, \\ -\tfrac{1}{2}(l+1)\hbar^2 & \text{for } j = l - \tfrac{1}{2} \text{ (“short” configuration).} \end{cases} \qquad (3.5.20)$$

Inserting this into the formula for the spin–orbit coupling and taking the difference between the two cases, we obtain the spin–orbit *splitting*:

$$\Delta V_{LS} = a_{LS}\,\frac{1}{r}\,\frac{\mathrm{d}V(r)}{\mathrm{d}r}\,\frac{2l+1}{2}\,\hbar^2 \qquad (l \neq 0). \qquad (3.5.21)$$

Let us make a few observations on this final expression:

- It is purely *ad hoc*, being constructed along the lines of the corresponding electromagnetic effect, whereas it should be *derived* from strong interaction theory and *chromo*magnetic interaction.
- The required sign is opposite to that observed in the electromagnetic case; there $\boldsymbol{l}$ and $\boldsymbol{s}$ prefer to *anti-align* whereas empirically the nuclear case indicates alignment: the long state has lowest energy.
- The “natural” magnitude would be roughly a factor 50 too small to explain the magic numbers.
- It predicts magic numbers at 14 (seen, but weakly) and 114 (not seen) while it does not reproduce the observed number 126; *i.e.* there are still problems for very large nuclei.

Angular momentum of the ground state

Satisfied then that such a description makes sense energetically and following along the lines of what is well-known in atomic physics, we can go on to describe the angular-momentum properties of nuclei. Since the shells are filled bottom–up, only the outermost shell has any role (the inner shells, being complete, must have overall $J^P = 0^+$). In particular therefore, a single, unpaired proton or neutron completely determines the spin–parity state of the nucleus.

Let us first examine case of the parity—recall that the parity of the spatial wave-function is simply $(-1)^l$. We need thus only consider the more general cases of even or odd numbers of protons and/or neutrons in the outer shell. There are then three distinct possibilities and including the total spin we have

even–even: empirically one finds that the pairing effect is such that protons and neutrons separately couple in such a way as to cancel both orbital and spin angular momentum leading to an overall system of $J^P = 0^+$ (there are no known exceptions to this rule);

even–odd: in this case the unpaired proton or neutron determines the spin–parity of the system and we have $J^P = j^p$;

odd–odd: here we must compose the two unpaired proton and neutron angular momenta in the usual way, $|j_p - j_n| \leq J \leq (j_p + j_n)$ and $P = (-1)^{l_p + l_n}$ (there exist empirical rules describing the observed combinations).

Here are two simple examples (the suffices on the state labels indicate j and the indices the degeneracy (or multiplicity), which is now $2j+1$):

	Z	N	$(1s_{1/2})^2$	$(1p_{3/2})^4$	$(1p_{1/2})^2$	$(1d_{5/2})^6$	$(2s_{1/2})^2$	$(1d_{3/2})^4$	J^P
^{17}O	8		2	4	2				
		9	2	4	2	1			$\frac{5}{2}^+$
^{39}K	19		2	4	2	6	2	3	$\frac{3}{2}^+$
		20	2	4	2	6	2	4	

Note however that for $N, Z > 50$, $(1g_{9/2})^{10}$, the level ordering is different for protons and neutrons (owing to the Coulomb effect). An example is the case of ^{95}Mo, which has $Z = 42$ and $N = 53$. The unpaired neutron sits in the $2d_{5/2}$ state and we therefore have $J^P = \frac{5}{2}^+$.

Magnetic moments of the ground state

The above understanding of the angular momentum of the ground state may be exploited to describe the magnetic moments. The situation here is a little more complicated since orbital and spin contributions are different owing to the non-trivial gyromagnetic ratios of the proton and neutron. The contribution of a single nucleon to the overall nuclear magnetic moment has two components:

$$\boldsymbol{\mu} \;=\; \left[g_l \frac{\langle \boldsymbol{l} \rangle}{\hbar} + g_s \frac{\langle \boldsymbol{s} \rangle}{\hbar} \right] \mu_N, \tag{3.5.22}$$

where $\mu_N \equiv \frac{e\hbar}{2m}$ (m the nucleon mass) is the *nuclear magneton* defined in analogy with case of the electron and the *Bohr magneton* $\mu_B \equiv \frac{e\hbar}{2m_e}$ (though it is clearly much weaker). The two quantities $g_{l,s}$ are respectively the orbital and spin gyromagnetic ratios. The first is simply due the equivalent circulating electric current, which naturally generates a magnetic field. The second is not calculable without a complete theory of the internal (quark) structure of the nucleon and must be therefore measured.* We thus have

$$g_l \;=\; \begin{cases} 1 & \text{for protons,} \\ 0 & \text{for neutrons,} \end{cases} \tag{3.5.23a}$$

$$g_s \;=\; \begin{cases} \;\;5.58 & \text{for protons,} \\ -3.83 & \text{for neutrons.} \end{cases} \tag{3.5.23b}$$

These should be compared with the electron case, for which $g_s = 2.0023$.†

Since the paired nucleons cancel their orbital and spin contributions, we again need only consider the unpaired nucleons or holes. Thus, for the case of a single unpaired nucleon, we need to evaluate the expectation value in Eq. (3.5.22). Consider the general definition of the nuclear gyromagnetic moment,

$$\boldsymbol{\mu}_{\text{nucleon}} \;:=\; \mu_N \, g_{\text{nucleon}} \frac{\langle \boldsymbol{j} \rangle}{\hbar}. \tag{3.5.24}$$

By projecting this and Eq. (3.5.22) onto the vector $\boldsymbol{j}$ (via the Wigner–Eckart the-

* In any case, the quark model only goes part way to explaining the precise values.

† This is one of the most accurately known quantities in experimental particle physics: the present value of the electron magnetic moment is $\mu_e = 1.00115965218076 \pm 0.00000000000027\,\mu_\text{B}$, see Patrignani *et al.* (2016).

orem) and comparing expressions we obtain (after elimination of common factors)

$$g_{\text{nucleon}}\langle \boldsymbol{j}^2\rangle \;=\; [g_l\langle \boldsymbol{l}\cdot\boldsymbol{j}\rangle + g_s\langle \boldsymbol{s}\cdot\boldsymbol{j}\rangle]\,. \tag{3.5.25}$$

Note that this is not a trivial operation and the full power of the Wigner–Eckart theorem (based on the tensorial properties of the general rotation operators) is necessary. Using a similar trick as used in evaluating the spin–orbit splitting, we can evaluate the right-hand side:

$$2\boldsymbol{l}\cdot\boldsymbol{j} \;=\; \boldsymbol{j}^2 + \boldsymbol{l}^2 - \boldsymbol{s}^2, \tag{3.5.26a}$$

$$2\boldsymbol{s}\cdot\boldsymbol{j} \;=\; \boldsymbol{j}^2 - \boldsymbol{l}^2 + \boldsymbol{s}^2. \tag{3.5.26b}$$

Whence we obtain (for nuclei with one unpaired nucleon)

$$\begin{aligned} g_{\text{nucleus}} &= g_{\text{odd nucleon}} \\ &= \frac{g_l\,[j(j+1)+l(l+1)-s(s+1)] + g_s\,[j(j+1)-l(l+1)+s(s+1)]}{2j(j+1)}, \end{aligned}$$

which, for the two possible cases $j = l \pm \frac{1}{2}$ results in

$$= \left[g_l \pm \frac{g_s - g_l}{2l+1}\right]. \tag{3.5.27}$$

Some examples of the resulting predictions are shown in Table 3.3. In all

Table 3.3: A collection of shell-model predictions for nuclear magnetic moments for cases where there is just a single unpaired nucleon or hole. The index -1 on the states indicates a hole.

nucleus	unpaired nucleon	state	J^P	μ/μ_N theor.	expt.
$^{15}_{7}$N	p	$1p_{1/2}^{-1}$	$\frac{1}{2}^-$	-0.264	-0.283
$^{15}_{8}$O	n	$1p_{1/2}^{-1}$	$\frac{1}{2}^-$	$+0.638$	$+0.719$
$^{17}_{8}$O	n	$1d_{5/2}$	$\frac{5}{2}^+$	-1.913	-1.894
$^{17}_{9}$F	p	$1d_{5/2}$	$\frac{5}{2}^+$	$+4.722$	$+4.793$

honesty, it must be remarked here that these are among the best agreements and that there are cases where the discrepancy is even as much as $\pm 1\,\mu_{\text{N}}$. The cause may be traced to the polarising effect that the single unpaired nucleon (especially if of high spin) has on the rest of the nucleus, clearly altering the potential in a

non-trivial way.

As a final remark, in order to be sure of avoiding possible confusion, let us make clear that here we are dealing with ordinary magnetic moments whereas the discussion on the problem of the spin–orbit coupling implicitly invokes a strong interaction or *chromomagnetic* moment (*i.e.*, due to QCD and *not* QED).

3.6 Bibliography

Bohr, N. and Wheeler, J.A. (1939), *Phys. Rev.* **56**, 426.

Goeppert Mayer, M. (1949), *Phys. Rev.* **75**, 1969; *see also*, *Phys. Rev.* **78**, 16; *Phys. Rev.* **78**, 22.

Haxel, O., Jensen, J.H.D. and Suess, H.E. (1949), *Phys. Rev.* **75**, 1766.

Jackson, J.D. (1975), *Classical Electrodynamics* (John Wiley & Sons), 2nd. edition.

Meija, J. *et al.* (2016), *Pure Appl. Chem.* **88**, 265.

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

von Weizsäcker, C.F. (1935), *Z. Phys.* **96**, 431.

Chapter 4

Scattering and Particle Interactions

4.1 Scattering

Perhaps the first important true scattering experiments are those performed by Geiger and Marsden* in 1909 under the guidance of Rutherford.† According to the Thomson so-called *plum-pudding* model of the atom, in which all the positive charge was taken to be distributed as a diffuse cloud and only the electrons were recognised as discrete entities, a massive, charged projectile (such as an α-particle) should only suffer minor deviations on passing through even the relatively dense material of a thin gold foil. However, in the experiments of Geiger and Marsden a surprisingly high number of α-particles were deflected through appreciable angles by the gold foil.

> "It was as if you fired a 15-inch shell at a sheet of tissue paper and it came back to hit you."
>
> Lord Rutherford

4.1.1 General scattering formalism

The process of diffusion is certainly the most ubiquitous of investigation methods available to the physicist. Indeed, almost any of the many techniques used for the study of matter may be viewed as some form of scattering: from the simple fairly non-invasive optical viewing of objects with the naked eye or microscope to the very destructive *deeply inelastic scattering* used to probe the interior structure of protons and neutrons.

* At the time Ernest Marsden was just a twenty-year-old undergraduate student.

† Rutherford was actually awarded the 1908 Nobel Prize for *Chemistry*, "for his investigations into the disintegration of the elements, and the chemistry of radioactive substances." He is quoted as commenting: "I must confess it was very unexpected and I am very startled at my metamorphosis into a chemist."

The basic elements of a scattering process are depicted schematically in Fig. 4.1. There are essentially three: the projectile, the target and the detector. Of course, if

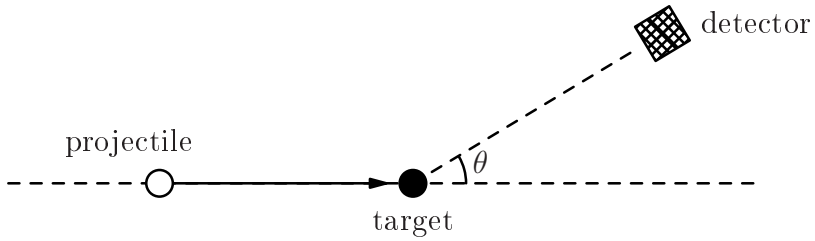


Figure 4.1: A schematic representation of a generic scattering process.

we are considering colliding-beam experiments then there will be two "projectiles" and no specific target. Nowadays there is usually rather more than one detector. The possibilities for these ingredients are manifold; the lists below contain many important examples.

Projectile: γ, $e^{\pm}$, $\mu^{\pm}$, $\nu(\bar{\nu})$, p, n, $\pi^{\pm}$, Λ^0, ^{4}He, ..., ^{197}Au, ^{207}Pb, ...

Target: $e^{\pm}$, ^{1}H, ^{2}D, ^{3}He, ^{4}He, ..., ^{197}Au, ^{207}Pb, ..., atoms, crystals, ...

Detector: photographic plate, emulsion, cloud chamber, wire chamber, calorimeter, silicon strips, CCD, ...

Now, by means of the various types of detectors we can measure the different kinematic variables (energy, momentum, direction) and in some cases even the polarisation (usually indirectly though) of a final-state particle. By combining such knowledge for an individual particle we may deduce its type (or *flavour*, photon, electron, proton, pion *etc.*). It may be important to study counts as a function of energy, momenta *etc.* or possible correlations with those of other particles. The sort of "news" one is seeking in such data is again manifold.

Distributions: in angle, energy, momentum *etc.*, provide information on the type and form of the potential at the heart of the scattering process;

Particle identification: provides information on the existence of new particles, conserved quantum numbers *etc.*;

Symmetries: direct observation of symmetries such as C, P, or T (or lack or violation thereof) may indicate the same at a fundamental level;

Probabilities or Rates: the measured rates provide direct indications of the strength of the underlying interaction and again the existence (or otherwise) of (partial) conservation laws.

Moreover, prior knowledge of the rates, *i.e.* capability of their calculation (including possible background), is indispensable to proper feasibility studies.

We should also mention here the *decay process* (a close relative of scattering). It has no real counterpart in classical physics and is intimately bound up with the quantum-mechanical nature of matter. However, decay measurements are, in general, a little less amenable to theoretical interpretation as there is no way to vary the input energies, momenta *etc.* and, of course, not all particles decay.

4.1.2 Definition of the cross-section

In order to formulate the cross-section it is easiest to turn back to the classical case, where the notion becomes that of an (effective) physical area presented to the projectile by the target. Thus, we shall consider the idealised setup shown in Fig. 4.2. We imagine that a beam of projectiles of uniform intensity is incident on

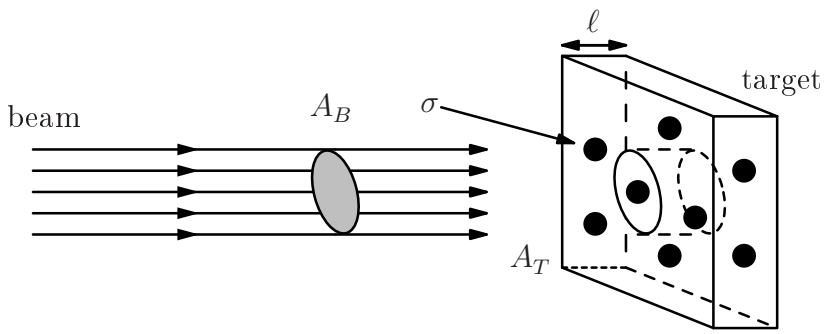


Figure 4.2: The idealised scattering process.

a fixed target of uniform density, thickness *etc.* If the beam intensity or current is I (particles/second) with cross-section area A_B (a real physical area), then the incident flux is simply

$$\Phi \equiv \frac{I}{A_B}. \tag{4.1.1}$$

The target may be characterised in terms of its transverse area A_T (presented to the beam), length l (along the beam axis) and the number N of scattering centres it contains. The target number density is clearly

$$\rho \equiv \frac{N}{A_T\, l}. \tag{4.1.2}$$

It is common to define a rather useful (more universal quantity) known as the target *thickness*:

$$t := \rho l = \frac{N}{A_T}. \tag{4.1.3}$$

By analogy with classical scattering, we can define an *effective* scattering cross-section σ (an area usually measured in barns: $1\,\mathrm{b} = 10^{-28}\,\mathrm{m}^2 = 100\,\mathrm{fm}^2$) via the probability P that any given *single* beam particle will strike or interact with a given *single* scattering centre on its passage through the target:

$$P \equiv \frac{\sigma}{A_T}, \tag{4.1.4a}$$

which for N scattering centres in the area A_T then becomes

$$P \equiv \frac{N\sigma}{A_T}. \tag{4.1.4b}$$

A consideration on the probability P is immediately in order: in general, we require $P \ll 1$ otherwise we must deal with the following possible complicating consequences:

- multiple scattering—if the probability of a single collision is close to unity then so too is that of multiple collisions;
- loss of beam density and/or uniformity—if the collisions significantly alter the beam (content, direction, energy) then the effective flux will not be constant throughout the target;
- shadowing (related to the previous point)—the particles at the back of the target are "hidden" by those at the front and therefore participate less in the scattering process, leading to an effective reduction in target density.

Thus, apart from exceptional circumstances, we shall always assume that $P \ll 1$, or equivalently $N\sigma \ll A_T$. Under circumstances of normal cross-sections, beam intensity and target density this is however not usually a problem.

What is actually measured experimentally is a counting rate, *i.e.*, a number of interactions per second:

$$\begin{aligned} R &= I \cdot P \\ &= \frac{IN\sigma}{A_T}. \end{aligned} \tag{4.1.5}$$

Now, assuming the ideal case of an incident flux distributed over the entire target

face, we may write $A_B = A_T$ and thus

$$\begin{aligned} R &= \frac{I}{A_B} \cdot N\sigma \\ &= \Phi N\sigma, \end{aligned} \tag{4.1.6a}$$

otherwise, we should write

$$R = It\sigma. \tag{4.1.6b}$$

More generically, defining $A := \max(A_B, A_T)$, we may simply write

$$R =: IN\sigma/A. \tag{4.1.7}$$

This then is to be taken as the *definition* of σ, the *effective* cross-section.

As defined here, this is a *total* cross-section; *i.e.*, it includes all final outcomes (angles, energies *etc.*) of the scattering process. However, in practice not only is it rather difficult (if not impossible) to make such an all-inclusive measurement, but it is also desirable to have more detailed information (*e.g.*, angular, energy *etc.* dependence). It is thus useful to divide the (single-particle) final-state phase space into infinitesimal cells $\mathrm{d}^3\boldsymbol{p}$ and thereby define a *differential* rate:

$$\frac{\mathrm{d}R}{\mathrm{d}^3\boldsymbol{p}} = \frac{\mathrm{d}R}{\mathrm{d}p_x\,\mathrm{d}p_y\,\mathrm{d}p_z} = \frac{\mathrm{d}R}{p^2\mathrm{d}p\,\mathrm{d}\cos\theta\,\mathrm{d}\phi} = \frac{\mathrm{d}R}{p^2\mathrm{d}p\,\mathrm{d}^2\Omega}. \tag{4.1.8}$$

In the same way, it is then natural to define a differential cross-section:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}^3\boldsymbol{p}}. \tag{4.1.9}$$

In the case of a number n of detected final-state particles this becomes

$$\frac{\mathrm{d}\sigma}{\mathrm{d}^3\boldsymbol{p}_1\,\mathrm{d}^3\boldsymbol{p}_2\ldots\mathrm{d}^3\boldsymbol{p}_n}. \tag{4.1.10}$$

Finally we may, of course, recover a total (or partial) cross-section by integrating (experimentally or theoretically over any or all of the final-state momentum variables.

4.1.3 The Rutherford formula

We shall now briefly examine the famous case of Rutherford scattering. The generic process is that of a projectile of mass m and charge z incident on a target (a nucleus) of mass M and charge Z. Since the scattering is elastic and two-body, there is a direct relationship between the final momentum (or energy) and the

angle of deviation with respect to the beam direction. For the present, we shall thus only consider the differential cross-section $\mathrm{d}\sigma/\mathrm{d}\Omega$ or $\mathrm{d}\sigma/\mathrm{dcos}\theta$.

The process we wish to examine is thus the elastic two-body scattering of a non-relativistic α-particle (mass m and momentum $\boldsymbol{p}_i$) off a stationary gold nucleus (mass M). As noted, two-body kinematics imposes a strict relation between the final energy and the scattering angle; it is thus only necessary to measure, say, the latter. In such a collision the maximum momentum transfer is

$$|\boldsymbol{p}_i - \boldsymbol{p}_f|_{\max} \;=\; \Delta p_{\max} \;=\; \frac{2M}{M+m}\,p \;>\; p \qquad (p \;=\; |\boldsymbol{p}_i|). \tag{4.1.11}$$

Since $m \ll M$, $\Delta p_{\max} \simeq 2p$ and we see that the effect observed and recounted above is certainly possible kinematically. Note too that by the same reasoning and substituting an electron for the gold nucleus, we see that the kinematical effect of scattering off (atomic) electrons is essentially negligible. Finally, if we consider scattering off a potential generated by a diffuse cloud of positive charge distributed over the entire *atomic* volume (the plum-pudding model), the effect is also negligible (recall that inside an atom the Coulomb potential due to such a distribution would behave as r and would thus be very weak).

The scattering of an α-particle by the Coulomb potential due to a nucleus is depicted in Fig. 4.3. The position of the particle along the trajectory may be

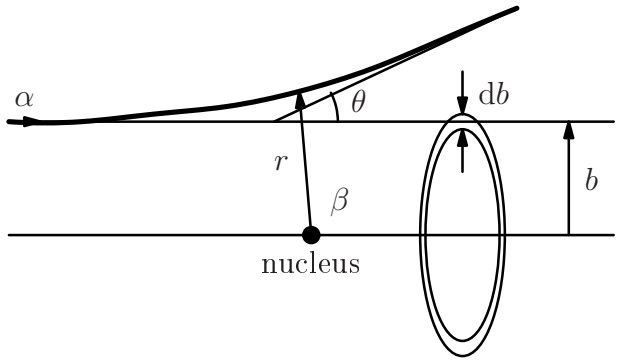


Figure 4.3: The scattering of an α-particle off the Coulomb potential due to a nucleus.

determined via the angle β relative to its initial direction of motion ($\beta = \pi$). We also define the *impact parameter* b as the separation between the parallel lines running along the projectile direction and through the scattering centre (which lies at the origin of the laboratory reference frame). The azimuthal angle ϕ plays no special role in the process and is therefore integrated out. The other angle we define is θ, the final direction of the scattered α-particle with respect to its initial

motion. We therefore clearly have the following limits:

$$0 \leq \theta \leq \pi \tag{4.1.12a}$$

and

$$\theta \leq \beta \leq \pi. \tag{4.1.12b}$$

The point $\theta = 0$ is called "forward scattering" (but does *not* necessarily imply *no interaction*) while $\theta = \pi$ is "backward scattering". Note that there is a one-to-one relationship between b and θ.

The *effective* differential cross-section is then simply given by the area of the ring corresponding to a given value of the impact parameter b:

$$\mathrm{d}\sigma(b) \;=\; 2\pi b\,\mathrm{d}b. \tag{4.1.13}$$

The element $\mathrm{d}\Omega \equiv \mathrm{d}\cos\theta\mathrm{d}\phi$ may be simplified since (assuming cylindrical symmetry) there can be no dependence on the azimuth ϕ; thus

$$\frac{\mathrm{d}\sigma(b)}{\mathrm{d}\cos\theta} \;=\; 2\pi b\,\frac{\mathrm{d}b}{\mathrm{d}\cos\theta}. \tag{4.1.14}$$

We therefore need to find the relationship between b and θ.

Let us now rapidly run through the solution of the problem in classical mechanics while noting immediately that the quantum solution does not alter the result in this very particular case (*i.e.* a Coulomb-like potential). The easiest approach is to first identify the conserved quantities, which for a conservative, central potential are the total energy E and the orbital angular momentum L of the α-particle:

$$\begin{aligned} E \;&=\; E_{\mathrm{kin}} + V_c(r) \;=\; \tfrac{1}{2}m\boldsymbol{v}^2 + \frac{zZ\alpha}{r} \\ &=\; \tfrac{1}{2}m\big[\dot{r}^2 + r^2\dot{\beta}^2\big] + \frac{zZ\alpha}{r} \;=\; \tfrac{1}{2}mv_\infty^2 \end{aligned} \tag{4.1.15a}$$

and

$$L \;=\; mr^2\dot{\beta} \;=\; mv_\infty b, \tag{4.1.15b}$$

where v_∞ is the asymptotic value of the projectile velocity and we have adopted the simplifying convention of setting $\hbar$ and c to unity. The fine-structure constant is $\alpha \simeq 1/137$; we shall comment on how to translate the final expressions into numerical values later.

Solving Eq. (4.1.15b) for $\dot{\beta}$, we have

$$\dot{\beta} \;=\; \frac{L}{mr^2} \;=\; \frac{bv_\infty}{r^2}. \tag{4.1.16}$$

Eq. (4.1.15a) may be rearranged to give

$$\dot{r}^2 \;=\; \frac{2}{m}\Big(E - V_c(r)\Big) - \dot{\beta}^2 r^2, \tag{4.1.17}$$

from which, using Eq. (4.1.16), we obtain

$$\mathrm{d}t \;=\; \mathrm{d}r\left[\frac{2}{m}\Big(E - V_c(r)\Big) - \frac{b^2 v_\infty^2}{r^2}\right]^{-1/2}. \tag{4.1.18}$$

The integral of Eq. (4.1.16) with respect to t then leads to

$$\begin{aligned} -\int_{\frac{1}{2}(\pi+\theta)}^{\theta} \mathrm{d}\beta \;&=\; b v_\infty \int_0^\infty \frac{\mathrm{d}t}{r^2} \\ &=\; b v_\infty \int_{r_{\min}}^\infty \mathrm{d}r\, \frac{1}{r^2}\left[\frac{2}{m}\Big(E - V_c(r)\Big) - \frac{b^2 v_\infty^2}{r^2}\right]^{-1/2}, \end{aligned} \tag{4.1.19}$$

where we should integrate from $t=0$ to ∞, the lower limit representing the halfway point or point of closest approach $r_{\min}$ to the nucleus, when $\beta = \frac{1}{2}(\pi+\theta)$. Integrating, one readily obtains

$$b \;=\; \frac{zZ\alpha}{m v_\infty^2}\,\cot\tfrac{\theta}{2}. \tag{4.1.20}$$

Exercise 4.1.1.

(a) Perform the above integration explicitly to obtain the answer given.

(b) An alternative approach is to consider the variation in linear momentum along the x-axis (parallel to the initial direction of motion).

For numerical values, we must correct the units with a factor $\hbar c \simeq 197$ on the right-hand side, whence the impact factor b must be expressed in fm and the asymptotic kinetic energy $\frac{1}{2}mv_\infty^2$ in MeV. A typical case has $\frac{1}{2}mv_\infty^2 \sim 6\,\mathrm{MeV}$ while $z=2$ and $Z=79$:

$$b \;=\; 197 \times \frac{2\cdot 79}{12\cdot 137}\,\cot\tfrac{\theta}{2} \;\simeq\; 19\,\cot\tfrac{\theta}{2}\ \text{fm}. \tag{4.1.21}$$

Thus, for example, if we set $b \simeq 1\,\text{Å} = 10^5\,\text{fm}$ (of the order of an *atomic* radius), we find $\theta \simeq 0.4\,\text{mrad}$.

Exercise 4.1.2. *By setting $\dot{r}=0$ in Eq. (4.1.15a) and solving for r, find the distance of minimum approach. By considering the case $\theta=\pi$ (backward scattering),*

obtain the absolute minimum distance of approach
Hint: *this case is simpler since also* $\dot{\beta}=0$.

This indicates that for such a target, ^{197}Au, the α-particle never comes into direct contact with the nucleus and that the scattering will therefore remain purely Coulomb and elastic.

Returning to the cross-section formula, we then finally obtain

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \frac{1}{16}\left[\frac{zZ\alpha}{E_\infty}\right]^2 \mathrm{cosec}^4\,\tfrac{\theta}{2}, \tag{4.1.22}$$

where E_∞ is the asymptotic kinetic energy of the α-particle. For 6 MeV α-particles on ^{197}Au, this gives

$$\simeq 0.89\,\mathrm{cosec}^4\,\tfrac{\theta}{2}\ \ \mathrm{b/sr}. \tag{4.1.23}$$

Analysis of the hundred thousand or more scattering events recorded for α-particles on gold fully confirmed the angular dependence predicted by the above analysis.

Exercise 4.1.3. *Demonstrate numerically that with reasonable (estimated) fluxes of* α*-particles on a gold foil, detector counts of the order of particles/second may be easily attained. The original experiment used* ^{222}Rn *as the alpha source with half-life approximately* 3.5×10^5 s.

Now, the formula did not however work for aluminium ($Z=13$). On replacing the gold foil by an aluminium foil (some years later), it turned out that while small-angle scattering still obeyed the above law, large-angle scattering did not. Rutherford correctly deduced that in large-angle scattering (which corresponds to closer approach to the nucleus) the α-particle was actually striking the nucleus. This meant that the size of the nucleus could be calculated by finding the maximum angle for which the inverse-square scattering formula worked and thus finding how close to the centre of the nucleus such an α-particle arrived. In this way, Rutherford estimated the radius of the aluminium nucleus to be approximately 10^{-14} m. Indeed, from the expression derived in Ex. 4.1.2, we find $r_{\min}\simeq 1$ fm for aluminium.

Exercise 4.1.4. *By setting* $\dot{r}=0$ *in Eq. (4.1.15a) and solving for* r*, find the distance of minimum approach for any general scattering angle.*

4.1.4 Small-angle scattering

Now, for small deviations the Rutherford formula diverges as θ^{-4} and would thus lead to an infinite interaction probability. This would clearly be unphysical and is

an indication that a hitherto neglected effect must come into action as $\theta \to 0$. It is not difficult to find: small angles correspond to large values of b. When b is of the same order as the atomic radius r_{atom}, the screening effect of the electrons must be taken into account and eventually, for $b > r_{\text{atom}}$, the α-particle will only see an overall neutral atom and therefore the cross-section must actually *vanish* as $\theta \to 0$.

We now wish to examine the passage of electrons and other charged particles inside matter and the possibility of describing the case of *multiple* scattering (through highly probable *small* angles). The results of this calculation will allow us to describe the energy deposition of charged particles as they pass through matter. Of course, we may take the Rutherford formula above as the starting point (merely changing z to reflect the charge of the particle under consideration). The situation we now wish to describe is that depicted in Fig. 4.4. The final scattering angle

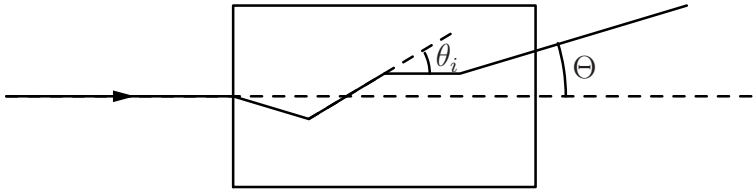


Figure 4.4: Multiple scattering of a charged particle in matter.

Θ is the result of a random walk in the transverse plane due to a (large) number n of scatterings through small angles θ_i. What we are interested in then is the distribution of the final overall angle Θ.

The small-angle approximation to the Rutherford formula (4.1.22) is

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \simeq \left[\frac{2zZ\alpha}{pv}\right]^2 \frac{1}{\theta^4}. \tag{4.1.24}$$

Note that this approximation $\sin\theta \sim \theta$ is good to within better than 30 % over the range $0 \leq \theta \leq \frac{1}{2}\pi$ and if larger angles were important, then the problem would in any case be intractable. However, the strong peaking in the forward direction means that the cross-section is indeed dominated by small angles. Clearly, the first problem with which we must deal is the (unphysical) singular behaviour as $\theta \to 0$.

As remarked earlier, the Rutherford formula is valid as long as the particle mainly sees only the nucleus, while for $b \sim r_{\text{atom}}$ the screening effect of the electrons must be taken into account. The Thomas–Fermi model (1927) describes well the

screening effect on the Coulomb potential:

$$V_c(r) \;\to\; V_{\rm TF}(r) \;\simeq\; \frac{zZ\alpha}{r}\,\exp(-r/r_{\rm atom}), \tag{4.1.25}$$

where experimentally we find $r_{\rm atom}\simeq 1.4a_0Z^{-1/3}$ and $a_0=\hbar^2/m_e\alpha\simeq 0.53\,\text{Å}$ is the Bohr radius. The exponential kills the potential beyond the confines of the atom (faster than any power) and thus protects the limit $\theta\to 0$.

Exercise 4.1.5. *Using this potential, show that, in the small-angle approximation, the integral may be equivalently obtained from*

$$\frac{{\rm d}\sigma}{{\rm d}\Omega} \;\simeq\; \left[\frac{2zZ\alpha}{pv}\right]^2 \frac{1}{(\theta^2+\theta_{min}^2)^2}, \tag{4.1.26}$$

where, from Eq. (4.1.20),

$$\theta_{min} \;=\; \frac{zZ\alpha}{\frac{1}{2}mv_\infty^2\, r_{atom}} \tag{4.1.27}$$

is therefore an effective cut-off on the integral, corresponding to $b=r_{atom}$.

The cut-off evaluated here arises from the simple classical finite size of the scattering potential and the screening effect of the electron cloud. However, given the physical dimensions of the objects and the momentum transfers involved, we should also consider possible quantum-mechanical effects. Rather than perform a full quantum mechanics calculation, we shall use simple reasoning to deduce a lower quantum limit to the scattering angle.

Now, the finite size of the target also implies localisation of the projectile in the transverse plane. Consequently, the Heisenberg uncertainty principle leads to a corresponding smearing of the transverse momentum. The implicit uncertainty on the transverse position is $\Delta x_\perp\sim r_{\rm atom}$, which induces a spread in the transverse momentum $\Delta p_\perp\sim\hbar/r_{\rm atom}$. This in turn, causes an uncertainty in the scattering angle $\Delta\theta\sim\Delta p_\perp/p\sim\hbar/pr_{\rm atom}$, which causes a smearing in the θ distribution. The result is again an effective lower cut-off equal to this uncertainty.

We thus have two possible lower limits to the θ integral:

$$\text{classical:}\quad \theta^{\rm c}_{\rm min} \;\simeq\; \frac{2zZ\alpha}{p\,v\,r_{\rm atom}} \;=\; \frac{2zZ\alpha}{\beta}\,\frac{\hbar}{p\,r_{\rm atom}} \tag{4.1.28a}$$

and

$$\text{quantum:}\quad \theta^{\rm q}_{\rm min} \;\simeq\; \frac{\hbar}{p\,r_{\rm atom}}, \tag{4.1.28b}$$

where $\beta\equiv v/c$ (we now drop the suffix ∞ and asymptotic values are implied).

Written in this way, comparison is straightforward:

$$\frac{\theta^{\rm c}_{\rm min}}{\theta^{\rm q}_{\rm min}} \simeq \frac{2zZ\alpha}{\beta} = \frac{2zZ}{137\beta}. \tag{4.1.29}$$

We immediately see that for $\beta \sim 1$ (*i.e.*, a relativistic projectile, which is usually the case) and Z not too large (we shall typically be considering $|z|=1$), $\theta^{\rm c}_{\rm min} < \theta^{\rm q}_{\rm min}$. Under normal circumstances it is thus the quantum mechanics limit that determines the actual cut-off. Note then that in this case there is no dependence on the mass of the projectile, only on its momentum p and charge z. Substituting for $r_{\rm atom}$, we may now rewrite the quantum limit as

$$\theta^{\rm q}_{\rm min} \simeq \frac{Z^{1/3}}{192}\frac{m_e c}{p}, \tag{4.1.30}$$

where the electron mass m_e merely appears owing to the formula for $r_{\rm atom}$ and has nothing to do with the projectile.

At this point it is important to note that wave mechanics also imposes an *upper* limit on the scattering angle due to the finite size of the *nucleus*. From simple optical arguments, see Fig. 4.5, we know that diffraction due to a black disk of radius R occurs mainly within an angular spread given by $\theta \lesssim \lambda\!\!\!^{-}/R$, where $\lambda\!\!\!^{-} \equiv \lambda/2\pi$. Outside this region the interference is, overall, destructive and the amplitude falls

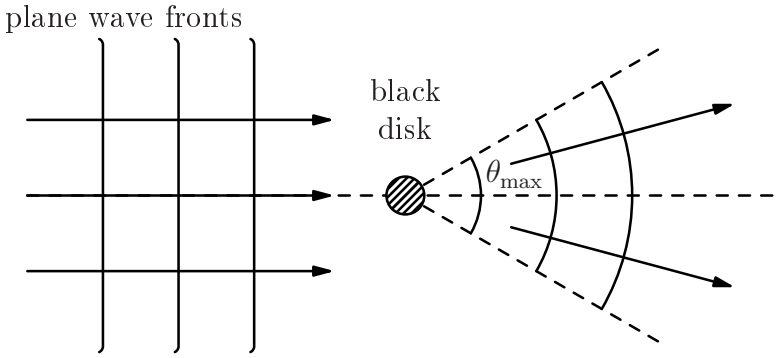


Figure 4.5: Optical black-disk diffraction.

rapidly to zero. Using the de Broglie relation, we thus have

$$\theta_{\rm max} \simeq \frac{\hbar}{p\, r_{\rm nucl}}, \tag{4.1.31}$$

where, empirically,

$$r_{\rm nucl} \simeq {\rm A}^{1/3} \times 1.4 \text{ fm}. \tag{4.1.32}$$

Putting everything together, we finally obtain

$$\theta_{\rm max} \simeq \frac{274}{A^{1/3}} \frac{m_e c}{p}. \tag{4.1.33}$$

We may now insert typical numerical values and estimate the upper and lower limits obtained: $p \simeq 50\,{\rm MeV}/c$ and $Z \simeq 20$, $A \simeq 40$ (calcium) lead to

$$\theta^{\rm q}_{\rm min} \simeq 10^{-4} \text{ rad} \tag{4.1.34a}$$

and

$$\theta_{\rm max} \simeq 0.8 \text{ rad}. \tag{4.1.34b}$$

As desirable, we thus find $\theta^{\rm q}_{\rm min} < \theta_{\rm max} < 1\,{\rm rad}$, but recall that the distribution is very strongly peaked for $\theta \to 0$.

We can now evaluate the mean-square value of θ (for θ small, ${\rm d}\cos\theta \simeq \theta {\rm d}\theta$):

$$\begin{aligned} \langle \theta^2 \rangle &\simeq \frac{\int_{\theta_{\rm min}}^{\theta_{\rm max}} {\rm d}\theta \; \theta^{-1}}{\int_{\theta_{\rm min}}^{\theta_{\rm max}} {\rm d}\theta \; \theta^{-3}} \\ &\simeq \frac{\ln[\theta_{\rm max}/\theta_{\rm min}]}{\frac{1}{2}[\theta_{\rm min}^{-2} - \theta_{\rm max}^{-2}]} \\ &\simeq 2\theta_{\rm min}^2 \ln[\theta_{\rm max}/\theta_{\rm min}]. \end{aligned} \tag{4.1.35}$$

Again, putting in the numbers, we obtain

$$\begin{aligned} \langle \theta^2 \rangle &\simeq 2\theta_{\rm min}^2 \ln\left[\frac{274 \times 192}{2^{1/3} Z^{2/3}}\right] \\ &\simeq 4\theta_{\rm min}^2 \ln\left[\frac{204}{Z^{1/3}}\right], \end{aligned} \tag{4.1.36}$$

where we have used the approximation $Z \simeq A/2$. We thus see that the mean value of θ^2 is very small (of order $\theta^2_{\rm min}$) and depends only very weakly on Z.

For incident low-energy particles (*i.e.*, non-relativistic), we find $\theta_{\rm max} \to 1$ and

$$\langle \theta^2 \rangle \simeq 2\theta_{\rm min}^2 \ln\left[\frac{192}{Z^{1/3}} \frac{p}{m_e c}\right]. \tag{4.1.37}$$

Note that this still depends only very weakly on Z, but now too on p.

4.1.5 Multiple scattering

We now wish to evaluate the mean final angle $\langle\Theta\rangle$ after multiple scattering. We shall thus also need to evaluate the mean number $\langle n\rangle$ of collisions. Successive collisions are independent and each scattering angle takes on a random value between $\theta_{\min}$ and $\theta_{\max}$ according to the distribution evaluated above. At this point, for $\langle n\rangle$ large enough we may apply the *central-limit theorem.*

Theorem 4.1.1 (The central-limit theorem)**.** *For m independent variables θ_i distributed randomly, each with both means $\langle\theta_i\rangle$ and variances $\langle(\theta_i-\langle\theta_i\rangle)^2\rangle$ finite, for $m\to\infty$ the distribution of $\Theta\equiv\sum_{i=1}^m\theta_i$ is a Gaussian with central value $\langle\Theta\rangle=\sum_{i=1}^m\langle\theta_i\rangle$ and mean-square value $\langle\Theta^2\rangle=\sum_{i=1}^m\langle\theta_i^2\rangle$.*

In particular, if the m independent variables all have the same mean and variance, then we have $\langle\Theta\rangle=m\langle\theta\rangle$ and $\langle\Theta^2\rangle=m\langle\theta^2\rangle$. Note that here, owing to the cylindrical symmetry of the problem, $\langle\theta\rangle=0$. The process we wish to describe may thus indeed be viewed as a random walk in the transverse plane.

From the definition of the cross-section in terms of collision probability, it immediately follows that* $\langle n\rangle=Nl\sigma$, where N is the number density of scattering centres in the target, l the length of the target (in the beam direction) and σ the cross-section. We therefore have

$$\langle n\rangle \;\simeq\; 2\pi Nl\left[\frac{2zZ\alpha}{p\beta}\right]^2\frac{}{2\theta_{\min}^2}. \tag{4.1.38}$$

And, from the central-limit theorem, we then have

$$\langle\Theta^2\rangle \;=\; 2\pi Nl\left[\frac{2zZ\alpha}{p\beta}\right]^2\ln\left[\frac{\theta_{\max}}{\theta_{\min}}\right]. \tag{4.1.39}$$

The appearance of p and β in the expression warns us that the formula will inevitably break down if the collisions are sufficient to significantly reduce the energy of the particle. Our final result is thus

$$\langle\Theta^2\rangle \;=\; 4\pi Nl\left[\frac{2zZ\alpha}{p\beta}\right]^2\ln\left[\frac{204}{Z^{1/3}}\right]. \tag{4.1.40}$$

The central-limit theorem also gives us the distribution: a Gaussian, independently of the exact angular distribution of the single-scattering process. The distribution we must consider is the projection onto some chosen longitudinal plane

* The number of collisions n obeys a Poisson distribution: $P(n\text{ events})=\frac{\lambda^n\,\mathrm{e}^{-\lambda}}{n!}$ (λ is the mean number).

(*e.g.*, x–z), for which we shall define the variable θ':

$$P_m(\theta')\,\mathrm{d}\theta' \;=\; \frac{1}{\sqrt{\pi\langle\Theta^2\rangle}}\,\exp\left[-\frac{\theta'^2}{\langle\Theta^2\rangle}\right]\mathrm{d}\theta'. \tag{4.1.41}$$

This is to be compared with the single-scattering formula, suitably rewritten:

$$P_s(\theta')\,\mathrm{d}\theta' \;=\; Nl\,\frac{\mathrm{d}\sigma}{\mathrm{d}\theta'}\,\mathrm{d}\theta'. \tag{4.1.42}$$

Then, since

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\theta'} \;=\; 2\pi\left[\frac{2zZ\alpha}{p\beta}\right]^2\frac{1}{\theta'^3}, \tag{4.1.43}$$

for single scattering we have

$$P_s(\theta')\,\mathrm{d}\theta' \;=\; 2\pi Nl\left[\frac{2zZ\alpha}{p\beta}\right]^2\frac{\mathrm{d}\theta'}{\theta'^3}. \tag{4.1.44}$$

From the dependence on θ', we see that for some large enough angle the single-scattering mechanism will be more probable and will thus eventually dominate over the Gaussian tail of the multiple-scattering process.

It is convenient to define a normalised angle, $\alpha := \theta'/\langle\Theta^2\rangle^{1/2}$, in terms of which the comparison is more transparent:

$$P_m(\alpha) \;=\; \frac{1}{\sqrt{\pi}}\,\mathrm{e}^{-\alpha^2} \tag{4.1.45a}$$

and

$$P_s(\alpha) \;=\; \frac{1}{\ln\!\left(204/\sqrt[3]{Z}\right)}\,\frac{1}{\alpha^3}. \tag{4.1.45b}$$

Typically, one finds that multiple-scattering dominates up to values of $\alpha \sim 2.5$, after which the single-scattering process takes over, thereby extending the tail of the distribution (see Fig. 4.6). While multiple scattering accounts for around 95 % of events, the single-scattering tail contains the remaining 5 %. The position of the crossover is determined by the value of $\langle n\rangle$ (which depends on the target thickness). For $\langle n\rangle \lesssim 100$, the distribution starts to peak more than a Gaussian and the crossover moves to $\alpha_c < 2.5$. On the other hand, if $\langle n\rangle$ grows too large, so that $\langle\Theta^2\rangle \sim \theta^2_{\mathrm{max}}$, then the single-scattering tail disappears.

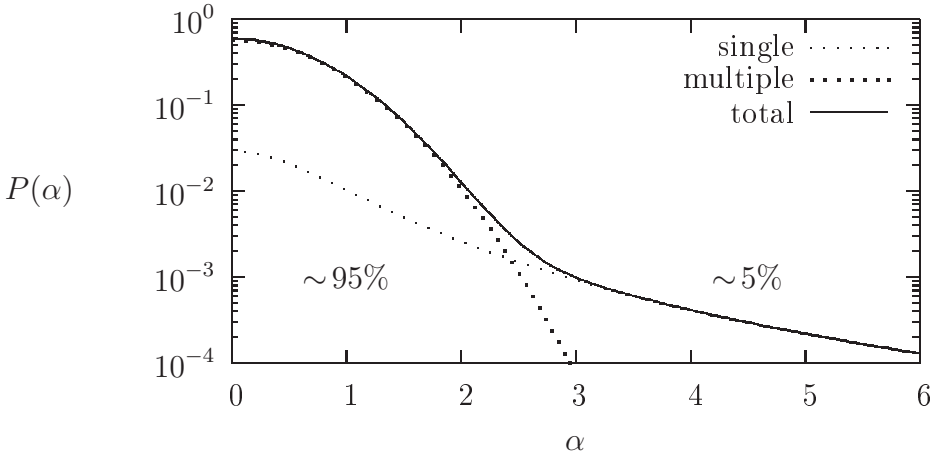


Figure 4.6: The overall distribution for multiple and single scattering (the distributions and the variable α are defined in the text).

4.2 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 fine detail of a small nuclear target obtained via such scattering will be 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 only exhibited a point-like behaviour (at least up to the highest energies presently available).

Thus, for low-energy scattering (where the weak force is particular suppressed) the interaction of the electron with a nucleon or nucleus is governed purely by QED and is therefore completely known.

4.2.1 Elastic scattering in quantum mechanics

The simplest example of scattering with electrons is the elastic case. Here energy and momentum are transferred from an electron to a nucleus via (single) photon exchange and the final nucleus remains intact, see Fig. 4.7. 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). Energy and momentum conservation

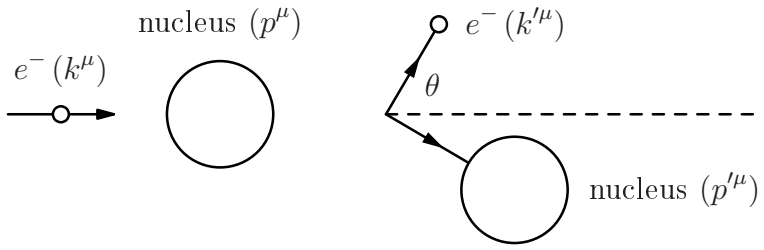


Figure 4.7: Elastic electron–nucleus scattering.

then require

$$k^\mu + p^\mu \;=\; k'^\mu + p'^\mu. \tag{4.2.1}$$

Since the momentum of the final nucleus is not usually measured, it is convenient to rewrite this as

$$p'^2 \;=\; M^2 \;=\; (k^\mu - k'^\mu + p^\mu)^2, \tag{4.2.2}$$

where M is the nuclear mass. By noting that in the laboratory frame $p^\mu = (M, \mathbf{0})$ it is then straightforward to derive the relation

$$EM \;=\; E'E(1-\cos\theta) + E'M, \tag{4.2.3}$$

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 (the electron is assumed to be highly relativistic). This may be rearranged to express E' as a function of E and θ:

$$E' \;=\; \frac{E}{1+(1-\cos\theta)E/M}, \tag{4.2.4}$$

which provides the well-known result that, for two-body *elastic* scattering, the final energy is determined by the angle (and *vice versa*).

Exercise 4.2.1. *Derive the previous two expressions explicitly.*
Hint: *start from Eq. (4.2.2).*

Let us now simplify to the non-relativistic limit for the expression of the cross-section (we shall treat the relativistic case later). The scattering cross-section for electrons will then simply be given by the Rutherford formula, with the obvious substitution $z = -1$:

$$\frac{\mathrm{d}\sigma^e}{\mathrm{d}\Omega} \;=\; \frac{1}{16}\left[\frac{Z\alpha}{E_\infty}\right]^2 \mathrm{cosec}^4\,\tfrac{\theta}{2}. \tag{4.2.5}$$

However, this is the cross-section for a point-like target while we wish to study the charge *distribution* inside the nucleus and so we shall now see how a distributed charge modifies the formula.

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).

Let us make a kinematical consideration relating to the first point. For elastic scattering of a light particle of total energy E, momentum p and mass m off a heavy stationary target of mass M (*i.e.* $m \ll M$), the maximum momentum transfer is trivially $\boldsymbol{q} \simeq 2\boldsymbol{p}$, which is then the maximum recoil momentum of the target particle. The maximum recoil kinetic energy $\nu_{\max} = E - E'_{\min}$ is thus $\nu_{\max} = \boldsymbol{q}^2/2M \simeq 4\frac{m}{M}E \ll E$; *i.e.* the energy loss for scattering off particles much *heavier* than E or m is negligible.

Exercise 4.2.2. *Find the equivalent relation in the relativistic case; i.e.* $E \gg m$

Our starting point will be Fermi's golden rule for the total transition rate W (Φ is the incident flux):

$$\Phi\,\sigma \;:=\; W \;=\; \frac{2\pi}{\hbar}\left|\langle\psi_f|\mathcal{H}_{\rm int}|\psi_i\rangle\right|^2 \frac{\mathrm{d}n}{\mathrm{d}E_f}, \tag{4.2.6}$$

where the third factor is the density of final states and E_f is the total final energy $K + m_A$ (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:

$$\psi_i \;=\; \frac{1}{\sqrt{V}}\,\mathrm{e}^{\frac{\mathrm{i}}{\hbar}\boldsymbol{k}\cdot\boldsymbol{x}}, \tag{4.2.7a}$$

$$\psi_f \;=\; \frac{1}{\sqrt{V}}\,\mathrm{e}^{\frac{\mathrm{i}}{\hbar}\boldsymbol{k}'\cdot\boldsymbol{x}}, \tag{4.2.7b}$$

where the normalisation is one particle in a volume V (this volume will, of course, disappear from the final answer for σ). The density of final states is

$$\mathrm{d}n(k) \;=\; V\,\frac{k^2\,\mathrm{d}k\;\mathrm{d}\Omega}{(2\pi\hbar)^3}, \tag{4.2.8}$$

where $k \equiv |\boldsymbol{k}|$. We therefore have, noting that the flux is $\Phi = v/V$,

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \;=\; \frac{V^2E'^2}{(2\pi)^2}\left|\,\mathcal{M}_{fi}\,\right|^2\frac{1}{\hbar^4 v} \tag{4.2.9}$$

and so we need to calculate the transition matrix element $\mathcal{M}_{fi} = \langle\psi_f|\mathcal{H}_{\rm int}|\psi_i\rangle$.

Now, for a non-relativistic electron (with charge e) the interaction is given in terms of the electromagnetic scalar potential ϕ: $\mathcal{H}_{\text{int}} = e\phi$. Thus, we have (suppressing $\hbar$)

$$\langle \psi_f | \mathcal{H}_{\text{int}} | \psi_i \rangle \;=\; \frac{e}{V} \int \mathrm{d}^3\boldsymbol{x} \;\, \mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}} \;\, \phi(\boldsymbol{x}), \tag{4.2.10}$$

where we have defined and already substituted $\boldsymbol{q} := \boldsymbol{k} - \boldsymbol{k}'$, the *momentum transfer*. It is immediately obvious that the above is none other than the Fourier transform of the potential $\phi(\boldsymbol{x})$, which is in turn determined by the charge-density distribution $\rho(\boldsymbol{x})$:

$$\boldsymbol{\nabla}^2 \phi(\boldsymbol{x}) \;=\; -\rho(\boldsymbol{x}). \tag{4.2.11}$$

The plane-wave form allows us to rewrite the expression for the matrix element directly in terms of $\rho(\boldsymbol{x})$ by noting that

$$\boldsymbol{\nabla}^2 \, \mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}} \;=\; -q^2 \, \mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}}, \tag{4.2.12}$$

where $q \equiv |\boldsymbol{q}|$. Green's theorem then leads to

$$\langle \psi_f | \mathcal{H}_{\text{int}} | \psi_i \rangle \;=\; \frac{e}{Vq^2} \int \mathrm{d}^3\boldsymbol{x} \;\, \mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}} \;\, \rho(\boldsymbol{x}). \tag{4.2.13}$$

It is convenient to define a normalised density,

$$\rho(\boldsymbol{x}) \;=:\; Ze\, f(\boldsymbol{x}), \tag{4.2.14}$$

such that

$$\int \mathrm{d}^3\boldsymbol{x} \;\, f(\boldsymbol{x}) \;=\; 1. \tag{4.2.15}$$

We thus obtain

$$\langle \psi_f | \mathcal{H}_{\text{int}} | \psi_i \rangle \;=\; \frac{Z\alpha}{Vq^2} \int \mathrm{d}^3\boldsymbol{x} \;\, \mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}} \;\, f(\boldsymbol{x}). \tag{4.2.16}$$

The integral on the right-hand side, the Fourier transform of the charge distribution, is called the *form factor*:

$$F(\boldsymbol{q}) \;\equiv\; \int \mathrm{d}^3\boldsymbol{x} \;\, \mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}} \;\, f(\boldsymbol{x}). \tag{4.2.17}$$

Putting everything together, for the differential cross-section we then finally have

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \;=\; \frac{Z^2\alpha^2 E'^2}{q^4} \left| F(\boldsymbol{q}) \right|^2. \tag{4.2.18}$$

It is easy to derive (neglecting the electron mass) that

$$\boldsymbol{q}^2 \;=\; (\boldsymbol{k}-\boldsymbol{k}')^2 \;=\; 4EE'\,\sin^2\frac{\theta}{2} \tag{4.2.19}$$

and thus we may write

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \;=\; \frac{Z^2\alpha^2}{16E^2\sin^4\frac{\theta}{2}}\,\big|F(\boldsymbol{q})\big|^2, \tag{4.2.20a}$$

or

$$\;=\; \big|F(\boldsymbol{q})\big|^2\,\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}^{\mathrm{Rutherford}}. \tag{4.2.20b}$$

In other words, the effect of the internal structure of the nucleon is to induce 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 (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. However, for q large 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.

4.2.2 Relativistic elastic scattering—the Mott formula

To resolve the internal structure of a nucleus (*i.e.*, to be sensitive to the energy dependence in the form factors), we require the wave-length of the exchanged photon to be small compared to the nuclear size. Then, since $\hbar c \simeq 200\,\mathrm{MeV}\cdot\mathrm{fm}$, we deduce that the energy required is of the order of hundreds of MeV. This implies, in particular, that the electrons will certainly be relativistic. In this case we should really perform a calculation based on the Dirac equation. A relativistic calculation for a point-like object leads to the Mott cross-section:

$$\frac{\mathrm{d}\tilde{\sigma}}{\mathrm{d}\Omega}^{\mathrm{Mott}} \;=\; \left(1-\beta\sin^2\frac{\theta}{2}\right)\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}^{\mathrm{Rutherford}}, \tag{4.2.21}$$

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 Mott cross-section falls off more rapidly than the Rutherford expresssion for large angles (indeed, for $\theta = \pi$ it vanishes). Note though again, that the effect conveniently factorises.

Let us just take a moment to try and understand this behaviour in physical terms. In the relativistic limit, $\beta \to 1$ (which is clearly equivalent to the limit $m \to 0$), the helicity or projection of the particle spin onto its direction of motion $h := \hat{\boldsymbol{p}} \cdot \hat{\boldsymbol{s}}$ is a conserved quantum number if the interactions are of a purely vector type (*e.g.*, via photon exchange). We thus speak of right- and left-handed fermions having $h = \pm 1$ respectively. Indeed, starting from the Dirac equation, it can be shown that helicity-flip amplitudes are proportional to m/E (where E is some typical interaction energy scale (*e.g.*, the centre–of–mass energy).

Consider now the extreme cases of forward and backward scattering, in which the incoming electron collides and either continues undisturbed or returns along the direction from where it came, see Fig. 4.8. If we assume a spin-zero nucleus, then

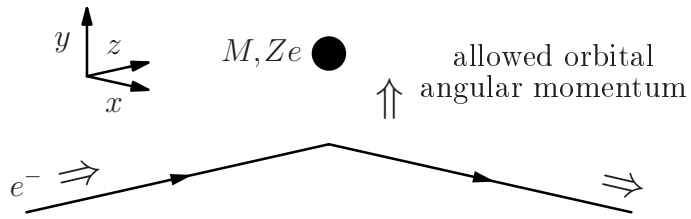


Figure 4.8: Helicity conservation in two-body scattering processes: the only possible orbital angular momentum is perpendicular (lying along the y axis) to the variation in the electron spin vector (lying in the x–z plane).

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 by itself. For forward scattering this is trivially the case while for backward scattering it is evidently impossible. For intermediate cases, one needs to understand how spin-projection eigenstates are constructed for arbitrary directions. A state of positive helicity travelling in a direction θ with respect to the chosen quantisation axis ($\hat{z}$ say) we find

$$|+,\theta\rangle \;=\; \cos\tfrac{\theta}{2}|+,z\rangle + \sin\tfrac{\theta}{2}|-,z\rangle. \tag{4.2.22}$$

Thus, the relevant amplitude is $\langle +,\theta|+,z\rangle = \cos\frac{\theta}{2}$, which on squaring gives just the Mott result. Of course, if the nucleus also possesses an intrinsic angular momentum (due to internal motion and/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 its spin.

Finally, we also should consider the recoil of the struck nucleus and consequent modification of the final-state phase space. The full Mott formula is then

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}^{\text{Mott}} \;=\; \frac{E'}{E}\frac{\mathrm{d}\tilde{\sigma}}{\mathrm{d}\Omega}^{\text{Mott}} \;=\; \frac{E'}{E}\left(1 - \beta\sin^2\frac{\theta}{2}\right)\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}^{\text{Rutherford}}. \tag{4.2.23}$$

4.2.3 Form factors

We have just seen that the effect of an extended charge distribution is factorisable into a form factor, which depends only on the momentum transfer and which, being a Fourier transform, contains in principle all necessary information on the charge distribution. That is, if we were able to measure $F(\boldsymbol{q})$, then by comparing data with the point-like Mott expression,

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}^{\mathrm{expt}} \qquad \textit{vis à vis} \qquad \left|F(\boldsymbol{q})\right|^2 \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}^{\mathrm{Mott}}, \tag{4.2.24}$$

over the *entire* range of $\boldsymbol{q}$ from zero to infinity we could 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 experiments on $F(\boldsymbol{q})$ is shown in Fig. 4.9. The pattern of maxima and minima, reminiscent of

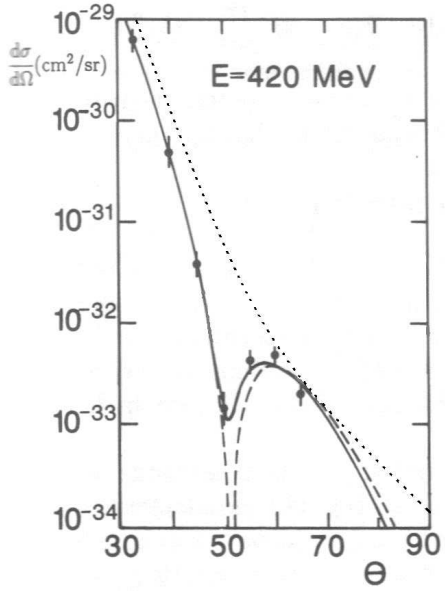


Figure 4.9: The angular dependence of elastic electron scattering off ^{12}C. The continuous line is a fit to experimental results, the dashed line a theoretical curve for a uniform sphere of charge and the dotted line the point-like Mott cross-section.

diffraction in classical optics, is clearly evident. Note that the rapid fall-off with angle severely limits the largest 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 $\lambda_{\mathrm{min}} \sim 0.2\,\mathrm{fm}$).

To extract the partial information available, there are various possibilities. One could, of course, perform an approximate Fourier transform of the measured function $F(\boldsymbol{q})$; however, it is more practical to adopt one of two standard approaches:

- comparison with parametrised trial functions;
- so-called the method of moments.

Fitting form factors to trial functions

The idea behind the fitting of form factors to trial functions is that one may find simple plausible parametrisations for charge-density distributions, which will in turn lead to parametrisations of the form factors. One then simply performs fits to the data and extracts the parameters. Note that the angular dependence at large angles provides information on the internal structure while as $\theta \to 0$ (for which eventually $\lambda > R_{\rm nucl}$) one should see a return to the typical q^{-4} behaviour.

Exploiting the typical spherical symmetry, we may integrate out the angular dependence and thus simplify the expression for $F(\boldsymbol{q})$:

$$F(q^2) \;=\; \frac{4\pi}{q}\int_0^\infty \mathrm{d}r\, r \sin qr\, f(r). \tag{4.2.25}$$

Note that the density $f(r)$ is then normalised as

$$4\pi \int_0^\infty r^2 \mathrm{d}r\, f(r) \;=\; 1. \tag{4.2.26}$$

In Table 4.1 we list some typical functional forms used. As an example, consider

Table 4.1: A collection of possible forms of charge distributions inside the nucleus, together with the corresponding form factors. In all cases R represents a notional 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

the case of a uniform sphere; the first minimum lies at $qR \simeq 4.5$. Thus, referring to the graph for ^{12}C shown in Fig. 4.9, we find $R \simeq 2.5\,$fm.

Experimentally in fact, it turns out that rather precise measurements can be performed and data may be collected over a large range of angles, covering a variation in cross-section of up to about seven orders of magnitude. One can thus find several minima and so obtain further information on the distributions. More simply, referring again to the graph for ^{12}C shown in Fig. 4.9, the fact that the minima are not as sharp as predicted for a uniform sphere indicates the existence of a "skin" of finite depth.

The moment method for fitting form factors

The bulk of the data lies in the region $q^2 \sim 0$ and thus it is tempting to try to exploit this region in particular by performing a power-series expansion around the point $q^2 = 0$. From Eq. (4.2.25) we obtain

$$\begin{aligned} F(q^2) &= 4\pi \int_0^\infty r^2 \mathrm{d}r\, f(r) \left[1 - \tfrac{1}{3!}(qr)^2 + \tfrac{1}{5!}(qr)^4 - \dots\right] \\ &= 1 - \tfrac{1}{6}\, q^2 \langle r^2 \rangle + \tfrac{1}{120}\, q^4 \langle r^4 \rangle - \dots \end{aligned} \tag{4.2.27}$$

The quantities $\langle r^2 \rangle$, $\langle r^4 \rangle$ *etc.* are just *moments* of the charge-density distribution. This may be inverted to give

$$\langle r^2 \rangle = -6 \left. \frac{\mathrm{d}F(q^2)}{\mathrm{d}q^2} \right|_{q^2=0} \quad \textit{etc.}, \tag{4.2.28}$$

i.e., the slope (in q^2) at the point $q^2 = 0$ directly provides the r.m.s. radius of the nucleus.* Of course, it is not quite so simple: the point $q^2 = 0$ corresponds to zero-angle scattering, which is impossible to measure directly and so extrapolation is necessary. Fortunately, the high statistics usually available for small angles allows considerable confidence in the final result.

As a final word, the two approaches may be combined: a suitable model parametrisation may be used to calculate the r.m.s. radius and higher moments as functions of the parameters. The various derivatives obtained from extrapolations to $q^2 = 0$ may then be used to fit the calculated moments and extract the parameters of the model.

* Note that for a uniform spherical distribution of radius R, we have $\langle r^2 \rangle = \frac{3}{5} R^2$.

Scattering results on nuclear structure

From the experimental data collected a picture of the larger nuclei emerges that is very well described by a two-parameter so-called *Fermi* function:

$$\rho(r) \simeq \frac{\rho_0}{1+\mathrm{e}^{(r-c)/a}}, \tag{4.2.29}$$

which has the same form as the Woods–Saxon potential already encountered.* The two free parameters are c (an *effective* nuclear radius) and a (a sort of skin depth). The quantity ρ_0, which is approximately the bulk nuclear density (since $c \gg a$), is fixed by normalisation to the mass number A. For large enough nuclei, typical numerical values of the parameters are

$$c \simeq 1.07\ \mathrm{A}^{1/3}\ \mathrm{fm}, \tag{4.2.30a}$$

$$a \simeq 0.54\ \mathrm{fm}. \tag{4.2.30b}$$

Note that $\rho(c) = \frac{1}{2}\rho(0)$ and $\rho(0) \simeq \rho_0$.

A series of experiment-based observations can be made:

- The r.m.s. nuclear radius is well described by the formula $\langle r^2 \rangle^{1/2} \simeq A^{1/3} r_0$ with $r_0 \simeq 0.94\,\mathrm{fm}$. For a uniform sphere of charge, using $\langle r^2 \rangle = \frac{3}{5}R^2$, this implies $R \simeq 1.21\mathrm{A}^{1/3}\,\mathrm{fm}$.

- The surface or skin thickness, typically defined by $t := r|_{\rho=0.1\rho_0} - r|_{\rho=0.9\rho_0}$, is roughly constant for heavy nuclei: $t = 2a\ln 9 = 2.4\,\mathrm{fm}$.

- The bulk *charge* density ρ_0 decreases slightly with increasing A. However, taking into account that the ratio Z/A also decreases, it is found that to a very good approximation the bulk nuclear density is constant. The value is $\rho_{\mathrm{nucl}} \simeq 0.166\,\mathrm{nucleons/fm}^3$.

- For light nuclei, the Fermi function is not a good description of the density and a Gaussian is closer.

- Clearly, in cases where there is a large spin angular momentum (common in very large nuclei), the rotational motion causes strong deformation and the approximation of spherical symmetry breaks down.

* Again, the similarity between the nuclear potential and density is indicative of the short-range nature of the strong nuclear force.

4.2.4 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 narrow 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. 4.10. There are essentially three

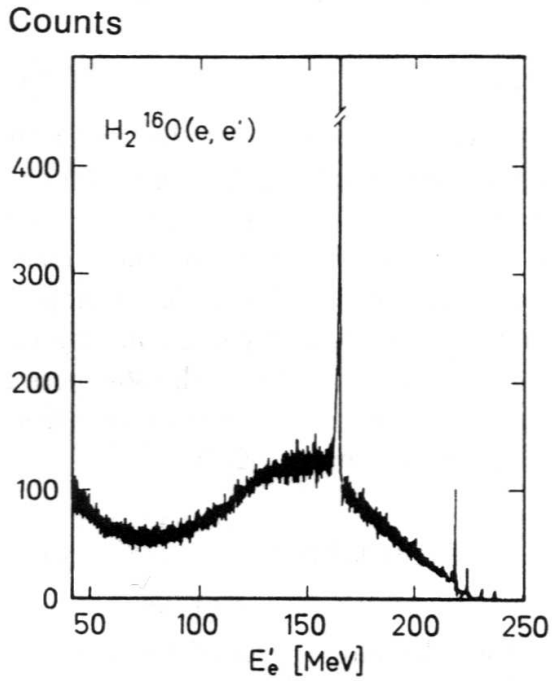


Figure 4.10: Quasi-elastic electron scattering off water as a function of outgoing electron energy (beam energy 246 MeV and scattering angle 148.5°).

prominent features to be found in the spectrum, two pronounced spikes and one broad underlying peak. The larger of the spikes is clearly due to elastic scattering off one of the hydrogen nuclei (this is determined simply from the position of the spike) while the smaller spike corresponds to the oxygen nucleus. The question remains then of the broader lower structure that never–the–less has a clear form of a peak. We shall now show that this may be attributed to scattering off a single proton bound inside an oxygen nucleus; the proton is expelled from the nucleus and this we call *quasi-elastic* scattering.

Exercise 4.2.3. *Invert the relation* (4.2.4) *derived earlier to provide a formula for the mass as a function of the electron energies and scattering angle. Thus, calculate the masses corresponding to the two peaks and deduce that their origins really are elastic scattering off a free proton and an oxygen nucleus.*

We have already seen that the Fermi-gas model makes rather precise predictions: the nucleons sit 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 the energy that must be supplied above and beyond the kinematical needs while the Fermi motion will cause smearing of the centre–of–mass energy leading to a broadening of the final-state spectrum. Let us now examine in detail how this works. The process we wish to study then has now a *three*-body final state, see Fig. 4.11. We begin by defining $\nu := E_e - E'_e$, the energy transfer in

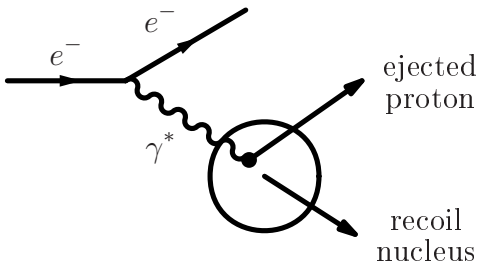


Figure 4.11: A schematic view of quasi-elastic scattering.

the laboratory frame. It is then a simple exercise to show that

$$\nu \;=\; \frac{|q^2|}{2M} + V_0 + \frac{2|\boldsymbol{q}||\boldsymbol{P}|\cos\alpha}{2M}, \tag{4.2.31}$$

where α is the angle between the vectors $\boldsymbol{q}$ and $\boldsymbol{P}$ (note that $q^2 < 0$).

Exercise 4.2.4. *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 evaluate $\langle\nu\rangle$ and σ_ν:

$$\begin{aligned}\langle\nu\rangle \;&=\; 2\pi\int_0^{p_\mathrm{F}} \boldsymbol{P}^2 \mathrm{d}|\boldsymbol{P}| \int_{-1}^{1} \mathrm{d}\cos\alpha \left[\frac{q^2}{2M} + V_0 + \frac{2|\boldsymbol{q}||\boldsymbol{P}|\cos\alpha}{2M}\right] \Bigg/ \left[\frac{4\pi p_\mathrm{F}^3}{3}\right] \\ &=\; \frac{q^2}{2M} + V_0 \end{aligned} \tag{4.2.32}$$

and

$$\begin{aligned}\sigma_\nu \;&\equiv\; \sqrt{\langle\nu^2 - \langle\nu\rangle^2\rangle} \;=\; \frac{q}{M}\sqrt{\langle\boldsymbol{P}^2\rangle \times \langle\cos^2\alpha\rangle} \;=\; \frac{q}{M}\sqrt{\tfrac{1}{3}\langle\boldsymbol{P}^2\rangle} \\ &=\; \frac{1}{\sqrt{5}}\frac{qp_\mathrm{F}}{M}, \end{aligned} \tag{4.2.33}$$

where we have used the result (for a uniform momentum distribution) $\langle\boldsymbol{P}^2\rangle = \frac{3}{5}p_\mathrm{F}^2$.

Exercise 4.2.5. *From the peak position and width as shown in Fig. 4.10, estimate the 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 were bound inside a potential well with strong spatial dependence then it would be permanently subject to forces that should be added to those operating during the scattering process. However, the data lead to a picture in which the nucleons sit in a potential well with a *flat* bottom (*i.e.*, they are not subject to any forces except for the brief moments in which they make contact with the boundary 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 of 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 a string-like interaction. Nevertheless, calculations (first performed by Feynman) in the impulse approximation describe the data perfectly 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.

4.3 The nucleon–nucleon interaction

To study the nucleon–nucleon interaction or potential, we have access to two different types of systems, both with important limitations however:

(i) *bound states*
in the case of the nuclear interaction there exists only one two-body bound state, namely the deuteron ($^2_1$D);

(ii) *nucleon–nucleon scattering*
the target can only consist of single protons while the projectile may be either a proton or neutron; so we may access the pp and pn potentials but not nn.

While the underlying theory for the strong interaction (QCD) does not allow us to calculate the nucleon–nucleon interaction from first principles or, *e.g.*, the

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

bound state of two nucleons (the deuteron), we can describe such an interaction via a model using our basic knowledge of quantum mechanics. We shall first examine the case of a bound state using a model potential and then show how we can describe simple low-energy nucleon–nucleon scattering.

4.3.1 The deuteron

The deuteron may be considered as the nuclear equivalent of the hydrogen atom or, better still, positronium in QED. A question immediately arises: why is there only one bound state, pn, and neither of the other two possibilities, pp ($^2_2$He) or nn? Note that, from the study of larger nuclei, we know that the potential is attractive and of equal strength between both nucleon species. We shall see that the answer lies in the complete quantum-mechanical description of fermion–fermion bound states.

Properties of the deuteron – data

Let us begin then with a list of (static) properties and parameters pertaining to the deuteron.

Binding energy: The deuteron is a relatively weakly bound nuclear state; that is, $E_B = 2.22452(20)\,\mathrm{MeV}$ (recall that the number in parenthesis represents the experimental error in the last significant figures quoted). The binding energy per nucleon is thus a little over 1.1 MeV, to be compared to that of, *e.g.*, ^{4}He, which is approximately 7 MeV. Indeed, the fusion of two deuterons to form a helium nucleus releases around 24 MeV.

Measurement of the binding energy (or equivalently mass) may be performed via study of the formation and/or decay processes:*

$$p + n \underset{\text{photodisintegration}}{\overset{\text{capture}}{\rightleftharpoons}} {}^2_1\mathrm{D} + \gamma \qquad \hookrightarrow E_{\mathrm{kin}} \simeq E_B. \tag{4.3.1}$$

A typical experimental setup might be as shown in Fig. 4.12. The quartz crystal exploits the standard property of Bragg refraction: $2d\sin\beta = n\lambda$, where d is the separation of the crystal planes, β the incident and diffracted angle with respect to the crystal planes and n is an integer.

* N.B. To be precise, $E_{\mathrm{kin}} = E_B - E_{\mathrm{recoil}}$, where in this case E_{recoil} is of order 1 keV and may therefore be neglected.

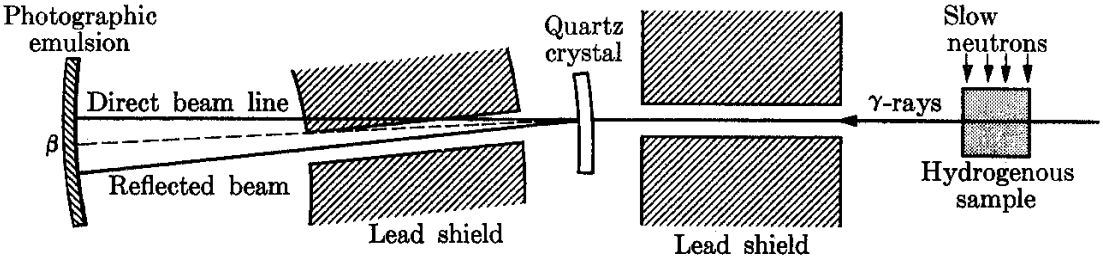


Figure 4.12: A typical experimental setup for measuring the deuteron binding energy.

Exercise 4.3.1. *For an initial "excited" deuteron nucleus of mass M^* say ($M^* = m_p + m_n$), calculate the photon energy E_γ in terms of the final deuteron mass M_D and the total energy release $Q = M^* - M_D$.*

Spin–parity: Via the study of resonance bands (optical, radio and microwave frequencies), the spin is easily deduced: $J = 1$. From the parities of the final states in decay process the intrinsic parity is found to be $P = +$. Put together, we have then $J^P = 1^+$ (to be compared with, *e.g.*, the photon $J^P_\gamma = 1^-$).

Isotopic spin: The isotopic-spin (or isospin) assignment of the deuteron is $I = 0$. Since it is composed of a proton and a neutron (both having isospin one-half) the possibilities *a priori* are either 0 or 1.

Magnetic moment: The most precise measurement of the deuteron magnetic moment is best reported in relation to that of the proton:

$$\frac{\mu_{\rm D}}{\mu_{\rm p}} = 0.307012192(15). \tag{4.3.2}$$

The comparison that is most informative is however between the single moments (in units of the nuclear magneton μ_N):

$$\begin{aligned} \mu_{\rm p} &\simeq 2.792847356\ \mu_{\rm N} \\ \mu_{\rm n} &\simeq -1.91304273\ \mu_{\rm N} \\ \mu_{\rm p} + \mu_{\rm n} &\simeq 0.87980463\ \mu_{\rm N}, \\ \mu_{\rm D} &\simeq 0.857393\ \mu_{\rm N} \end{aligned} \tag{4.3.3}$$

where the difference, while small, is significant.

Electric quadrupole moment: The electric quadrupole moment of the deuteron is $Q = 0.00286 \pm 0.00015\,{\rm e\,b}$ (where e is the charge of the proton). This is

possible even for a system composed of one charged and one neutral particle, since the components themselves are also composite and have a highly non-trivial internal structure (as witnessed by the non-zero magnetic moment of the neutron).

Radius: The radius of the deuteron is approximately 2.1 fm, which should be compared to that of a single proton or neutron (~ 0.8 fm) and other small nuclei such as ^{4}He (~ 1.7 fm). Such a relatively large size again suggests a loosely bound system.

Properties of the deuteron – theoretical deductions

From the previously listed physical properties of the deuteron we can make a series of theoretical deductions as to its nature. From the spin–parity assignment $J^P = 1^+$ and the magnetic moment $\mu_{\rm D} \simeq \mu_{\rm p} + \mu_{\rm n}$, we can immediately infer the orbital and spin structure of the proton–neutron pair.

First of all, the parity of such a composite object is given by the product of the parities of its components (including the spatial wave-function). That is,

$$\psi_D \;=\; \psi_p \;\psi_n \;\psi(\boldsymbol{x}_{pn}), \tag{4.3.4}$$

where $\psi_{p,n}$ are the wave-functions describing the internal structure of the proton and neutron, and $\psi(\boldsymbol{x}_{pn})$ describes their relative spatial motion. And so

$$P_D \;=\; P_p \;P_n \;P_{\rm spatial}, \tag{4.3.5}$$

where $P_{p,n}$ are the intrinsic parities of the proton and neutron, and $P_{\rm spatial}$ is the parity of the spatial wave-function. Now, from the Dirac equation we know that fermion and antifermions have opposite intrinsic parity and by convention we choose to assign even parity to fermions and odd parity to their antiparticles. In any case, here we have two fermions and so $P_p P_n = +$. Since $P_D = +$, this then implies that $P_{\rm spatial} = +$ too.

From the solution of the Schrödinger equation for a two-body system we know that $P_{\rm spatial} = (-)^L$, where L is the orbital angular momentum quantum number. Thus, L must be even here. That is, the most probable spatial state of the deuteron is an s-wave although some small admixture of d-wave might be possible.

Indeed, if we now consider that the deuteron magnetic moment almost corresponds to the sum of the proton and neutron moments, then we see that the spins of the two nucleons must be aligned and the total spin S will be unity, to which should be added the orbital angular momentum $L = 0, 2, \ldots$. For $L = 0$ there is then only one possibility: $J = 1$ while in the case $L = 2$ there two: $J = 1$ or 3. Clearly then, both $L = 0$ and 2 are possible. Indeed, the presence of a small admixture of

d-wave would explain the magnetic moment: the d-wave still has the two nucleon spins aligned but they are overall anti-aligned with the total spin of the deuteron and therefore partially cancel the contribution to the magnetic moment. A numerical estimate of the contribution would suggest that deuteron is approximately 4% d-wave. Such a value is also consistent with the quadrupole measurement.

The deuteron – theory

The two-body Schrödinger equation: We have still to understand the question of the observed number of nucleon–nucleon bound states (just one, the deuteron). To do so we shall need to examine in some detail the quantum mechanical description of nucleon–nucleon systems. We start then from the reduced time-independent Schrödinger equation for a two-body system:

$$E\,\psi(\boldsymbol{x}) \;=\; -\frac{\hbar^2}{2m}\,\boldsymbol{\nabla}^2\psi(\boldsymbol{x}) + V(\boldsymbol{x})\,\psi(\boldsymbol{x}). \tag{4.3.6}$$

where, recall, m is the reduced mass $m = m_1 m_2/(m_1+m_2)$, which in this case is approximately half that of the nucleon. Here the wave-function only refers to the relative position; the full wave-function is $\Psi(\boldsymbol{x},\boldsymbol{x}_{\rm CM}) = \psi(\boldsymbol{x})\,\psi_{\rm CM}(\boldsymbol{x}_{\rm CM})$, where $\boldsymbol{x}$ is the relative position of the two nucleons and $\boldsymbol{x}_{\rm CM}$ of the centre–of–mass system.

Now, while we know that the ground-state energy is $E = -2.225\,\text{MeV}$, we do not know the form of the potential $V(\boldsymbol{x})$, which may also be a function of $\boldsymbol{S}$, $\boldsymbol{L}$ *etc.* We may however assume that to a reasonable approximation the potential is spherically symmetric, so that $V(\boldsymbol{x}) \to V(r)$, r is then just the distance between the centres of the two nucleons, and the Schrödinger equation may be separated into radial and angular parts via a corresponding factorisation of the wave-function:

$$\psi(\boldsymbol{x}) \;=\; \frac{u_l(r)}{r}\,Y_{lm}(\theta,\phi). \tag{4.3.7}$$

The equation for the radial part is then

$$0 \;=\; \frac{\mathrm{d}^2 u_l(r)}{\mathrm{d}r^2} + \frac{2m}{\hbar^2}\left[E - V(r) - \frac{l(l+1)\hbar^2}{2mr^2}\right] u_l(r). \tag{4.3.8}$$

We see that there is an effective contribution to the potential from the orbital angular momentum, which is often called the centrifugal potential:

$$V(r) \;=\; \frac{l(l+1)\hbar^2}{2mr^2} \;=\; \tfrac{1}{2}\frac{\langle \boldsymbol{L}^2\rangle}{I}, \tag{4.3.9}$$

which is just the energy associated with the corresponding classical orbital angular

momentum. This is equivalent to an effective *repulsive* force for $l>0$, thus justifying the assumption that the ground-state corresponds to $l=0$ (assuming V to be spherically symmetric).

We can construct a simple model for the nucleon–nucleon potential, which should contain the important details, as a finite-depth well with a hard-core repulsive force representing the finite and impenetrable size of the nucleons themselves. The radial direction is divided into three regions:

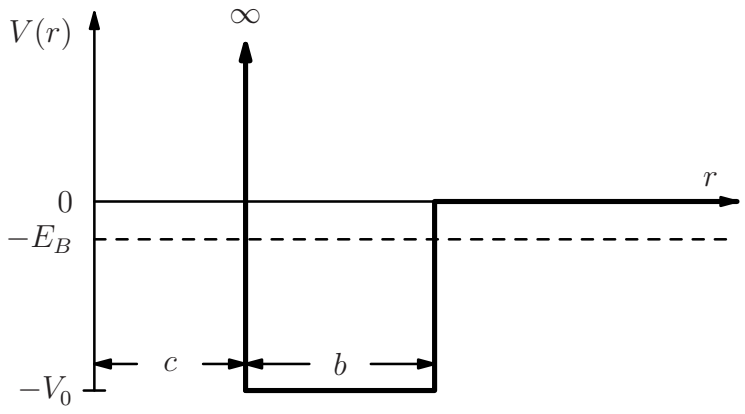


Figure 4.13: A model potential well for the nucleon–nucleon interaction in the deuteron and other possible two-nucleon systems. The region $0<r<c$ (r is the distance between the centres of the two nucleons) represents a hard core and the well, of depth V_0, lies in the region $c<r<c+b$.

(*I*) $0<r<c$ – here the potential is infinite, thus describing the hard impenetrable core of the nucleons themselves;

(*II*) $c<r<c+b$ – the well is flat of depth V_0;

(*III*) $c+b<r$ – outside the finite range of the strong nuclear interaction the potential is immediately zero.

The form of the radial wave-function in each region is then easy to write down:

(*I*) $u(r)=0$;

(*II*) $u(r)=\mathrm{e}^{\pm \mathrm{i}kr}$, with $k=\sqrt{2m(V_0-E_B)}/\hbar$;

(*III*) $u(r)=\mathrm{e}^{-\kappa r}$, with $\kappa=\sqrt{2mE_B}/\hbar$.

Continuity of the wave-function and its first derivative at the boundary then leads to the following relation between k and κ:

$$k\cot kb \;=\; \kappa. \tag{4.3.10}$$

In region (II) we have $u(r) = A\sin k(r-c)$ and in (III) $u(r) = B\exp(-\kappa r)$, where the coefficients A and B are fixed by the overall normalisation and the continuity condition $A\sin kb = B\exp\big(-\kappa(b+c)\big)$.

There are thus three free parameters that must be fixed by experimental observation: b, c, V_0; a fourth E_B is the eigenvalue corresponding to the known binding energy in the case of the deuteron. From high-energy scattering we find $c \simeq 0.4\,\mathrm{fm}$. The wave-function above may be used to calculate the mean-square radius of the deuteron:

$$\begin{aligned}\langle r_D^2\rangle &= \int \mathrm{d}^3\boldsymbol{x}\, r^2\,|\psi(\boldsymbol{x})|^2 \\ &= \int \mathrm{d}r\, r^2\,|u(r)|^2 \\ &= \frac{1}{8\kappa^2} - \frac{1}{8k^2} + \frac{(2c+b)(1+\kappa b)}{8\kappa} + \frac{c^2}{4} + \frac{\kappa b^3}{24(1+\kappa b)} + \langle r_N^2\rangle, \qquad (4.3.11)\end{aligned}$$

where the last term takes into account the finite nucleon size: $\langle r_N^2\rangle^{1/2} \simeq 0.8\,\mathrm{fm}$. Now, since $V_0 \gg E_B$, we find that the first term dominates while the second is decidedly small. Taking $V_0 \simeq 40\,\mathrm{MeV}$ and $b \simeq 1.9\,\mathrm{fm}$ gives $\langle r_D^2\rangle^{1/2} \simeq 2.2\,\mathrm{fm}$, which is close but not equal to the experimental value of 2.1 fm. Enforcing the experimental value, leads to $V_0 \simeq 73\,\mathrm{MeV}$ and $b \simeq 1.3\,\mathrm{fm}$, which is rather too deep. We note finally that in three dimensions such a potential well has a critical minimum depth $V_{\min}$, such that for $V_0 < V_{\min}$ there are *no* bound-state solutions and for only slightly deeper wells there will just a single bound state. We thus begin to understand the paucity of nucleon–nucleon bound states—indeed, for the possible parameter sets listed, we do find just the one, ground-state, solution.

Spin dependence: We have already observed that the only known bound two-nucleon state has $J = 1$. It should be stressed, however, that the non-existence of $J = 0$ states does not necessarily imply that the overall nuclear force is not attractive in such cases; it simply requires $E_B \leq 0$, a *negative* binding energy. Now, the condition $E_B = 0$ is equivalent to $kb = \pi/2$ and $V_0 = K = \hbar^2 k^2/2m$, the kinetic energy inside the well. This in turn implies then that to avoid bound states, we require

$$V_0 b^2 \;\leq\; 102\ \mathrm{MeV}\cdot\mathrm{fm}^2. \qquad (4.3.12)$$

So that, in fact, it is sufficient that there be some extra contribution that reduces the effective potential depth. Such an extra contribution may easily arise from an appreciable spin–spin interaction between the two nucleons. The known state has spins parallel and therefore we may deduce that the force is attractive in this case and repulsive for the antiparallel case.

Now, we have still not completely explained the absence of ${}^{2}_{2}$He and ${}^{2}_{0}$?. Note however, that in both cases the nucleons involved are identical (the system is either pp or nn respectively) and therefore the Pauli exclusion principle requires that the overall wave-function be antisymmetric. The s-wave spatial state (or $L=0$) is symmetric and therefore the antisymmetry must come from the spin part. This then *constrains* the two nucleons to combine into a singlet (or spin zero) combination, which, as we have just hypothesised, presumably leads to a repulsive contribution to the potential and therefore no bound state.

4.3.2 Nucleon–nucleon scattering

Partial-wave expansion

In order to obtain further information on the nucleon–nucleon interaction, we must resort to scattering experiments. Here we shall use the method of partial waves, which is, in general, useful for studying low-energy scattering. Recall once again that a two-body process may be cast into the form of a Schrödinger equation for a single particle scattering off a fixed potential, where the coordinate is the separation of the two particles and the mass used is again the reduced mass, which in this case will be approximately half the nucleon mass.

The initial state is a free particle in motion along the beam direction with energy E and momentum p and therefore may be represented naturally as a plane wave in, say, the z direction:

$$\psi_{\rm in}(\boldsymbol{x},t) \;=\; \mathrm{e}^{\mathrm{i}\,(kz-\omega t)}, \tag{4.3.13}$$

with frequency $\omega=E/\hbar$ and wave-number $k=p/\hbar$. Since the scattering will be assumed elastic (*i.e.* $E_{\rm in}=E_{\rm out}$), the time dependence will be the same in all cases and is therefore irrelevant. We shall hence just consider

$$\begin{aligned} u_{\rm in}(\boldsymbol{x}) \;&=\; \mathrm{e}^{\mathrm{i}kz} \;=\; \mathrm{e}^{\mathrm{i}kr\cos\theta} \\ &=\; \sum_{l=0}^{\infty} \mathrm{i}^{\,l}\,\sqrt{4\pi(2l+1)}\,j_l(kr)\;Y_{lm}(\theta)\big|_{m=0}\,. \end{aligned} \tag{4.3.14}$$

The functions $j_l(kr)Y_{l0}(\theta)$ represent the set of Bessel (j_l) functions* and spherical-harmonic (Y_{lm}) solutions to the free-particle Schrödinger equation and the sum is just the expansion of a plane-wave solution over such a basis set. Note that the simplification $m=0$ is obtained by choosing the z-axis (the beam direction) as the angular-momentum quantisation axis. Thus, since the motion is along the

* Recall that there is another set of possible solutions, using the so-called Neumann functions $n_l(kr)$. These are however singular at the origin and so are automatically excluded here.

z direction (and there is therefore no motion in the plane normal to $\hat{z}$), there can be no orbital angular momentum projection along the z direction and hence only the $m=0$ spherical waves contribute.

The use of such a basis set is an important simplification considering the conservation of angular momentum implicit in a spherically symmetric potential scattering. If we study the scattering functions wave–by–wave, then the coefficient of each contribution must remain unaltered in magnitude and may thus change by at most an imaginary phase. This condition is known as "*partial-wave unitarity*".

Now, although the presence of an infinite sum would appear to render calculation somewhat difficult, consideration of the actual physical system will allow us an important simplification. We are interested in low-energy scattering, say $E_{\mathrm{lab}} \sim 100\,\mathrm{MeV}$, and we also know that the nuclear interaction has a finite range of around 1 fm, which is then effectively the upper limit to the typical impact parameter b. These two values place an upper limit on the angular momentum that may be generated between the target and projectile. A kinetic energy of, say, 10 MeV implies a momentum equivalent (using the reduced mass) of about 100 MeV/c. The maximum orbital angular momentum is just the product of this with the maximum impact parameter just given. We thus obtain the limit $|\boldsymbol{L}| \lesssim 100\,\mathrm{MeV/c \cdot fm}$, which is only ~ 0.5 in units of $\hbar$. In other words, in most cases of interest only the s-wave actually contributes appreciably and the sum is effectively truncated very early. Of course, for higher energies the limit is higher: $E_{\mathrm{kin}} = 100\,\mathrm{GeV}$ leads to $|\boldsymbol{L}| \lesssim 500\,\hbar$. Partial-wave analysis is very useful in low-energy (nuclear) studies but entirely impracticable for modern-day high-energy physics (unless, that is, the form of the interaction in question does not enforce, say, purely s-wave scattering).

We shall now make a rather informal derivation of the so-called partial-wave expansion formula. Restricting our attention for the moment to $l=0$, the part of the unperturbed wave-function that contributes is just

$$u_0(\boldsymbol{r}) \;=\; \frac{\sin kr}{kr} \;=\; \frac{\mathrm{e}^{\mathrm{i}kr} - \mathrm{e}^{-\mathrm{i}kr}}{2\mathrm{i}kr}. \tag{4.3.15}$$

The two terms in this expression have a particular interpretation: reinstating the temporal dependence $\mathrm{e}^{-\mathrm{i}\omega t}$, we see that the first represents an outgoing spherical wave while the second is incoming. We now wish to calculate the effect of "switching on" a finite-range spherically symmetric (or central) scattering potential. Clearly, only the outgoing wave has actually "seen" the potential (naturally placed at the origin) and moreover flux conservation implies that its action cannot change the magnitude of its coefficient. Therefore, the only effect can be to introduce a *phase-shift* into the outgoing wave, thus far *outside* the range R of the

scattering potential we may write

$$u_0(\boldsymbol{r}) \xrightarrow{V\neq 0} \frac{\mathrm{e}^{\mathrm{i}(kr+2\delta_0)} - \mathrm{e}^{-\mathrm{i}kr}}{2\mathrm{i}kr}, \tag{4.3.16}$$

where the factor 2 multiplying the phase-shift δ_0 is for later convenience. Note that δ_0 will in general be a function of energy. Note also that the new wave-function too is a solution of the Schrödinger equation for a free particle outside the range of the scattering potential. Note too the k cannot change as the scattering is elastic.

The difference between these two solutions is then just the scattered wave:

$$u_{\mathrm{scatt}}(\boldsymbol{r}) \;=\; \frac{\mathrm{e}^{\mathrm{i}(kr+2\delta_0)} - \mathrm{e}^{\mathrm{i}kr}}{2\mathrm{i}kr} \;=\; \sin\delta_0\, \frac{\mathrm{e}^{\mathrm{i}(kr+\delta_0)}}{kr}. \tag{4.3.17}$$

Thus, in the presence of the scattering potential, but far *outside* its range (*i.e.* for $r \gg R$), the full solution to the Schrödinger equation is of the form

$$u(\boldsymbol{x}) \xrightarrow{r\to\infty} \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} + f(\theta,\phi)\, \frac{\mathrm{e}^{\mathrm{i}kr}}{r}, \tag{4.3.18}$$

where θ and ϕ are the polar scattering angles with respect to $\boldsymbol{k}$, which we have taken to be along the z-axis, while $r = |\boldsymbol{x}|$. That is, we have the original incident plane wave (the first term) plus an outgoing spherical wave centred on the origin (the second term), with possible angular modulation $f(\theta,\phi)$.

Now, to calculate the partial-wave cross-section σ_0 we need the ratio between the incoming flux and that scattered. The scattered flux may be calculated via the standard quantum mechanical formula, or more simply as the density multiplied by the velocity. The total scattered flux (integrating over the surface of a sphere of radius r) is thus

$$\Phi^{\mathrm{tot}}_{\mathrm{scatt}} \;=\; |u_{\mathrm{scatt}}|^2 \cdot v \cdot 4\pi r^2 \;=\; \left[\frac{\sin\delta_0}{kr}\right]^2 v \cdot 4\pi r^2 \;=\; 4\pi v\, \frac{\sin^2\delta_0}{k^2}, \tag{4.3.19}$$

where we have exploited the angular independence of s-wave scattering. Adopting the same normalisation, the incident flux is just v. And so the s-wave partial cross-section is

$$\sigma_0 \;=\; 4\pi\, \frac{\sin^2\delta_0}{k^2}. \tag{4.3.20}$$

Moreover, since s-wave scattering is isotropic, we can write

$$\frac{\mathrm{d}\sigma_0}{\mathrm{d}\Omega} \;=\; \frac{\sin^2\delta_0}{k^2}. \tag{4.3.21}$$

Exercise 4.3.2. *Derive the above formulæ using the standard form in quantum mechanics of the current density for a given wave-function.*

For the general partial wave l, the angular dependence is encoded in the full Bessel and spherical harmonic functions giving the expansion

$$u(\boldsymbol{x}) \;=\; \sum_{\ell=0}^{\infty}(2\ell+1)\,\mathrm{i}^{\,\ell}\frac{1}{2\,\mathrm{i}\,kr}\left[\mathrm{e}^{+\,\mathrm{i}\,(kr-\frac{1}{2}\ell\pi+2\delta_\ell)}-\mathrm{e}^{+\,\mathrm{i}\,(kr-\frac{1}{2}\ell\pi)}\right]P_\ell(\cos\theta), \qquad (4.3.22)$$

where recall θ is the centre–of–mass scattering angle and P_l are the Legendre polynomials; moreover, $\mathrm{i}^{\,\ell}\equiv\mathrm{e}^{\,\mathrm{i}\,\frac{1}{2}\ell\pi}$. We thus obtain

$$f(\theta,\phi) \;=\; \frac{1}{k}\sum_{\ell=0}^{\infty}(2\ell+1)\,\mathrm{e}^{\,\mathrm{i}\delta_\ell}\sin\delta_\ell\,P_\ell(\cos\theta) \qquad (4.3.23)$$

and therefore

$$\frac{\mathrm{d}\sigma_l}{\mathrm{d}\Omega} \;=\; (2l+1)\,\frac{\sin^2\delta_l}{k^2}\,\Big|P_l(\cos\theta)\Big|^2. \qquad (4.3.24)$$

In the case of s-wave scattering, transformation to the laboratory frame gives

$$\frac{\mathrm{d}\sigma_0}{\mathrm{d}\Omega_{\mathrm{lab}}} \;=\; \frac{\sin^2\delta_0}{k^2}\,4\cos\theta_{\mathrm{lab}}. \qquad (4.3.25)$$

Exercise 4.3.3. *Derive explicitly the laboratory-frame expression for the s-wave partial cross-section.*
Hint: *show that for equal-mass particles $\theta=2\theta_{lab}$.*
Confirm moreover that whereas the differential cross-section is different in the two frames, its total integral is invariant, as it should be.

The partial-wave expansion formula provides then an effective and convenient parametrisation of low-energy elastic scattering cross-sections. For each partial wave there is just one parameter, the phase-shift $\delta_l(E)$, which enters as precisely a shift in the imaginary phase of the wave-function (owing to the presence of a perturbing potential. If therefore, as we have seen, only a very limited number of partial waves can contribute, only a limited number of parameters is necessary. Note moreover, from Eq. (4.3.21), that δ_l also determines the modulation of the partial contribution via the coefficient $\sin\delta_l$ in the partial amplitude.

Going back to our model for the deuteron, we can examine the solutions for the positive energy case $E>0$ (*i.e.* a free-particle). The form of the radial wave-function in each region is now:

(I) $u(r)=0$;

(II) $u(r)=\mathrm{e}^{\pm \mathrm{i}kr}$, with $k=\sqrt{2m(V_0+E)}/\hbar$;

(III) $u(r)=\mathrm{e}^{\pm \mathrm{i}\kappa r}$, with $\kappa=\sqrt{2mE}/\hbar$.

In region (II) we have $u(r)=A\sin k(r-c)$ and in (III) $u(r)=B\sin(\kappa r+\delta_0)$, where again the coefficients A and B are fixed by the overall normalisation and the continuity condition at the boundary. Continuity of the derivative leads to the following relation between k and κ:

$$k\cot kb \;=\; \kappa\cot[\kappa(c+b)+\delta_0]. \tag{4.3.26}$$

By solving this equation for δ_0 (as a function of E) and using the values of the parameters given, we can calculate the partial-wave cross-section for s-wave scattering, see Fig. 4.14. For $E_{\mathrm{lab}} \gtrsim 5\,\mathrm{MeV}$ the agreement is very good, but for lower

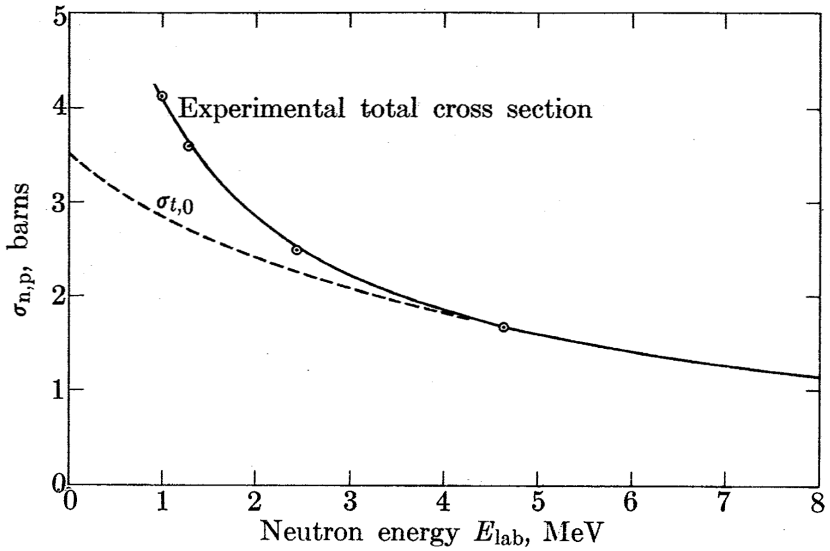


Figure 4.14: Plots of the experimental (solid curve) and theoretical (dashed curve) cross-sections for np scattering as a function of neutron energy E_{lab}.

energies the theoretical curve is systematically lower than the data.

The difference is easily explained. The data used to fix the parameters here come from the study of the deuteron, which, we recall, is a spin-one combination of the two nucleon spins (*i.e.* a spin triplet state). However in a scattering process all spin combinations may and indeed do contribute. The multiplicity is, as always, $2j+1$ and so there are three spin-one combinations and just one for spin-zero

(singlet). The physical scattering cross-section is then a weighted superposition:

$$\sigma^{\text{expt}} \;=\; \tfrac{1}{4}\sigma_0^{\text{spin-0}} + \tfrac{3}{4}\sigma_0^{\text{spin-1}}. \tag{4.3.27}$$

Evidently then, we may deduce that at lower energies the spin-zero potential is stronger than that of the spin-one state.

Scattering lengths

We are examining the case of low-energy scattering, where the partial-wave expansion is particularly useful owing to the natural simplifications in this limit. So, let us now consider the extreme case of so-called *threshold* scattering, *i.e.* the zero-energy or infinite wave-length limit. For $E\to 0$, $\kappa\to 0$ and $l=0$, outside the nucleon radius ($r>R=b+c$) the radial part of the Schrödinger equation reduces to

$$\frac{\mathrm{d}^2u(r)}{\mathrm{d}r^2} \;=\; 0, \tag{4.3.28}$$

to which the solution is simply

$$u(r) \;\propto\; (r-a_0), \tag{4.3.29}$$

i.e. a straight line. Consider then the limit of such a wave form as $\kappa\to 0$, or equivalently $\lambda\to\infty$:

$$u(r) \;\propto\; \lim_{\kappa\to 0}\sin(\kappa r+\delta_0). \tag{4.3.30}$$

To eliminate the unknown constant of proportionality, consider the normalised derivative ratio (the prime denotes the spatial derivative):

$$\frac{u'(r)}{u(r)} \;=\; \kappa\cot(\kappa r+\delta_0),$$

which, from Eq. (4.3.29), must also give

$$\xrightarrow{\kappa\to 0} \;\frac{1}{r-a_0}. \tag{4.3.31}$$

Note that, since we are dealing with the solution *outside* the potential well, the boundary condition at the origin cannot be applied and thus a_0 is not yet determined. However, in order to determine the constant a_0, we may take r with any value; it is most convenient to choose $r=0$ and thus

$$\lim_{\kappa\to 0}\kappa\cot\delta_0 \;=\; -\frac{1}{a_0}. \tag{4.3.32}$$

The parameter a_0 is known as the *scattering length*. Now, we have that

$$\sigma_0 \;=\; 4\pi\frac{\sin^2\delta_0}{\kappa^2} \;=\; 4\pi\frac{1}{\kappa^2|\cot\delta_0 - \mathrm{i}|^2}. \tag{4.3.33}$$

In the limit we therefore find

$$\lim_{\kappa\to 0}\sigma_0 \;=\; 4\pi a_0^2. \tag{4.3.34}$$

One might imagine that since a_0 has dimensions of a length that its magnitude should be roughly determined by the range R of the potential, especially since it seems to be equivalent to a size of the *effective* scattering centre. However, it turns out that, in general, no such relation exists and that in the case of an attractive potential a_0 may even be orders of magnitude larger than R. Let us now try to give a physical meaning to a_0. From the definition (4.3.29), it is clear that a_0 is none other than the intercept of the limiting straight-line tangent to the wave-function outside the well region in the case $E\to 0$, see Fig. 4.15.

Indeed, from Fig. 4.15a, we see that in the specific case of a repulsive potential the intercept must fall inside the potential region and therefore $a_0 < R$. However, the other two figures demonstrate the cases of attractive potentials without and with bound states respectively. To understand this behaviour, consider first the case of a non-binding potential well: the wave-function can have *no* nodes inside or outside the well and therefore must have positive slope at the outside boundary. Thus, a non-binding potential is characterised by a *negative* scattering length.

Now, imagine increasing the depth of the well: the curvature of the wave-function will increase, the slope will decrease and thus (the negative) a_0 will grow in magnitude until it reaches the limiting point of zero slope. This situation corresponds precisely to a well depth that supports just a single bound state with *zero* binding energy. Henceforth, as the well depth increases, since $u(r)$ must tend monotonically to zero as $r\to\infty$, $u(r)$ and $u'(r)$ will always have *opposite* signs at the boundary and thus the slope there will always be such that $a_0 \geq R$. The behaviour at the point where a_0 becomes infinite (or, rather, ill-defined) can be understood as a *resonance* condition (we shall examine this problem in detail shortly): the energy of the corresponding bound state is *zero*, as too is the energy of the incident particle in the $k\to 0$ limit. Thus, it is possible for the particle to actually become trapped by the scattering potential.

The connection between the scattering length and the bound-state energy can be made more quantitative. Examining Fig. 4.15 for the case $V_0 < 0$, we see that if a_0 is large and positive then, since $\mathrm{e}^{-\kappa r}$ is also approximately a straight line just outside the potential range (*i.e.* for $r \gtrsim R$), it is very similar to $u(r) \propto r - a$ and $\mathrm{e}^{-\kappa r}$, with $\kappa \simeq 0$, is indeed the wave-function for a bound state of energy $E = 0^-$

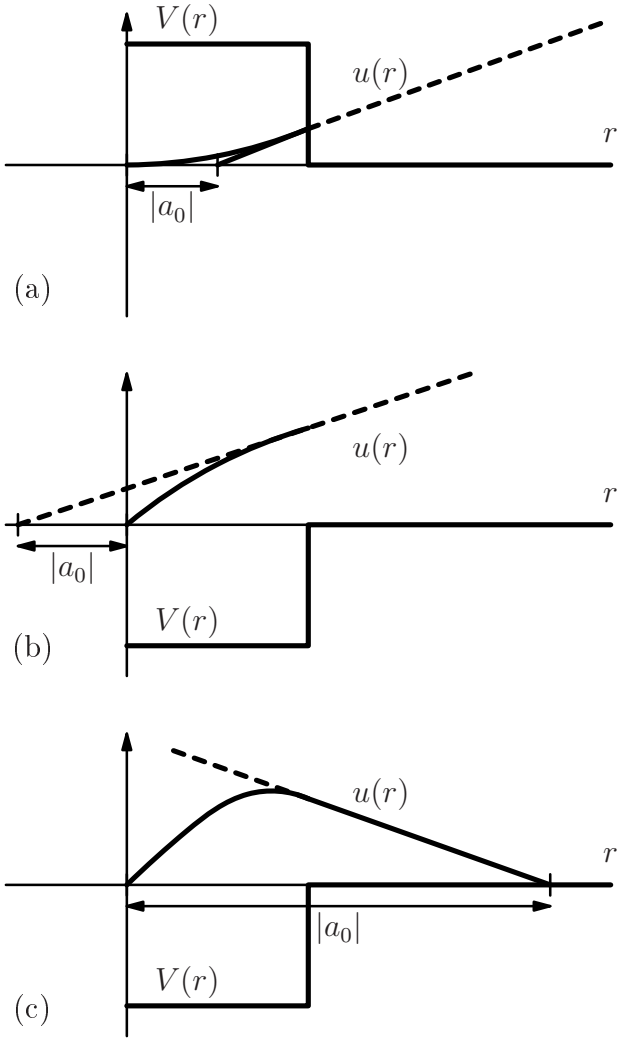


Figure 4.15: A geometric description of the scattering length as the intercept of the limiting straight-line behaviour outside the potential range for $k \to 0$, for: (a) a repulsive potential, (b) an attractive potential that does not support bound states and (c) a potential that supports at least one bound state.

for $r > R$. Now, the wave-function inside for the two cases $E = 0^+$ (zero-energy scattering) and $E = 0^-$ (bound state of zero binding energy) are very similar since in both cases the wave-number k is determined by the equation

$$k = \frac{\sqrt{2m(E - V_0)}}{\hbar} \simeq \frac{\sqrt{2m|V_0|}}{\hbar}. \tag{4.3.35}$$

Matching the logarithmic derivatives for the two cases (bound state and threshold

scattering), we have

$$-\left.\frac{\kappa\,\mathrm{e}^{-\kappa r}}{\mathrm{e}^{-\kappa r}}\right|_{r=R} = \left.\frac{1}{r-a_0}\right|_{r=R}, \tag{4.3.36}$$

which for $a_0 \gg R$ leads to

$$\kappa \simeq \frac{1}{a_0} \tag{4.3.37}$$

Now, since the binding energy B is given by

$$B = \frac{\hbar^2\kappa^2}{2m} \simeq \frac{\hbar^2}{2ma_0^2}, \tag{4.3.38}$$

we therefore have a direct connection between the scattering length a_0 and the energy level of the bound state. In other words, providing $a_0 \gg R$, it is possible to directly deduce the binding energy (for a weakly bound state) by performing scattering measurements at very low energies.

Moreover, it can be shown that in general, for $k \to 0$,

$$\delta_l \sim n_l\,\pi - a_l\,k^{2l+1} + c_l\,k^{2l+3}, \tag{4.3.39}$$

where we have explicitly taken into account the general ambiguity in δ_l up to additive factors of integer multiples of π. Indeed, it can also be shown (Levinson's theorem) that the integer n_l is just the number of bound states that the potential supports for a given angular momentum l. The low-energy expansion of $\cot\delta_0$ then becomes

$$k\,\cot\delta_0 \xrightarrow{k\to 0} -\frac{1}{a_0} + \frac{r_0k^2}{2}, \tag{4.3.40}$$

where a_0 has the same meaning as before and the new constant r_0 introduced is known as the *effective range*—one can show that it roughly corresponds to the distance at which the potential becomes small. The limiting zero-energy s-wave cross-section now becomes

$$\sigma_0^{\mathrm{tot}} = 4\pi \lim_{k\to 0}\left|\frac{1}{k\cot\delta_0 - \mathrm{i}\,k}\right|^2 \simeq \frac{4\pi a_0^2}{\left(1-\frac{1}{2}r_0k^2a_0\right)^2 + k^2a_0^2}. \tag{4.3.41}$$

4.3.3 The optical theorem

We can now provide a simplified proof of a very important and general theorem in scattering. Let us first calculate the total cross-section:

$$\sigma^{\mathrm{tot}} = \int \mathrm{d}\Omega\,\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \int \mathrm{d}\Omega\,\left|f(\theta,\phi)\right|^2$$

$$= \int \mathrm{d}\Omega \left| \frac{1}{k} \sum_{l=0}^{\infty} (2l+1)\ \mathrm{e}^{\mathrm{i}\delta_l} \sin\delta_l\, P_l(\cos\theta) \right|^2,$$

which, using the orthonormality of the Legendre polynomials, reduces to

$$= \frac{4\pi}{k^2} \sum_{l=0}^{\infty} (2l+1)\ \sin^2\delta_l. \tag{4.3.42}$$

Consider now the forward amplitude, *i.e.* for $\theta = 0$ (or $\cos\theta = 1$):

$$f(0) \ = \ \frac{1}{k} \sum_{l=0}^{\infty} (2l+1)\ \mathrm{e}^{\mathrm{i}\delta_l} \sin\delta_l, \tag{4.3.43}$$

where we have used the fact that $P_l(1) = 1$ for all l. The imaginary part of this last expression is just the total cross-section (4.3.42), up to a factor $4\pi/k$, thus

$$\sigma^{\mathrm{tot}} \ = \ \frac{4\pi}{k}\ \mathrm{Im}\, f(0). \tag{4.3.44}$$

This result is precisely the optical theorem: the *total cross-section* is proportional to the *forward scattering amplitude*. Its validity actually extends far beyond the simple proof given here and does not, for example depend on a non-relativistic approach nor indeed on the partial-wave expansion.

There is an apparent contradiction here: the left-hand side is, by definition, proportional to an amplitude *squared* while the right-hand side is *linear* in the amplitude. This may be reconciled by carefully considering the physical origin of a scattering cross-section. In Eq. (4.3.18) we see that the general form of the wave-function is a sum of two terms:

$$u(\boldsymbol{x}) \ \propto \ \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} + f(\theta,\phi)\, \frac{\mathrm{e}^{\mathrm{i}kr}}{r}. \tag{4.3.45}$$

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 corresponding phenomenon has long been well known: a bright central spot is seen behind a black disc diffracting a light source of suitable wavelength (Sellmeier,

1871; Rayleigh, 1871).

Exercise 4.3.4. *Derive the optical theorem using the approach outlined above, i.e. by considering the total flux along the z axis.*

4.4 Resonances and the Breit–Wigner form

4.4.1 Resonances in classical mechanics

In classical mechanics the equation of motion for a forced oscillator subject to friction is

$$m\ddot{x} + \gamma\dot{x} + kx \;=\; F\cos\omega t, \tag{4.4.1}$$

Defining the natural frequency of the oscillator as $\omega_0 := \sqrt{k/m}$, the solution is

$$x \;=\; x_0\cos(\omega t + \phi_0), \tag{4.4.2}$$

where the phase ϕ_0 is given by

$$\tan\phi_0 \;=\; \frac{-\gamma\omega}{m(\omega^2 - \omega_0^2)} \tag{4.4.3}$$

and the oscillation *amplitude* is

$$x_0 \;=\; \frac{F}{[m^2(\omega^2-\omega_0^2)^2 + \gamma^2\omega^2]^{1/2}}. \tag{4.4.4}$$

The total energy of the oscillator is thus

$$E \;=\; \frac{kF^2}{[m^2(\omega^2-\omega_0^2)^2 + \gamma^2\omega^2]}. \tag{4.4.5}$$

Notice how the presence of dissipation (in the form of friction, $\gamma \neq 0$) tames the otherwise divergent behaviour for $\omega = \omega_0$.

4.4.2 Breit–Wigner resonances in quantum mechanics

A similar behaviour occurs in quantum mechanics for the production of intermediate so-called resonant (virtual and unstable) states when the natural energy of the resonant state is near the actual energy of the system. A formal description may be given by considering the variation of phase shifts in the partial-wave decomposition of scattering amplitudes. We shall start from the standard form of

the scattering amplitude in quantum mechanics, taken for some particular partial wave l:

$$f_l \;=\; \frac{\hbar}{2\mathrm{i}p}\,(2l+1)\left[a_l\;\mathrm{e}^{2\mathrm{i}\delta_l}-1\right]P_l(\cos\theta), \tag{4.4.6}$$

where p is the centre–of–mass initial-state momentum, a_l is the amplitude of the of the l-th scattered partial wave (l is just the quantum number for the orbital angular momentum involved) and δ_l is the phase-shift, which contains all relevant information on the scattering potential.

In the purely elastic case with no absorption (*i.e.* $a_l=1$; $a_l<1$ implies absorption) the corresponding partial cross-section is then

$$\sigma_{\mathrm{el}} \;=\; \frac{\pi\hbar^2}{p^2}\,(2l+1)\left|\mathrm{e}^{2\mathrm{i}\delta_l}-1\right|^2. \tag{4.4.7}$$

This expression has a maximum whenever $\delta_l(E)=(n+\frac{1}{2})\pi$ with n integer. The maximum value is

$$\sigma_{\mathrm{el}} \;=\; (2l+1)\,\frac{4\pi\hbar^2}{p^2}. \tag{4.4.8}$$

Note that for the case of *total* absorption (*i.e.* $a_l=0$), the cross-section is just a quarter of this and does not depend on δ_l.

Experimentally, cross-sections are often observed with a well-pronounced peak at some particular centre–of–mass energy. The question then is how to describe such effects in the absence of a complete theory of the interaction involved. From the above formula, we see 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 $\cot\delta_l$ 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 encoded into the spherical harmonics) so each phase-shift δ_l can only be a function of E. One thus says that those *energies* for which $\delta_l=(n+\frac{1}{2})\pi$, for which the cross-section is therefore maximal, correspond to *resonances*. If we assume that such resonances are sufficiently well separated from each other then, for any given (resonant) energy, only one dominates the cross-section. It is instructive to study the behaviour of the amplitude in the neighbourhood of such a point, $E\approx E_0$ say.

The identity (already encountered)

$$\frac{\mathrm{e}^{2\mathrm{i}\delta}-1}{2\mathrm{i}} \;\equiv\; \frac{1}{\cot\delta-\mathrm{i}} \tag{4.4.9}$$

may be used to rewrite the elastic partial-wave amplitude as

$$f_l \;=\; \frac{\hbar/p}{(\cot\delta_l - \mathrm{i})}\,(2l+1)P_l(\cos\theta). \tag{4.4.10}$$

At the resonance we have $\cot\delta_l = 0$ and we may therefore perform a Taylor expansion about this point in the energy E. Retaining only the leading term, linear in $E - E_0$, we may write

$$\cot\delta_l(E) \;\simeq\; -\frac{2}{\Gamma}\,(E - E_0), \tag{4.4.11}$$

where the choice of sign is conventional. Suppose now that the phase-shift *grows* with energy in the neighbourhood of the resonance. Therefore, $\cot\delta_l(E)$ *decreases* and the parameter Γ introduced above is *positive*. Inserting this into Eq. (4.4.10) one obtains the standard Breit–Wigner (BW) form for the amplitude:

$$f_l(E) \;\simeq\; -\frac{\hbar}{p}\,\frac{\Gamma/2}{(E - E_0) - \mathrm{i}\Gamma/2}\,(2l+1)\,P_l(\cos\theta). \tag{4.4.12}$$

And the elastic cross-section then becomes

$$\sigma_l \;\simeq\; \frac{4\pi\hbar^2}{p^2}\,(2l+1)\,\frac{(\Gamma/2)^2}{(E - E_0)^2 + (\Gamma/2)^2}. \tag{4.4.13}$$

More simply, we may rewrite this as

$$\sigma_l \;\simeq\; \sigma_{\max}\,\frac{(\Gamma/2)^2}{(E - E_0)^2 + (\Gamma/2)^2}. \tag{4.4.14}$$

For any production process that passes through a (well-isolated) resonant channel, one thus finds that the spectrum or cross-section takes the BW form shown in Fig. 4.16 (Breit and Wigner, 1936). Note that, in practice, the peak behaviour is superimposed over the energy dependence of $\sigma_{\max}$, which from Eq. (4.4.8) 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 cross-section for particles 1 and 2 scattering via formation of a resonance R may be expressed as

$$\sigma_R \;\simeq\; \frac{4\pi\hbar^2}{p^2}\,\frac{(2J+1)}{(2s_1+1)(2s_2+1)}\,\frac{(\Gamma/2)^2}{(E - E_R)^2 + (\Gamma/2)^2}, \tag{4.4.15}$$

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 possible multiplicity factors for any final-state spins are subsumed in the numerator Γ.

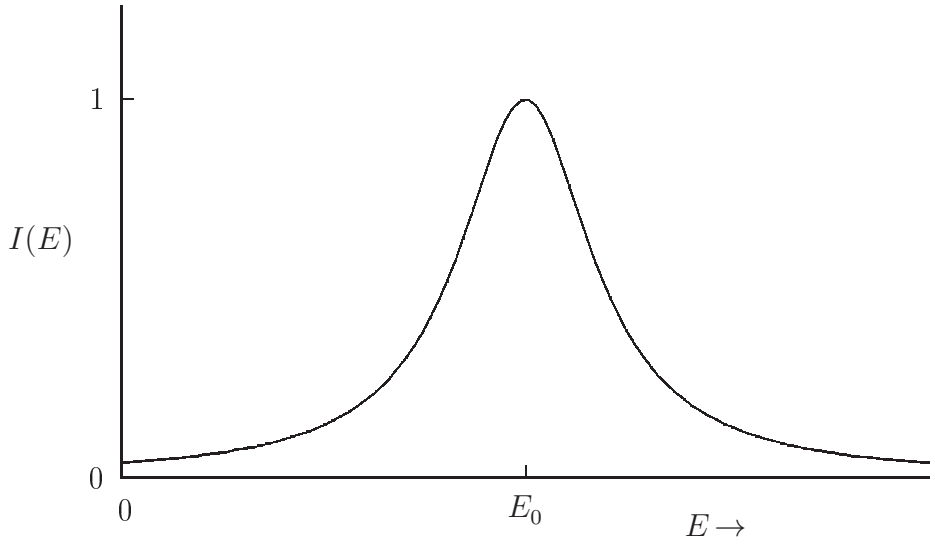


Figure 4.16: The standard Breit–Wigner resonance form; the underlying behaviour of $\sigma_{\max}$ has been factored out.

4.4.3 Breit–Wigner 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 lifetime (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:

$$\mathcal{P}(t) \propto \mathrm{e}^{-\Gamma t}. \tag{4.4.16}$$

In quantum mechanics this should simply be the squared modulus of the wave-function describing the state. A *plausible* form for the time dependence of the wave-function is then

$$\psi(t) \propto \mathrm{e}^{-\mathrm{i}Et}\,\mathrm{e}^{-\frac{1}{2}\Gamma t} = \mathrm{e}^{-\mathrm{i}\left(E-\mathrm{i}\frac{1}{2}\Gamma\right)t}. \tag{4.4.17}$$

Considering a particle of physical mass m_0 in its rest-frame, E may be replaced by m_0. Such a 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 final *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 $\Gamma \ll m_0$):

$$\frac{1}{p^2 - m^2} \rightarrow \frac{1}{p^2 - m_0^2 + \mathrm{i} m_0 \Gamma}. \tag{4.4.18}$$

One can show, by explicit calculation, that the effect of the self-interaction induced by the decay channel (*i.e.* the possibility of temporary spontaneous fluctuations into the decay final state) 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 $e^+e^- \rightarrow \gamma, Z^0 \rightarrow \mu^+\mu^-$ (see Fig. 4.17). At low-

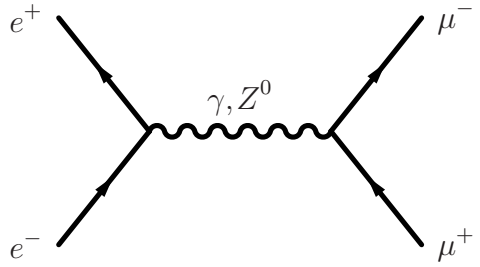


Figure 4.17: An example Feynman graph for the $e^+e^- \rightarrow \mu^+\mu^-$ scattering process. Time runs from left to right while the vertical axis schematically represents the three spatial coordinates.

est order in perturbation theory the propagator associated with the intermediate Z^0 state is normally:

$$\frac{1}{q^2 - m_Z^2}, \tag{4.4.19}$$

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:

$$\frac{1}{s - m_Z^2 + \mathrm{i} m_Z \Gamma_Z}, \tag{4.4.20}$$

where use has been made of the Mandelstam variable

$$s \equiv (p_1 + p_2)^2 \equiv q^2 \equiv E_{\mathrm{CM}}^2. \tag{4.4.21}$$

The interaction probability is proportional to the modulus squared of the amplitude and thus we should really examine

$$\left| \frac{1}{s - m_Z^2 + \mathrm{i} m_Z \Gamma_Z} \right|^2 = \frac{1}{(s - m_Z^2)^2 + m_Z^2 \Gamma_Z^2}. \tag{4.4.22}$$

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. (4.4.22) is relevant for relativistic field-theory calculations; noting that for energies near the *pole* mass ($E_{\rm CM}\sim m_Z$)

$$\begin{aligned} s-m_Z^2 &= E_{\rm CM}^2-m_Z^2 \\ &= (E_{\rm CM}+m_Z)(E_{\rm CM}-m_Z) \simeq 2m_Z\,(E_{\rm CM}-m_Z), \end{aligned} \tag{4.4.23}$$

one readily obtains the standard BW form:

$$\propto \frac{1}{(E_{\rm CM}-m_Z)^2+\frac{1}{4}\Gamma_Z^2}. \tag{4.4.24}$$

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 event-by-event true energy of the intermediate resonance and thus on the final-state momentum $\boldsymbol{p}$ (evaluated in the rest-frame). One can show, by explicit calculation, that for a resonance of nominal mass M the generally correct form is

$$\Gamma(s) = \left|\frac{\boldsymbol{p}}{\boldsymbol{p}_0}\right|^{2J+1}\Gamma(M^2), \tag{4.4.25}$$

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 J is the intrinsic spin of the resonance. The effect of this is 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 basis in a treatment of all processes as involving only elementary particles 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.

4.5 Bibliography

Breit, G. and Wigner, E.P. (1936), *Phys. Rev.* **49**, 519.

Fermi, E. (1927), *Rend. Acc. Naz. Lincei* **6**, 602.

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

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

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

Rayleigh (1871), *Phil. Mag.* **41**, 107; 274.

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

Sellmeier, W. (1871), *Annalen Phys.* **143**, 271; *Annalen Phys.* **145** (1872) 399; *Annalen Phys.* **145** (1872) 520; *Annalen Phys.* **147** (1873) 386; *Annalen Phys.* **147** (1873) 525.

Thomas, L.H. (1927), *Math. Proc. Cambridge Phil. Soc.* **23**, 542.

Chapter 5

Detectors and Accelerators

In this chapter we shall examine the more experimental side of nuclear and particle physics and take a look at particle detection and acceleration. These two areas have undergone tremendous development over the last few decades and a serious treatment would require a separate dedicated course (or courses). The intention here is thus merely to provide the student with a basic idea of the techniques used. We shall follow the historical development to some extent and try to motivate the particular choices made, as dictated by technical capabilities and experimental requirements of the day.

5.1 Particle detection

The aim of particle detection is to record information on the products of nuclear and subnuclear collisions. One can broadly divide the desired ends into two categories: *measurement* and *identification*. As we shall see though, the latter can only be achieved through the former. We shall therefore be mainly examining how one measures various quantities, such as: position, energy, momentum *etc.* Once again, the principal and simplest mechanism by which particles may interact and therefore be probed or detected is the electromagnetic interaction. Note though that in the case of hadrons the strong nuclear interaction also plays an important role while, of course, neutrinos may only detected through their weak interaction.

We shall thus be particularly interested in the interactions of charged particles with matter. The main effect of such interactions is a loss of energy through ionisation of the detector material and it is this process that is actually measured. In this section then we shall first examine the theoretical description of energy loss through ionisation, *i.e.* the celebrated Bethe–Bloch formula, and then go on to discuss the various types of detectors used in nuclear and particle physics.

As an aside, although we shall not touch on the subject here, it should not

be forgotten that an important and non-trivial area of study directly related to particle detection is that of the statistical analysis of data. Almost all nuclear- and particle-physics measurements are statistical in nature and therefore require a deep understanding of this aspect of data analysis.

5.1.1 Supplementary reading

The books by Kleinknecht (1990) and Knoll (1989) are particularly well suited for this part of the course. Another useful resource (though rather encyclopaedic) is *The Particle Detector BriefBook* by Bock and Vasilescu (1998), which can be found in html version at:

`http://www.cern.ch/Physics/ParticleDetector/BriefBook/`

A fairly complete coverage of the basics may also be found in Povh *et al.* (1995). While, as always, there is a specific up-to-date review article in the PDG publication (Patrignani *et al.*, 2016).

5.1.2 Particle detectors

A particle detector usually has a role in the measurement of energy, momentum and/or position of a particle passing through or being stopped inside it and may thus contribute to particle identification. The identification of a particle consists essentially of measuring any two of its energy, momentum and velocity, from which the standard energy–momentum or velocity relations (in the non-relativistic case $E = \frac{1}{2}mv^2 = p^2/2m$ and relativistically $E^2 - p^2c^2 = m^2c^4$ or $\gamma = E/m$) are sufficient to determine the mass. Provided the precision is sufficient, a mass determination usually identifies a particle unambiguously while particle and antiparticle may be distinguished via their opposite charges and therefore opposite deflection by electromagnetic fields.

The general manner in which particles interact with matter is via electromagnetic forces and, in particular, their ability to cause ionisation of the medium in which they are travelling. Indeed, it is the ionisation process that is mainly responsible for energy losses as particles pass through matter and the Bethe–Bloch formula (which we shall now study) describes such a process. Note though that for electrons and photons there are other mechanisms whereby they may lose energy and which may become more important according to the particular energy regime in question.

5.1.3 EM interaction of particles and matter

The interaction of charged particles with matter has already been studied to some extent in earlier chapters: the main underlying process is that generically described as Rutherford scattering, see Fig. 5.1. Assuming M to be sufficiently large, as too

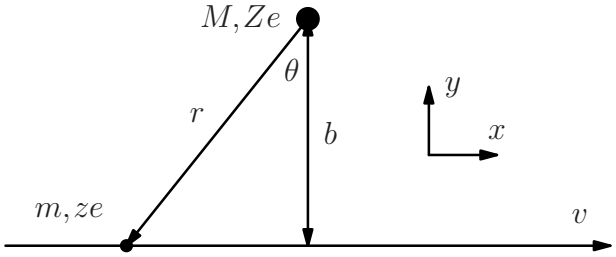


Figure 5.1: The interaction between a charged particle (mass m, charge ze) and a nucleus (mass M, charge Ze) in the impulse approximation.

the typical distance of closest approach, we may take the nucleus as stationary (formally, the approximation requires $t_{\rm int} v_{\rm recoil} \ll b$). For high-energy particles we may also assume that the deviation is very small and that the trajectory is approximately a straight line.

In this case the average force along the x-axis cancels and we are left with an instantaneous transverse force given by*

$$\begin{aligned} F_y &= \frac{zZ\alpha}{r^2}\cos\theta && \text{(neglecting a factor } \hbar c) \\ &= \frac{zZ\alpha}{b^2}\cos^3\theta && \text{(using } r = b/\cos\theta). \end{aligned} \tag{5.1.1}$$

Integrating with respect to time, we obtain the momentum transfer,

$$\Delta p_y = \int_{-\infty}^{\infty} {\rm d}t\, F_y. \tag{5.1.2}$$

Now, $x = b\tan\theta$ and therefore ${\rm d}x = b\sec^2\theta\,{\rm d}\theta$, from which we obtain

$${\rm d}t = \frac{{\rm d}x}{v} = \frac{b\,\sec^2\theta\,{\rm d}\theta}{v}. \tag{5.1.3}$$

* For clarity of the expressions, we shall henceforth set $\hbar = 1$ and $c = 1$.

And for the momentum transfer we therefore have

$$\begin{aligned}\Delta p_y &= \frac{zZ\alpha}{b^2}\int_{-\pi/2}^{\pi/2} \mathrm{d}\theta\, \cos^3\theta\, \frac{b}{v\cos\theta} \\ &= \frac{2zZ\alpha}{bv}. \end{aligned} \tag{5.1.4}$$

Note that although we have derived the above in the non-relativistic approximation, it actually still holds in the relativistic limit. The reason is simple: the direction of motion (relativistic or not) is orthogonal to the momentum transfer, which is thus not affected by a boost in the x direction. To write this in a more obviously covariant manner, we simply substitute β for v:

$$\Delta p_y = \frac{2zZ\alpha}{b\beta}. \tag{5.1.5}$$

It is often useful to reason in terms of a maximum or average force $\Delta F_y^{\max}$ applied for some limited period of time τ: $\Delta p_y = \Delta F_y^{\max}\tau$. Thus,

$$\Delta F_y^{\max} = \frac{zZ\alpha}{b^2}\gamma \qquad \Leftrightarrow \qquad \tau = \frac{2b}{\beta}\frac{1}{\gamma}, \tag{5.1.6}$$

where the factor γ in the first equation takes into account the effective relativistic mass increase and in the second the time-dilation effect.

If the target nucleus remains non-relativistic, the energy transfer is

$$\Delta E = \frac{(\Delta p)^2}{2M} = \left(\frac{2zZ\alpha}{b\beta}\right)^2 \frac{1}{2M} \propto \frac{Z^2}{M}. \tag{5.1.7}$$

Considering the passage of particles through matter made up of atoms containing A nucleons, of which Z protons, and therefore Z electrons, we may estimate the ratio of energy loss to electrons over that to nuclei:

$$\frac{\Delta E_{e^-}}{\Delta E_{\mathrm{nucl}}} \simeq \frac{Z\times 1^2/m_e}{1\times Z^2/Am_p} \simeq \frac{Z/m_e}{Z/2m_p} = \frac{2m_p}{m_e} \simeq 4000, \tag{5.1.8}$$

where we have used the approximation $A\sim 2Z$. We thus see that the major energy loss will be to atomic electrons:

$$\Delta E_{e^-} = \frac{2z^2\alpha^2}{b^2\beta^2 m_e}. \tag{5.1.9}$$

Let us now consider the energy loss on passing through some detector material:

$$P(b)\,\mathrm{d}b \;=\; n_e\,2\pi b\,\mathrm{d}b\,, \tag{5.1.10}$$

where we have used $P(b)=2\pi b$ per electron with n_e electrons per unit area. Taking the density of the material to be ρ, the thickness Δx, the mass number A (and Avogadro's number N_A), we have

$$n_e \;=\; \frac{Z\rho N_A \Delta x}{A}. \tag{5.1.11}$$

Using the above relation between ΔE and b, we thus obtain

$$P(\Delta E) \;=\; P(b)\,\frac{\mathrm{d}b}{\mathrm{d}\Delta E} \;=\; \frac{\pi Z\rho N_A \Delta x}{A}\,\frac{2z^2\alpha^2}{\beta^2 m_e}\,\frac{1}{\Delta E^2}. \tag{5.1.12}$$

We now need to integrate over ΔE to calculate the mean energy loss during the passage through a length Δx of the material:

$$\overline{\Delta E} \;=\; \int_{\Delta E_{\min}}^{\Delta E_{\max}} \mathrm{d}\Delta E\,\Delta E\,P(\Delta E) \;=\; \frac{\pi Z\rho N_A \Delta x}{A}\,\frac{2z^2\alpha^2}{\beta^2 m_e c^2}\,\ln\frac{\Delta E_{\max}}{\Delta E_{\min}}, \tag{5.1.13}$$

where, for completeness, we have reinstated the factors of c. It is common practice to gather related parameters into a single constant:

$$\begin{aligned} C \;&\equiv\; 2\pi N_A \left(\frac{\alpha}{m_e c^2}\right)^2 \\ &\simeq\; 0.3\;\left(\mathrm{g}\cdot\mathrm{cm}^{-2}\right)^{-1}. \end{aligned} \tag{5.1.14}$$

We thus have

$$\overline{\frac{\mathrm{d}E}{\mathrm{d}x}} \;=\; C\rho\,\frac{m_e}{\beta^2}\,\frac{z^2 Z}{A}\,\ln\frac{\Delta E_{\max}}{\Delta E_{\min}}, \tag{5.1.15}$$

which may be rewritten in terms of $b \propto 1/\sqrt{\Delta E}$ as

$$\overline{\frac{\mathrm{d}E}{\mathrm{d}x}} \;=\; C\rho\frac{m_e}{\beta^2}\,\frac{z^2 Z}{A}\,2\ln\frac{b_{\max}}{b_{\min}}. \tag{5.1.16}$$

The remaining task is to identify the limits of integration (either in ΔE or b).

$\Delta E_{\min}$ and $b_{\max}$: The lower energy limit is attained in the soft or adiabatic regime; *i.e.* when the forces are too weak to cause an electronic transistion. This effectively occurs when $\tau_{\rm int} > \tau_{\rm orbit}$, which leads to the following upper limit on the impact parameter b:

$$\frac{2b}{\beta\gamma} \lesssim \frac{h}{I_0}, \tag{5.1.17}$$

where $I_0 \sim 16\,\mathrm{Z}^{0.9}\,\mathrm{eV}$ is the mean ionising potential (for $Z > 1$).

$\Delta E_{\max}$ and $b_{\min}$: The upper energy limit is given by the lower limit on b, which arises owing to quantum mechanical effects (the uncertainty principle) discussed earlier. As usual, the effective resolution is limited by the wavelength of the photon exchanged and thus we have

$$b_{\min} \approx \frac{\hbar}{m_e\beta\gamma}. \tag{5.1.18}$$

The final form for the mean energy loss per unit distance traversed is then

$$\overline{\frac{\mathrm{d}E}{\mathrm{d}x}} = C\rho\,\frac{m_e c^2}{\beta^2}\,\frac{z^2 Z}{A}\,2\ln\frac{\pi\beta^2\gamma^2 m_e c^2}{I_0}, \tag{5.1.19}$$

where, for completeness, we have reinstated the factors of c. Note that although this is, by construction, only the mean value, (statistical) fluctuations are essentially negligible as the number of interactions is typically very large.

5.1.4 Bethe–Block formula

The first complete relativistic treatment is due to Bethe (1932), who obtained

$$\overline{\frac{\mathrm{d}E}{\mathrm{d}x}} = C\rho\frac{m_e c^2}{\beta^2}\,\frac{z^2 Z}{A}\,2\left[\ln\frac{2\beta^2\gamma^2 m_e c^2}{I_0} - \beta^2\right]. \tag{5.1.20}$$

Later Bloch (1933) further refined this and the standard form adopted today is

$$\overline{\frac{\mathrm{d}E}{\mathrm{d}x}} = C\rho\frac{m_e c^2}{\beta^2}\,\frac{z^2 Z}{A}\,2\left[\frac{1}{2}\ln\frac{2\beta^2\gamma^2 m_e c^2 W_{\max}}{I_0^2} - \beta^2 - \frac{\varepsilon}{Z} - \frac{\delta(\beta)}{2}\right], \tag{5.1.21}$$

where $W_{\max}$ is the maximum energy transfer kinematically permitted in a single collision between a particle of mass m and an electron:

$$W_{\max} = \frac{2\beta^2\gamma^2 m_e c^2}{1 + 2\gamma m_e/m + (m_e/m)^2}. \tag{5.1.22}$$

The two additional corrections present in the full Bethe–Block formula (5.1.21) are as follows.

ε/Z: This, the so-called "shell-correction" term, corrects the approximation that the velocity of the ionising particle is much larger than that of the target or struck electron. This contribution reduces the stopping power by up to 6 % (it is, however, often neglected).

$\delta(\beta)/2$: This, the so-called "density-effect" term, is a correction for polarisation or dielectric effects inside the medium, which again may reduce the stopping power since the effective electromagnetic fields of the ion will be reduced (or screened) by the dielectric effect. This is important for large b and therefore large γ and is only present in solid media (*i.e.* not in gases or liquids).

Exercise 5.1.1. *Derive formula* (5.1.22).

Note that in the literature there exist many ways of presenting the Bethe–Bloch formula (typically by redefining the constant C). However, it is clear that as far as the projectile parameters are concerned, it depends on both z and β but not m. For β small and neglecting the (slow) logarithmic dependence, it behaves as $1/\beta^2$ until it reaches a minimum for $p \approx 4mc$. It then grows logarithmically for relativistic velocities. For $\gamma \gtrsim 100$ (*i.e.* ultrarelativistic) the curve eventually more or less reaches a plateau, reflecting the density effect. At the minimum we have the following value:

$$\frac{1}{\rho}\frac{\mathrm{d}E}{\mathrm{d}x} \sim 1{-}2 \text{ MeV}/\left(\mathrm{g}\cdot\mathrm{cm}^{-2}\right), \tag{5.1.23}$$

where the units are defined in terms of typical so-called target thickness. In Fig. 5.2 a typical curve is shown for the range of validity. Note that for $\beta\gamma$ small, $\beta\gamma \to \beta$ while for $\beta\gamma$ large, $\beta\gamma \to \gamma$. Other materials have very similar curves, simply shifted up or down; *e.g.* the curve for lead (Pb) lies below that for copper (Cu).

5.1.5 Electrons, positrons and photons

For the lightest electromagnetically interacting particles (electrons, positrons and photons) there are other forms of energy loss. We shall first examine the cases of electrons and positrons, turning finally to photons.

Electrons and positrons

Owing to their particularly small mass, electrons and positrons may also lose energy via *bremsstrahlung*. That is, the rapid changes of velocity that they can

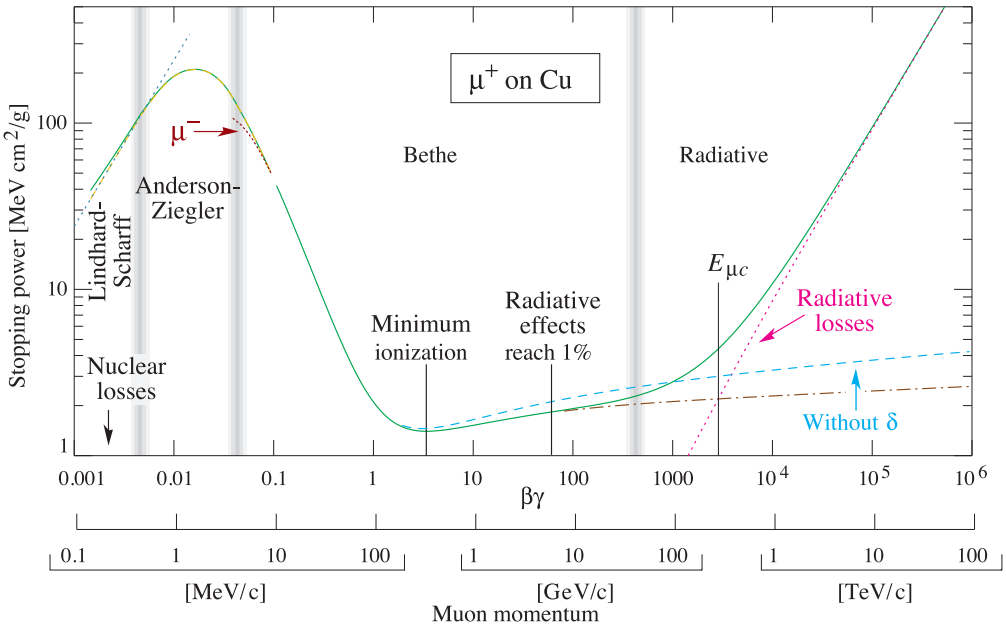


Figure 5.2: The Bethe–Bloch stopping power ($\rho^{-1}dE/dx$) for positive muons in copper as a function of $\beta\gamma \equiv p/mc$.

undergo in the presence of the electromagnetic fields of the nucleus provoke γ emission. The rate of energy loss of electrons and positrons is proportional to their energy and depends strongly on the type of material, in particular, it is roughly proportional to E and Z^2. This form of energy loss dominates over ionisation for $E \gtrsim 600\,\mathrm{MeV/Z}$. For $e^{\pm}$ it is then usual to define the *radiation length* X_0 such that:

$$E \propto \mathrm{e}^{-x/X_0} \qquad \text{since} \qquad \frac{\mathrm{d}E}{\mathrm{d}x} \propto E. \tag{5.1.24}$$

As a typical example radiation length, that of lead is 0.56 cm. It is more or less constant with energy and only depends on A and Z of the material. For $Z > 23$ to a fair approximation $X_0 \simeq 120\,\mathrm{Z}^{-2/3}\,\rho^{-1}$ (with the density ρ in $\mathrm{g/cm^3}$).*

Photons

For a single photon, energy loss is not a gradual or continuous process, since the interaction involves its complete absorption. At the level of radiation flux intensity, however, the attenuation is clearly apparently continuous since the quantum of radiation energy ($\hbar\omega$) is very small. One thus talks of an attenuation or absorption

* The radiation length X_0 is sometimes given in units of $\mathrm{g/cm^2}$, being defined as the X_0 used here multiplied by the density of the material.

coefficient μ:

$$I \propto \mathrm{e}^{-\mu x}, \tag{5.1.25}$$

where I is the radiation intensity. The coefficient μ depends on both the photon energy (or frequency) and on the absorbing medium.

There are three principal mechanisms whereby photons interact with matter. In order of growing importance with increasing energy they are: the photoelectric effect, the Compton effect* and (e^+e^-) pair production (also known as conversion).

Photoelectric effect: Here we are again dealing with an ionisation effect—on absorbing a photon an atomic electron is simply excited out of its orbit into the continuum. Such a mechanism is important for energies up to a few keV.

Compton effect: Compton scattering is the process whereby a real photon is absorbed and immediately re-emitted by an electron;† it is thus effectively an elastic two-body scattering process, see Fig. 5.3. The natural energy regime here

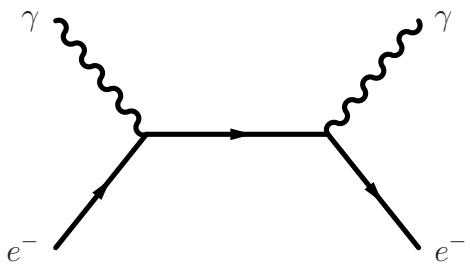


Figure 5.3: One of the Feynman graphs representing the Compton scattering process.

is that for which the photon wavelength is comparable with the so-called electron *Compton wavelength* $\lambda_{\mathrm{Compton}} \equiv h/m_e c$ and thus for energies up to the order of the electron rest mass 512 keV. It is generally only dominant for a small window of energies around 1 MeV.

Pair production: For energies above a few MeV, the dominant process becomes that of electron–positron pair production, also known as the Bethe–Heitler process (Bethe and Heitler, 1934), This is very closely related to the *bremsstrahlung* process mentioned above (see Fig. 5.4) and also occurs owing to electromagnetic

* The 1927 Nobel Prize in Physics was divided equally between Arthur Holly Compton "for his discovery of the effect named after him" and Charles Thomson Rees Wilson "for his method of making the paths of electrically charged particles visible by condensation of vapour."

† Note that, strictly speaking, the final-state photon may be emitted before absorption of the incident photon occurs.

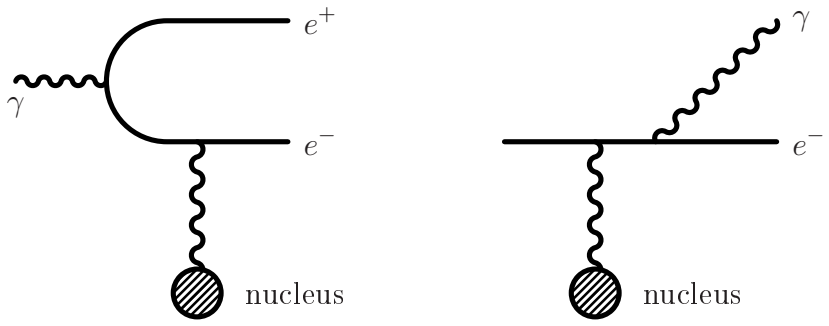


Figure 5.4: Examples of the Feynman graphs for the related Bethe–Heitler (left) and *bremsstrahlung* (right) processes (the blob represents a charged nucleus).

interaction with the charges (nuclei or electrons) in the medium. The resulting so-called *conversion length* is given by

$$\lambda = \tfrac{9}{7} X_0. \tag{5.1.26}$$

Note that such a process is possible only in the presence of electromagnetically interacting matter, owing to the requirement of energy–momentum conservation.

Exercise 5.1.2. *Calculate the threshold photon energy for electron–positron pair-production. Examine its dependence on the mass of the source of the electromagnetic field: consider the two case of a heavy nucleus and a single electron.*

A comparison of the cross-sections for these three dominant processes, together with some other less important contributions, is shown in Fig. 5.5. The total cross-section, calculated as the sum of all processes considered, coincides with the experimental curve, also shown. Note, in particular, that the Compton process is almost completely obscured by the other two dominant processes for heavier nuclei.

Vavilov–Čerenkov and transition radiation

A further important form of interaction of charged particles with matter is the so-called *Vavilov–Čerenkov*[*] effect (Čerenkov, 1934; Vavilov, 1934). A charged particle will cause radiation when its velocity exceeds that of light *inside* the medium it is traversing, or if it moves across the boundary between one medium and another with differing refractive indices (then known as *transition* radiation).

[*] The 1958 Nobel Prize in Physics was awarded jointly to Pavel Alekseyevich Čerenkov, Il´ja Mikhailovich Frank and Igor Yevgenyevich Tamm "for the discovery and the interpretation of the Čerenkov effect."

[*] Such an effect had already been predicted by Heaviside (1888, 1889).

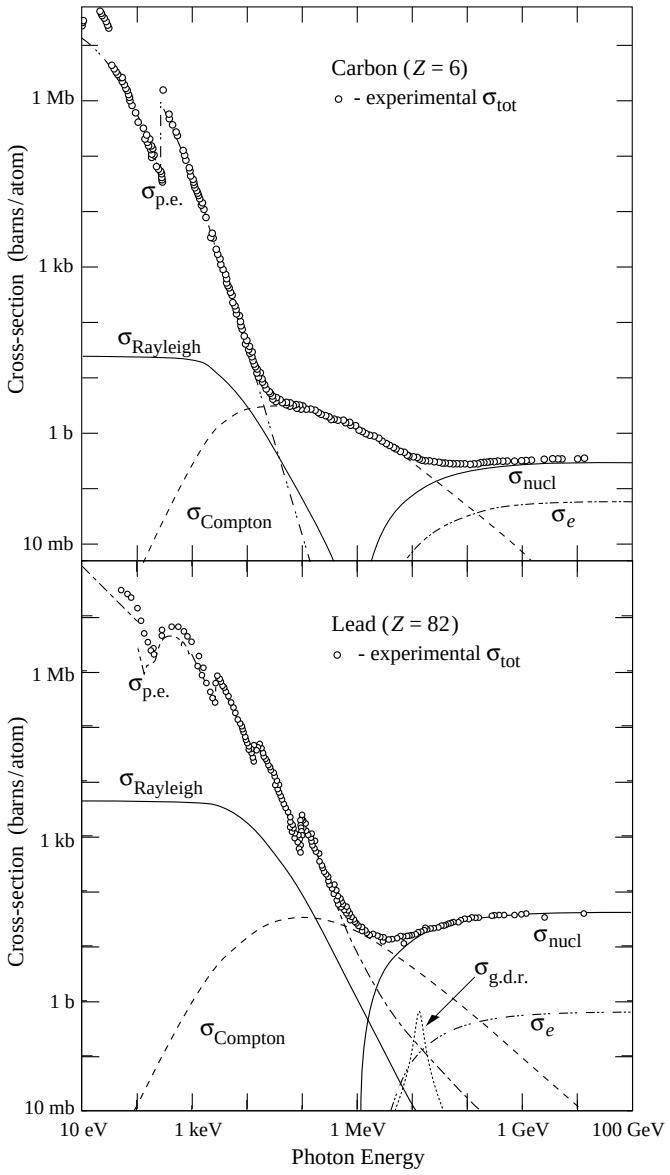


Figure 5.5: A comparison, as functions of photon energy, of contributions for $\sigma_{\text{p.e.}}$ the photoelectric effect, σ_{Compton} the Compton effect and $\sigma_{\text{nucl,e}}$ the Bethe–Heitler process on nuclei and electrons in carbon and lead. Other processes shown include Rayleigh scattering and giant dipole-resonance nuclear interactions. The experimental curve, also shown, coincides with the sum of all processes considered.

Although such processes are irrelevant for energy loss, they are used in high-energy physics experiments as a means to measure the particle velocity (other detectors are typically sensitive to either energy or momentum). Note that it is actually the

electrons of the medium that radiate under influence of the passage of the particle of interest.

Recall that the refractive index, $n \geq 1$, of a transparent medium determines the velocity of light inside the medium $c' = c/n$. If therefore $\beta > 1/n$, we have $\beta > c'$ *i.e.* the particle may be considered as superluminal while *inside* the medium. Čerenkov radiation is then the equivalent of the supersonic bang produced by an aircraft travelling at a speed in excess of that of sound or of the bow wave produced by a speedboat. The point is that the forward parts of successive spherical wave fronts lie outside the previous fronts, as depicted in Fig. 5.6, and thus the effective

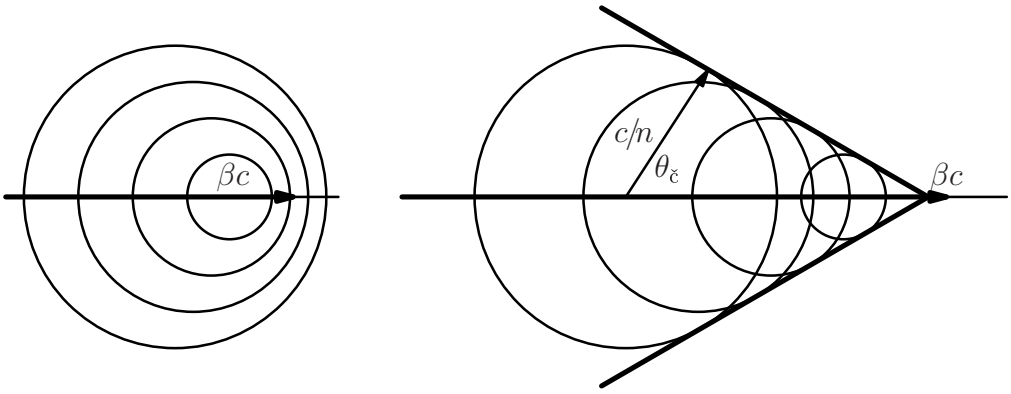


Figure 5.6: A schematic depiction of Čerenkov radiation and the wavefront angle in the ideal case of a non-dispersive medium. The left-hand figure shows the case of a slowly moving particle while that on the right represents the effect of an in-medium superluminal velocity.

wave front takes on a conical shape and is in effect amplified. The angle $\theta_{\check{c}}$ of propagation of the radiation with respect to the particle motion is given by*

$$\cos\theta_{\check{c}} \;=\; \frac{1}{\beta n}. \tag{5.1.27}$$

The threshold velocity $\beta_t = 1/n$ and therefore we have

$$\frac{p_t}{m} \;=\; \beta_t\gamma_t \;=\; \frac{1}{\sqrt{n^2-1}}. \tag{5.1.28}$$

Frank and Tamm (1937) derived the explicit formula for the spectrum of Čerenkov radiation. Expressed as the number of photons per unit path length $\mathrm{d}x$ for a particle of charge z and per unit frequency interval $\mathrm{d}\omega$ of the emitted photons,

* In a dispersive medium (a frequency-dependent refractive index) the situation is somewhat more complicated, as the wavefront is then not orthogonal to the direction of propagation.

the formula is

$$\frac{\mathrm{d}^2 N_\gamma}{\mathrm{d}x\,\mathrm{d}\omega} = \alpha z^2 \sin^2\theta_{\check{c}} = \alpha z^2 \left[1 - \frac{1}{\beta^2 n^2(\omega)}\right] \simeq 370\, z^2 \sin^2\theta_{\check{c}}\ \mathrm{eV}^{-1}\,\mathrm{cm}^{-1}. \tag{5.1.29}$$

Note that the integral over frequencies does not diverge for $\omega \to \infty$ as both β and $n(\omega) \to 1$ at high energies and the emission is therefore cut off.

In detector use there are essentially two possibilities: either simple threshold Čerenkov detectors or so-called differential Čerenkov detectors using mirrors. In the first case one simply exploits the threshold behaviour by varying the gas pressure to vary the refractive index. If simultaneously the particle momenta are measured (via some other method), one can then separate particles lighter and heavier than a certain mass threshold. Already in the 1950's Čerenkov counters were regularly used experimentally: *e.g.*, in the discovery of the antiproton* (Chamberlain *et al.*, 1955) and antineutron (Cork *et al.*, 1957).

In the so-called **R**ing **I**maging **Č**erenkov **H**odoscope (RICH) detector, first proposed by Séguinot and Ypsilantis (1977), the resulting cone of emitted light is collected on a position-sensitive planar photon detector. This allows the reconstruction of a ring or disc, the radius of which is related to the Čerenkov emission angle. In a focusing RICH detector the photons are collected by a spherical mirror, with focal length f, and focused onto a photon detector placed at the focal plane. The result is a circle of radius $r = f \sin^2\theta_{\check{c}}$, independently of the emission point along the particle track (see Fig. 5.7). Since this allows the collection of many photons, such a scheme is very useful for low refractive-index radiators (*e.g.*, gases), where a greater radiation length is required to create a sufficient number of photons.

Transition radiation: When a particle with charge z crosses the boundary between the vacuum and a medium with plasma frequency ω_p, so-called transition radiation is emitted. The energy radiated has intensity

$$I = \tfrac{1}{3}\,\alpha z^2 \gamma\, \hbar\omega_p, \tag{5.1.30}$$

where ω_p is the plasma frequency, given by

$$\hbar\omega_p = \sqrt{\frac{4\pi N_e \alpha}{m_e}} \tag{5.1.31}$$

* The 1959 Nobel Prize in Physics was awarded jointly to Emilio Gino Segrè and Owen Chamberlain "for their discovery of the antiproton."

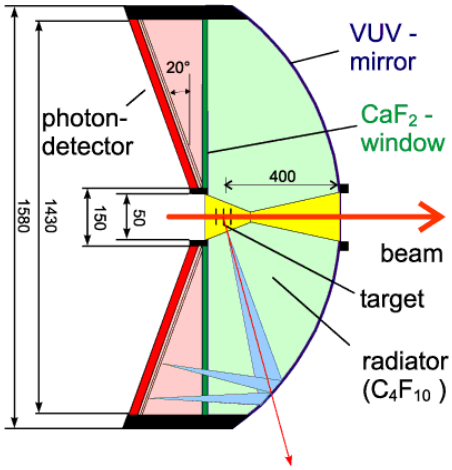


Figure 5.7: A Ring Imaging Čerenkov (RICH) detector (from the HADES experiment at the GSI in Darmstadt—the main purpose here is to identify e^+e^- pairs); dimensions are given in mm.

and N_e is the electron number density. This translates into

$$\hbar\omega_p \;=\; 28.81\sqrt{\rho\;\overline{Z/A}}\;\;\text{eV}, \tag{5.1.32}$$

where the material density ρ is taken in g/cm^3. Typically then, only electrons have sufficiently high velocities: $\gamma \gtrsim 1000$. The X-rays so produced have energies ranging from a few keV to a few tens of keV and are emitted principally within an angle $1/\gamma$ from the original direction (relativistic beaming).

5.1.6 Detectors

The earliest detectors aimed at producing and revealing tracks, *i.e.* for position (and thus momentum) measurement therefore. Ionising particles may leave tracks via various effects. The simplest form is a photographic plate or emulsion (often silver bromide): on development of the emulsion one obtains an image of the track, where the amount of localised "blackening",then reflects the rate of energy loss of the ionising particle at a given point. The discovery of the pion was made by Lattes *et al.* (1947) via the study of cosmic rays in photographic emulsions at altitude (in the Pyrenees).*

* The 1950 Nobel Prize in Physics was awarded to Cecil Powell "for his development of the photographic method of studying nuclear processes and his discoveries regarding mesons made with this method."

Cloud chambers

Later tracking devices exploited condensation or boiling in supersaturated materials (gases or liquids) triggered by the presence of charged ions. The first such detector is the cloud chamber, developed by C.T.R. Wilson* (invented in 1899 while the first functioning example dates back to 1911). It is essentially a sealed container filled with a supersaturated vapour of water or alcohol. When a charged particle traverses the vapour–liquid mixture, it causes ionisation and the resulting ions then act as seeds, around which water droplets will form since the mixture is beyond the condensation point.

The high energies of α- and β-particles cause tracks to be created, due to the many ions being produced along the trajectories of such charged particles. These tracks have distinctive shapes (*e.g.*, the track due to an α-particle is broad and shows evidence of deflection by collisions while that of an electron is thinner and straighter). If a uniform magnetic field is applied across the cloud chamber, positively and negatively charged particles will curve in opposite directions.

Cloud chambers played a prominent role in experimental particle physics from the 1920s up until the 1950s, when the bubble chamber (see later) came into preeminence. In particular, the discoveries of the positron (Anderson, 1933b) and the kaon (Rochester *et al.*, 1947) were both made with cloud chambers.

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, via 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 when a charged particle passes through supersaturated water vapour it provokes local condensation. The cloud chamber consists of a container fitted with a piston and 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 becomes supersaturated. At this point the chamber is photographed (possibly from two different angles so as to permit a stereo image) and any charged particle passing through it in that moment leaves a track of fine condensation droplets. The presence of a magnetic field reveals the sign of the charge.

In Fig. 5.8 we see one of Anderson's positron events. The track here enters from

* See footnote on page 127.

† 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."

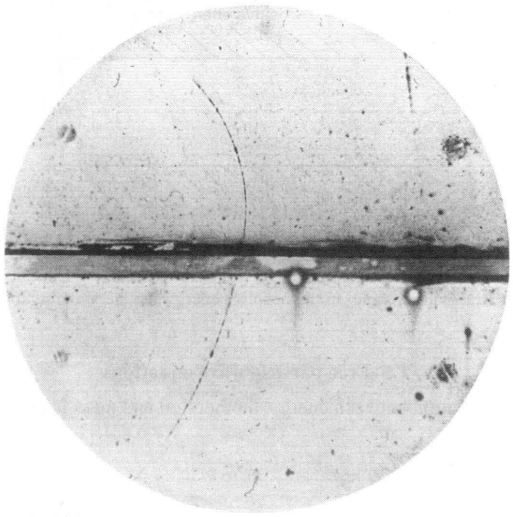


Figure 5.8: A Wilson cloud-chamber photograph showing the passage of a positron. The length of the track for a relatively low-energy particle allows one to distinguish between a positron and a proton. The figure is taken from Anderson (1933a).

the 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 down 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.*

Bubble chambers

A similar basic principle lies at the heart of the bubble chamber, which was invented in 1952 by Glaser.† ‡ A bubble chamber is a large spherical or cylindrical container filled with a transparent superheated liquid (usually hydrogen). The pressure (typically around ten atmospheres) is very suddenly lowered for a brief period (again, via rapid movement of a piston). An ionising particle traversing it

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

† The 1960 Nobel Prize in Physics was awarded to Donald A. Glaser "for the invention of the bubble chamber."

‡ The 1968 Nobel Prize in Physics was awarded to Luis Alvarez "for his decisive contributions to elementary particle physics, in particular the discovery of a large number of resonance states, made possible through his development of the technique of using hydrogen bubble chamber and data analysis."

at the same instant again creates ions, which now form the seeds for local boiling. Within a time scale of about 3 ms small bubbles are generated along the particle trajectory.* The track may be photographed from different angles to allow more or less full three-dimensional reconstruction. The time scale of the bubble formation permits a cycle frequency of about 10 Hz. However, the data analysis requires much longer times: track reconstruction is a complex procedure (see Fig. 5.9). The total number of reconstructed events per experiment is then typically at most

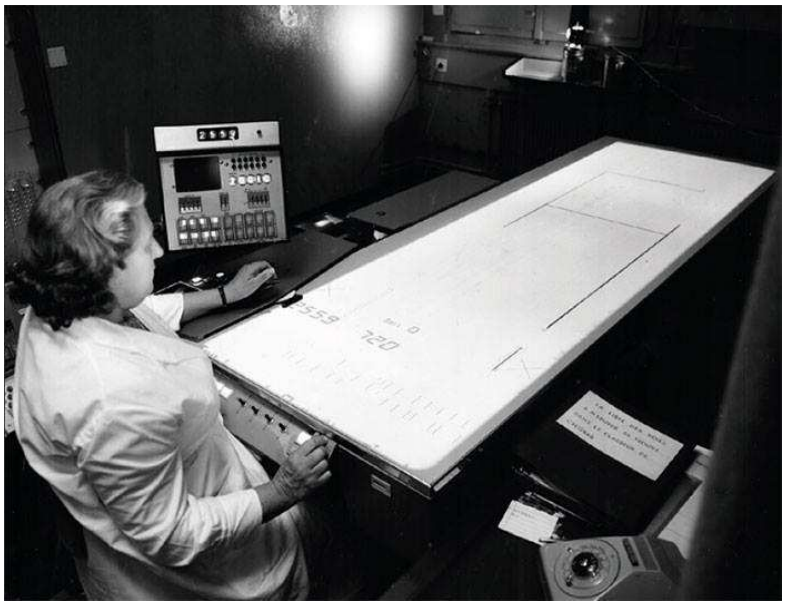


Figure 5.9: Studying bubble chamber film on a scanning table: the operator identifies events of interest and records details such as vertex position and track co-ordinates onto magnetic tape—figure reproduced from the CERN Courier, June 1973.

of the order of 10^5–10^6. A further limitation of this type of detector is the fact that, owing to its large volume (of the order of several cubic metres), it must also double as the target. That is, its use imposes fixed, hydrogen-target experiments.

Ionisation chamber

When gas in a chamber between electrodes, held at a significant potential difference, is ionised by the passage of, *e.g.*, α-particles, β-particles, X-rays, γ-rays or other sources ionisation, the ions and dissociated electrons so created move towards the electrodes of opposite polarity, thus creating an ionisation current, which may

* It is often recounted that Glaser was inspired by the bubbles in a glass of beer. However, in a 2006 talk he denied this; He did note though that while beer may not have been the inspiration for the bubble chamber, he did in fact perform early experiments using beer to fill prototypes.

be measured by a galvanometer or electrometer: the accumulated charge is simply proportional to the number of ions created and thus to the energy deposited. Typical average values for the ionising energies per pair are 40 eV for helium and 26 eV for argon.*

A voltage potential, which may range from a few volts to many kilovolts, depending on the particular application, is applied between the electrodes. The applied voltage also allows the device to work continuously by removing free electrons and preventing the device from becoming saturated. The current so-generated is called the bias current, and prevents the device from reaching the point where no more ions can be collected.

As such, these types of counters have little practical use in nuclear and particle physics. However, an example application is for smoke detection. In a smoke detector, the gap between the electrode plates is exposed to the ambient atmosphere. The chamber contains a small amount of americium($^{241}_{95}$Am, half-life 432.2 yr), which is an α-emitter. The α-particles carry a sizable amount of energy, and when they collide with gas in the ionisation chamber (atmospheric air—mostly nitrogen and oxygen) the momentum transferred ionises the gas molecules; *i.e.*, the neutral gas molecules will lose one or more electrons and become charged ions.

Since the plates are at different voltages (in a typical smoke detector the voltage difference is a few volts), both the ions and electrons will be attracted to the plates. This small but continuous flow of ions between the plates represents a measurable electric current. If smoke enters the detector, it disrupts the current because the ions collide with the smoke particles and are thus neutralised. The consequent fall in current triggers the alarm.

Geiger–Müller counter

The precursor of many types of modern ionisation radiation detectors is the Geiger–Müller counter. We now recall briefly the basic workings of this device: it consists of a tube filled with an inert or noble gas (such as helium, neon or argon) held at low pressure ($\sim$0.1 atm). The walls of the container are typically coated with a conductor to form the cathode while the anode is a wire running along the centre of the tube. These are then held at a potential difference of several hundred volts.

The passage of ionising radiation or particles, as usual, produces positively charged ions, and negative electrons. The electric field generated by the electrodes accelerates the ions towards the cathode and the electrons towards the anode. For sufficiently high voltage, the resulting acceleration then causes further (cascade) ionisation through collisions. As the voltage increases the number of avalanches

* Note that the mean ionising energy is always rather larger than the first ionising energy (*i.e.* the energy required to remove the first electron), which, *e.g.*, for argon is 15.7 eV.

created increases and so too the anode current until the saturation point (or so-called Geiger plateau) is reached. Indeed, Geiger–Müller counters are operated at the Geiger plateau and so the output is a short, intense pulse of current, which is then measured or simply counted, the rate being proportional to the radiation intensity.

Most tubes can detect α, β and γ radiation. Moreover, a neutron detector may be constructed either by coating the tube with boron or by adding boron trifluoride or ^{3}He to the gas. Neutrons interact with the boron nuclei to produce α-particles, or with the ^{3}He to produce hydrogen and tritium ions plus electrons; the charged particles so emitted trigger the usual avalanche process.

Proportional counters

A proportional counter is again a type of gaseous ionisation detector. Its operation is similar to that of a Geiger–Müller counter. However, it operates at a lower voltage (below the Geiger plateau) and is therefore capable of energy measurement. An inert gas (such as argon or xenon) is used to fill the tube, but a quench gas is added to avoid spurious extra discharges. A typical mixture is 90 % argon and 10 % methane. Again a sufficiently energetic incoming ionising particle, excites electrons from their atomic orbits, thus producing a positive ion and a free electron (often called an ion pair). A charged particle traversing the chamber leaves a track of such ion pairs. The electrons so created then drift towards the anode, while the positive ions drift towards the cathode, much more slowly; the electrode currents can then be read out. Typical drift times are of the order of microseconds and milliseconds, respectively.

In a proportional counter the operating voltage is high enough to accelerate the drifting electrons to an energy whereby they too become ionising particles and thus each one may generate a so-called *Townsend avalanche*. By choosing an appropriate operating voltage, these avalanches are independent and therefore the total charge created is proportional to the energy released in the initial ionising collision. Therefore, by measuring the total charge (the time integral of the measured current) collected at the electrodes, the kinetic energy of the initiating particle may be determined.

In fact the geometry of the electrodes and potential differences are such that the electric field is not, in general, high enough to produce a Townsend avalanche. Only around the anode is there a strong enough field, which then causes avalanche amplification to occur, seeded by each single original electron and the multiplication factor (up to approximately 10^6) is thus independent of where it was produced.

Multiwire and drift chambers

A multiwire proportional chamber (MWPC), or more simply wire chamber, is an improvement over the simple single-wire proportional chamber described above, which allows simultaneous combined determination of both energy and position.*

By measuring the arrival time of the current pulses in the wires and taking into account that the ions need some time to drift to the nearest wire, the distance at which the particle passed the wire may be deduced. This greatly increases the accuracy of the path reconstruction and the device is then known as a *drift chamber*. The accuracy achievable is around 50 μm. By arranging two drift chambers with the wires placed orthogonally (in the plane normal to the beam direction) the position may be detected with greater precision.

A further simple detector may be used to detect the particle at a fixed distance from the wire planes, a tri-dimensional reconstruction can then be performed. This allows the velocity of the particle to be calculated from the difference in time of the passage of the particle in the different parts of the detector. Such a device is known as a time projection chamber (TPC), see Fig. 5.10. An

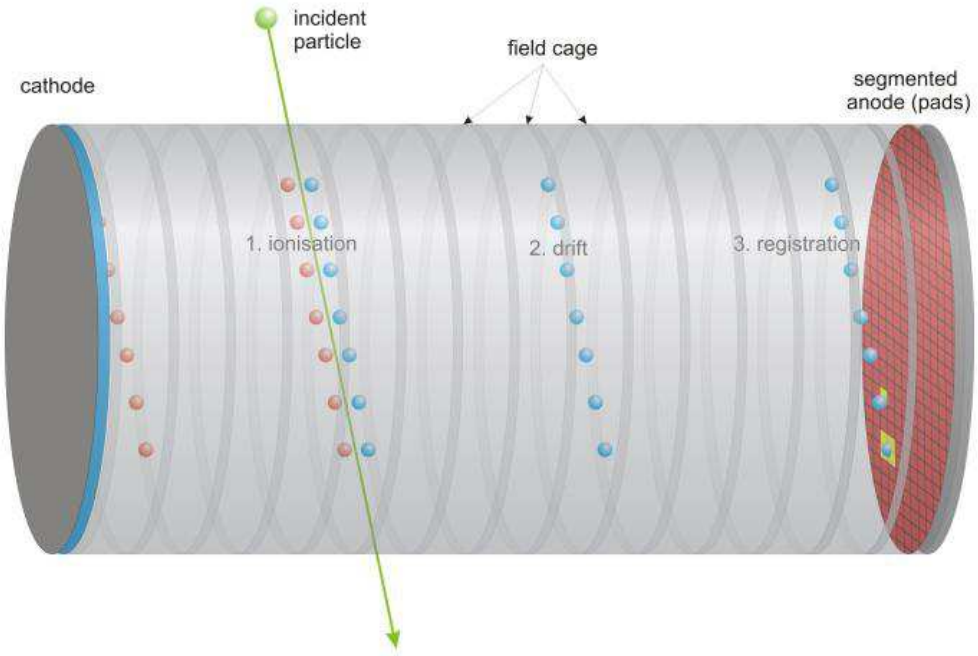


Figure 5.10: A schematic view of a typical time projection chamber (TPC).

example is the TPC used in ALEPH (one of the experiments conducted in the the

* The 1992 Nobel Prize in Physics was awarded to Georges Charpak "for his invention and development of particle detectors, in particular the multiwire proportional chamber."

Large Electron–Positron Collider (LEP) ring at the Centre Europée de Rechèrche Nucleaire (CERN—in Geneva). This was a very large volume cylindrical detector: 2.2 m in length and 1.8 m radius. An electric field of strength 100 kV/m drives the ions and electrons towards MWPC's at the ends of the chamber, which then provide measurement of the x and y position; the z position being inferred from the drift-time measurement—this requires an outer detector. The resolution there was of order 60 μm. A 1.5 T magnetic field allows determination of the particle's charge and momentum while also helping to avoid dispersion.

Silicon-strip detectors

Starting in the early 1980s detector devices were developed based on semiconductors (usually silicon or germanium) for the detection of charged particles or photons. In such detectors, the passage of ionising radiation creates free electron–hole pairs. As in the previously discussed detectors, the application of an electric field causes the electrons and holes to travel towards the electrodes, where they appear as electrical pulses to be measured. The minimum energy required to create an electron–hole pair is known, depends only on the material and is independent of the incident-radiation energy (*e.g.* for silicon it is 3.62 eV and for germanium 2.98 eV, *cf.* argon 26 eV). The number of such pairs is therefore proportional to the energy deposited by the radiation in the semiconductor and so measurement of the number of pairs allows the energy of the incident radiation to be determined.

The energy required to produce an electron–hole pair is rather lower than the energy required for ion–electron pairs in the previously discussed gas detectors. In particular, this means that fluctuations in the pulse heights are reduced and that the energy resolution is better. The electron and hole velocities are also very high and so the time resolution is also very good. Moreover, as compared to gaseous ionisation chambers, the density of the material in a semiconductor detector is very high and they may therefore be much smaller in size.

The principle of most silicon particle detectors is to dope narrow silicon strips or crystals (having width of order 100 μm), thus forming diodes. If these are then reverse biased, when charged particles pass through the strips, small ionisation currents are generated, which can then be detected and measured. Typically then, thousands of such detectors are arranged around the interaction point to give an accurate picture of the paths followed by the final-state particles produced. In particular, this is then used to reconstruct decay vertices and thus infer the nature of the initial particle. For example, the decay $D^+ \to K^+K^+\pi^-$ produces three (charged) tracks, all pointing back to the same vertex.

Charge-coupled devices

The charge-coupled device (CCD) was first proposed and constructed in 1970 at the Bell Laboratories.*

It usually consists essentially of some 30 k to 200 k potential wells, in which charges created by the passage of photons (or other ionising particles) can be trapped and then manipulated. Such devices may thus be read out electronically, providing precise, two-dimensional, positional information. Indeed, the great advantage is the creation of a true two-dimensional image and not the product of two one-dimensional projections (with consequent degradation of the information), as in all other modern detectors.

The area of such an object is of order of $1\,\mathrm{cm}^2$, with a thickness of around $20\,\mu\mathrm{m}$, while the individual pixels are about $30\,\mu\mathrm{m}$ square. They presently offer a resolution of around $5\,\mu\mathrm{m}$, but future capabilities may achieve even down to $1{-}2\,\mu\mathrm{m}$ precision.

A particular example of their use is in lifetime measurement: the NA32 experiment of the ACCMOR collaboration at CERN performed a measurement of the Ξ_c^0 lifetime (a baryon consisting of ssc quarks): $\tau \simeq 0.8 \times 10^{-13}\,\mathrm{s}$, which is probably not achievable with other techniques.

Momentum measurement

The momentum $\boldsymbol{p}$ of a particle may be deduced from the curvature of the track in a magnetic field $\boldsymbol{B}$. If $\boldsymbol{p}$ and $\boldsymbol{B}$ are orthogonal then we have

$$p \;=\; QBR, \tag{5.1.33}$$

where Q is the charge of the particle and R is the radius of curvature of the track. If there is an angle $\theta \neq 90^\circ$ between $\boldsymbol{p}$ and $\boldsymbol{B}$ then the particle follows a spiral path and $p^\perp = p\sin\theta = QBR$.

What is typically measured are the two parameters L (chord length) and s (sagitta) pertaining to an arc of the trajectory, see Fig. 5.11. From the geometry, we then trivially have

$$R^2 \;=\; (R-s)^2 + (L/2)^2. \tag{5.1.34}$$

If, as is usually the case, $s \ll L$ then we may approximate the solution for R:

$$R \;=\; \frac{L^2 + 4s^2}{8s} \;\simeq\; \frac{L^2}{8s}, \tag{5.1.35}$$

* One half of the 2009 Nobel Prize in Physics was awarded jointly to Willard S. Boyle and George E. Smith "for the invention of an imaging semiconductor circuit – the CCD sensor."

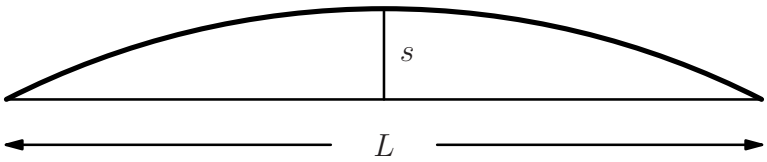


Figure 5.11: The parameters L (chord length) and s (sagitta) of an arc, providing a determination of R (radius): $R^2=(R-s)^2+(L/2)^2$.

from which we have

$$p_\perp \simeq \frac{L^2QB}{8s}. \tag{5.1.36}$$

To study the question of the measurement error, consider

$$\left|\frac{\mathrm{d}p_\perp}{\mathrm{d}s}\right| \simeq \frac{L^2QB}{8s^2} = \frac{p_\perp}{s}. \tag{5.1.37}$$

We thus have that

$$\frac{\delta p_\perp}{p_\perp} \simeq \frac{\delta s}{s}; \tag{5.1.38}$$

that is, the absolute error in $p_\perp$ increases with $p_\perp$. Note, however, that since $p_\perp$ is *inversely* proportional to s the fractional error also grows, as we shall now show. Assuming δs to be constant, as s becomes smaller ($p_\perp$ becomes large) there will clearly come a point for which $s=\delta s$ and thus $\delta p_\perp = p_\perp$; this will then be the maximum measurable value of $p_\perp$:

$$p_\perp^{\max} \simeq \frac{L^2QB}{8\delta s} \tag{5.1.39}$$

and therefore

$$\frac{\delta p_\perp}{p_\perp} \simeq \frac{p_\perp}{p_\perp^{\max}}. \tag{5.1.40}$$

Energy measurement—calorimetry

Now, while from the preceding discussion it should be clear that, momentum measurements do not significantly alter the values being measured (the magnitude of the momentum measured is essentially constant), the same is not true for typical true energy measurements. Indeed, calorimetry requires significant degradation of the original energy. This requires that calorimeters be the outer or final stage of any chain of detectors. The only exception is muon detection (see later).

Note here that the (statistical) precision improves with growing energy, simply because the number of photons produced increases proportionally with energy.

Note also that energy measurement is possible for neutral particles (*e.g.*, photons).

Electromagnetic calorimeters: Each single electromagnetic interaction with the calorimeter material leads to a *shower* of electrons and positrons: the ionising interaction creates a free electron, which then produces photons, which in turn produce e^+e^- pairs and these may continue the process. The shower will stop when the single photon energy falls below the pair-production threshold. Both the number of charged particles and the length of the shower track are proportional to the energy of the primary.

The length scale for the showering process (and thus too for the calorimeter itself) is set by the radiation length X_0 of the material (recall that the photon conversion length is given by $\lambda = \frac{9}{7}X_0$). Therefore, in order to limit the dimensions of the object, materials of low X_0 (*i.e.*, high Z) and density are required; typical depths used are 15–30 radiation lengths. There are two principal types of electromagnetic calorimeter.

Homogeneous: Such a calorimeter consists of a single material, which must then act as both a shower generator and detector. There are then two possible modes of functioning:

Čerenkov emission: The standard material is lead glass (55 % Pb and 45 % SiO_2), for which the radiation length is $X_0 = 2.36\,\mathrm{cm}$. The energy resolution is given by $\delta E/E \simeq 5/\sqrt{\mathrm{E}}\,\%$ (with E in GeV).

Scintillation: The standard material is NaI crystals (doped with Tl), for which the radiation length is $X_0 = 2.6\,\mathrm{cm}$. The energy resolution is given by $\delta E/E \simeq 1.5/\mathrm{E}^{1/4}\,\%$ (E in GeV).

Heterogeneous or sampling: In this case the detector takes on a sandwich form of convertor layers of a high-Z material, such as lead, interleaved with detector layers, which may be:

- scintillators (usually plastic),
- ionisation chambers (*e.g.*, liquid argon),
- proportional counters.

The typical resolution is $\delta E/E \simeq 5\text{–}15/\sqrt{\mathrm{E}}\,\%$ (E in GeV).

Hadronic calorimeters: Protons, neutrons, pions *etc.* interact strongly with nuclear matter and thereby create showers similar to the electromagnetic case. However, the uncertainties here are much larger as up to 30 % of the energy is lost through low-energy, slow or weakly interacting particles that "escape" detection. Nuclear interaction lengths are, moreover, rather longer: *e.g.*, for iron $X_{\mathrm{nucl}} \simeq 17\,\mathrm{cm}$

and they are therefore always of the heterogeneous type and typical depths are 15−30 nuclear interaction lengths.

Muon detectors

Muons are highly penetrating particles: they are much heavier than electrons but do not have the strong interactions of pions, kaons and protons. One therefore simply assumes that in a large detector such as those in ATLAS or CMS a particle that has traversed that large amount of intermediate material (in the hadronic and electromagnetic calorimeters) must be a muon. At the Large Hadron Collider (LHC) the muon detectors are tracking detectors (*e.g.* wire chambers) that represent the outermost layer of the experimental setups. They are, of course, not only sensitive to muons, but to any charged particle that may have reached them (including *e.g.* cosmic radiation products).

Particle identification

Partial but useful information is provided by a knowledge of the charge, which may be obtained by examining the effect of a magnetic field. Pulse shapes may also help somewhat. However, the only method that can uniquely determine the particle identity is to measure the mass, though this is not at all easy.

For suitable lifetimes (neither too short nor too long), vertex reconstruction with energy and momentum detection of the final-state particles allows the invariant mass of the final-state system to be calculated and this gives directly the mass of the decaying object.

Typically, though a set of complementary measurements are made: firstly to distinguish between electrons, photons and hadrons; secondly, in the case of hadrons, to distinguish between those few (π, K and p) that may actually reach the detectors before decay. For these, the method depends on the energy:

- for $E \lesssim 1\,\mathrm{GeV}$ time-of-flight allows sufficiently accurate velocity measurement,
- for $E \lesssim 100\,\mathrm{GeV}$ Čerenkov detectors are the best means to measure velocity.
- for $\gamma \gtrsim 1000$ one has to adopt transition radiation detectors, again for velocity measurement.

These combined with an energy measurement can then provide a reliable estimate of the mass.

5.2 Particle acceleration

5.2.1 Supplementary reading

In the case of accelerators, a modern and fairly complete but technical treatment may be found in the book by Lee (1999) although very good introductions are provided by Conte and MacKay (1991) and by Wilson (2001). As always, the basics can also be found in Povh *et al.* (1995). A complete list of past and present accelerator colliders in the world, together with a technical discussion, may be found in the PDG compendium Patrignani *et al.* (2016).

5.2.2 Particle accelerators

Before embarking on our discussion of *man-made* accelerators, let us first recall that there also exist *natural* sources of high-energy particles. One can basically identify two: the galactic electromagnetic fields and radioactive nuclei. The enormous electromagnetic fields present in the galaxies are capable of producing tremendously high-energy particles, ranging from protons up to nuclei of nickel. These are mainly stable nuclei: $\sim 89\,\%$ hydrogen (p), $\sim 10\,\%$ helium (α-particles), and $\sim 1\,\%$ heavier elements. Such nuclei then constitute $\sim 99\,\%$ of the so-called "cosmic radiation" while electrons constitute most of the remaining $1\,\%$.

The energies of these nuclei lie mainly between 100 MeV and 10 GeV; the highest energy recorded for cosmic particles is of the order of 3×10^{11} GeV. Such an energy is, of course, much higher than anything that may (or probably will ever) be produced in the laboratory. Unfortunately, such events are also exceedingly rare and thus of minimal use experimentally. They can however, for example, be used to estimate the proton–proton total cross-section for very high (and otherwise inaccessible) energies and thus check the predictive power of models for hadronic interactions. Radioactivity, on the other hand, can lead to quite intense production but is, of course, limited to rather lower energies. If we include in this category nuclear power plants, then a typical use in modern physics has been for producing high-intensity beams of low-energy neutrinos.

It is thus clear that, in order to obtain intense beams of high-energy particles, some form of artificial acceleration is needed. The only interaction that can be harnessed in the laboratory for acceleration purposes is the electromagnetic force and thus only charged particles can be considered. Moreover, if the beam is to have an appreciable lifetime, then the particles should be stable. This narrows the field down to electrons, protons and their respective antiparticles, together with heavier ions.

Note though that secondary beams of various types have been produced and used to great effect. Here, however, we wish to concentrate on the particular issue

of acceleration and shall thus consider only the electron (and positron) and the proton and antiproton; the case of heavy ions is not so different to that of the proton. The major difference between the electron and the proton is, of course, their rest mass and, as we shall see, this has important implications for acceleration to high energies.

It should be mentioned (though only in passing) that there are now also projects investigating the possibility of producing and using very high-energy colliding muon beams. For a given energy, synchrotron radiation (one of the main energy limitations in circular colliders) is proportional to the inverse fourth power of the mass and therefore, since $m_e/m_\mu \simeq 1/207$, there is a relative suppression factor for a muon machine of about 0.5×10^{-9}. Although the muon is unstable, its lifetime of 1.6 μs allows it to complete approximately $150\,B$ revolutions in a ring of which about half contains bending magnets with an average field of B tesla.

Modern particle accelerators constitute the largest and by far the most expensive instruments used by physicists. They can all be described in terms of three basic elements:

1. a source of charged particles,
2. a vacuum tube to permit collision-free motion,
3. the accelerator mechanism itself.

An electrostatic field is sufficient to accelerate charged particles. For example, by creating a large potential difference between the ends of an evacuated tube; Cockcroft and Walton (1932) first accelerated protons to 250 keV. In a similar manner but exploiting the same principles as the *Van de Graaff generator*, in the early thirties Van De Graaff himself designed an accelerator that builds up a potential between two electrodes via a moving belt. Van de Graaff accelerators can accelerate particles up to energies of around 15 MeV.

Electrostatic accelerators

The Cockcroft–Walton accelerator: In 1932 Cockcroft and Walton (graduate students working under the supervision of Rutherford) achieved 400 keV with protons, which were then made to collide with lithium nuclei ($^7_3$Li), forming two helium nuclei ($^4_2$He) as the products (see Fig. 5.12).* This is, indeed, the first example of man-made or induced fission (the energy release is 17 MeV). In order to reach such energies, since the highest voltage typically available then was only about 300 kV, Cockcroft and Walton had to design a new type of high-voltage

* The 1951 Nobel Prize in Physics was awarded jointly to Sir John Douglas Cockcroft and Ernest Thomas Sinton Walton "for their pioneer work on the transmutation of atomic nuclei by artificially accelerated atomic particles."

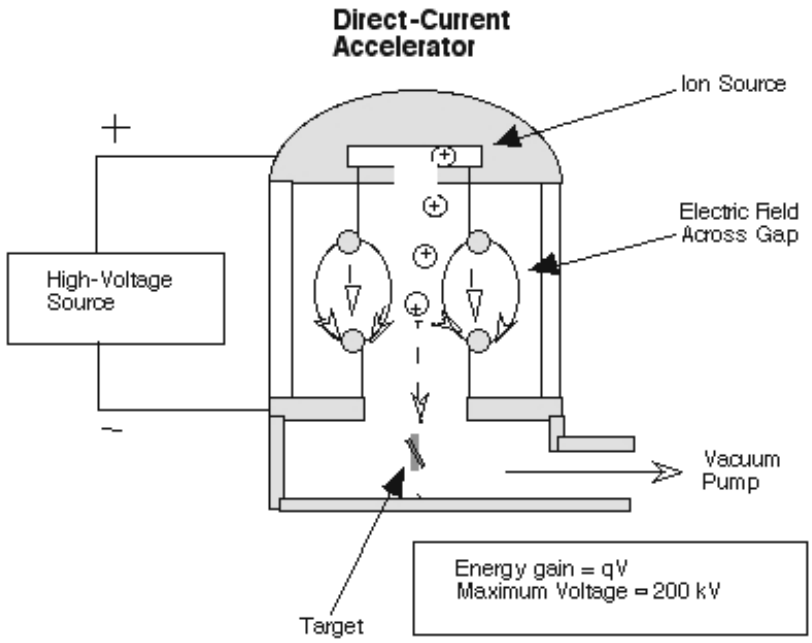


Figure 5.12: A schematic sectional view of an early Cockcroft–Walton accelerator (the construction has a cylindrical symmetry about the vertical axis).

generator or multiplier (see Fig. 5.13). The upper limit to the energy that may be

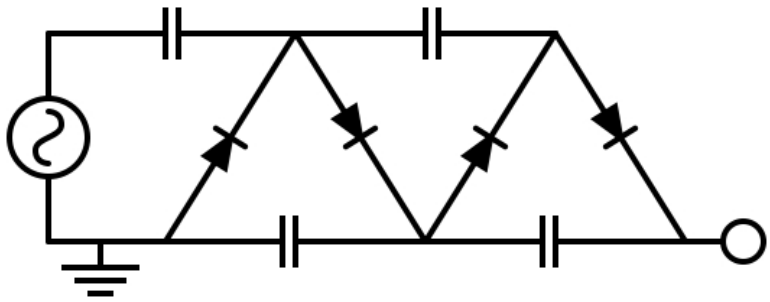


Figure 5.13: A circuit diagram for the Cockcroft–Walton voltage multiplier.

transferred via this method is about 1 MV; for such voltages air begins to conduct. This type of very basic machine is still used today as a pre-accelerator, particularly for ions, to more powerful machines.

The Van de Graaff and tandem accelerators: By means of a rotating belt a large electrostatic charge and thus (via the relation $V = Q/C$) high voltage may be built up, usually on a metallic sphere. This is the basis of the *Van De Graaff* generator (patent filed in 1931). By placing the apparatus inside a tank filled with an insulating gas (such as SF_6)the rapid discharge that normally takes place in air can be avoided. In this way potential differences of up to around 15 MV may be produced. Using such a system as part of an accelerator naturally results in, for example, 15 MeV protons (see Fig. 5.14).

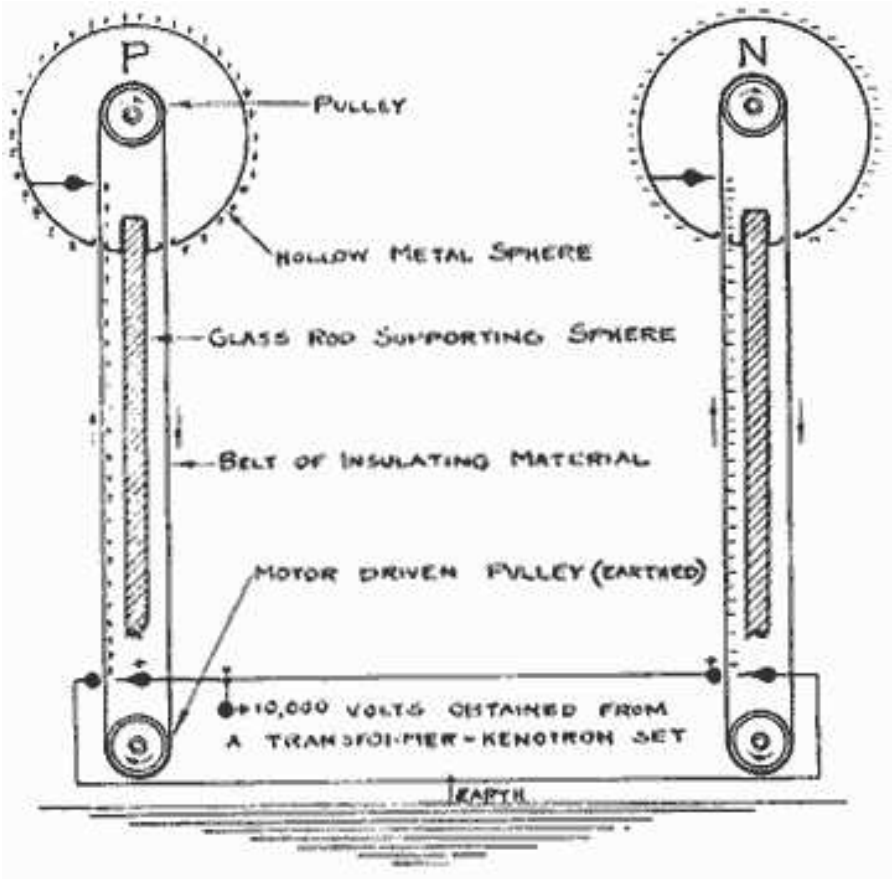


Figure 5.14: The Van de Graaff generator.

The final energy may be doubled by using a back-to-back arrangement, known as the *tandem* Van de Graaff accelerator (see Fig. 5.15). The idea, first applied in the 1950s, is to initially accelerate negative ions entering through a standard beam pipe towards a positively charged terminal. At this point, by placing a thin foil across the beam pipe inside the terminal (where the maximum energy has already been reached), electrons will be stripped from the ions thus rendering them

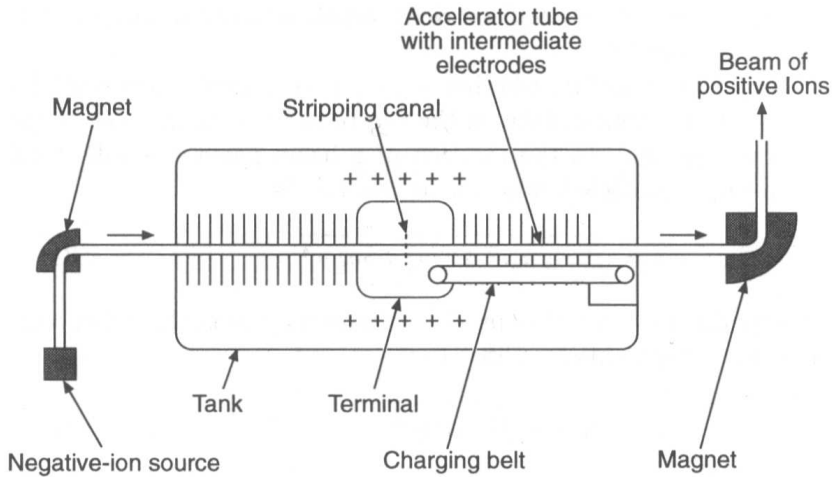


Figure 5.15: The tandem Van de Graaff accelerator.

positive. On leaving the terminal they are then repelled by the positive charge and again accelerated; a final energy of around 30 MeV may thus be attained. Of course, if the ions are multiply charged then the energy will be correspondingly higher.

A problem with the tandem accelerator is the need for negative ions, which, being difficult to produce, are available only for a limited number of elements whereas positive ion sources exist for a larger variety of nuclei. Moreover, positive ions may be produced with charges greater than one, which is however the maximum for negative ion sources.

On the other hand, modern tandem machines can reliably produce continuous currents of up to 100 μA and as such are useful as basic accelerators for nuclear physics. The energies attainable are sufficient for precision study of nuclear reactions and spectroscopy.

Linear accelerators

An obvious improvement, in retrospect, of the single accelerator is to place a number of them in series. In 1928 Widerøe demonstrated that electrons may be accelerated through an evacuated cylinder by applying a radio frequency (RF) alternating voltage to separated sections of a cylinder so that the electrons always see an accelerating electric field as they pass from one section to another. The frequency is determined by the velocity, tube length and separation in order that the electrons arrive at each stage in phase with the RF voltage, they are then accelerated on every passage. This is the basic principle of the linear accelerator, or *linac* (see Fig. 5.16). In order that the time between tubes remain constant,

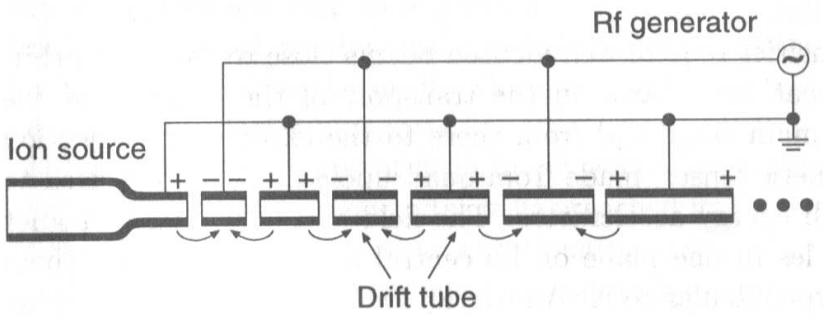


Figure 5.16: A schematic view of the Widerøe linac.

the length of the sections (or *drift tubes*) must progressively increase to match the increasing velocity of the accelerating particles.

A further improvement on this so-called π-mode structure is the so-called Alvarez 2π-mode accelerator. In this form, by exploiting electromagnetic inductance in a cavity, the free charges on the drift tubes are made to oscillate back and forth; it is then these that, simultaneously, produce the same field between each pair of tubes. A big advantage of such a system is that no current actually flows between the separate tubes themselves.

The modern-day linear particle accelerator is essentially the extension of the Widerøe scheme to a long linear array of accelerating "cells" powered by a RF source operating in the GHz frequency range and providing megawatts of power. In addition to varying the length of successive cells to account for the increasing particle velocity, there are improvements obtained by matching the relative phase of the electromagnetic wave and the particle velocity in the accelerator. Thus, particles that lag behind the main bunch should receive a bigger "kick" to allow then to catch up while those that drift ahead should receive less acceleration and thus fall back. The form of the waveguide cavity in the accelerator is thus such that the phase velocity of the electromagnetic wave will match the particle velocity only at those points in which acceleration occurs.

In principle, a linac of any length, and therefore any energy, can be built. Note that as $v \to c$, the length of the accelerating sections becomes constant, thus simplifying the construction of very high-energy linacs. The largest linac in the world, at SLAC, is 3.2 km long, with around 100 000 stages (see Fig. 5.17). It has the capacity to accelerate both electrons and positrons to energies of around 50 GeV. The SLAC linac was originally designed to collide a beam of electrons against a fixed target. It was then modified to accelerate both electrons and positrons along the same beam line. The electrons and positrons were then transported around opposite semicircular arcs to meet in head-on collisions. Such a scheme provides sufficient centre–of–mass energy to produce the Z^0 boson when an e^-e^+

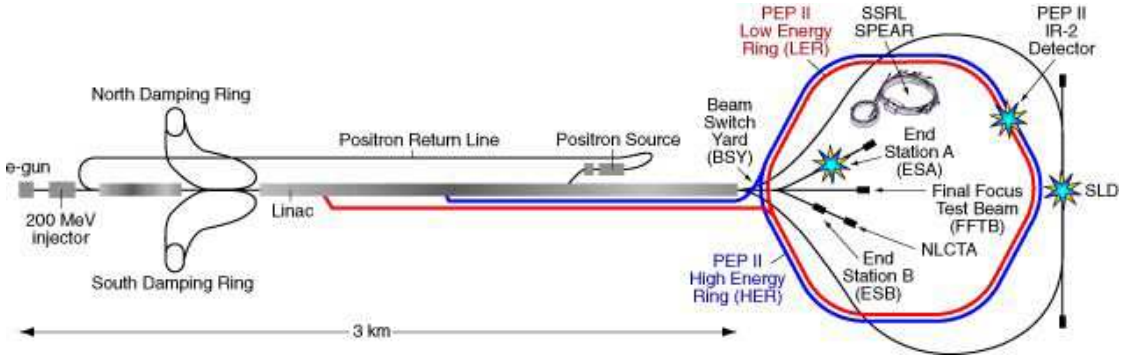


Figure 5.17: The SLAC linac.

pair annihilate.

The cyclotron

In 1932 Lawrence (Lawrence *et al.*, 1931, 1932) developed and first operated the *cyclotron*, the first circular accelerator, with which 1.25 MeV was achieved for protons. * The cyclotron may be considered as being effectively a linac rolled up into a spiral. Such a machine consists of two D-shaped vacuum chambers (called D's or *dees*) placed back-to-back. A magnetic field, produced by a large electromagnet, maintains the particles in circular motion. At each passage through the gap between the D's, a charged particle is accelerated. From the Lorentz force on an ion with charge Z, one has

$$R = \frac{p}{ZeB}, \tag{5.2.1}$$

where R is the radius of the path followed, p the particle momentum and B the magnetic field applied. As a particle gains energy, it therefore spirals outwards until it has sufficient energy to finally leave the accelerator.

Exercise 5.2.1. *Explain why the particles are accelerated as they move across the gap but move at constant radius in either* D.

Exercise 5.2.2. *Show that the maximum velocity a proton can reach in a* D *of radius R and magnetic field strength B is given by (ignoring relativistic effects)*

$$v_{max} = \frac{BeR}{m_p}.$$

* The 1939 Nobel Prize in Physics was awarded to Ernest Lawrence "for the invention and development of the cyclotron and for results obtained with it, especially with regard to artificial radioactive elements."

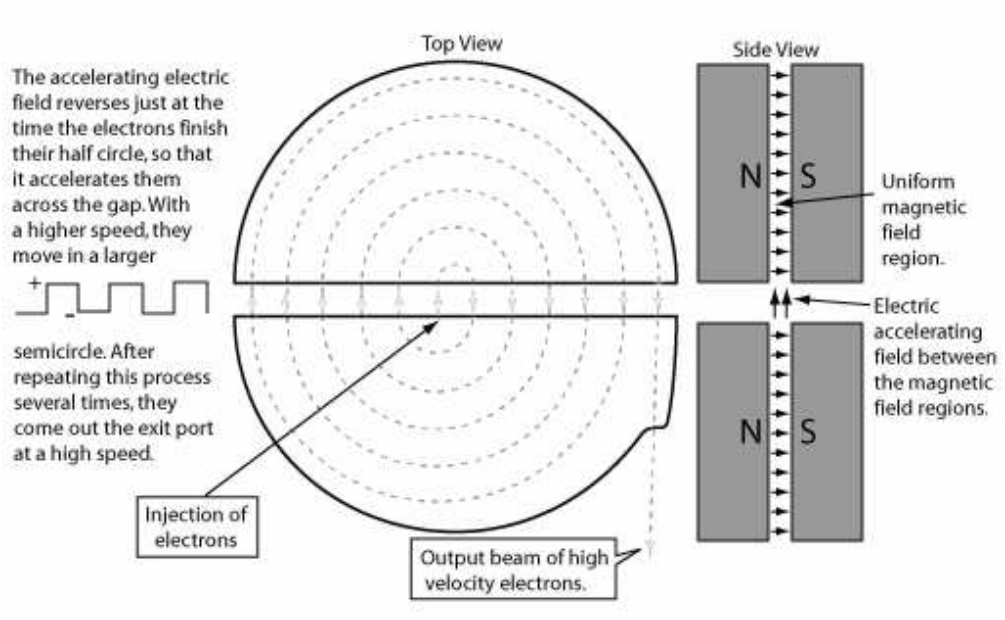


Figure 5.18: The cyclotron: (left) top view, (right) side view.

Exercise 5.2.3. *Evaluate the previous velocity limit for protons in a cyclotron of* 1.20 m *diameter and magnetic field strength* 0.50 T. *Hence, show that the frequency of the alternating potential difference must be* 7.61 MHz.

The cyclotron frequency is given by

$$\omega_c = \frac{v}{R} = \frac{\beta c}{R} = \frac{pc^2}{RE}, \tag{5.2.2}$$

which, combined with the expression in Eq. (5.2.1), then leads to

$$\omega_c = \frac{eB}{\gamma m}. \tag{5.2.3}$$

The RF must be an integer multiple of this quantity. For non-relativistic particles $\gamma \sim 1$ and the frequency is approximately constant. For relativistic particles it is necessary to vary ω_{RF} and/or B in order to remain "in tune" with the particle motion. One way to simplify this procedure for relativistic particles is to arrange for $\Delta E = m$, so that $\Delta\gamma = 1$ for each revolution. Thus, during the n-th revolution one has a cyclotron frequency of

$$\omega_n = \frac{eB}{nm}. \tag{5.2.4}$$

Note that for highly relativistic particles $\beta \sim 1$ and the frequency becomes approximately constant, but the magnetic field must still be varied.

When heavier particles in a cyclotron gain an energy of more than a few MeV, their mass grows appreciably, according to the relativistic effect. This causes a corresponding reduction in their acceleration and they therefore fall out of phase with the acceleration pulses in the gap between the D's. The *synchrocyclotron* (sometimes also called the frequency-modulated cyclotron) solves the problem. In this apparatus, the frequency of the RF generator accelerating the particles is automatically adjusted to stay in tune with the motion; as the particles gain mass, the frequency of accelerations is gradually lowered. Note that as the maximum energy of a synchrocyclotron increases, for fixed maximum magnetic field, so must its size, since the radius of curvature increases.

The betatron

When electrons are accelerated, they undergo a large increase in *inertial* mass already at a relatively low energies. Thus, for a total energy of 1 MeV, an electron has twice as much mass as an electron at rest. Synchrocyclotrons cannot be adapted to make allowance for such large increases in mass.

Exercise 5.2.4. *Using the relativistic formula for a particle mass, calculate the velocity of an electron that has inertial mass twice its rest mass.*

Therefore, another type of cyclic accelerator, the *betatron* (initially developed in 1935 by Steenbeck and improved into a working version by Kerst in 1940), was invented to accelerate electrons (Kerst, 1940, 1941; Kerst *et al.*, 1941). The betatron consists of an evacuated toroidal chamber, placed between the poles of an electromagnet. The electrons are kept in a circular path by a magnetic field called a guide field, see Fig. 5.19. By applying an alternating current to the electromagnet, the electromotive forceinduced by the changing magnetic flux through the circular orbit accelerates the electrons. Faraday's law states that the electromotive force $\mathcal{E}$ induced by a changing magnetic flux Φ_B through a surface $\mathcal{S}$ enclosed by a path $\mathcal{C}$ is given by

$$\oint_{\mathcal{C}} \boldsymbol{\mathcal{E}} \cdot d\boldsymbol{s} \;=\; \dot{\Phi}_B, \tag{5.2.5}$$

where

$$\Phi \;=\; \oint_{\mathcal{S}} \boldsymbol{B} \cdot d\boldsymbol{S}. \tag{5.2.6}$$

During operation, both the guide field and the magnetic flux are varied to keep the radius of the orbit of the electrons constant. It turns out that the field strength at the radius of the orbit must be exactly half the average field linking the same orbit. As the magnetic field increases the change provokes acceleration, the faster

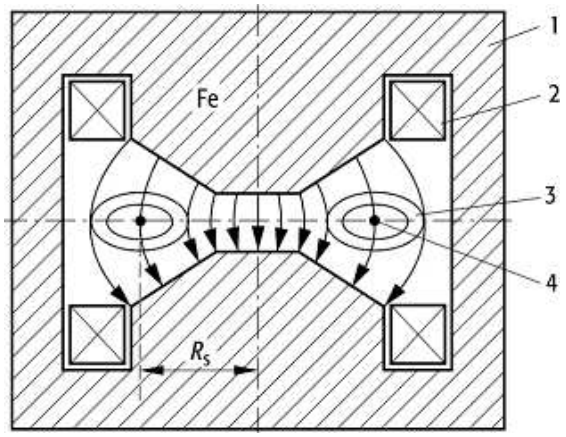


Figure 5.19: A cross-section view of a betatron: 1) magnetic material (iron), 2) coils, 3) toroidal beam pipe and 4) electron beam.

electrons then require a larger magnetic field to maintain the same orbit. Since the variations are linear, the same field can do both jobs.

In 1949 such a machine was built and used in Chicago. The essential parameters were: $R \sim 1.2\,\mathrm{m}$, $B \sim 1.8\,\mathrm{T}$ (the magnet weighed around 275 tons!), $I \sim 13\,\mathrm{A}$ and the energy reached was 315 MeV.

The synchrotron

The *synchrotron* is the most recent and most powerful member of the accelerator family. A synchrotron consists of a vacuum tube in the shape of a large ring through which the particles travel; the tube is surrounded by magnets to keep the particles in circular motion inside the tube, see Fig. 5.20. The particles only enter the tube after already having been pre-accelerated to several MeV in a smaller machine. Particles are accelerated at one or more points (usually placed symmetrically) along the ring during their orbits around the accelerator. To maintain the particles in circular orbits, the strength of the magnets in the ring must be increased as the particles gain energy. In only a few seconds, the particles reach energies greater than 1 GeV and may produce a variety of elementary particles when struck by the accelerated particles. The synchrotron principle can be applied to either protons or electrons, although most of the very large machines are proton synchrotrons.

Exercise 5.2.5. *Show that the radius of curvature of the path of particles of momentum p and charge Q in a synchrotron is given by the formula*

$$R = \frac{p}{QB}, \tag{5.2.7}$$

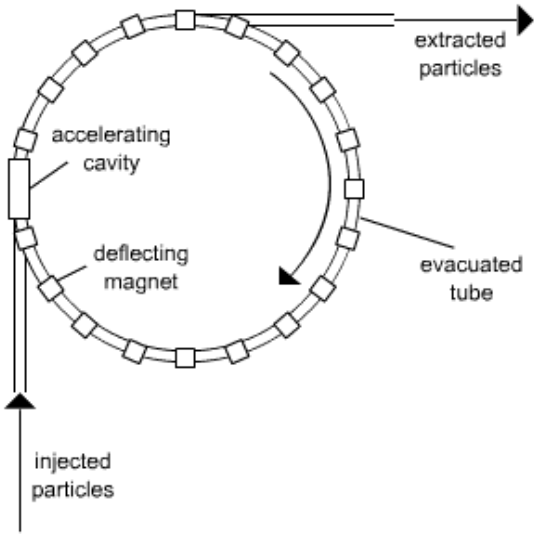


Figure 5.20: A schematic view of a synchrotron.

where B is the magnetic field strength.

Exercise 5.2.6. *A synchrotron is composed of four curved sections having radius R connected by four straight sections each of length L. If the period of the RF oscillator corresponds to exactly one revolution, show that*

(a) *the velocity of the particles is*

$$v = (2\pi R + 4L)f \tag{5.2.8}$$

(b) *and, by considering the relativistic momentum of particles of mass m, that the magnetic field strength of the synchrotron is given by*

$$B = \frac{m}{eR}(2\pi R + 4L)\, f \left[1 - \frac{(2\pi R + 4L)^2 f^2}{c^2}\right]^{-1/2}. \tag{5.2.9}$$

Note that in synchrotrons a computer is used to maintain this relation between magnetic field and oscillator frequency.

A serious limitation to the energy achievable in such a design is due to what is known as *synchrotron radiation*. This is none other than the better known *bremsstrahlung*; however, it is not caused by the linear acceleration of the particles, rather by the centripetal acceleration implicit in the circular motion. For relativ-

istic particles one has an energy loss per revolution given by

$$\Delta E \;=\; -\frac{4\pi}{3}\,\frac{\alpha}{R}\,\beta^3\,\gamma^4, \tag{5.2.10}$$

where $\beta = v/c$ is just the velocity and $\gamma = E/m$ is the usual Lorentz factor; that is, the energy loss for relativistic particles is proportional to E^4. Note too that it therefore varies as m^{-4}; for a given energy, the loss for electrons is thus roughly a factor 10^{13} larger than it is protons. The highest beam energy reached in an electron synchrotron is about 100 GeV (at LEP) while the planned final energy of the protons in the LHC is about 7 TeV.

For protons then synchrotron radiation is not the major difficulty. High energies and momenta require large magnetic fields. At the Tevatron and LHC the fields in the bending magnets are 4.4 T and 8.3 T respectively, both use superconducting coils to avoid excessive energy loss due to the high currents necessary and therefore also require very low temperatures.

Storage rings

A *storage-ring* collider accelerator is a synchrotron that produces more energetic collisions between particles than a conventional synchrotron, which only collides accelerated particles with a fixed target. A storage-ring collider accelerates two sets of particles in rotation in opposite directions in the ring and then allows these two sets to collide. CERN's LEP and LHC are examples of storage-ring colliders. In 1987, the Fermi National Accelerator Laboratory—near Chicago converted the Tevatron into a storage-ring collider and installed a three-storey-high detector that observed and measured the products of such head-on particle collisions. In this machine the centre–of–mass energy was sufficient to produce a small number of *top* quark–antiquark pairs.

In the early eighties, the two largest proton–antiproton synchrotrons were a 500-GeV ring at CERN and a similar device at the Fermi National Accelerator Laboratory. The capacity of the latter, called the Tevatron, was increased to a potential 1 TeV in 1987 by installing superconducting magnets, making it the most powerful accelerator in the world. In 1989, CERN began operating the LEP, a 27-km ring that accelerated electrons and positrons to an energy of 50 GeV (the LEP II stage reached 100 GeV). This machine was taken out of operation at the end of 2000 to make way for the construction (in the same tunnel) of the LHC. This last (and final in the series) is designed to reach proton beam energies of around 7 TeV (in heavy-ion operation it will provide 2.8 TeV per nucleon) and came into operation in the year 2008. We might also mention the Superconducting Super-Collider proposal in the U.S.A. in the nineties, which would have brought into collision protons with energies of up to 20 TeV each. The project was unfortunately

abandoned as too expensive although a large amount of civil-engineering work had already been undertaken.

Exercise 5.2.7. *Compare the centre–of–mass energies available at the Tevatron for fixed-target (assume a proton target) and beam–beam collisions.*

5.3 Bibliography

Anderson, C.D. (1933a), *Phys. Rev.* **43**, 491.

Anderson, C.D. (1933b), *Science* **77**, 432; *Phys. Rev.* 491.

Bethe, H. (1932), *Z. Phys.* **76**, 293.

Bethe, H. and Heitler, W. (1934), *Proc. Royal Soc. (London)* **A146**, 83.

Blackett, P.M.S. and Occhialini, G.P.S. (1933), *Proc. Royal Soc. (London)* **A139**, 699.

Bloch, F. (1933), *Annalen Phys.* **16**, 285; *Z. Phys.* 363.

Bock, R.K. and Vasilescu, A. (1998), *The Particle Detector BriefBook* (World Sci.).

Čerenkov, P.A. (1934), *Dokl. Akad. Nauk. Ser. Fiz.* **2**, 451.

Chamberlain, O., Segrè, E., Wiegand, C. and Ypsilantis, T. (1955), *Phys. Rev.* **100**, 947.

Cockcroft, J.D. and Walton, E.T.S. (1932), *Proc. Royal Soc. (London)* **A136**, 619.

Conte, M. and MacKay, W.W. (1991), *An Introduction to the Physics of Particle Accelerators* (World Sci.).

Cork, B., Lambertson, G.R., Piccioni, O. and Wenzel, W.A. (1957), *Phys. Rev.* **104**, 1193.

Frank, I.M. and Tamm, I. (1937), *C. R. Acad. Sci. URSS* **14**, 109.

Glaser, D.A. (1952), *Phys. Rev.* **87**, 665.

Heaviside, O. (1888), *The Electrician*, 23; 83; 147.

Heaviside, O. (1889), *Phil. Mag.* **27**, 324.

Kerst, D.W. (1940), *Phys. Rev.* **58**, 841.

Kerst, D.W. (1941), *Phys. Rev.* **60**, 47.

Kerst, D.W. and Serber, R. (1941), *Phys. Rev.* **60**, 53.

Kleinknecht, K. (1990), *Detectors for Particle Radiation* (Cambridge U. Press).

Knoll, G.F. (1989), *Radiation Detection and Measurement* (John Wiley & Sons).

Lattes, C.M.G., Muirhead, H., Occhialini, G.P.S. and Powell, C.F. (1947), *Nature* **159**, 694.

Lawrence, E.O. and Livingston, M.S. (1931), *Phys. Rev.* **38**, 834.

Lawrence, E.O. and Livingston, M.S. (1932), *Phys. Rev.* **40**, 19.

Lee, S.Y. (1999), *Accelerator Physics* (World Sci.).

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

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

Rochester, G.D. and Butler, C.C. (1947), *Nature* **160**, 855.

Séguinot, J. and Ypsilantis, T. (1977), *Nucl. Instrum. and Meth.* **142**, 377.

Van De Graaff, R.J. (1931), in proc. of the *Schenectady Meeting* (Schenectady, Sept. 1931), ed. American Phys. Soc.; *Phys. Rev.* **38**, 1919.

Vavilov, S. (1934), *Dokl. Akad. Nauk. Ser. Fiz.* **8**, 457.

Widerøe, R. (1928), *Arch. Elektrotech.* **21** [4], 387.

Wilson, E. (2001), *An Introduction to Particle Accelerators* (Clarendon Press).

Chapter 6

Nuclear Decay

A naturally occurring process that allows the study of the nuclear potential and interactions is that of decay. While we do not, of course, have any control over the kinematic variables or quantum numbers involved, the large number of different decay processes observed in Nature to some extent mitigates such a limitation. The study of the various types of decays then provides us with information on the nuclear structure and interaction and allows us to improve our description of the nucleus.

Besides the case of fission (examined earlier), there are three classes of nuclear decay: α, β and γ; these involve, respectively, the emission of a ^{4}He nucleus, an electron–neutrino pair and a photon. The fundamental interactions responsible are, respectively, the strong nuclear force, the weak nuclear interaction and electromagnetism. In this chapter we shall deal in some detail with first two while the latter will only be mentioned briefly here. Note that in all cases while (for historical reasons) one often talks of "radiation", this is only strictly correct for the last.

The emission of γ-rays, which clearly cannot alter the basic nature of the nucleus, simply implies transitions between an excited state and another or the ground state of a given nucleus. Such excited states are then the single nucleon levels available above the Fermi level. Typically, they will involve protons, via its charge, but may also, in principle involve neutrons, via its large anomalous magnetic moment (recall though that magnetic interactions are relatively suppressed by a factor v/c). The basic information that the spectroscopic study of γ-transitions then supplies is on the available nuclear energy levels in much the same way as in atomic physics. We shall thus not pursue the question further.

On the other hand, α-decay opens a window onto the strong nuclear potential and, as we shall see, also provides an example of the tunnel effect at work in real physics. The β-decay process introduces an entirely new force, the so-called weak nuclear (or more simply weak) interaction. This is an amazingly rich area of

nuclear and particle physics, leading as it does, to the discovery of the neutrino, parity violation and many other surprising and unexpected phenomena.

6.1 Alpha decay

As noted earlier, the binding energy of the ^{4}He nucleus (or α-particle) is particularly large (it is a doubly magic nucleus):

$$B \;=\; 28.3 \text{ MeV}, \tag{6.1.1}$$

or slightly more than 7 MeV per nucleon. This immediately implies that any nucleus with $A>4$ (and $Z,N\geq 2$) and B/A less than 7 MeV per nucleon should be unstable against α-decay. In other words, we might imagine that two protons and two neutrons inside a large nucleus, finding themselves in close proximity, may spontaneously form a helium nucleus and that the surplus binding energy thus liberated will be sufficient to expel the particle so-formed from the parent. This condition is satisfied for $A\gtrsim 140$ and therefore all nuclei with larger values of A are unstable and undergo α-decay. An example of such a decay is

$$^{232}_{92}\text{U} \;\rightarrow\; ^{228}_{90}\text{Th} + \alpha + 5.43 \text{ MeV}, \tag{6.1.2}$$

the energy being released as the kinetic energy of both the α-particle and the recoiling nucleus. For a recoil momentum $\boldsymbol{p}$ in the rest frame of the decaying particle, we thus have

$$E \;=\; \frac{\boldsymbol{p}^2}{2m_\alpha} + \frac{\boldsymbol{p}^2}{2m_{A'}} \;=\; E_\alpha\left[1+\frac{m_\alpha}{m_{A'}}\right] \;\approx\; E_\alpha\left[1+\frac{4}{A-4}\right] \;=\; \frac{E_\alpha}{1-4/A}. \tag{6.1.3}$$

We should note in passing that such an energy implies non-relativistic kinematics.

One might expect that such decays, begin energetically favourable, be very fast. This is, however, not so: measured lifetimes of known decays range from around 0.1 μs up to 10^{17} yr. There must therefore be some hinderance or obstacle—the so-called Coulomb barrier (similar to that already discussed in conjunction with fission, see Fig. 3.4). Such decays obey an exponential law due to the constant fractional rate:

$$\frac{\mathrm{d}N}{\mathrm{d}t} \;=\; -\lambda N, \tag{6.1.4}$$

which leads to the standard exponential solution for the number of surviving nuclei

$$N(t) \;=\; N_0\,\mathrm{e}^{-\lambda t}. \tag{6.1.5}$$

The decay rate λ is then the inverse of the mean lifetime τ. Note that nuclear

physicists often prefer to use the half-life $\tau_{1/2}$, the time in which half an initial sample will decay. Recall that the relation between the two is $\tau = \tau_{1/2}/\ln 2$.

Already in 1911 Geiger and Nuttall had derived an empirical law describing the rate as a function of the energy released E; the modern version of the Geiger–Nuttall law is

$$\ln \lambda \;=\; a - b\,\frac{Z}{\sqrt{E}}, \tag{6.1.6}$$

where typical values, for λ measured in seconds^{-1} and E in MeV, are $a \sim 107$ and $b \sim 3.4$ (the best-fit values vary slightly for different decay series). Most experimental values for E lie close around 5 MeV and so sharp a concentration is explained by the very strong energy dependence. Compare these with typical β-decay lifetimes (where the energy release is similar and the underlying interaction is weak. Indeed, this also explains the non-observation of the following energetically possible decay of bismuth:

$$^{209}_{83}\mathrm{Bi} \;\rightarrow\; ^{205}_{81}\mathrm{Tl} + \alpha + 3.11\ \mathrm{MeV}. \tag{6.1.7}$$

Exercise 6.1.1. *Using the Geiger–Nuttall formula, Calculate the expected lifetime for the α-decay of $^{209}_{83}$Bi ($Q = 3.11$ MeV) and thus explain its non-observation.*

6.1.1 Tunnelling in alpha decay

The theory of alpha decay as a (quantum) tunnelling effect was developed very early on by Gamow (1928), see also Gurney and Condon (1929). We shall now provide a simple derivation based on an approach similar to the description of electrical transmission lines of varying impedance—a more precise theoretical approach is provided by the so-called WKB approximation. The problem is to solve the Schrödinger equation for the case of a spatially varying potential. The approximation we shall adopt (and which will be seen, *a posteriori*, to be valid for α-decay) is of a slowly varying potential (this is just the WKB condition).

The starting point is the wave-function describing the motion of an α-particle in the presence of the nuclear potential. We are only interested in the spatial dependence; moreover, only the radial motion will be of importance and so a one-dimensional treatment will suffice; we thus write $u(x) = \mathrm{e}^{\pm \mathrm{i}kx}$, where the wave-number k is given by $\hbar k = \sqrt{2mE_{\mathrm{kin}}}$. Inside a potential barrier the kinetic energy (calculated as the difference between the total and potential energies) becomes negative and the solution is an evanescent wave (*i.e.*, non-propagating) given by $u(x) = \mathrm{e}^{\pm \kappa x}$, with $\hbar\kappa = \sqrt{2m|E_{\mathrm{kin}}|}$.

Consider now the step barrier shown in Fig. 6.1, with an incoming wave approaching from the left. In the two regions the wave-function for a particle of total

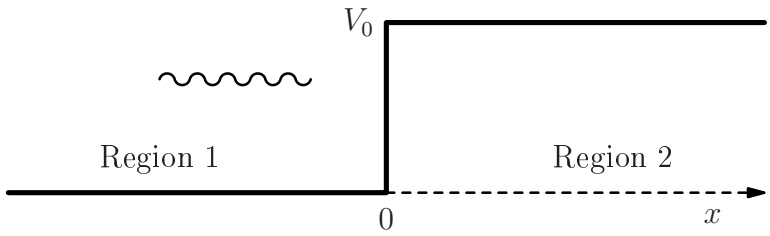


Figure 6.1: A simple step barrier for an incoming wave approaching from the left.

energy E will be

$$u_1(x) \;=\; \mathrm{e}^{+\mathrm{i}k_1 x} + R\,\mathrm{e}^{-\mathrm{i}k_1 x} \tag{6.1.8a}$$

and

$$u_2(x) \;=\; \begin{cases} T\,\mathrm{e}^{+\mathrm{i}k_2 x} & (E \;>\; V_0), \\ T\,\mathrm{e}^{-\kappa x} & (E \;<\; V_0), \end{cases} \tag{6.1.8b}$$

where the wave-numbers are given by

$$\hbar k_1 \;=\; \sqrt{2mE}, \qquad \hbar k_2 \;=\; \sqrt{2m(E-V_0)} \quad \text{and} \quad \hbar\kappa \;=\; \sqrt{2m(V_0-E)}. \tag{6.1.9}$$

The coefficients R and T are to be determined by the boundary conditions at $x=0$; that is, $u(x)$ and its first derivative must be continuous there (or, equivalently, the density and flux must be continuous). One then finds

$$T \;=\; \frac{2k_2}{k_1+k_2} \quad \text{and} \quad R \;=\; \frac{k_1-k_2}{k_1+k_2} \qquad (E \;>\; V_0). \tag{6.1.10}$$

The case $E<V_0$ is obtained from the above via the substitution $k_2 \to \mathrm{i}\kappa$. In this case we have $|R|=1$; *i.e.*, total reflection.

Consider now the double step barrier shown in Fig. 6.2, with an incoming wave

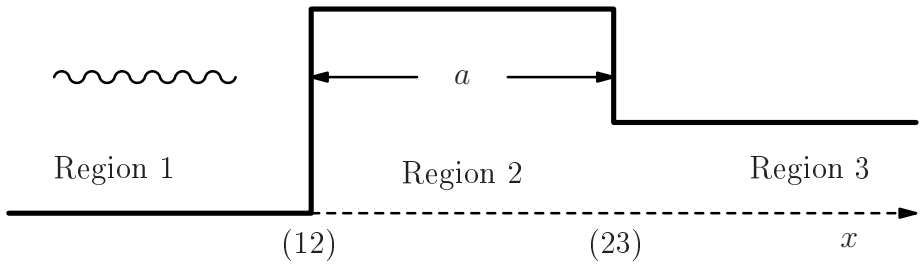


Figure 6.2: A double step barrier for an incoming wave approaching from the left.

again approaching from the left. The standard quantum mechanical treatment would require the simultaneous solution of the boundary conditions for the both the 1−2 and 2−3 interfaces. One can equivalently consider the problem as continued

reflection and transmission at each boundary. That is, in first approximation, we calculate the transmission from region 1 to 3 as the product of transmission at the 1−2 interface followed by propagation and finally transmission at the 2−3 interface:

$$T \approx T_{12}\, \mathrm{e}^{\mathrm{i}k_2 a}\, T_{23}. \tag{6.1.11}$$

The full (exact) solution then requires the addition of one, two and more possible reflections back and forth between the two boundaries:

$$\begin{aligned} T &= T_{12}\, \mathrm{e}^{\mathrm{i}k_2 a}\left[1 + R_{23}\, \mathrm{e}^{\mathrm{i}k_2 a}\, R_{21}\, \mathrm{e}^{\mathrm{i}k_2 a} + (R_{23}\, \mathrm{e}^{\mathrm{i}k_2 a}\, R_{21}\, \mathrm{e}^{\mathrm{i}k_2 a})^2 + \ldots\right] T_{23}. \\ &= \frac{T_{12}\, \mathrm{e}^{\mathrm{i}k_2 a}\, T_{23}}{1 - R_{23}\, \mathrm{e}^{\mathrm{i}k_2 a}\, R_{21}\, \mathrm{e}^{\mathrm{i}k_2 a}} \\ &= \frac{4k_1 k_2\, \mathrm{e}^{\mathrm{i}k_2 a}}{(k_1 + k_2)(k_2 + k_3) - (k_3 - k_2)(k_1 - k_2)\, \mathrm{e}^{2\mathrm{i}k_2 a}}. \end{aligned} \tag{6.1.12}$$

In the case that the height of the middle barrier is greater than the total energy then again we must substitute $k_2 \to \mathrm{i}\kappa$:

$$T = \frac{4\mathrm{i}k_1\kappa\, \mathrm{e}^{-\kappa a}}{(k_1 + \mathrm{i}\kappa)(\mathrm{i}\kappa + k_3) - (k_3 - \mathrm{i}\kappa)(k_1 - \mathrm{i}\kappa)\, \mathrm{e}^{-2\kappa a}}. \tag{6.1.13}$$

Note that this is the exact solution and that no approximation has yet been made.

We now apply the slowly varying potential approximation (a sort of adiabatic approximation) and thus assume that $\kappa a \gg 1$. We may therefore neglect the second term in the denominator of Eq. (6.1.13). Defining then B as the penetration probability for the barrier, we have

$$\begin{aligned} B &= \frac{\text{transmitted flux}}{\text{incident flux}} = \frac{|T|^2 \cdot v_3}{1 \cdot v_1} = |T|^2 \frac{k_3}{k_1} \\ &\approx \frac{16k_1^2\kappa^2\, \mathrm{e}^{-2\kappa a}}{(k_1^2 + \kappa^2)(\kappa^2 + k_3^2)} \frac{k_3}{k_1} \\ &= \frac{16}{(k_1/\kappa + \kappa/k_1)(\kappa/k_3 + k_3/\kappa)}\, \mathrm{e}^{-2\kappa a}. \end{aligned} \tag{6.1.14}$$

Remembering that the parameters $k_{1,3}$ and κ all depend on the energy, we see that the dominant energy dependence lies in the exponential damping factor $\mathrm{e}^{-2\kappa a}$. We shall find that κa is just a first glimpse of what is known as the Gamow factor.

We now have the necessary formulæ to attack the more general problem of a continuously varying potential, see Fig. 6.3. The penetration probability is the

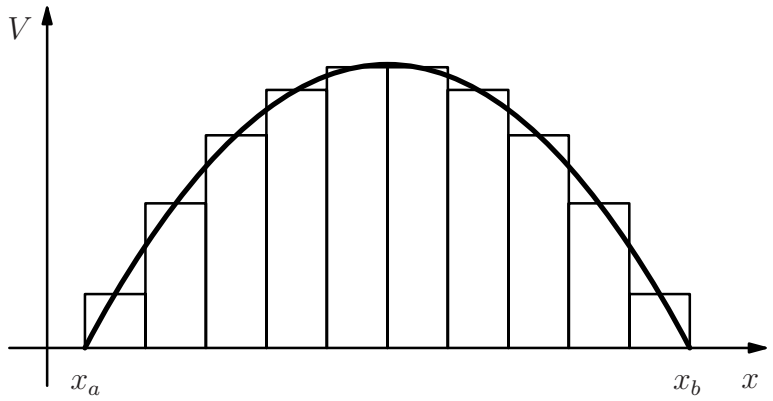


Figure 6.3: A continuously varying potential divided into infinitesimal barriers.

product of the probabilities for each of N infinitesimal barriers in the limit $N \to \infty$:

$$B \;=\; B_1 \cdot B_2 \cdots B_N. \tag{6.1.15}$$

Taking first the exponential factors, as given by Eq. (6.1.14), we have

$$\begin{aligned} B \;&=\; \mathrm{e}^{-2\kappa_1 \delta x}\; \mathrm{e}^{-2\kappa_2 \delta x}\; \cdots\, \mathrm{e}^{-2\kappa_N \delta x} \\ &=\; \exp\bigl[-2\,(\kappa_1 + \kappa_2 + \,\cdots + \kappa_N)\,\delta x\bigr]\,, \\ &\xrightarrow{N\to\infty}\; \exp\left[-2\int_{x_a}^{x_b} \mathrm{d}x\,\kappa(x)\right] \end{aligned} \tag{6.1.16}$$

where, recall, $\hbar\kappa(x) = \sqrt{2m[V(x) - E]}$. The above integral is then just the full Gamow factor.

We still have to deal with the prefactors and also the question of reflection within each of the infinitesimal barriers. Now, each transmission prefactor is of the form

$$\frac{2k_{n-1}}{k_{n-1} + k_n} \tag{6.1.17a}$$

while those for reflection are

$$\frac{k_{n-1} - k_n}{k_{n-1} + k_n}. \tag{6.1.17b}$$

The assumption of slow variation, which implies $|k_{n-1} - k_n| \ll |k_{n-1} + k_n|$, thus allows us to neglect reflection. The transmission prefactor may now be slightly

manipulated:

$$\left(\frac{2k_{n-1}}{k_{n-1}+k_n}\right)^2 = \left(\frac{2k_{n-1}}{2k_{n-1}+\Delta k_{n-1}}\right)^2$$
$$\simeq \left(1-\frac{\Delta k_{n-1}}{2k_{n-1}}\right)^2$$
$$\simeq \left(1-\frac{\Delta k_{n-1}}{k_{n-1}}\right) \simeq \exp\left\{-\frac{\Delta k_{n-1}}{k_{n-1}}\right\}, \qquad (6.1.18)$$

where $\Delta k_{n-1} := k_n - k_{n-1}$. And the N products become

$$\exp\left\{-\sum_{n=1}^{N}\frac{\Delta k_{n-1}}{k_{n-1}}\right\} \xrightarrow{N\to\infty} \exp\left\{-\int_{x_a}^{x_b}\frac{\mathrm{d}k}{k}\right\}$$
$$= \exp\left\{-\ln\left[\frac{k_b}{k_a}\right]\right\} = \frac{k_a}{k_b}, \qquad (6.1.19)$$

where $k_{a,b} = k(x_{a,b})$.

We can now put everything together to obtain the final expression. First of all, note that a result of our approximation is that in all regions where $E > V_0$ the transmission is total (*i.e.*, there is no reflection). We thus need only consider our expressions above for the region in which $E < V_0$; that is, between the two classical turning points, say x_a and x_b, for which $E = V_0$ and therefore $k = 0$. The transmission probability is then just the product of Eqs. (6.1.16) and (6.1.19) multiplied by the ratio of the velocities v_b/v_a (which is just k_b/k_a):

$$T_{a\to b} \simeq \exp\left[-2\int_{x_a}^{x_b}\mathrm{d}x\,\kappa(x)\right]\cdot\frac{k_a}{k_b}\cdot\frac{k_b}{k_a} = \exp\left[-2\int_{x_a}^{x_b}\mathrm{d}x\,\kappa(x)\right]. \qquad (6.1.20)$$

This is then essentially just the result obtained by Gamow (1928).

We have, however, been a little cavalier in dealing with the integration $\mathrm{d}k/k$ in Eq. (6.1.19), which passes through the singular points $x_{a,b}$ where $k \to 0$. In fact, the integration proceeds from some point outside the barrier to the left, where k is real, through into the barrier, where k becomes imaginary, and then out again on the other side. The situation in the complex k plane is depicted in Fig. 6.4. We may deform the contour away from the origin following an arc (of radius ε) provided we can show that the wave-function does not change appreciably. Now, for $k \approx 0$ the Schrödinger equation reduces to simply

$$\frac{\mathrm{d}^2u(x)}{\mathrm{d}x^2} \approx 0 \qquad \text{so} \qquad \frac{\mathrm{d}u(x)}{\mathrm{d}x} \approx \text{constant and finite} \qquad (6.1.21)$$

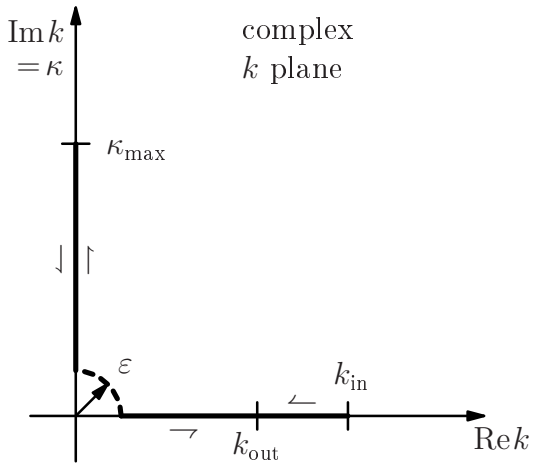


Figure 6.4: The contour for the k integral in the complex k plane. The dotted arc around the origin represents a permitted infinitesimal deformation of the contour.

and therefore $u(x)$ will, indeed, not vary appreciably. To perform the integrals along the arcs, we write $k = \varepsilon \mathrm{e}^{\mathrm{i}\phi}$. We then have

$$-\frac{1}{2}\int_{\text{arc}} \frac{\mathrm{d}k}{k} \;=\; -\frac{1}{2}\int_0^{\pi/2} \frac{\mathrm{i}\,\mathrm{d}\phi\,\varepsilon\,\mathrm{e}^{\mathrm{i}\phi}}{\varepsilon\,\mathrm{e}^{\mathrm{i}\phi}} \;=\; -\frac{\mathrm{i}\pi}{4}. \tag{6.1.22}$$

On returning, the sign changes and so the contribution actually cancels overall. The effect then is simply an irrelevant phase shift by $-\pi/4$ in the wave-function inside the barrier (the classically *unphysical* region).

The calculation to be performed in order to obtain the barrier penetration probability is therefore simply

$$B \;\approx\; \exp\left[-\frac{2}{\hbar}\int_{x_a}^{x_b} \mathrm{d}x\,\sqrt{2m[V(x)-E]}\right]. \tag{6.1.23}$$

The question now is the form of the potential to be adopted. This should be the potential as experienced by an α-particle; the following parametrisation (valid only for $r \gtrsim 10\,\text{fm}$; *i.e.* effectively *outside* the nucleus) is obtained from data on Rutherford scattering with energetic α-particles:

$$V(r) \;\simeq\; \frac{2.88Z}{r} - 1100\exp\left\{\frac{1.17A^{1/3}-r}{0.574}\right\}\text{MeV} \qquad (\text{with } r \text{ in fm}). \tag{6.1.24}$$

An example plot is shown in Fig. 6.5. The first term (which dominates at large distances) is just the Coulomb repulsion expected for a particle of (positive) charge

two.

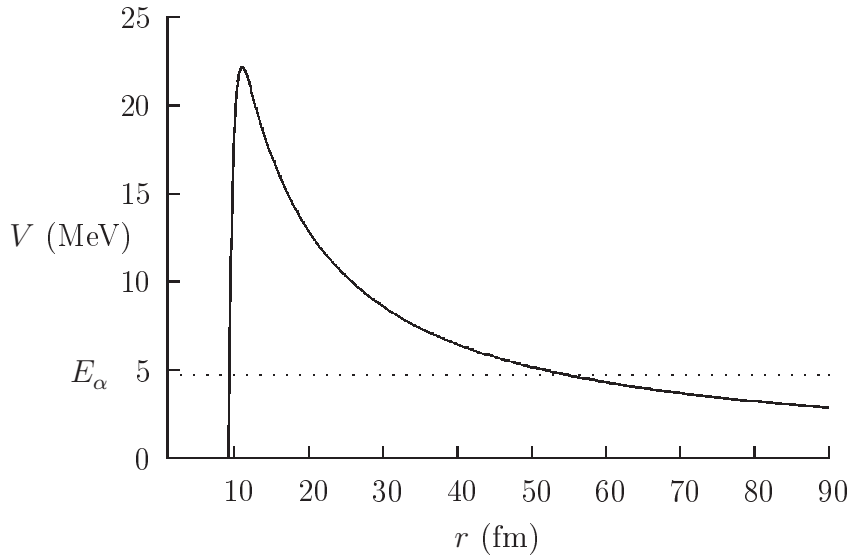


Figure 6.5: The nuclear potential (Coulomb and strong nuclear interaction) as seen by an α-particle in the case of $^{230}_{90}$Th, the decay energy is $E_\alpha = 4.76\,\mathrm{MeV}$.

The full problem is three-dimensional and so the barrier-penetration probability is the flux ratio integrated over the full solid angle:

$$B \;=\; \frac{k_b \times \int r_b^2 \mathrm{d}\Omega\, |\psi(\boldsymbol{x})|^2 \;|_{|\boldsymbol{x}|=r_b}}{k_a \times \int r_a^2 \mathrm{d}\Omega\, |\psi(\boldsymbol{x})|^2 \;|_{|\boldsymbol{x}|=r_a}}. \tag{6.1.25}$$

Now, $\psi(\boldsymbol{x})$ is given in terms of the reduced wave-function $u(r)$ and the spherical harmonics $Y_{lm}(\Omega)$ by

$$\psi(\boldsymbol{x}) \;=\; \frac{u(r)}{r}\, Y_{lm}(\Omega) \tag{6.1.26}$$

and thus B indeed reduces to the same form as used in the discussions above. For simplicity, we do not consider orbital angular momentum (*i.e.* $l=0$ is assumed as $l>0$ would imply an even higher barrier) and therefore the calculation to be performed is just as already described.

In order to have some idea of the final form for the probability (or Gamow factor), we shall now adopt a simplified model for the nuclear potential as seen by an α-particle. We shall assume a square well of radius r_a and for $r>r_a$ the Coulomb form a/r, as shown in Fig. 6.6. We thus have that r_b is given by $E_\alpha = a/r_b$,

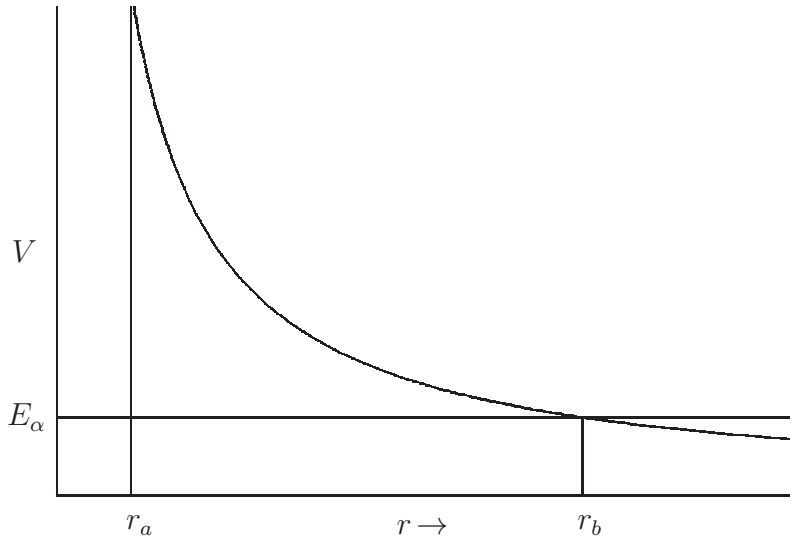


Figure 6.6: A schematic approximation to the nuclear potential as seen by an α-particle.

where $a = 2(Z-2)\alpha$. The Gamow factor is then

$$G = \frac{1}{\hbar}\int_{r_a}^{r_b} \mathrm{d}r\, \sqrt{2m[V(r) - E_\alpha]},$$

on substituting $r = \frac{a}{E_\alpha}\cos^2\theta$ this becomes

$$\begin{aligned} &= \frac{a}{\hbar}\sqrt{\frac{2m}{E_\alpha}}\int_{r_a/r_b}^{1} \mathrm{d}\cos^2\theta\, \sqrt{\frac{1}{\cos^2\theta} - 1} \\ &= \frac{a}{\hbar}\sqrt{\frac{2m}{E_\alpha}}\; 2\int_0^{\cos^{-1}\sqrt{r_a/r_b}} \mathrm{d}\theta\, \sin^2\theta \\ &= \frac{a}{\hbar}\sqrt{\frac{2m}{E_\alpha}}\left[\pi/2 + O(r_a/r_b)\right] \\ &\simeq \sqrt{2}\,\pi\,(Z-2)\,\alpha\sqrt{\frac{mc^2}{E_\alpha}}. \end{aligned} \tag{6.1.27}$$

Note that the energy dependence is precisely that of the Geiger–Nuttall law and for Z large the Z dependence is also the same.

6.1.2 Spontaneous α-particle formation probability

The final step is to estimate the rate at which α-particles may spontaneously form and thus be candidates for the tunnelling process. Once an α-particle has formed, given the natural mean kinetic energy inside the nuclear potential of approximately $E_{\rm kin} = 30\,{\rm MeV}$ and the consequent velocity $v_\alpha = \sqrt{2E_{\rm kin}/m}$, we expect an approximate frequency of collisions with the barrier $v_\alpha/2r_a$, where we take r_a to be the approximate radius of the nucleus.

For the formation rate, we need to estimate the fraction of time f that two protons and two neutrons will be sufficiently close to form a helium nucleus; *i.e.* within a sphere of radius r_α, the helium nucleus radius. Given some initial nucleon (given the more or less equal and large numbers of protons and neutrons, which type is unimportant), then another must find itself with a distance $2r_\alpha$, which thus has probability $(2r_\alpha/r_a)^3$. Each of the remaining two then has approximate probability $(r_\alpha/r_a)^3$ of being inside the resulting sphere of radius r_α. One thus finds

$$f \approx 8(r_\alpha/r_a)^9. \tag{6.1.28}$$

A more careful estimate arrives at $f \approx 25(r_\alpha/r_a)^9 \approx 25(4/A)^3$. For $A = 230$ this gives $f \approx 1.3 \times 10^{-4}$. Taking now $r_a \simeq 8\,{\rm fm}$, one obtains a so-called reduced rate

$$\lambda_0 \simeq f\,\frac{v_\alpha}{2r_a} \simeq 6 \times 10^{17}\ {\rm s}^{-1}. \tag{6.1.29}$$

Inverting the logic and using the calculated Gamow factor, the experimental data indicate values typically of order 2×10^{18} – $3 \times 10^{19}\,{\rm s}^{-1}$. While not perfect, the fact that such a crude estimate lies within a little more than one order of magnitude should be considered as a success. Indeed, it is rather easy to find possible sources of this discrepancy:

- more than one α-particle may be formed simultaneously,
- local density fluctuations may favour α-particle formation,
- large nuclei are not usually symmetric and the (flattened) polar regions present a reduced width for the barrier.

The important point to appreciate is that vast number of orders of magnitude between the largest and smallest decay rates are already well explained by the exponential Gamow factor. A couple of examples of greatly differing rates should suffice to underline this observation, see Table 6.1.

Table 6.1: A comparison of two very different α-decay rates.

	$\tau_{1/2}\,(\mathrm{s})$	B	$\lambda_0\,(\mathrm{s}^{-1})$	E_α/MeV
$^{144}_{60}\mathrm{Nd}$:	1.58×10^{23}	2.18×10^{-42}	2.02×10^{18}	1.90
$^{212}_{84}\mathrm{Po}$:	3.04×10^{-7}	1.32×10^{-13}	1.73×10^{19}	8.78

6.2 Beta decay and the weak interaction

The early observations of β-decay witnessed the change by one unit of electric charge of the parent nucleus accompanied by the emission of an electron (later also positron emission became known to be possible) with energies varying from the lowest detectable up to a few tens of MeV. This is already in sharp contrast with all other known two-body decays (α and γ emission), which all have necessarily mono-energetic spectra. Moreover, it was evident from knowledge of the spin quantum numbers involved that angular momentum could not be conserved if this were a purely two-body decay. The explanation, first proposed by Pauli in 1930, is that this is, in fact, a three-body decay, the third particle being the neutrino.*†

> "Dear Radioactive Ladies and Gentlemen!
> I have hit upon a desperate remedy to save... the law of conservation of energy.
> ...there could exist electrically neutral particles, which I will call neutrons, in the nuclei...
> The continuous beta spectrum would then make sense with the assumption that in beta decay, in addition to the electron, a neutron is emitted such that the sum of the energies of neutron and electron is constant....
> But so far I do not dare to publish anything about this idea, and trustfully turn first to you, dear radioactive ones, with the question of how likely it is to find experimental evidence for such a neutron...
> I admit that my remedy may seem almost improbable because one probably would have seen those neutrons, if they exist, for a long time.
> But nothing ventured, nothing gained...
> Thus, dear radioactive ones, scrutinise and judge."
>
> Wolfgang Pauli

In order to satisfy angular momentum conservation, the neutrino should be a spin-½ fermion. Moreover, the observed electron-spectrum end-point corresponds

† Pauli initially called the new particle a "neutron". When what is now known as the neutron was later discovered, Fermi suggested the name neutrino for Pauli's particle, being a "small" (light) neutral particle.

† Niels Bohr, for example, openly opposed this interpretation of β-decay, preferring to entertain the possibility that energy, momentum and angular momentum might not be conserved.

to the electron carrying away all of the available energy (within experimental precision), which indicates that the neutrino should have negligible mass. It was deemed therefore, in first approximation, to have precisely zero mass.

As early as 1934 Fermi had already devised a theory of the weak nuclear interaction*, in which such decays were described via a four-point local interaction transmuting a neutron inside the nucleus into a proton, with the simultaneous emission of an electron and (what is now known as) an electron antineutrino ($\bar{\nu}_e$). This is what is called β^- decay. It is also possible for a free neutron:

$$n \;\to\; p + e^- + \bar{\nu}_e. \tag{6.2.1}$$

A further possibility is for a proton to transmute into a neutron, emitting a positron and an electron neutrino (ν_e). This is called β^+ decay. Since the proton is lighter than the neutron, such a process is only possible inside a nucleus, where the final-state neutron may occupy a lower energy level than the original proton:

$$p + \text{energy} \;\to\; n + e^+ + \nu_e. \tag{6.2.2}$$

Finally, a process known as electron capture or K-capture is possible. Here an atomic electron from the lowest (K) shell is captured by a proton (inside the nucleus), which then transmutes into a neutron emitting a neutrino:

$$p + e^- + \text{energy} \;\to\; n + \nu_e. \tag{6.2.3}$$

Some examples of these various forms of β-decay are

$$^{137}_{55}\mathrm{Cs} \;\to\; ^{137}_{56}\mathrm{Ba} + e^- + \bar{\nu}_e \qquad (\beta^-\text{-decay}), \tag{6.2.4a}$$

$$^{22}_{11}\mathrm{Na} \;\to\; ^{22}_{10}\mathrm{Ne} + e^+ + \nu_e \qquad (\beta^+\text{-decay}), \tag{6.2.4b}$$

$$^{22}_{11}\mathrm{Na} + e^- \;\to\; ^{22}_{10}\mathrm{Ne} + \nu_e \qquad (\text{K-capture}). \tag{6.2.4c}$$

Note that in all cases the mass number (A) does not change, but the charge (Z) does. We have seen earlier that for constant A the nuclear mass is a quadratic function of $(Z-N)$, due to the neutron–proton symmetry term, see *e.g.* Eq. (3.2.8), and so one may expect cascades or chains of $\beta^\pm$-decays down either side of the valley so-formed until the lightest isobar is reached, which must then be stable against this type of decay.

Now, since the mass formula also contemplates a nucleon-pairing term, see *e.g.* Eq. (3.2.9), the β-decay chains for odd and even A must be distinguished. In the first case (A odd) one of N or Z is odd and the other even and so the pairing

* An initial paper by Fermi was rejected for publication by Nature as "it contained speculations which were too remote from reality."

contribution is constant. In the second though (A even) either both are odd or both even and the contribution continually varies; or, more pictorially, there are two separate mass parabolæ (the series for even–even lying below the odd–odd), see Fig. 6.7.

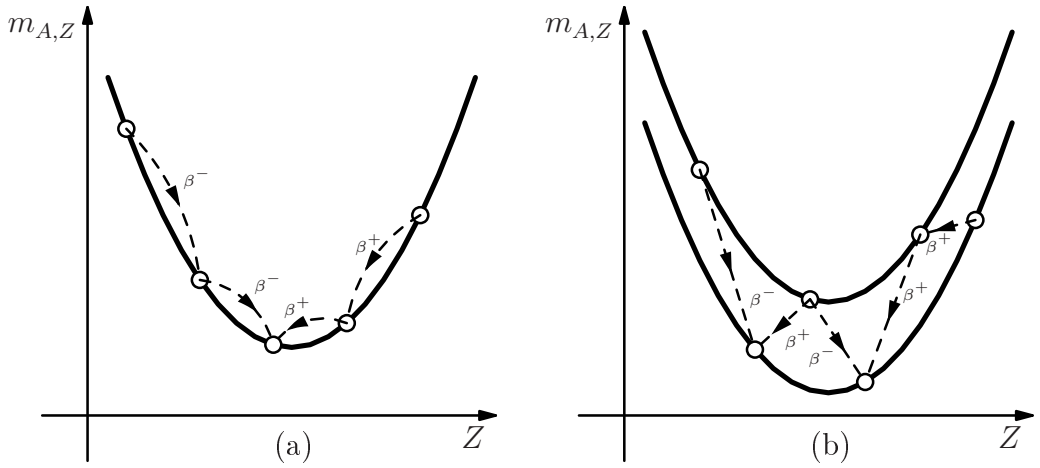


Figure 6.7: The two possible types of β-decay chains for (a) A odd and (b) A even.

If the two lowest-lying isobars in a given chain are too close in mass for the heavier to decay, then both are β-stable. Indeed, there may be up to three stable isobars; *e.g.*, ${}^{96}_{40}\mathrm{Zr}$, ${}^{96}_{42}\mathrm{Mo}$ and ${}^{96}_{44}\mathrm{Ru}$. Moreover, so-called double β-decay is also possible (Goeppert-Mayer, 1935), in which Z changes by two units and two electron–neutrino pairs are emitted. Such a process is clearly doubly weak and therefore highly suppressed (all measured lifetimes are in excess of $10^{19}\,\mathrm{yr}$). It is therefore only visible if the single decay mode is prohibited. For example, in Fig. 6.7b the two lowest-lying states are both stable against single β-decay but the higher-mass nucleus may decay via the double β-decay mode into the lower; there are just 35 such decays possible.

6.2.1 Fermi theory

As already mentioned, the theory explaining β-decay (and, more in general, the weak nuclear interaction) was devised by Fermi (1934). Let us first examine the kinematics: since it is relatively very massive, the final-state nucleus may be taken to have negligible recoil energy (though the recoil momentum may be appreciable) and therefore the electron–neutrino pair have energy

$$E_e + E_\nu \simeq Q, \tag{6.2.5}$$

where Q is the energy released (or so-called Q-value).

Exercise 6.2.1. *The decay* $^{210}\mathrm{Bi} \to {}^{210}\mathrm{Po} + e^- + \bar{\nu}_e$ *has a* Q*-value of* 1.16 MeV*; calculate the maximum recoil kinetic energy of the daughter nucleus. Do the same for* $^{22}_{11}\mathrm{Na} \to {}^{22}_{10}\mathrm{Ne} + e^+ + \nu_e$*,* $Q = 4.38\,\mathrm{MeV}$ *(take* $m_\nu = 0$*). What is the minimum recoil energy?*

In quantum mechanics the rate of such a decay may be obtained using Fermi's golden rule, Eq. (4.2.6), which here reads

$$\Gamma \;=\; \frac{1}{\tau} \;=\; \frac{2\pi}{\hbar} \int \mathrm{d}n_f \, |\mathcal{M}_{fi}|^2, \tag{6.2.6}$$

where τ is the mean lifetime for this decay. The transition matrix element is given by

$$\mathcal{M}_{fi} \;=\; \int \mathrm{d}^3\boldsymbol{x} \;\; \psi_{\mathrm{f}}^*(\boldsymbol{x})\, \mathcal{H}_{\mathrm{int}}\, \psi_{\mathrm{i}}(\boldsymbol{x}), \tag{6.2.7}$$

where $\psi_{\mathrm{i,f}}$ are the initial- and final-state wave-functions and $\mathcal{H}_{\mathrm{int}}$ is the relevant interaction Hamiltonian. The initial-state wave-function is just that of the parent nucleon (Ψ) while that of the final state will be the product of the wave-functions describing the daughter nucleus (Ψ') and those of the emitted electron (ψ_e) and antineutrino (ψ_ν). The problem then is the choice of operator to represent the interaction Hamiltonian. The solution adopted by Fermi was simply the identity operator, thus:

$$\mathcal{M}_{fi} \;=\; \int \mathrm{d}^3\boldsymbol{x}\, \Psi'^*(\boldsymbol{x})\, \psi_e^*(\boldsymbol{x})\, \psi_\nu^*(\boldsymbol{x}) \cdot G_{\mathrm{F}} \cdot \Psi(\boldsymbol{x}), \tag{6.2.8}$$

where G_{F} is a constant defining the strength of the interaction (now known as the Fermi constant). Such a form is significantly different to that used for the electromagnetic interaction, which is viewed as mediated by the exchange of a photon propagating between two distinct space–time points. However, at the distance scale to which β-decay is sensitive, there is no experimental evidence for similar propagation and the interaction appears to take place at a single point in space–time.

Exercise 6.2.2. *We now know that the weak interaction is in fact mediated by the bosons* $W^\pm$ *and* Z^0*; the former having a mass around* 80 GeV *and the latter* 91 GeV*. Using the uncertainty principle, estimate the maximum distance that such particles could propagate as virtual intermediate states. Compare this to the wavelength associated with the typical momenta involved.*

The wave-functions for the electron–neutrino pair should naturally be taken as

plane waves (describing particles of well-defined linear momenta)

$$\psi_{e,\nu}(\boldsymbol{x}) \;=\; \frac{e^{\mathrm{i}\boldsymbol{k}_{e,\nu}\cdot\boldsymbol{x}}}{\sqrt{V}}, \tag{6.2.9}$$

where the standard normalisation of one particle in a volume V has been adopted (V will, of course, disappear from the final answer). One thus has

$$\mathcal{M}_{fi} \;=\; \frac{G_{\mathrm{F}}}{V}\int \mathrm{d}^3\boldsymbol{x}\;\; \mathrm{e}^{-\mathrm{i}(\boldsymbol{k}_e+\boldsymbol{k}_\nu)\cdot\boldsymbol{x}}\;\; \Psi'^{*}(\boldsymbol{x})\,\Psi(\boldsymbol{x}), \tag{6.2.10}$$

where $\boldsymbol{k}_{e,\nu}$ are the electron and neutrino wave-numbers. We may now make a further simplifying approximation: the momentum of the electron and neutrino are limited to a few MeV and therefore the corresponding wavelengths are much larger than the size of the nucleus, which is the integration region determined by the presence of $\Psi(\boldsymbol{x})$ and $\Psi'(\boldsymbol{x})$. The exponential will thus not vary appreciably under the integral and may be taken as constant. One then has

$$\mathcal{M}_{fi} \;\simeq\; \frac{G_{\mathrm{F}}}{V}\int \mathrm{d}^3\boldsymbol{x}\;\; \Psi'^{*}(\boldsymbol{x})\,\Psi(\boldsymbol{x}), \tag{6.2.11}$$

which just describes the overlap between the initial- and final-state nuclear wave-functions. In certain cases, where the final-state proton (or neutron) occupies the same nuclear level as the initial-state neutron (or proton), these are (almost) identical and then, to a good approximation, the integral is unity. Otherwise, with some knowledge of the approximate wave-functions, it may be estimated rather well and, in any case, to a very good approximation is a constant, independent of the electron and neutrino energies.

For a particle of mass m and momentum $\boldsymbol{p}$, with the standard normalisation to a box of volume V, the density of states $\mathrm{d}n/\mathrm{d}E$ is derived by considering

$$\mathrm{d}^3\boldsymbol{n} \;=\; \frac{V}{(2\pi\hbar)^3}\,\mathrm{d}^3\boldsymbol{p} \;=\; \frac{V}{(2\pi\hbar)^3}\,p^2\mathrm{d}\Omega\,\mathrm{d}p\,. \tag{6.2.12}$$

Assuming spherical symmetry, the angular integrals may be performed giving $\int \mathrm{d}\Omega = 4\pi$. The single-particle density of states is therefore

$$\frac{\mathrm{d}n}{\mathrm{d}E} \;=\; \frac{V}{(2\pi\hbar)^3}\,4\pi p^2\,\frac{\mathrm{d}p}{\mathrm{d}E}. \tag{6.2.13}$$

For the two-particle density of states in consideration here, one then has

$$\frac{\mathrm{d}n(Q,E_e)}{\mathrm{d}E_e} \;=\; \left[\frac{V}{(2\pi\hbar)^3}\,4\pi\right]^2 p_e^2\,\frac{\mathrm{d}p_e}{\mathrm{d}E_e}\;p_\nu^2\,\frac{\mathrm{d}p_\nu}{\mathrm{d}E_\nu}\bigg|_{E_e+E_\nu=Q}. \tag{6.2.14}$$

Using relativistic kinematics, one has the mass-shell relation $E^2=p^2c^2+m^2c^4$ and therefore $E\,\mathrm{d}E=pc^2\mathrm{d}p$. Thus, for a given final-particle energy–momentum, one has

$$p^2\,\frac{\mathrm{d}p}{\mathrm{d}E} \;=\; \frac{E\,p}{c^2} \;=\; \frac{E\,\sqrt{E^2-m^2c^4}}{c^3}. \tag{6.2.15}$$

Finally, assuming a zero-mass neutrino and using $E_\nu=Q-E_e$,

$$\frac{\mathrm{d}n(Q,E_e)}{\mathrm{d}E_e} \;=\; \left[\frac{V}{(2\pi\hbar)^3}\,4\pi\right]^2 \frac{E_e\,\sqrt{E_e^2-m_e^2c^4}}{c^3}\;\frac{(Q-E_e)^2}{c^3}. \tag{6.2.16}$$

This expression must then be integrated over E_e from the rest-mass energy m_ec^2 up to the Q-value. It is customary here to normalise the energies to the electron rest-mass energy and thus the integral to be performed is

$$f(Q) \;=\; \int_1^{x_{\max}} \mathrm{d}x\,x\,\sqrt{x^2-1}\,(x_{\max}-x)^2, \tag{6.2.17}$$

where $x_{\max}=Q/m_ec^2$. Now, for $x_{\max}\gg 1$, $f(Q)\simeq x_{\max}^5/30$. And so the final result for the decay rate is

$$\Gamma \;=\; \frac{1}{\tau} \;\simeq\; \frac{G_\mathrm{F}^2\,Q^5\,K}{60\pi^3\hbar^7c^6}, \tag{6.2.18}$$

where K represents the nuclear wave-function overlap integral (6.2.11). The Q^5 dependence (arising from the phase-space integral) is known as Sargent's rule (Sargent, 1933), originally a simple empirical observation, which is thus explained.*

There is an important correction to the above expression, due to the Coulomb interaction between the charge of the emitted electron or positron and that of the daughter nucleus (Fermi, 1934). Put simply, one can imagine that a positron will be ejected with greater energy (the force is repulsive) than an electron (attractive). Calculation in a non-relativistic approximation leads to the following multiplicative correction to the rate, the Coulomb factor or Fermi function (Mott and Wilson, 1933):

$$F(Z_f,Q) \;\simeq\; \frac{\eta}{1-\mathrm{e}^{-\eta}}, \quad \text{with} \quad \eta \;=\; \pm 2\pi Z_f\,\alpha/\beta \tag{6.2.19}$$

* It is then a general rule for all three-body decays.

where the plus (minus) sign applies to electrons (positrons) and $\beta = v/c$ is the asymptotic velocity of the emitted electron.

6.2.2 The Kurie plot

The simple form of the (uncorrected) spectrum,

$$P(E_e) \;=\; E_e \sqrt{E_e^2 - m_e^2 c^4}\,(Q - E_e)^2, \tag{6.2.20}$$

suggests a rather useful method of plotting the data (Kurie *et al.*, 1936). Let us define the following function:

$$K(E_e) \;:=\; \sqrt{\frac{P(E_e)}{F(Z_f, Q) \cdot E_e \cdot \sqrt{E_e^2 - m_e^2 c^4}}}, \tag{6.2.21}$$

where $P(E_e)$ is the measured electron spectrum. Clearly, using Eq. (6.2.20), this should give

$$K(E_e) \;\propto\; Q - E_e. \tag{6.2.22}$$

That is, one expects a straight line intersecting the abscissa at $E_e = Q$, see Fig. 6.8.

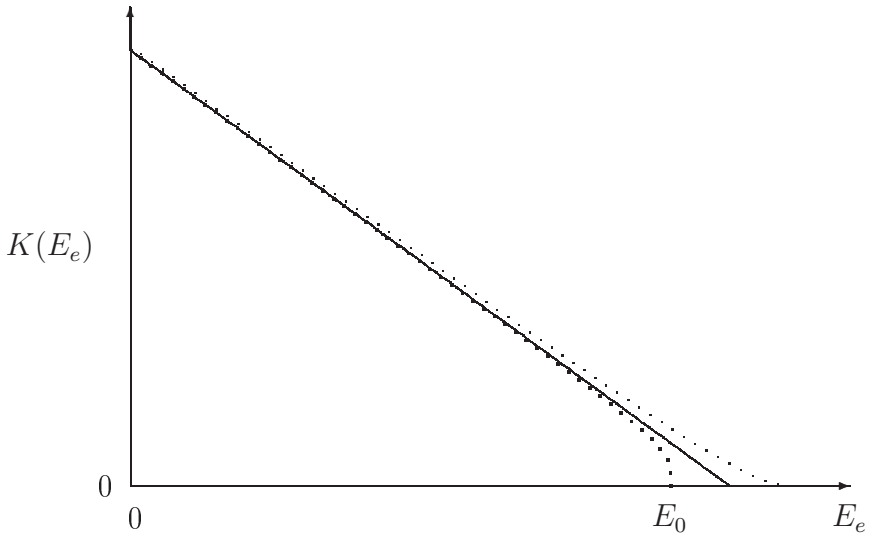


Figure 6.8: Example Kurie plots of the β-decay spectrum: the ideal zero-mass neutrino case (continuous line), ideal non-zero mass (heavy dotted line) and realistic non-zero mass (light dotted line).

Such a plot has particular use in the search for a non-zero neutrino mass. So far the standard assumption of a zero-mass electron neutrino has been made; if one instead inserts a mass term for the neutrino, then the expected form of the Kurie plot becomes

$$K(E_e) \propto \sqrt{(Q-E_e)\sqrt{(Q-E_e)^2 - m_\nu^2 c^4}}. \tag{6.2.23}$$

This now intersects the abscissa at $E_e = Q - m_\nu c^2$ and one would thus expect to see a curve of the form shown by the heavy dotted line in Fig. 6.8. Unfortunately, the effects of finite experimental resolution and Fermi motion of the decaying nucleus *etc.* all combine to smear the measured distribution into the form shown by the light dotted line in Fig. 6.8. To date there exists only an upper limit of about 2 eV on the electron-neutrino mass coming from such measurements on tritium decay (see Patrignani *et al.*, 2016).

Exercise 6.2.3. *Derive expression* (6.2.23).

6.2.3 The Gamow–Teller extension to Fermi theory

While Fermi's theory works well in certain cases, many nuclear decays require a more detailed description. In particular, no account has been made of the possible spin correlations. Indeed, the construction so-far adopted is explicitly scalar in its interaction. As such, it only applies directly to decays involving spin-zero nuclei (both parent and daughter) or possibly where the spin does not change. There are a number of such decays and we shall discuss them in more detail shortly, when we examine the problem of cleanly extracting the value of G_F.

In order to take into account the dependence on the spins of the electron and neutrino, Gamow and Teller (1936) provided a more complete description by inclusion of possible spin correlations via introduction of the Pauli spin matrices and two-component spinors into the matrix element $\mathcal{M}_{fi}$. Without entering into the details of the construction, we shall simply note that, *a priori*, there are various possibilities leading to different angular correlations between final-state momenta, which may be tested experimentally.

What emerges from the data is a classification into just two types of decay. The distinction is made via the spin state of the electron–neutrino pair: being both spin-½ particles, the total may give either a spin-zero singlet (so-called Fermi transitions) or spin-one triplet (so-called Gamow–Teller transitions). The former can clearly only contribute when the spins of the parent and daughter nuclei are the same while the latter may also contribute to spin-changing decays (though not in the case of both having spin zero).

For a full understanding of the spin and momentum correlations, the Dirac theory of fermions is necessary. Here, the dynamics is described as current–current interactions and, in analogy with the electromagnetic spin-one photon mediated force, we find a vector-like coupling, but we also find that a pseudovector or axial-vector coupling is necessary. For elementary particles, Dirac theory prescribes precisely the relative strengths; however, the proton and neutron are both composite objects and their spins are the result of the underlying (quark) spins and orbital motions. It is then customary to express the strengths of the two couplings as $g_V G_{\rm F}$ and $g_A G_{\rm F}$ respectively. The vector current corresponds to a conserved weak charge while the axial current represents a sort of weak magnetic effect, which has no corresponding conservation law. So, while we may appeal to charged vector-current conservation to fix g_V (as unity) or equivalently absorb it into the definition of $G_{\rm F}$, the ratio g_A/g_V is then an entirely free parameter to be measured experimentally.

6.2.4 Neutron β-decay

The case of neutron β-decay is rather special. First of all, owing to the small Q-value, it has a particularly long lifetime: $\tau_n = 880.2 \pm 1.0\,\mathrm{s}$. Moreover, since the initial and final nuclear states are just a free neutron and proton respectively, the wave-function overlap integral (K above) is exactly unity.* One might therefore imagine that, in principle at any rate, this decay would allow a precise measurement of the Fermi constant $G_{\rm F}$. However, as we have just observed, there are two possible couplings involved since both Fermi and Gamow–Teller transitions (or equivalently vector and axial-vector couplings) may contribute here since both the proton and neutron have non-zero spins.

The total decay rate is then given by Eq. (6.2.18) multiplied by a factor

$$g_V^2 + 3g_A^2, \tag{6.2.24}$$

where g_V and g_A are the vector and axial-vector couplings. The factor three (and implicit factor one) represents the triplet (singlet) nature of the Gamow–Teller (Fermi) transition. As noted, the vector constant corresponds to a conserved current and maybe absorbed into the definition of $G_{\rm F}$; this however leaves g_A, or rather the ratio g_A/g_V, as a further unknown parameter, but of the internal nucleon structure rather than of the weak-interaction theory.

We should also add that, owing to the very small Q-value, the earlier approximation to the phase-space integral is rather poor and that the exact result here is $f(Q) = 0.47(Q/m_e)^5/30$. Besides the total decay rate, various angular correlations

* Actually, the small mass difference does imply a small (essentially negligible) correction.

between particle momenta may be measured, which lead to the possibility of extracting the ratio g_A/g_V separately. The world average value for this quantity has approximately the same 2‰ precision as the lifetime measurement and is found to be

$$g_A/g_V \;=\; -1.2723\pm0.0023, \tag{6.2.25}$$

see Patrignani *et al.* (2016). And thus a reliable determination of G_F from is neutron β-decay is possible. We shall examine its precise value after the next section.

6.2.5 Measuring G_F in nuclear β-decay

In order to measure the Fermi constant G_F in nuclear β-decay accurately, it is clearly necessary to eliminate the uncertainty arising from the axial-vector contribution. As already noted, the Gamow–Teller coupling cannot contribute to transitions between two spin-zero nuclei. These then provide a reliable method for determining G_F.

The Fermi integral (6.2.17), including also the Fermi function (6.2.19) with the integrand, can then be simply tabulated for the various known decays. It is thus customary to define the so-called ft value:

$$ft_{1/2} \;:=\; 0.693\,\frac{2\pi^3\hbar^7c^6}{G_\mathrm{F}^2(m_ec^2)^5|\mathcal{M}_{fi}|^2}, \tag{6.2.26}$$

where the coefficient $0.693 = \ln 2$ is the conversion factor for the use of the half-life. The ft value varies experimentally over all β-decays from about 10^3 to 10^{20} s and we see that it mainly reflects variations in the nuclear wave-function overlap integral implicit in $|\mathcal{M}_{fi}|^2$.

The shortest are typically the so-called *super-allowed* $0^+ \to 0^+$ transitions just mentioned. Of these there are 18 known and measured decays (13 to high precision), for which the ft values are all around 3000 s; an example is the decay

$$^{14}\mathrm{O}(\beta^+)^{14}\mathrm{N}. \tag{6.2.27}$$

Note that these are all β^+ decays: given that the initial- and final-state neutron and proton occupy the same state, the Coulomb repulsion active in the proton case implies that the higher-mass nucleus is indeed that with one more proton. However, one also often finds the two exceptional cases of free neutron decay and $^3\mathrm{H}(\beta^-)^3\mathrm{He}$ included in this category; both are β^- decays.

6.2.6 The τ–θ puzzle

We now turn to one of the most spectacular discoveries in weak-interaction phenomenology: namely, parity violation. Until the 1950s all of the three discrete symmetries of Nature C, P and T where considered inviolable.* However, within the growing phenomenology of weakly interacting particles certain surprising emerged behaviour that seemed to indicate a possible violation of spatial-inversion invariance.

Let us first remind the reader that the laws of classical physics appear to respect the symmetry of spatial (and temporal) inversion usually denoted P (and T). With the discovery of antimatter we should add the symmetry under exchange of matter and antimatter C. These three discrete symmetries formed then a cornerstone of classical physics. The experimental study of the weak interaction, however, soon demonstrated that none of these is an exact symmetry of the particle world.

A puzzle involving two new subatomic particles, $\tau^{\pm}$ and $\theta^{\pm}$, arose in the early fifties (Dalitz, 1953).† Both were members of the newly found family of so-called "*strange*" particles, relatively long-lived objects which were being produced in the new accelerator experiments and which seemed to favour decays and transitions among themselves—suggestive of a new (partially) conserved quantum number: *strangeness*. Where the mass differences permitted, such particles decayed strongly into other strange states. In those cases where energy conservation did not permit such final states (as, for example, in the case of the decay of the lightest such states), the decay was always weak and strangeness was never conserved. Already then the weak interaction appears different: the newly invented quantum number strangeness was found to be exactly conserved by the strong and the electromagnetic interactions but always violated in weak processes.

The τ^+ and θ^+ were almost identical in terms of their masses ($\simeq 494\,\mathrm{MeV}$) and lifetimes ($\simeq 1.24\times10^{-8}\,\mathrm{s}$); the corresponding negatively charged antiparticles also existed with identical properties and behaviour. The two particles, however, decayed differently: 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 suggested the need to distinguish two separate particles. The fact that their lifetimes were very similar rendered the idea of two distinct but identical states (one might say "twins") very puzzling—no other pair of particles displayed such striking similarity.

Note that through the much earlier work of Noether (1918) it was understood

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

† Note that the τ described here is not to be confused with the heavy charged lepton discovered in 1977 and now also called τ. Note too that Dalitz refers to the two new particles as τ and χ.

that conservation and invariance were closely associated. Indeed, Noether's theorem states that for every continuous symmetry of a system their corresponds an *additively* conserved quantum number. In fact, this also applies to discrete symmetries, such a parity, where the quantum number is then conserved *multiplicatively*.

The need to assume two distinct states lies in the fact that the parity of the two final states had to be different (we shall see why shortly). This observation led Lee and Yang in 1956 to suggest the possibility that parity invariance or conservation may be violated and that there was only one state, which today is known as the charged kaon ($K^{\pm}$). They also suggested that if the answer to the τ–θ puzzle was indeed parity violation, then such effects might also be observable, *e.g.*, in the final-state spatial distributions for the β-decay of polarised nuclei. They proposed studying the dependence on the pseudoscalar quantity $\boldsymbol{p}\cdot\boldsymbol{s}$, where $\boldsymbol{p}$ is the final electron momentum and $\boldsymbol{s}$ the spin of the decaying nucleus.

Let us first examine just why the final-state parity assignments must differ. Experiments showed that the spin of both objects was zero; thus, the question of angular-momentum conservation is rather simple since pions too have spin zero: the total orbital angular momentum of the final state must then be zero. With no orbital contribution to the total angular momentum of the system, the overall parity is just the product of the intrinsic parities of the particles in the final state—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 overall final-state parities then depend on the *intrinsic* parity assignment of the pion.

Exercise 6.2.4. *Consider the Q-value of such a decay and show that the linear momentum available is insufficient to generate one unit ($\hbar$) of orbital angular momentum.*
Hint: *the typical maximum distance involved in an interaction between hadronic particles is* 1 fm.

It turns out that the pion is a so-called pseudoscalar, having therefore negative intrinsic parity. Consequently, in τ–θ decays the two-pion final-state parity is positive, $(-1)^2$, whereas for three pions it is negative, $(-1)^3$.

6.2.7 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 fermion and antifermion have *opposite* parities. Since fermion number is conserved (and thus only fermion–antifermion pairs may be created or annihilated), the absolute value

of, say, the electron parity is undetectable and therefore also irrelevant. By convention, the parity of fermions (antifermions) is positive (negative) although, we stress, 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 the product of the parities of the parts. The parity of a pion is therefore the product of the intrinsic parities of the quark and antiquark of which it is composed and the parity of the spatial wave-function describing their 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 (with presumably therefore zero internal orbital angular momentum and the quark and antiquark spins antiparallel), should have *negative* intrinsic parity; *i.e.* it is pseudoscalar. This is indeed experimentally verified.

The measurement of $P_\pi = -1$ is conceptually rather simple: consider associated production (via the strong interaction) of a neutron pair from a low-energy collision between a negatively charged pion and a deuteron:

$$\pi^- d \;\rightarrow\; nn. \tag{6.2.28}$$

Experimentally, the deuteron is determined to be 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, with respect to parity, the nucleons are all merely spectators in the process. What is important is the relative initial- and final-state orbital angular momenta. Now, the process is actually that of K-capture: the pion is first trapped, forming an excited *pionic* atom, and then rapidly cascades down to the lowest Bohr orbit, *i.e.* an s-wave. The total angular momentum of the initial state is thus one, but with $L = 0$ and therefore positive spatial parity. The parity of the initial and (so too) the final states is therefore precisely the pion parity. 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 overall angular momentum of the nn system must then be just one unit of $\hbar$ and this may be formed by either the sum of the two spins giving zero (singlet) and orbital motion $L = 1$ or the sum of spins giving one (triplet) and orbital motion $L = 0$, 1 or 2. For the $S = 0$ spin-singlet state, we have the following spin wave-function:

$$\tfrac{1}{\sqrt{2}} \left(|{\uparrow\downarrow}\rangle - |{\downarrow\uparrow}\rangle \right). \tag{6.2.29a}$$

Note that it is antisymmetric under interchange of the neutrons. Since they are

identical fermions, the overall wave-function must be antisymmetric and so the spatial part must be symmetric, which requires L even. The total spin can then only be even and so $S=0$, $L=1$ is excluded. For the $S=1$ triplet state, on the other hand, we have the following three possible spin wave-functions:

$$|\uparrow\uparrow\rangle, \qquad \tfrac{1}{\sqrt{2}}\left(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle\right), \qquad |\downarrow\downarrow\rangle. \tag{6.2.29b}$$

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 give total spin one. We must therefore have $S=1$, $L=1$, negative spatial parity and finally $P_\pi=-1$.

Exercise 6.2.5. *By considering the intrinsic parities of the proton and neutron, together with their known orbital and spin alignments inside the deuteron, explain why theoretically we expect* $P_d=+1$.

6.2.8 The physical consequences of parity violation

The proposal then was to measure the dependence of a decay rate on a *pseudo-scalar* quantity such as $\boldsymbol{p}\cdot\boldsymbol{s}$. The reason for this is quite simple: suppose that the transition matrix element or quantum amplitude for some decay rate (or indeed any production process) takes the general form ($\boldsymbol{p}$ and $\boldsymbol{s}$ are any pair of relevant momentum and spin vectors)

$$\mathcal{M}(\boldsymbol{p},\boldsymbol{s}) \;\propto\; a + b\,\boldsymbol{p}\cdot\boldsymbol{s}, \tag{6.2.30}$$

with a and b scalar quantities not depending (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 while $\boldsymbol{s}$, a pseudovector or axial vector, does not. The b term will thus change sign with respect to the a term under $\mathcal{P}$ and so

$$\left|\,\mathcal{M}^{\mathcal{P}}\,\right| \;\neq\; \left|\,\mathcal{M}\,\right|, \tag{6.2.31}$$

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

Now, although it is clearly not possible to apply $\mathcal{P}$, as such, experimentally, the presence of such a term would be detectable from the spatial dependence implied. For example, for a fixed spin $\boldsymbol{s}$, we have similarly that

$$\left|\mathcal{M}(-\boldsymbol{p})\right| \;\neq\; \left|\mathcal{M}(+\boldsymbol{p})\right|. \tag{6.2.32a}$$

And so it is sufficient to compare, *e.g.*, the decay rates for final-state electron

momenta ($\boldsymbol{p}$) parallel and antiparallel to the polarisation axis ($\boldsymbol{s}$) of the decaying nucleus. Note that the two statements are entirely equivalent. A final-state electron observed moving, say, parallel to the nuclear spin, moves in the opposite direction in the parity-inverted experiment. Note finally that analogously, for fixed $\boldsymbol{p}$, we also have

$$\left|\mathcal{M}(-\boldsymbol{s})\right| \neq \left|\mathcal{M}(+\boldsymbol{s})\right|. \tag{6.2.32b}$$

6.2.9 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.* at Columbia University in New York. The basic requirement of polarising the initial-state decaying nucleus then led Wu to select ^{60}Co as it has a nuclear spin of $J=5$ in natural units (with also a large magnetic moment). The experiment was by no means simple; in order to substantially polarise the cobalt specimen and avoid depolarisation by thermal motion, very low temperatures and thus new refrigeration techniques were necessary.

Measurement of 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 difficulty in control over systematics would make any difference found highly suspect. The observation rendered in Eq. (6.2.32b) was thus exploited and the polarisation axis was 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

$$^{60}\mathrm{Co} \;\rightarrow\; ^{60}\mathrm{Ni}^{*} + e^{-} + \bar{\nu}_e, \tag{6.2.33}$$

where the nickel daughter ($J=4$) is produced in an excited state. The subsequent transition to the ground state is a two-step process involving double γ emission. Both transitions are of the electric-quadrupole (E4) type with a 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 (*i.e.* horizontal) directions; the emitted photons prefer to lie in the equatorial plane rather than along the polarisation axis.

The procedure thus essentially consisted in cooling the cobalt 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. 6.9 and can be summarised via the following angular decay distribution for the electrons with respect to the polarisation direction of the ^{60}Co nuclei (as suggested by

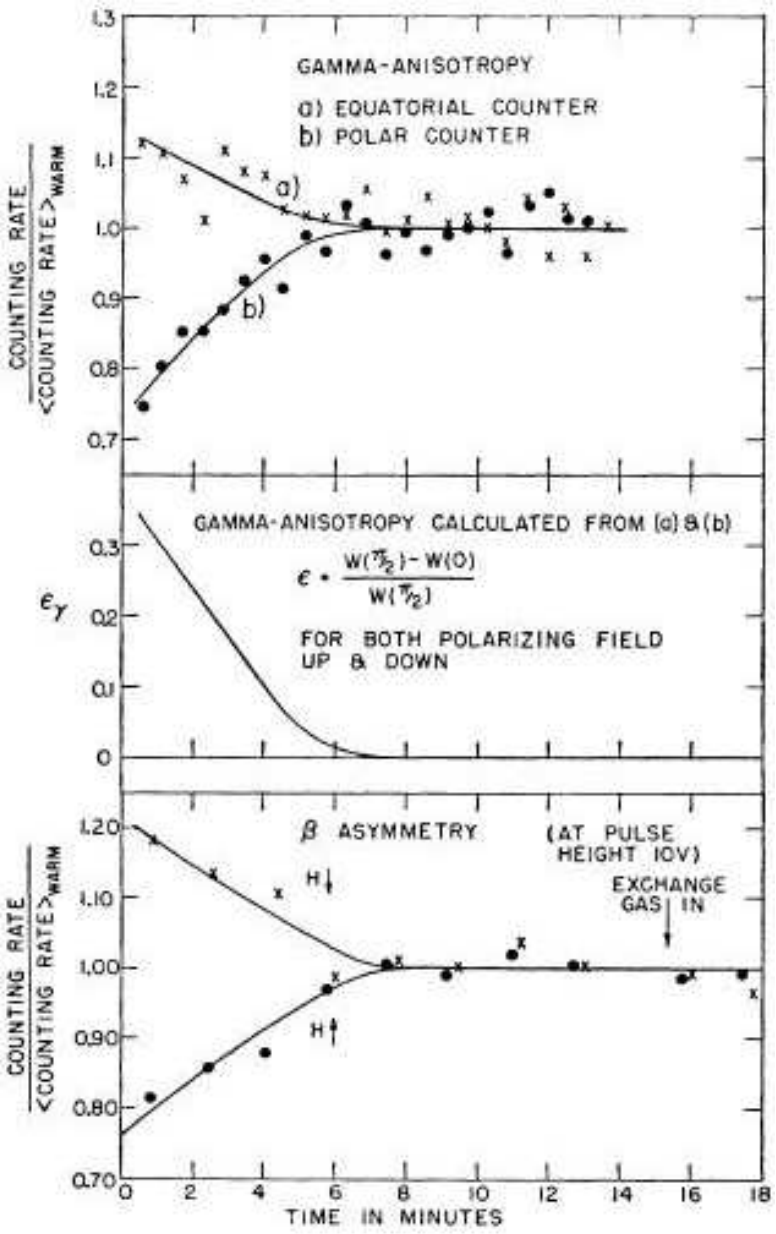


Figure 6.9: Results of the parity violation experiment performed by Wu *et al.* (1957). The γ anisotropy and the β asymmetry with the polarising magnetic field pointing up and down as functions of the time after switching off the cryostat.

Lee and Yang):

$$I(\theta) \propto 1 + \alpha \frac{v}{c} \cos\theta, \tag{6.2.34}$$

where, given $\cos\theta = \pm 1$, the asymmetry parameter α is effectively measured dir-

ectly. 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 show that, in general, $|\alpha| = 1$; *i.e.* parity is *maximally* violated in β-decay.

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

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

Note that these findings were confirmed by almost simultaneous experiments on muon decay, carried out at the same time Garwin, Lederman and Weinrich (1957). We shall discuss muon decay shortly.

> "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 have been performed thirty years earlier and it is probable that the effect had actually been observed as early as the twenties but not recognised. In fact, it had been only in the thirties with the work of Wigner that the role of parity had become to be fully appreciated.

6.2.10 The helicity of the neutrino

The experiments performed by Wu *et al.* and by Garwin *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. 6.10), combining as it does a number of non-trivial physical effects and phenomena.

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

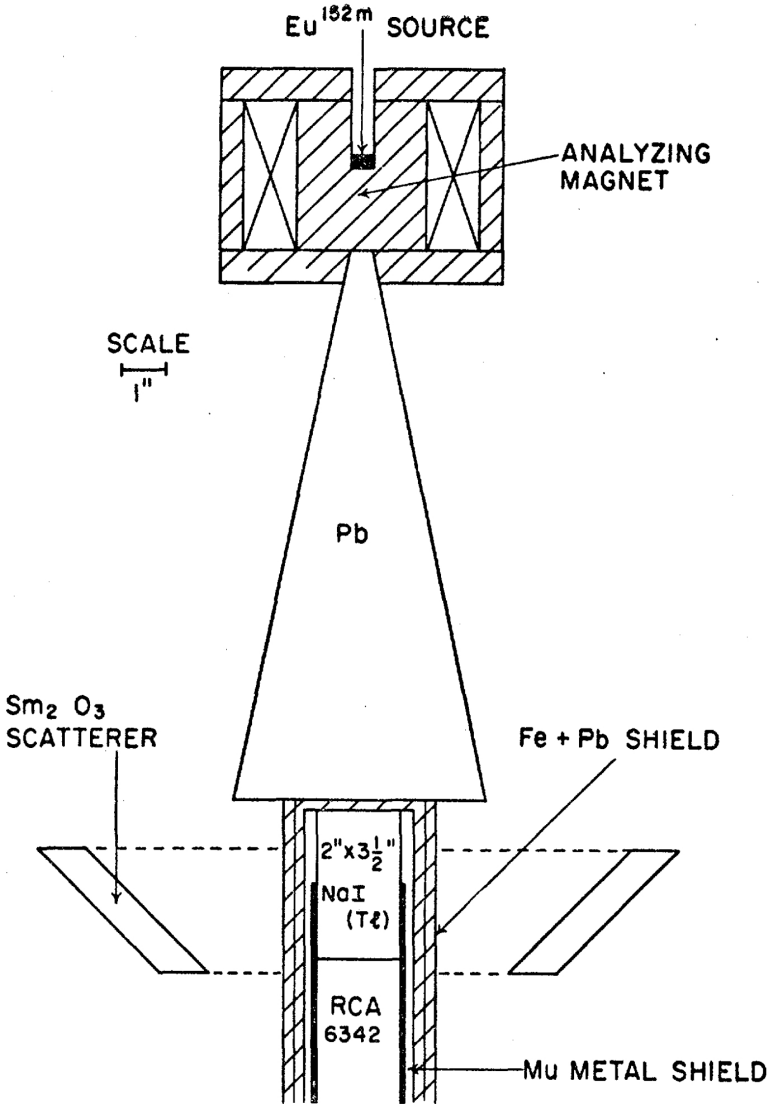


Figure 6.10: Apparatus for analysing the circular polarisation of γ-rays in the experiment to measure neutrino helicity performed by Goldhaber *et al.* (1958).

The process used was K-capture:

$$\begin{array}{rcccc} & & ^{152}\mathrm{Eu} + e^- & \rightarrow & ^{152}\mathrm{Sm}^* + \nu_e \\ & & & & \hookrightarrow\ ^{152}\mathrm{Sm} + \gamma, \\ \mathrm{J} & = & 0 & & 1 \qquad 0 \end{array} \tag{6.2.35}$$

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

The spin actually measured was that of the emitted photon. 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.

$$\begin{array}{ccccc} \longrightarrow & & \overset{\gamma}{\Rightarrow} + \overset{e^-}{\leftarrow} & & \overset{\gamma}{\Rightarrow} + \overset{e^-}{\rightarrow} \quad \ldots \\ \text{photon direction} & & \text{absorbable} & & \text{unabsorbable} \end{array} \tag{6.2.36}$$

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 (again orbital angular momentum is excluded), 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* absorption. 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 already 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).

Exactly why this should be so is not yet known. The construction of the decay matrix elements afforded by the Dirac description allows us to *describe* the effect: essentially the two terms in Eq. (6.2.30) represent the Fermi and Gamow–Teller transition amplitudes respectively and it is the interference between the two that causes the difference between the two momentum–spin alignments.* However,

* Incidentally, this indicates that, in order to detect parity violation, we must study decays in

quite why the weak interaction behaves this way is unexplained.*

6.2.11 Cabibbo theory

The final topic of this section concerns the attempts to measure the Fermi coupling constant with precision and the reconciliation of the apparent discrepancies so uncovered. As already noted, there are a variety nuclear decays that permit rather precise extraction of the value for $G_{\rm F}$. What we have not yet discussed is the broader phenomenology of the weak interaction. Just as nuclear β decay couples a hadron transition current ($n \leftrightarrow p$) to that of an electron–neutrino, other decays and interactions are observed (with comparable rates) involving, *e.g.*, a muon–muon-neutrino current and also transitions between p, n and strange baryons, such as Λ^0, $\Sigma^{0,\pm}$ and $\Xi^{0,-}$ (these last contain two strange quarks). In particular, the muon decays via

$$\mu^- \;\rightarrow\; e^- + \bar{\nu}_e + \nu_\mu. \tag{6.2.37}$$

All particles involved here are elementary; this process is therefore a perfect candidate for a precise determination of $G_{\rm F}$ and measurement of the muon lifetime leads to (Patrignani *et al.*, 2016)

$$G_{\rm F} \;=\; 1.166364(5) \times 10^{-5}\ \text{GeV}^{-2}. \tag{6.2.38}$$

Now, already at the percent precision level a discrepancy becomes manifest: the nuclear-decay data are systematically about 5% lower than rates calculated using the μ-decay value for $G_{\rm F}$. The discrepancy still stands after all possible sources of correction are taken into account. On the other hand, the measured rate for the strangeness-changing decay

$$\Lambda^0 \;\rightarrow\; p + e^- + \bar{\nu}_e \tag{6.2.39}$$

appears to be only around 5% of its predicted value—other such decays display similar values.† This observation pointed Cabibbo (1963) to the solution.

The central point here is the question of interaction basis states: the masses of the various hadronic states involved are certainly not generated by the weak interaction but are most likely a result of (yet to be understood) strong-interaction

which *both* Fermi and Gamow–Teller transitions contribute.

* We note too in passing that the opposite behaviours of the electron and positron (or neutrino and antineutrino) indicate that there is a parallel and also maximal violation of the charge-conjugation symmetry C.

† We note in passing that the energetically allowed $\Lambda^0 \rightarrow p + \mu^- + \bar{\nu}_\mu$ also occurs with a compatible rate (*i.e.* taking into account the reduced phase-space), thus providing a demonstration of the universality of the weak interaction between the leptons e and μ.

effects. However, the mass eigenstates, with which we then identify the physical particles that we detect and that propagate in the laboratory, need not necessarily coincide with the eigenstates of the weak interaction. Comparing the two decays above (n and Λ^0) we see that the final states are identical. Now, both n and Λ^0 are neutral and so may *mix* without violating charge conservation (or indeed any other well-respected conservation law) and so we may hypothesise a basis set of weak-interaction states ($\psi^{\rm W}$), in terms of which, the mass eigenstates (ψ) are

$$\psi_n \;=\; \cos\theta_{\rm C}\,\psi_n^{\rm W} + \sin\theta_{\rm C}\,\psi_\Lambda^{\rm W}, \tag{6.2.40a}$$

$$\psi_\Lambda \;=\; -\sin\theta_{\rm C}\,\psi_n^{\rm W} + \cos\theta_{\rm C}\,\psi_\Lambda^{\rm W}. \tag{6.2.40b}$$

The angle introduced is the Cabibbo angle and the construction is so designed as to maintain the correct overall normalisation. If we thus assume that it is only the *weak* neutron state that couples to a proton, then in nuclear β-decay the corresponding matrix element will be multiplied by a factor $\cos\theta_{\rm C}$, which becomes $\cos^2\theta_{\rm C}$ in the decay rate, whereas the Λ^0-decay rate will receive a factor $\sin^2\theta_{\rm C}$. So, if the value of the Cabibbo angle is such that $\sin^2\theta_{\rm C}\simeq 0.05$, then the problem is solved. The current value for the angle is (Patrignani *et al.*, 2016)

$$\sin\theta_{\rm C} \;=\; 0.2253 \pm 0.0008. \tag{6.2.41}$$

With this simple *quantum mechanical* mixing effect, the three different decay types ($n\leftrightarrow p$, strangeness-changing and purely leptonic) are reconciled with a single, universal, weak coupling constant $G_{\rm F}$. We should note that Occam is kept happy as the same angle also describes well similar decays of the pseudoscalar mesons, the pion and the kaon. Why this should be so can only be understood with advent of the quark model.

6.2.12 The extension of Cabibbo theory

The modern language for hadronic interactions uses quarks as the elementary fields. A proton is composed of two up quarks (charge $\frac{2}{3}$) and one down (charge $-\frac{1}{3}$) whereas a neutron has one up and two down; the Λ^0 hyperon contains one each of up, down and strange (also charge $-\frac{1}{3}$): thus,

$$|p\rangle \;=\; |uud\rangle, \qquad |n\rangle \;=\; |udd\rangle, \qquad |\Lambda^0\rangle \;=\; |uds\rangle. \tag{6.2.42}$$

The process of β-decay then becomes the transmutation of a d- or s-quark inside the parent hadron into a u-quark:

$$d \;\to\; u + e^- + \bar{\nu}_e \qquad\qquad \text{(neutron } \beta\text{-decay)}, \tag{6.2.43a}$$

$$s \;\to\; u + e^- + \bar{\nu}_e \qquad (\Lambda^0\ \beta\text{-decay}). \tag{6.2.43b}$$

And the Cabibbo mixing becomes quark mixing, which we now write in a more compact form:

$$\begin{pmatrix} d \\ s \end{pmatrix} = \begin{pmatrix} \cos\theta_{\rm C} & \sin\theta_{\rm C} \\ -\sin\theta_{\rm C} & \cos\theta_{\rm C} \end{pmatrix} \begin{pmatrix} d^{\rm W} \\ s^{\rm W} \end{pmatrix}. \tag{6.2.44}$$

Again, it is assumed that only the weak d-quark state couples to the u-quark.

Two questions now arise: firstly, if $d^{\rm W}$ couples to u, to what does $s^{\rm W}$ couple? The answer, provided by Glashow, Iliopoulos and Maiani (1970), is the c-quark, which was duly discovered four years later (Aubert *et al.*, 1974; Augustin *et al.*, 1974; Bacci *et al.*, 1974). The second question regards the specific (simple) form of the mixing and the possibility of violating CP (or, equivalently, T). It was, in fact, soon discovered that CP is also violated (Christenson *et al.*, 1964). Given that our method of constructing theories via current–current interactions guarantees the maintenance of CPT, a violation of CP implies a corresponding (compensating) violation of T.* Recall that in quantum mechanics the operation of temporal inversion involves complex conjugation (the relevant operator is *antilinear*) and therefore the presence of an imaginary phase in the mixing matrix would lead to T violation and could thus explain the observed CP violation.†

So, can we include an imaginary phase in the Cabibbo mixing matrix (which in principle in quantum mechanics may be complex)? The answer is no, for a 2×2 mixing matrix that is. The reason is that there are constraints (unitarity) and also phase ambiguities (of the external quark states), which reduce the number of true degrees of freedom to just the one, real, Cabibbo mixing angle. In short, if we call the matrix U, then unitarity requires $U^\dagger U = \mathbb{1}$ (which represents four constraint equations: three on Euler angles and one on a phase). Moreover, the global phases of each of the four external states (d, s, $d^{\rm W}$ and $s^{\rm W}$) are irrelevant and may be redefined at will—one, however, is a reference point; so that only three are truly free. This all leaves just one (real) Euler angle: $\theta_{\rm C}$.

The general result, well-known in matrix theory, for an $n\times n$ matrix is that the number of independent parameters is given as:

$$\tfrac{1}{2}n(n-1) \quad \text{real (Euler) angles} \tag{6.2.45a}$$

and

$$\tfrac{1}{2}(n-1)(n-2) \quad \text{imaginary phases.} \tag{6.2.45b}$$

* Compare this to what we discovered in β-decay: there both P and C are maximally violated, but in just such a manner that the product CP is preserved.

† The effect of the imaginary phase here is rather similar to that introduced earlier to describe a metastable state, where we wrote $m \equiv m_0 - {\rm i}\Gamma/2$. Clearly, the general decay process does itself provide an *arrow of time*.

For $n=2$, we see that the Cabibbo formulation is correct and minimal. However, for $n=3$, as shown by Kobayashi and Maskawa* in 1973, precisely one non-trivial imaginary phase may be introduced and thus T and consequently CP may be violated. Note that this prediction came before even the c-quark had actually been discovered. It implies though the existence of two further quarks: beauty or bottom (b) and top or truth (t), both of which have since been observed experimentally.†

6.2.13 First detection of neutrinos

In order to detect neutrinos, a reaction should be sought that is initiated by a neutrino, producing a final state that would otherwise be inaccessible. Just such a process is

$$\bar{\nu}_e + p \rightarrow n + e^+. \tag{6.2.46}$$

Note, of course, that it is actually an antineutrino that is involved here. The idea of Cowan and Reines in 1956 was then to exploit antineutrinos produced via β-decay in a nuclear reactor (typical fluxes are of order 10^{12}–$10^{13}\,\mathrm{s}^{-1}\mathrm{cm}^{-2}$). The positron so produced annihilates almost immediately with an electron, leading to two γ-rays, whose combined spectrum starts at twice the electron mass. The neutron may be detected via subsequent nuclear capture (Cowan and Reines used Cadmium), which is also accompanied by γ-emission. The near coincidence of the two events (the neutron-absorption photon is emitted within $5\,\mu\mathrm{s}$ of the positron-annihilation signal) then provides a clear and unambiguous signature for the antineutrino interaction.

* 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."

† The more commonly adopted terms are beauty and top although bottom is also often used.

† The 1995 Nobel Prize in Physics was awarded "for pioneering experimental contributions to lepton physics" jointly with one half to Martin L. Perl "for the discovery of the tau lepton" and with one half to Frederick Reines "for the detection of the neutrino." Clyde Cowan died in 1974.

6.3 Bibliography

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.

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

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

Cowan, C.L., Reines, F., Harrison, F.B., Kruse, H.W. and McGuire, A.D. (1956), *Science* **124**, 103.

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

Fermi, E. (1934), *Z. Phys.* **88**, 161; *transl.*, Wilson, F.D., *Am. J. Phys.* **36**, 1150; *see also*, *Nuovo Cim.* **11**, 1.

Gamow, G. (1928), *Z. Phys.* **51**, 204.

Gamow, G. and Teller, E. (1936), *Phys. Rev.* **49**, 895.

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

Geiger, H. and Nuttall, J.M. (1911), *Phil. Mag.* **22**, 613.

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

Goeppert-Mayer, M. (1935), *Phys. Rev.* **48**, 512.

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

Gurney, R.W. and Condon, E.U. (1929), *Phys. Rev.* **33**, 127.

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

Kurie, F.N.D., Richardson, J.R. and Paxton, H.C. (1936), *Phys. Rev.* **49**, 368.

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

Mott, N.F. and Wilson, M.H.S. (1933), *The Theory of Atomic Collisions* (Oxford U. Press).

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

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

Sargent, B.W. (1933), *Proc. Royal Soc. (London)* **A139**, 659.

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

Chapter 7

Nuclear Fission, Fusion and Energy

7.1 Nuclear fission

We have already provided a simple description of nuclear fission via the process of quantum tunnelling. Here we shall present the phenomenological details relevant to energy production and discuss some of the technical problems involved.

7.1.1 Fission and decay chains

What was discussed earlier was nuclear decay via *spontaneous* fission; however, in many cases an otherwise energetically disfavoured fission process may be assisted by the absorption of a low-energy neutron. A typical decay via *induced* nuclear fission is the following:

$$^{235}_{92}\mathrm{U} + n \;\rightarrow\; ^{93}_{37}\mathrm{Rb} + ^{141}_{55}\mathrm{Cs} + 2n. \tag{7.1.1}$$

This process is exothermic and therefore energetically allowed even for thermal neutrons. There are various combinations of possible decay products; the example exhibits the typical asymmetry in mass number A of the two daughter nuclei. In fact the fraction of fission decays producing equal or very similar mass numbers is experimentally found to be much suppressed with respect to to such a final state (by a factor of around 600).

Now, the above process actually occurs via three distinct stages. First of all, the neutron is absorbed to produce an excited state of $^{236}_{92}\mathrm{U}$. Note that the ground state $^{236}_{92}\mathrm{U}$ is actually more stable than $^{235}_{92}\mathrm{U}$ (*i.e.*, $m_{236} < m_{235} + m_n$). The excited state is however highly unstable and immediately undergoes the fission process

$$^{236}_{92}\mathrm{U}^{*} \;\rightarrow\; ^{94}_{37}\mathrm{Rb} + ^{142}_{55}\mathrm{Cs}. \tag{7.1.2}$$

Since now A is much smaller, both the daughter nuclei find themselves to be neutron rich; that is, $N/A \sim 0.61$, whereas typical values for stable nuclei lie around 0.59 for these values of A. Finally therefore, within a time period of order 10^{-16} s, the two daughter nuclei emit a variable number of so-called *prompt* neutrons (this process is described as *evaporation*). Typical mean values for the number of prompt neutrons are shown in Table 7.1. Note that while the mean number varies with

Table 7.1: Typical mean numbers ν of neutrons emitted in nuclear fission.

	^{233}U	^{235}U	^{239}Pu
ν	2.48	2.42	2.86

both A and Z, the shape of the distribution and its variance are roughly constant. Very often the decay products are also unstable and the process continues (also via β and γ decay), thus forming a so-called *decay chain*. This then can also give rise to so-called *retarded* neutrons emitted much later. Two typical examples of such chains are

$$^{93}\mathrm{Rb} \xrightarrow{6\ \mathrm{s}} {}^{93}\mathrm{Sr} \xrightarrow{7\ \mathrm{min}} {}^{93}\mathrm{Y} \xrightarrow{10\ \mathrm{hr}} {}^{93}\mathrm{Zr} \xrightarrow{10^6\ \mathrm{yr}} {}^{93}\mathrm{Nb}, \tag{7.1.3a}$$

$$^{141}\mathrm{Cs} \xrightarrow{25\ \mathrm{s}} {}^{141}\mathrm{Ba} \xrightarrow{18\ \mathrm{min}} {}^{141}\mathrm{La} \xrightarrow{4\ \mathrm{hr}} {}^{141}\mathrm{Ce} \xrightarrow{33\ \mathrm{days}} {}^{141}\mathrm{Pr}. \tag{7.1.3b}$$

The energy release or Q-value is very large and depends on the precise final state. For example, the ^{235}U fission described above (7.1.1) has $Q = 181\,\mathrm{MeV}$, which is essentially all transferred to the emitted neutrons. The average fission Q-value for ^{235}U is about 200 MeV. Recall that this energy is actually generated via the Coulomb repulsion between the daughter nuclei, which comes into play once the barrier has been overcome. The initial form is in the kinetic energy of the neutrons, but it finally emerges as both heat and radiation. Moreover, the emitted neutrons may then, in principle, provoke further fission, leading to the possibility of a *chain reaction*.

7.1.2 Controlled chain reactions

Natural uranium occurs as 99.28 % ^{238}U ($\tau_{1/2} = 4.468 \times 10^9\,\mathrm{yr}$) with only 0.72 % ^{235}U ($\tau_{1/2} = 7.038 \times 10^8\,\mathrm{yr}$). The mean number of neutrons produced in the induced fission of ^{235}U is 2.5, which means that a chain reaction is, in principle, sustainable. In principle, if the number of neutrons produced per fission is greater than one, then a chain reaction is possible.

Exercise 7.1.1. *Using the given lifetimes calculate how long ago the two isotopes would have been in equal concentrations. Of what is this figure an estimate? How*

much of the ^{238}U *will have decayed after the formation of the Earth's crust since that time? How much energy is liberated per nucleus of uranium in the* $^{238}_{92}\mathrm{U} \rightarrow {}^{206}_{82}\mathrm{Pb}$ *chain? Hint: the decays are only of the* α *and* β *type; use the formula of von Weizsäcker for the atomic masses given in Table 3.1.*

Now, the neutrons produced in nuclear fission are *fast* (they are not thermal) and, since the cross-section for induced fission by neutron absorption is inversely proportional to the velocity, this implies a very low probability of inducing further fission. That is, in a finite block of matter they would be more likely to simply escape. One thus needs a so-called *moderator*, *i.e.* a material capable of slowing down the neutrons to thermal energies (a process known as thermalisation) via elastic scattering. During this process some neutrons will still inevitably be lost and if the effective number falls below one, the chain reaction will terminate. The important properties of the moderator are that it should:

- contain relatively light nuclei to maximise the energy absorption in collisions;
- have a low cross-section for neutron capture;
- be very dense (*i.e.* solid) so as to maximise the collision probability.

A material that well fulfills these criteria is graphite(beryllium has been used, also water and heavy water): the carbon nucleus has $A = 12$, the cross-section for neutron capture is about 3.4 mb and it is a densely packed solid. This then is the basis for the first so-called *atomic pile* constructed by Fermi and colleagues in 1942 at Chicago University (in a disused rackets court); see Fig. 7.1. Control rodsmade

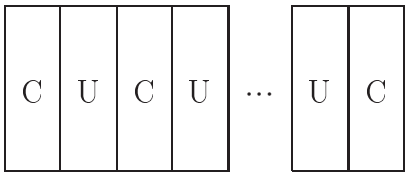


Figure 7.1: A schematic view of the first atomic pile.

of cadmium, indium, and silver were used to absorb the neutrons and to measure their flux. While cadmium and indium absorb neutrons; silver becomes radioactive when irradiated by neutrons and can therefore be used for flux measurement, *i.e.* to monitor the fission rate. The chain reaction could thus be controlled by inserting or withdrawing the rods. Note that, gambling on the validity of his calculations, Fermi had made absolutely no provision for no radiation shielding or cooling of the reactor.

Fermi's four-factor formula

It is usual to define the multiplication factor k_∞ as the number of neutrons per fission effectively available to provoke further fission, assuming an infinite-volume

sample. We then have the following three cases:

$$k_\infty \begin{cases} < 1 & \text{subcritical,} \\ = 1 & \text{critical (a nuclear reactor),} \\ > 1 & \text{supercritical (meltdown or a nuclear bomb).} \end{cases} \tag{7.1.4}$$

The effective k_∞ is a product of various factors contributing to a reduction in the number of neutrons finally available. First of all, some neutrons are eliminated via other processes, such as $^{235}\mathrm{U}(n,\gamma)^{236}\mathrm{U}$. Defining σ_A the overall cross-section for non-fission absorption and σ_F that for fission, for ν the mean number of initially produced neutrons per fission, we have the so-called *reproduction factor*

$$\eta \;=\; \nu\,\frac{\sigma_F}{\sigma_F+\sigma_A}. \tag{7.1.5}$$

For pure $^{235}\mathrm{U}$, $\sigma_A \simeq 97\,\mathrm{b}$ and $\sigma_F \simeq 584\,\mathrm{b}$ and thus for $\nu \simeq 2.5$, we have $\eta \simeq 2.08$. However, as we have seen, naturally occurring uranium is a mixture containing $99.28\,\%$ $^{238}\mathrm{U}$, for which $\sigma_A \simeq 2.75\,\mathrm{b}$ and σ_F is zero. Therefore, we should use

$$\overline{\sigma}_i \;=\; 0.0072\,\sigma_i^{235} + 0.9928\,\sigma_i^{238} \quad \text{with } i \;=\; F \text{ or } A, \tag{7.1.6}$$

which leads to

$$\overline{\sigma}_F \;\simeq\; 4.20\,\mathrm{b} \quad \text{and} \quad \overline{\sigma}_A \;\simeq\; 3.43\,\mathrm{b}. \tag{7.1.7}$$

The resulting reproduction factor is now only $\eta \simeq 1.33$. This may be improved by *enriching* natural uranium: for example, if the fraction of $^{235}\mathrm{U}$ is raised to just $3\,\%$, then we obtain $\eta \simeq 1.84$.

Next we must consider the efficiency of the thermalisation process. Now, some fast neutrons may cause fission (with cross-section $\sigma \simeq 1\,\mathrm{b}$). This provides a small increase in the reproduction factor $\varepsilon \simeq 1.03$. In graphiteabout one hundred collisions are necessary to thermalise a fast neutron. During this process, in which the energy drops from several MeV to essentially zero, the neutrons must pass through the region $10-100\,\mathrm{eV}$. Here many resonances occur, for which the capture cross-section is enormous: $\sigma \simeq 1000\,\mathrm{b}$. Therefore, in order to maximise the efficiency of the moderator, the standard geometry is one of successive layers of uranium interleaved with graphite, which does not capture neutrons. The distance over which neutrons thermalise in graphite is about 19 cm and so the typical thickness of the layers is just that.

If we define now the fractions of fast and thermal neutrons that survive capture f and p respectively, both factors turn out to be about 0.9. Putting all this

together, we finally obtain the so-called four-factor formula:

$$k_\infty \;=\; \eta\,\varepsilon\,p\,f \tag{7.1.8}$$

and inserting the numbers already given, we find $k_\infty \simeq 1.54$ (1.11) for 3 % enriched (natural) uranium.

The *six-factor* modification

One final consideration must be made before obtaining a practically useful figure: the material we are considering will not have infinite spatial extension and therefore some neutrons will simply be lost through the outer surface. If the loss fractions for thermal and fast neutrons are $l_{\rm therm,fast}$, then we have two extra depletion factors (leading to the six-factor formula), which we shall however combine here into one:

$$(1 - l_{\rm therm}) \times (1 - l_{\rm fast}) \;\sim\; (1 - l_{\rm therm} - l_{\rm fast}). \tag{7.1.9}$$

Now, the total loss $l_{\rm therm} + l_{\rm fast}$ should clearly decrease with growing linear dimension of the uranium–graphite pile and increase with the total distance that the neutrons typically travel before inducing fission. This distance (which we denote M, the so-called *migration* length) is the combined effect of simple diffusion inside the uranium layers and the distance required to thermalise inside the graphite layers. Each being a series of stochastic steps, the two contributions should be added in quadrature:

$$M \;=\; \sqrt{L_{\rm diff}^2 + L_{\rm therm}^2}\;. \tag{7.1.10}$$

For graphite, $L_{\rm therm} = 18.7\,{\rm cm}$ and $L_{\rm diff} = 50.8\,{\rm cm}$. From a dimensional analysis, we expect the neutron loss for escape to be a growing function of M/R. Numerical simulations show that for a spherical geometry (of radius R), which should minimise losses, a reasonable empirical formula for this loss factor is

$$k_\infty - k \;\simeq\; \pi^2\,\frac{M^2}{R^2}, \tag{7.1.11}$$

where k is then the reproduction factor corrected for finite-volume effects.

We may now invert the above formula to obtain an estimate of the critical dimension; that is, the size required for such a nuclear pile or core in order to sustain the chain reaction, in other words, for which $k = 1$:

$$R_c \;\simeq\; \frac{\pi M}{\sqrt{k_\infty - 1}}. \tag{7.1.12}$$

Inserting the value calculated for k_∞, one finds $R_c \sim 5\,{\rm m}$. This then is the typical

scale of the nuclear core in a fission reactor.

In order to turn this into a useful and safe power-generating device, there is one more question to be addressed: the time scale, which is an indicator of the stability of such a system. There are two basic processes that the neutrons undergo: moderation and diffusion. These have the following typical time scales:

$$\tau \sim \begin{cases} 10^{-6}\ \mathrm{s} & \text{moderation,} \\ 10^{-3}\ \mathrm{s} & \text{diffusion.} \end{cases} \tag{7.1.13}$$

Clearly, the longer of the two times (diffusion) determines the overall time scale for the chain reaction (*i.e.*, from one generation to the next). Now, at each step the differential equation governing the change in neutron number is

$$\mathrm{d}N \;=\; (kN - N)\,\frac{\mathrm{d}t}{\tau}, \tag{7.1.14}$$

for which the solution is

$$N(t) \;=\; N_0\ \mathrm{e}^{(k-1)t/\tau}\,. \tag{7.1.15}$$

Thus, if $k<1$, the number of neutrons decays exponentially and the chain reaction dies out. If $k=1$, the number is constant and we have a controlled, steady-state, self-sustaining chain reaction. However, if $k>1$ the number of neutrons (causing fission) increases *exponentially*!

Consider, *e.g.*, the case $k=1.01$, *i.e.* just a 1 % increase at each step in the number of useful neutrons produced. The exponential time constant is then $\tau/(k-1)\sim 0.1\,\mathrm{s}$ and thus after just one second the number N of neutrons is multiplied by a factor $\mathrm{e}^{10}\sim 22{,}000$. This multiplication would occur *each* and *every* second! The system as such is therefore highly unstable: just a very small upward fluctuation in the number of neutrons produced will rapidly send the reactor into a supercritical state. The solution is to provide *controllers* or *absorbers* to be inserted as and when necessary into the pile. These usually take the form of cadmium rods, which may be moved in or out of the pile to decrease or increase respectively the reaction rate.

This then is the basis of a nuclear reactor for the generation of power. The energy emerges in the form of heat (the temperature of the pile rises), which may then be converted into mechanical energy and finally into electrical power for domestic and industrial use.

7.2 Fusion

Just as fission releases energy for $A > 56$ (iron—the peak of the binding-energy curve, see Fig. 3.2), there is also an energy gain when two nuclei fuse together to produce a nucleus with $A < 56$. For example,

$$^{20}\mathrm{Ne} + {}^{20}\mathrm{Ne} \rightarrow {}^{40}\mathrm{Ca} + 20.7 \text{ MeV}. \tag{7.2.1}$$

This is equivalent to $\sim 0.5\,\mathrm{MeV}$/nucleon, which is similar to the fission energy yield. As always, it is necessary to overcome the Coulomb barrier (though now from *outside*), which in the above case is about 21.2 MeV. In other words, we must supply 21.2 MeV, for which the return is 41.9 MeV; this implies a gain of approximately a factor two.

The problem is though to overcome the barrier. One possible method might be to use an accelerator (which would then be presumably powered by part of the energy yield). Typical accelerators can produce a beam current of around 10^{-6} A. Therefore, assuming that all the beam particles actually produce a fusion reaction (which is, of course, improbable), we would generate a feeble 2 W. Another possibility is to use thermal energy: *i.e.*, one might imagine heating up a gas to a sufficient temperature. Again, using the above example we require a (thermal) kinetic energy of $\frac{1}{2} \times 21.2\,\mathrm{MeV}$, which is then given by $\frac{3}{2}kT$. We are thus required to reach $kT \simeq 7\,\mathrm{MeV}$. Translating this into a temperature, we find $T \simeq 10^{11}\,\mathrm{K}$, which is much hotter than the Sun's core and certainly difficult to attain in the laboratory.

Let us now take a look at the simplest processes. First of all, the process

$$p + n \rightarrow {}^{2}\mathrm{He} + e^{-} + \bar{\nu}_e \tag{7.2.2}$$

although having no Coulomb barrier, is not useful as ^{2}He is not stable. Another possibility might be*

$$^{2}\mathrm{H} + {}^{2}\mathrm{H} \rightarrow {}^{4}\mathrm{He} + \gamma. \tag{7.2.3}$$

Note that the helium nucleus does not have excited states and therefore the excess energy (or Q-value) must be carried away immediately and this can only be in the form of γ emission. However, here $Q = 23.8\,\mathrm{MeV}$, which is greater than either the neutron or proton separation energy and so the following processes are far more probable

$$^{2}\mathrm{H} + {}^{2}\mathrm{H} \rightarrow \begin{cases} {}^{3}\mathrm{He} + n + 3.3 \text{ MeV}, \\ {}^{3}\mathrm{H} + p + 4.0 \text{ MeV}. \end{cases} \tag{7.2.4}$$

* Here we shall use standard nuclide notation, for which deuterium is ^{2}H and tritium ^{3}H.

Clearly, the more stable the final states, the greater the energy released. And thus, energetically speaking, a better process is

$$^2\mathrm{H} + {}^3\mathrm{H} \rightarrow {}^4\mathrm{He} + n + 17.6 \text{ MeV}. \tag{7.2.5}$$

However, tritium is very active, undergoing β^- decay with a half-life of 12.3 yr. The maximum β^- energy is 18 keV with a mean value of 6 keV. As such it does not pose a serious threat to living beings; however, being a gas, it may be easily inhaled, in which case the damage it may cause is greatly enhanced. Moreover, the energy in reaction (7.2.5) is released as fast neutrons.

As already underlined, the ^{4}He nucleus is particularly stable and has a much higher binding energy per nucleon than other nearby nuclei. Indeed, as we shall see shortly, the main process in the Sun is the fusion of four protons to produce, via a series of intermediate steps and together with other products, helium:

$$4p \rightarrow {}^4\mathrm{He} + \cdots + 26 \text{ MeV}. \tag{7.2.6}$$

Similarly, the fusion of two deuterons would also provide a sizeable energy:

$$^2\mathrm{H} + {}^2\mathrm{H} \rightarrow {}^4\mathrm{He} + 24 \text{ MeV}, \tag{7.2.7}$$

the difference between the two cases being just twice the deuteron binding energy less the energy carried away by neutrinos. The remaining two possibilities are $^3\mathrm{H} + p$ and $^3\mathrm{He} + n$; however, both involve more exotic and hard-to-create initial states. In the Sun further processes are possible, the first being ^{4}He fusion:

$$^4\mathrm{He} + {}^4\mathrm{He} \rightarrow {}^8\mathrm{Be}. \tag{7.2.8}$$

This, however, is essentially undetectable since ^{8}Be decays immediately into $^4\mathrm{He} + {}^4\mathrm{He}$ ($\tau \sim 10^{-16}$ s). Moreover, the higher Coulomb barrier present for helium fusion requires even higher temperatures. A further possible process is

$$^4\mathrm{He} + {}^4\mathrm{He} + {}^4\mathrm{He} \rightarrow {}^{12}\mathrm{C}. \tag{7.2.9}$$

This however would require the proximity of three initial particles and is thus highly improbable. In the sun there is though a small dynamical equilibrium concentration of ^{8}Be and so the process

$$^4\mathrm{He} + {}^8\mathrm{Be} \rightarrow {}^{12}\mathrm{C} \tag{7.2.10}$$

also occurs, albeit to a very small degree.

7.2.1 Coulomb barrier penetration

Again, the kinematics is rather simple: for the processes described in Eqs. (7.2.4) and (7.2.5), we have a two-body non-relativistic final state and so the ratio of the kinetic energies is simply

$$\frac{K_1}{K_2} \;=\; \frac{p_1^2}{2m_1} \Big/ \frac{p_2^2}{2m_2} \;=\; \frac{m_2}{m_1}, \tag{7.2.11}$$

where we have used the fact that $p_1 = p_2$. The lighter object thus carries away the larger fraction of the energy released. Thus for the deuterium–tritium and deuterium–deuterium fusion processes described above, the final state nucleon carries away 80 % and 75 % respectively.

Using the same approximate form for the potential (or Coulomb barrier) *outside* the nucleus already used to calculate the Gamow factor in α-decay at the end of Sec. 6.1.1, we have that the height of the Coulomb barrier is just

$$V_{\rm C} \;=\; V_{\rm max}(r) \;=\; \alpha\,\frac{Z_1 Z_2}{R_1 + R_2}, \tag{7.2.12}$$

where $Z_{1,2}$ and $R_{1,2}$ are respectively the charges and radii of the two initial objects. The deuterium–tritium case will clearly have the lowest barrier, which is found to be about 0.4 MeV (note that this is still much larger than the room-temperature value of kT). The Gamow factor is then

$$G \;\simeq\; \sqrt{\frac{2m}{\hbar^2}} \int_{r_a}^{r_b} {\rm d}r\, \sqrt{V(r) - Q}, \tag{7.2.13}$$

where m is the usual reduced mass, r_a is determined from $V(r_a) = V_{\rm max}$ and r_b from $Q = Z_1 Z_2 \alpha / r_b$. We thus obtain

$$G \;\simeq\; Z_1 Z_2 \alpha \sqrt{\frac{2m}{Q}} \left[\cos^{-1} \sqrt{\frac{Q}{V_{\rm C}}} - \sqrt{\frac{Q}{V_{\rm C}} \left(1 - \frac{Q}{V_{\rm C}} \right)} \right]. \tag{7.2.14}$$

Assuming $Q \ll V_{\rm C}$, we may approximate the factor inside the square brackets by $\pi/2$. $Q = \frac{1}{2} m v^2$ and so

$$G \;\simeq\; \pi Z_1 Z_2 \alpha / (v/c), \tag{7.2.15}$$

where v is the relative velocity in the centre–of–mass system.

From the usual dimensional considerations, or from the partial-wave expansion formula *e.g.* Eq. (4.3.42), we then expect a fusion cross-section of the following

form:

$$\sigma_F \propto \frac{1}{v^2}\, e^{-2G}. \tag{7.2.16}$$

And the reaction rate is proportional to $v\sigma_F$. For so-called *thermonuclear* fusion, we need to consider the statistical velocity distribution:

$$n(v) \propto e^{-\frac{1}{2}mv^2/kT}. \tag{7.2.17}$$

The fusion rate R will then be given by

$$R = \langle v\sigma_F \rangle \tag{7.2.18}$$

We therefore have

$$R \propto \int_0^\infty v^2 \mathrm{d}v\; e^{-\frac{1}{2}mv^2/kT}\; \frac{e^{-2G}}{v}. \tag{7.2.19}$$

The first factor in the integrand strongly favours low values of v whereas the second is sharply peaked for large values—the overall effect is strong suppression at low temperatures. Moreover, evaluation of the integral shows that for mean kinetic energies less than about 1 MeV (equivalent to temperatures of around 10^{10} K) the deuterium–tritium process is favoured by a factor of about 100.

7.2.2 Plasma confinement

As already discussed, the principal difficulty in artificially inducing any such fusion process lies in the Coulomb repulsion (or barrier) between the two protons. That is, the protons and neutrons must all be brought into close proximity in order to sufficiently increase the tunnelling probability. Moreover, once formed, the helium nucleus must find a non-destructive way to release the excess energy.

The main method being researched today is that of plasma confinement* via very powerful electromagnetic fields inside a machine called a *tokamak*. The form of such a device is usually a toroidal container having dimensions ranging from the order of a very few tens of centimetres up to the 6.2 m of the International Thermonuclear Experimental Reactor (ITER)† (spherical designs are also being experimented). The plasma is then maintained in a circular orbit via magnetic fields similar to those used in the ring colliders discussed earlier (in ITER superconducting magnets will provide 11.8 T). The technological problems are enormous:

* Plasma is a very hot and dense phase of matter in which the atoms are all fully ionised.

† The ITER Tokamak, housed in laboratories sited near Marseille, will be nearly 30 metres high, and weigh 23,000 tons. The project is run and funded as a multinational collaboration involving China, the European Union, India, Japan, Korea, Russia and the United States. The current plan is to achieve the first plasma in 2025.

the temperatures and pressures needed are close to those found in the sun! In fact, no machine has yet been constructed that is capable of sustaining (for a significant period) the fusion process beyond the so-called *break-even* point (the point at which the energy produced exceeds that consumed).

Other possible methods, as already mentioned, include the acceleration of protons and/or deuterons via traditional techniques. The energetic particles are then made to collide with a suitable target where the fusion process may take place. Such projects are, however, still at a very early experimental stage. It has been suggested that deuterium absorbed into pellets of palladium might be subject to very intense local electric field capable of squeezing two such nuclei close enough to provoke so-called *cold* fusion. Unfortunately, the reproducibility of the claimed experimental results has been thrown into serious doubt and as yet no group has produced significant data.

7.3 The Sun

7.3.1 Basic parameters

The Sun, the centre of the Solar System, is the nearest star to the Earth, from where we observe it at a distance of very nearly $1.5 \times 10^{12}\,\mathrm{m}$ (or equivalently 499 light-seconds). It has a radius of $6.96 \times 10^{8}\,\mathrm{m}$ (about 109 times that of Earth) and mass $1.99 \times 10^{30}\,\mathrm{kg}$, (330000 times that of Earth), which accounts for about 99.86 % of the total mass of the Solar System. The composition of the Sun by mass is about 73.5 % hydrogen, 24.9 % helium and the remaining 1.7 % consists of heavier elements, including oxygen, carbon, neon, iron *etc.* The mean density is thus about $1.4\,\mathrm{g/cm^3}$, which is not so different to that of water at standard temperature and pressure, but only a quarter of that of Earth (the core though is certainly much more dense). Two data that will be important for our discussion are the temperature and so-called *luminosity*. What can be measured is the surface temperature $\simeq 5.8 \times 10^{3}\,\mathrm{K}$ (though again the core temperature is certainly much higher) and the apparent (*i.e.* as measured from Earth) luminosity (or flux) $F \simeq 1.4\,\mathrm{kW/m^2}$.

Of course, the quantity that we need is the absolute luminosity L, which is just the total irradiated power. This may however be easily calculated from the apparent luminosity:

$$L \;=\; 4\pi R^2\, F \;=\; 3.86 \times 10^{26}\ \mathrm{W}, \tag{7.3.1}$$

where R is the Earth–Sun distance. This implies (via the Einstein relation) a

continuous loss of mass given by

$$\frac{\mathrm{d}m}{\mathrm{d}t} = -4.2 \times 10^9 \text{ kg/s}. \tag{7.3.2}$$

While this may seem enormous, written in fractional form it is actually seen to be minute:

$$\frac{1}{m}\frac{\mathrm{d}m}{\mathrm{d}t} = -2 \times 10^{-21} \text{ s}^{-1}. \tag{7.3.3}$$

In other words the maximum lifetime of the Sun is of order 10^{21} s, which is very much longer than the present age of the universe ($\sim 4.3 \times 10^{17}$ s).

Exercise 7.3.1. *Show that, with such an energy output, it is impossible that the Sun be fuelled by chemical reactions.*

The power output is thus about 2.4×10^{45} eV/s. If we now assume that a nuclear reaction typically releases around 1 MeV, then such an output requires of order 2×10^{39} reactions/s. It is reasonable to assume that each nucleus in the Sun may partake in one such reaction, after which it should be considered spent. We may estimate the number of nuclei available as follows:

$$\frac{M_\odot \times N_\mathrm{A}}{\langle A \rangle}, \tag{7.3.4}$$

where $M_\odot$ is the mass of the Sun (given above), $N_\mathrm{A} = 6 \times 10^{26}$ is Avogadro's number and $\langle A \rangle \simeq 2$ is an approximate average atomic mass number for nuclei in the sun. We thus find that there are about 5×10^{56} nuclei available, which would imply a maximum lifetime for energy production in the Sun of order 10^{10} yr, a much smaller but still acceptable value.

A star such as the Sun is formed owing to the gravitational attraction acting on a cloud of (hydrogen) gas, which is thus unstable against collapse. As the gas collapses potential energy is converted into kinetic energy and the system heats up. In the case of the sun a point is reached where the temperature and density of the plasma formed is sufficient to initiate and maintain fusion processes, which then further heat the system and thus provide sufficient pressure to contrast the gravitational collapse and allow the creation of an equilibrium state. Note, however that the condition is not one of thermodynamic equilibrium as the centre is much hotter than the surface. The energy continuously produced in the core of the Sun propagates via radiation, conduction and convection towards the surface, where it is liberated as the radiation that we see and feel from our viewpoint on Earth.

7.3.2 Thermonuclear cycles

There are two energy-producing cycles in the Sun (Bethe, 1939a,b):* the principal process is the so-called $p-p$ cycle. This is composed of three distinct steps:

$$p + p \rightarrow {}^2\mathrm{H} + e^+ + \nu_e \qquad Q = 0.42 \text{ MeV}, \tag{7.3.5a}$$

$$p + {}^2\mathrm{H} \rightarrow {}^3\mathrm{He} + \gamma \qquad Q = 5.49 \text{ MeV}, \tag{7.3.5b}$$

$${}^3\mathrm{He} + {}^3\mathrm{He} \rightarrow {}^4\mathrm{He} + 2p \qquad Q = 12.86 \text{ MeV}. \tag{7.3.5c}$$

To this must be added the rapid annihilation of the positron produced in the first step with an electron, producing two photons and additional energy:

$$e^+ + e^- \rightarrow 2\gamma \qquad Q = 1.02 \text{ MeV}. \tag{7.3.5d}$$

Taking into account the final two protons (thus returning to the system), we see that the overall effect can be summarised as follows:

$$4p \rightarrow {}^4\mathrm{He} + 2\nu_e + 6\gamma \qquad Q = 26.73 \text{ MeV}. \tag{7.3.5e}$$

The two neutrinos emitted at the first step each carry away an average 0.26 MeV, an energy which is then essentially lost as neutrinos can typically traverse the entire universe without interacting at all. This leaves an effective 6.55 MeV produced per proton consumed. The above is known as the PP-I chain and is dominant for temperatures in the range $(10-14)\times10^6$ K; it is therefore the principal source of energy produced by the Sun. However, in general, the last step may be replaced by others, leading to another two observed branches and one possible, but unobserved, final state.

The PP-II branch, dominant for temperatures in the range $(14-23)\times10^6$ K, proceeds as follows:

$${}^3\mathrm{He} + {}^4\mathrm{He} \rightarrow {}^7\mathrm{Be} + \gamma \qquad Q = 1.59 \text{ MeV}, \tag{7.3.6a}$$

the beryllium nucleus may then capture a free electron in the plasma,

$${}^7\mathrm{Be} + e^- \rightarrow {}^7\mathrm{Li} + \nu_e \qquad Q = 0.86 \text{ MeV}, \tag{7.3.6b}$$

* The 1967 Nobel Prize in Physics was awarded to Hans Bethe "for his contributions to the theory of nuclear reactions, especially his discoveries concerning the energy production in stars."

lithium-7 soon breaks up via proton absorption,

$$^7\mathrm{Li} + p \;\rightarrow\; ^4\mathrm{He} + ^4\mathrm{He} \qquad\qquad Q \;=\; 17.35\ \mathrm{MeV}. \tag{7.3.6c}$$

Here 90 % (10 %) of the neutrinos produced in the second step carry away an energy of 0.861 MeV (0.383 MeV), depending on whether the lithium-7 is left in the ground state or an excited state, respectively.

The PP-III branch, dominant for temperatures above about 23×10^6 K, is as follows:

$$^7\mathrm{Be} + p \;\rightarrow\; ^8\mathrm{B} + \gamma \qquad\qquad Q \;=\; 0.14\ \mathrm{MeV}, \tag{7.3.7a}$$

$$^8\mathrm{B} \;\rightarrow\; ^8\mathrm{Be}^* + e^+ + \nu_e \qquad\qquad Q \;=\; 14.02\ \mathrm{MeV}, \tag{7.3.7b}$$

$$^8\mathrm{Be}^* \;\rightarrow\; ^4\mathrm{He} + ^4\mathrm{He} \qquad\qquad Q \;=\; 3.03\ \mathrm{MeV}, \tag{7.3.7c}$$

Again, we must add the annihilation of the positron produced in the second step with an electron, producing two photons and the usual energy:

$$e^+ + e^- \;\rightarrow\; 2\gamma \qquad\qquad Q \;=\; 1.02\ \mathrm{MeV}. \tag{7.3.7d}$$

The PP-IV (or HEP) branch is theoretically predicted as possible, but has yet to be observed:

$$^3\mathrm{He} + p \;\rightarrow\; ^4\mathrm{He} + e^+ + \nu_e \qquad\qquad Q \;=\; 18.80\ \mathrm{MeV}. \tag{7.3.8}$$

To this we may add another very rare process: *pep* (a form of electron capture) for the first step in the production of deuterium:

$$p + e^- + p \;\rightarrow\; ^2\mathrm{H} + \nu_e \qquad\qquad Q \;=\; 1.44\ \mathrm{MeV}. \tag{7.3.9}$$

Now, returning to the PP-I cycle, we note that the first step (7.3.5a) is a weak process (in addition to having the Coulomb barrier to overcome) while the others are all due to the strong interaction. The time scale of the single cycle is thus determined by that of this step. The cycle frequency may be most easily (and reliably) estimated as the total power output divided by the energy released per cycle:

$$f_{pp} \;\simeq\; P_\odot^{\mathrm{tot}} / Q_{pp}^{\mathrm{tot}} \;\simeq\; 10^{38}\ \mathrm{Hz}. \tag{7.3.10}$$

Assuming the Sun to be initially more or less all protons, we can estimate the true

expected lifetime:

$$\tau \sim \frac{M_\odot}{4m_p}\frac{1}{f} \simeq 3\times 10^{18}\ \text{s} \simeq 10^{11}\ \text{yr}. \tag{7.3.11}$$

This rather long time is due to the one very slow weak-decay step in the PP-I chain, which acts as a bottle neck to the entire process—fortunately for us on Earth!

There is one further chain, which is responsible for approximately 2 % of the energy production in the Sun, the so-called carbon–nitrogen–oxygen (CNO) cycle. Theoretical models show that this should dominate for slightly higher temperatures than those found in the Sun ($\gtrsim 17\times 10^6\,\text{K}$). As with the PP cycles, there are different possible branches of the CNO cycle too; here we shall just discuss the main CNO-I branch. In this cycle the carbon, nitrogen and oxygen components remain constant on average and essentially merely act as catalysts for the continuous conversion of protons into helium. The six steps, thus forming a continuous loop, are:

$$^{12}\text{C} + p \rightarrow {}^{13}\text{N} + \gamma, \tag{7.3.12a}$$
$$^{13}\text{N} \rightarrow {}^{13}\text{C} + e^+ + \nu, \tag{7.3.12b}$$
$$^{13}\text{C} + p \rightarrow {}^{14}\text{N} + \gamma, \tag{7.3.12c}$$
$$^{14}\text{N} + p \rightarrow {}^{15}\text{O} + \gamma, \tag{7.3.12d}$$
$$^{15}\text{O} \rightarrow {}^{15}\text{N} + e^+ + \nu, \tag{7.3.12e}$$
$$^{15}\text{N} + p \rightarrow {}^{12}\text{C} + {}^{4}\text{He}. \tag{7.3.12f}$$

The nett effect is

$$4p \rightarrow {}^{4}\text{He} + 2e^+ + 2\nu_e + 3\gamma. \tag{7.3.13}$$

Note, of course, that the neutrinos produced have different energies to those of the PP cycles.

As we have already noted, in order to fuse, the initial-state nuclei must overcome the Coulomb barrier, requiring that they move into very close proximity. This necessitates large energies and therefore, for thermally driven fusion, very high temperatures—note that, owing precisely to the presence of the Coulomb barrier, the cross-section is indeed a growing function of energy in the relevant region. It is estimated that the core temperature of the sun is of order $10^7\,\text{K}$, which is equivalent to $kT \sim 1\,\text{keV}$. such an energy is still clearly far below the Coulomb barrier, but the tails of the statistical distribution evidently contain enough high-energy nuclei to sustain the fusion processes.

7.3.3 Stellar models

As already indicated, stellar models assume a natural equilibrium between gravitational forces and thermodynamic pressure:

- Suppose that, via natural fluctuations, the nuclear processes were to slow down, the reduced energy production would then allow the temperature to fall and thus too the pressure; the gravitational attraction would then compress the core causing the pressure and temperature to rise, which would lead to increased kinetic energies and thus increased fusion rates.
- If instead the nuclear processes were to accelerate, the extra heating would cause the temperature to rise and thus too the pressure; the gravitational attraction would then be partially overcome and the core would expand, lowering the pressure and temperature, which would in turn lead to decreased kinetic energies and decreased fusion rates.

A complete stellar model must clearly necessarily exploit and combine our knowledge of:

- thermodynamics,
- nuclear and particle physics,
- fluid dynamics.

Put all together, these lead to what is known as the standard solar model (SSM). While we cannot provide a complete description in this brief section, we shall explain the underlying ideas.

We shall start with an estimate of the equilibrium energies: the easiest method is based on the virial theorem,* according to which the average potential and kinetic energies are roughly equal: thus,

$$E_{\text{grav}} \approx 2E_{\text{kin}}. \tag{7.3.14}$$

Assuming the Sun to have a uniform spherical mass distribution, the gravitational potential energy is given by

$$E_{\text{grav}} = \frac{3}{5}\frac{GM_\odot^2}{R_\odot} \sim 10^{41}\ \text{J}. \tag{7.3.15}$$

* The virial theorem actually states that for a potential of the form r^p, the average potential and kinetic energies are related by $pE_{\text{pot}} = 2E_{\text{kin}}$.

Exercise 7.3.2. *Derive the above formula.*
Hint: *consider adding an initially infinitely separated spherical shell of mass of thickness* $\mathrm{d}r$ *to an existing uniform spherical distribution of radius* r.

The total kinetic energy, assuming the Sun to be mainly protons, is

$$E_{\text{kin}} \;=\; N_p \,\langle \tfrac{1}{2} m_p v_p^2 \rangle. \tag{7.3.16}$$

Using the fact that $M_\odot \equiv N_p m_p$, we may thus deduce the mean velocity of protons within the Sun:

$$\langle v_p^2 \rangle^{1/2} \;\sim\; 3 \times 10^5 \;\text{ms}^{-1}, \tag{7.3.17}$$

which implies that they are non-relativistic. On the other hand, if we consider the mean kinetic energy, we find

$$\langle \tfrac{1}{2} m_p v_p^2 \rangle \;\sim\; 0.5 \;\text{keV}. \tag{7.3.18}$$

Such a high kinetic energy as compared to typical ionisation energies implies that the protons in the Sun's core are almost entirely ionised and thus we have a plasma. Assuming further that

$$\langle \tfrac{1}{2} m_p v_p^2 \rangle \;\sim\; \tfrac{3}{2} kT, \tag{7.3.19}$$

we find $T \sim 5 \times 10^6\,\text{K}$ as the mean temperature inside the Sun's core.

7.3.4 The solar neutrino problem

From the SSM, one can calculate the solar temperature as a function of r and deduce the surface temperature from the Stefan–Boltzmann law

$$W \;=\; \sigma T^4, \tag{7.3.20}$$

where, we recall, Stefan's constant is

$$\sigma \;=\; \frac{2\pi^5 k^4}{15 c^2 h^3} \;=\; 5.67 \times 10^{-8} \;\text{J s}^{-1}\text{m}^{-2}\;\text{K}^{-4}. \tag{7.3.21}$$

From these data, the SSM leads to a temperature at the centre of the Sun

$$T_\odot^{\text{core}} \;=\; 1.5 \times 10^7 \;\text{K} \tag{7.3.22}$$

The importance of this result is to indicate the dominant process in the energy production of the Sun, which is thus confirmed as the PP-I cycle. Once the production mechanism is known, the neutrino flux is then rather precisely determined by the measured power output. We just need to add a final observation on the

origins of the different parts of the neutrino spectrum:

$$\text{low } E \leq 0.42 \text{ MeV}: \qquad \text{(PP-I cycle)} \tag{7.3.23a}$$

and

$$\text{high } E = \text{a few MeV}: \qquad \text{(CNO and other cycles).} \tag{7.3.23b}$$

The first measurements of the solar-neutrino flux date back to 1964 and the experimental work of Davis*at the Homestake Gold Mine in Lead, South Dakota, see also the related theoretical calculations of Bahcall (1964). Using two 500-gallon tanks of perchloroethylene (a fluid commonly used for dry-cleaning) as detectors, Davis measured the rate of the production of ^{37}Ar from the neutrino-capture reaction

$$\nu_e + {}^{37}\text{Cl} \rightarrow {}^{37}\text{Ar} + e^-. \tag{7.3.24}$$

The findings were

$$\Phi_\nu^{\text{expt}} \sim 0.4\,\Phi_\nu^{\text{th}}. \tag{7.3.25}$$

It should be noted, however, that the neutrinos detected by Davis were of high energy ($>0.814\,$MeV, the threshold for the reaction) that is, those generated by the CNO cycle. For these the temperature dependence of the predicted flux is particularly strong:

$$\Phi_\nu^{\text{th}} \propto T_{\text{core}}^{18}. \tag{7.3.26}$$

Such sensitivity to the model details made it necessary to carry out experiments on the lower-energy neutrinos; *i.e.* those produced in the PP-I cycle. One such measurement was performed by the GALLEX (GALLium EXperiment) collaboration at the INFN Gran Sasso National Lab., which ran between 1991 and 1997. The detector consisted of 30 t of gallium, which is sensitive to very low-energy neutrinos, and the process exploited was

$$\nu_e + {}^{71}\text{Ga} \rightarrow {}^{71}\text{Ge} + e^-, \tag{7.3.27}$$

for which the threshold is 0.23 MeV. It is therefore sensitive to neutrinos produced in the PP-I chain, for which the upper energy limit is 0.42 MeV.

The PP-I solar neutrinos, via the above weak interaction, transformed less than one gallium nucleus per day into germanium, which after chemical extraction is detected via observation of its radioactive decay products. The theory in this case is much less susceptible to uncertainty. Most of the energy irradiated by the Sun is produced by the PP-I cycle, in which each step produces about 25 MeV and

* The 2002 Nobel Prize in Physics was awarded one half jointly to Raymond Davis Jr. and Masatoshi Koshiba "for pioneering contributions to astrophysics, in particular for the detection of cosmic neutrinos." and the other half to Riccardo Giacconi "for pioneering contributions to astrophysics, which have led to the discovery of cosmic X-ray sources."

exactly two electron neutrinos of 0.41 MeV. Since, as already calculated, the power output of the Sun is 2.4×10^{45} eV/s, the fusion process must have a rate of about 10^{38} cycles/s and so we may expect a neutrino production rate $R_\nu^\odot \simeq 2\times10^{38}\,\nu_{\rm e}$/s. The expected solar-neutrino flux at Earth is then

$$\Phi_\nu^{\rm Earth} \simeq \frac{R_\nu^\odot}{4\pi r_{S-E}^2} \simeq 6\times10^{10}\ \nu/{\rm cm}^2/{\rm s}. \tag{7.3.28}$$

The GALLEX collaboration recorded just around two thirds of this figure (see, eg, Hampel *et al.*, 1999), thus confirming Davis' and Bahcall (1964)'s results. That only one third of the neutrinos are "lost" can be explained via mixing angles, which may reduce the oscillation effects.

The true significance of Davis' findings was long in being understood owing to the various theoretical problems surrounding its interpretation. First of all, the question of the SSM is central: the comparison is always with a theoretically estimated flux based on our description of what happens inside the Sun. Clearly, the GALLEX results essentially sweep away such doubts, as the sensitivity of the PP-I dominance to the temperature is not so delicate.

There is, however, one further uncertainty: our poor knowledge of the nuclear neutrino-capture cross-section for these energies. Owing to the very low cross-sections, neutrino measurements are always fraught with difficulties and so the cross-sections used to estimate the neutrino fluxes (for all experiments) are again essentially based on theoretical estimates, which may or may not be correct, depending, as they do, on our (poor) knowledge of the nuclear wave-function. For a long period there was thus a tendency to dismiss the results as being due to inaccuracies in the description of either the Sun or the low-energy nuclear neutrino cross-sections.

Somewhat after the GALLEX results were published the Sudbury Neutrino Observatory (SNO) produced more convincing evidence for oscillation effects. The SNO detector consisted of 1,000 tonnes of heavy water, held in a 6-metre radius container situated in the Creighton Mine in Sudbury, Ontario, Canada. The deuterium in the target can interact with neutrinos via both charged-current ($W^\pm$) and neutral-current (or Z^0) exchange, which is then sensitive to all neutrino types. This experiment was thus able to confirm the overall solar neutrino flux while at the same time measuring the electron-neutrino flux.

If the results are taken at face value, one is forced to seek a particle-physics explanation. The most natural (since already observed in other systems) is the phenomenon of (quantum) oscillation.* Put simply, the idea (already proposed in

* 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."

1957 by Pontecorvo) is that a quantum-mechanically mixed state (*i.e.* via superposition of states from a different basis) may transform itself into another over some time interval. Since the experimental apparatus used by Davis was typically only sensitive to electron neutrinos, if the transformation is into, say, muon neutrinos, these will then go undetected and the measured flux will be reduced. A necessary condition for such oscillation is that the neutrino have a non-zero mass. Such an explanation then subverts one of the basic assumptions of the so-called standard model (SM) of particle physics: namely, that the neutrino be massless. Of course, there is actually no experimental result that requires the neutrino to be massless and we otherwise only had (albeit extremely low) upper limits on its mass.

Further experimentally, confirmation that this might indeed be the explanation came from the Kamiokande and Super-Kamiokande experiments in Japan, see for example Shiozawa (1999) other experiments produced similar results. The idea there was to measure the flux of neutrinos produced by cosmic-ray collisions in the atmosphere (so-called atmospheric neutrinos) and compare the results for those produced immediately overhead with those coming from the other side of the Earth. The time difference is sufficient to allow those travelling further to undergo the transmutation mentioned and thus at least partially "disappear". We shall now briefly explain the quantum mechanics of such oscillation phenomena.

7.3.5 Quantum oscillation

The double potential well in quantum mechanics nicely demonstrates one of the more surprising phenomena associated with quantisation: namely, 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 region of the wells is very close 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. 7.2.

The time-dependent solutions may thus be represented generically as

$$\psi_{0,1}(x,t) \;=\; u_{0,1}(x)\;\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_{0,1}t}, \tag{7.3.29}$$

where the precise form of $u_{0,1}(x)$ is entirely irrelevant for the present purposes.

These then represent the "unperturbed" energy 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 a pure eigenstate. In other words, a different basis

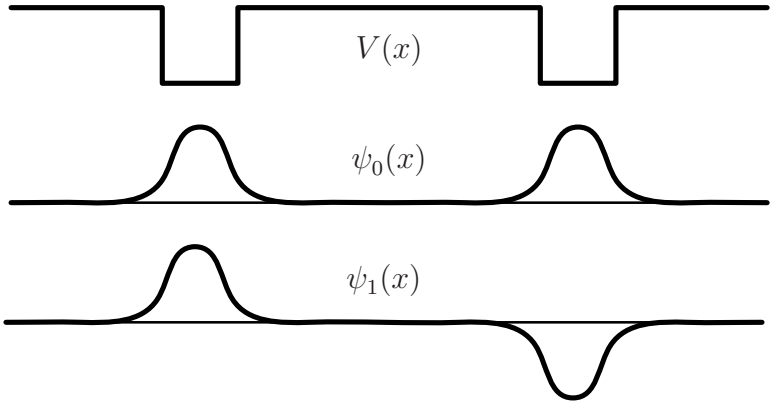


Figure 7.2: The double potential well and the x dependence of the corresponding two lowest-energy eigenstates.

(it might be called the "occupation" basis) is necessary to describe such an external interaction:

$$u_{L,R}(x) \simeq \tfrac{1}{\sqrt{2}}\Big[u_0(x) \pm u_1(x)\Big]\,. \tag{7.3.30}$$

Such a state describes the system at the instant $t=0$, when the particle is introduced into the left- or right-hand well. This cannot now be a stationary solution and at later times the evolution will be given by (we shall take the left-hand case for definiteness)

$$\psi_L(x,t) = \tfrac{1}{\sqrt{2}}\Big[u_0(x)\,\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_0 t} + u_1(x)\,\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_1 t}\Big]\,. \tag{7.3.31}$$

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:

$$\begin{aligned}\psi_L(x,t) &= \tfrac{1}{2}\Big[\Big(u_L(x)+u_R(x)\Big)\,\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_0 t} + \Big(u_L(x)-u_R(x)\Big)\,\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_1 t}\Big] \\ &= \tfrac{1}{2}\Big[u_L(x)\Big(\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_0 t}+\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_1 t}\Big) + u_R(x)\Big(\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_0 t}-\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_1 t}\Big)\Big]\,.\end{aligned} \tag{7.3.32}$$

Let us now introduce the average energy $E \equiv \frac{1}{2}(E_0+E_1)$ and the energy difference $\Delta E \equiv E_1 - E_0$. We then finally have

$$\psi_L(x,t) = \Big[u_L(x)\cos\tfrac{1}{2\hbar}\Delta E t - \mathrm{i}\,u_R(x)\sin\tfrac{1}{2\hbar}\Delta E t\Big]\,\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}Et}\,. \tag{7.3.33}$$

The coefficients of $u_{L,R}(x)$, which determine the probabilities of finding the particle in the left- and right-hand wells, therefore oscillate:

$$P_L = \cos^2\tfrac{1}{2\hbar}\Delta E t = \tfrac{1}{2}\big[1+\cos\tfrac{1}{\hbar}\Delta E t\big]\,. \tag{7.3.34a}$$

$$P_R \;=\; \sin^2 \tfrac{1}{2\hbar}\Delta E t \;=\; \tfrac{1}{2}\left[1+\cos \tfrac{1}{\hbar}\Delta E t\right]. \qquad (7.3.34b)$$

Note that the frequency is thus $\frac{1}{\hbar}\Delta E$ (and not half that). In other words, the particle effectively oscillates between the two potential wells with a frequency equal to the difference of the natural frequencies associated with the two states.

To conclude this excursus, let us underline a common aspect of such phenomena: the original system possessed a symmetry (in this case under parity reversal), which the external interaction (*i.e.*, the process by which the physical state is created) does *not* respect. This is a sufficient (and indeed necessary) condition for the induced mixing, which lies at the heart of the phenomenon of quantum oscillation. Note too that this is basically the same phenomenon known as "*beating*", which occurs when two oscillating systems with slightly differing naturally frequencies (or modes) are coupled.

In order to translate the above formalism to the phenomenon of particle oscillation, one may think of the distinct occupation of the two potential wells as representing the two different weak-interaction eigenstates (ν_e and ν_μ, say*), while the two superpositions (ψ_0 and ψ_1) represent the mass eigenstates. Now, the real neutrino situation is slightly more complicated: two such bases may, in general, be related by a mixing angle (analogous to that of Cabibbo). In the above simple example the two basis sets used are actually precisely equivalent to a rotation through 45° ($\sin\frac{\pi}{4}=\cos\frac{\pi}{4}=\frac{1}{\sqrt{2}}$) and is thus a very particular case. In general therefore, the oscillation probability is multiplied by a factor $\sin^2 2\theta$, where θ is the mixing angle between the two basis states. It also turns out that, given the very small mass of neutrinos compared to their kinetic energy in typical experimental situations, a better description is obtained by considering momentum eigenstates and not those with respect to energy as here. Finally, just as in the quark mixing case, there are actually three basis states and so the mixing requires a three-by-three matrix, which also allows for CP violation in the neutrino (or lepton) sector.

Applying all this to the neutrino case (and simplifying to just two neutrino species we find that in the limit $E_\nu \gg m$ (taking $v_\nu \simeq c$) we may rewrite the time-dependent transition probability as

$$P(\nu_e \to \nu_\mu; t) \;\simeq\; \sin^2 2\theta \, \sin^2\!\left(\frac{1.27\Delta m^2 L(t)}{E_\nu}\right), \qquad (7.3.35)$$

where $\Delta m^2 = m_2^2 - m_1^2$ is in eV^2, $L(t)$ is the distance travelled in metres, E_ν is the neutrino energy in MeV and θ is the relevant mixing angle.

* Note that what we define to be, say, a muon neutrino is the state that is produced in association with a muon; *i.e.*, it is defined by the weak interaction and therefore need not necessarily be a mass eigenstate.

Exercise 7.3.3. *Using the earlier formula for the oscillation probability, but with a non-trivial mixing angle, derive the above formula.*
Hint: *consider the two neutrino states as eigenstates of momentum ($\boldsymbol{p}$ say) and use the Einstein relation $E^2 = \boldsymbol{p}^2 + m^2$ (in the limit $E \gg m$), to rewrite the energy difference as a mass-squared difference.*

7.4 Bibliography

Bahcall, J.N. (1964), *Phys. Rev. Lett.* **12**, 300.

Bethe, H.A. (1939a), *Phys. Rev.* **55**, 103.

Bethe, H.A. (1939b), *Phys. Rev.* **55**, 434.

Davis, J., Raymond (1964), *Phys. Rev. Lett.* **12**, 303.

Hampel, W. *et al.*, GALLEX Collab. (1999), *Phys. Lett.* **B447**, 127.

Pontecorvo, B. (1957), *Zh. Eksp. Teor. Fiz.* **33**, 549; *transl.*, *Sov. Phys. JETP* **6**, 429.

Shiozawa, M., Super-Kamiokande and Kamiokande Collabs. (1999), *Nucl. Instrum. and Meth.* **A433**, 307.

Chapter 8

The Standard Model

In this final, very brief, chapter we shall simply outline the constituents and general form of the so-called *SM* of particle physics. While there is insufficient time in such a course as the present to delve into detailed explanations or derivations, it is hoped to provide the student with a meaningful glimpse of the picture we now have of the fundamental interactions in physics and at least a knowledge of the elementary particles comprising the SM.

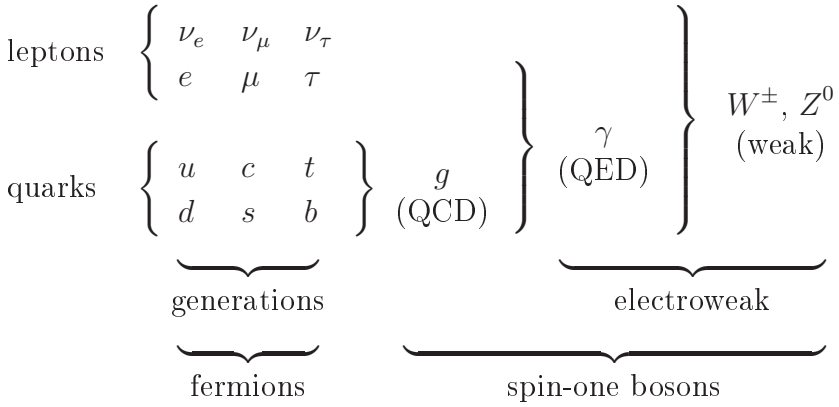
8.1 Fundamental forces and particles

8.1.1 The table of forces and particles

The building blocks we now have in our hands consist in a number of quarks and leptons, interacting via the exchange of several different types of spin-one bosons. Leaving aside gravity and not yet wishing to comment on the last state to be experimentally identified (the Higgs boson), these are grouped according to type and generation (or family) in Table 8.1.

Each generation comprises four fermions (and their antiparticles): a down-type quark (electric charge $-1/3$), an up-type quark (charge $+2/3$), a charged lepton (charge -1) and a zero-charge neutrino. Each lepton and quark pair are to be considered as interaction partners in the sense that the corresponding weak states transform into one another under the effect of the weak interaction. However, as we have seen earlier, the states listed (with the exception of the neutrinos) actually correspond to mass eigenstates while the weak-interaction basis involves a rotation or mixing (CKM) matrix. As we move down the table the mass increases, as too as we progress from left to right across the generations (although little is yet known of the neutrino masses). The origin of such a hierarchy is still a complete mystery and is presumably bound up in a more complete theory, which has still to be

Table 8.1: 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.



developed.

A further remarkable fact is the nature of the spins of the various fields involved: all matter fields are fermionic. This is vital for the nuclear and atomic structures we observe (and of which we are made); *e.g.*, without the Pauli exclusion principle all atoms would be chemically very similar and the biological richness of our world would be impossible. Moreover, all (exchange) forces are mediated by spin-one or vector bosons, which in the basic formulation at least are massless. This leads to the gauge principle (or symmetry), which has the important merit of guaranteeing that quantum corrections do not radically change the appearance of the particles and their interactions (masses, charges *etc.*). Indeed, all attempts to construct field-theoretical models not possessing such a symmetry are plagued with ambiguities and essentially uncontrollable so-called *renormalisation* (or quantum-correction) effects.

8.1.2 The three interactions

In physics then there are just four fundamental forces: gravity, electromagnetism, the strong nuclear force and the weak nuclear force. In particle and nuclear physics phenomenology the first has little direct importance (*e.g.*, the Bohr radius of a *gravitationally* bound hydrogen atom would be much larger than the radius of the universe) and so, although it should be said that in a complete theory it must be included, we shall only discuss the other three here. We shall, in addition, attempt to present the current state of affairs and briefly describe how a more complete theory might unify these interactions.

The strong nuclear force (QCD)

The strong nuclear force is described in terms of the quarks and antiquarks, which combine to form all the known hadrons (qqq baryons and $q\bar{q}$ mesons), and the gluons that they exchange (just as the photon is exchanged in the electromagnetic interaction)—the other particles listed in the above table do not participate in the strong interaction. The theory is known as quantum chromodynamics (QCD), the motivation for the colour connection being the fact that there are actually three distinct (but equivalent) strong charges (red, green and blue, say), with symmetry group $\mathrm{SU}(3)_{\mathrm{col}}$.

The need for three "colours" arises from the observation that, for example, the spin-$\frac{3}{2}$ baryon resonance Δ^{++} with $s_z = \frac{3}{2}$ must be composed of three u quarks all in an s-wave and all having spin up. Such a totally symmetric wave-function would be in complete contrast with the Pauli exclusion principle or Fermi–Dirac statistics. The simplest way to avoid this conflict is to assign a new quantum number to the quarks, having at least three distinct values, over which the wave-function may then be antisymmetrised. A three-quark baryon state may thus be represented schematically as:

$$|B\rangle \;=\; \varepsilon_{ijk}\, q_i\, q_j\, q_k, \tag{8.1.1}$$

where i, j and k represent the three colour charges, while a quark–antiquark meson is simply

$$|M\rangle \;=\; q_i\, \bar{q}_i. \tag{8.1.2}$$

The idea to adopt the colour analogy arises from the requirement that such a state be colour neutral (just as atoms are electrically neutral) and the observation that the three primary colours combine to form *white* (note that the above baryon state, being antisymmetric in ijk, is effectively colour neutral while the neutrality of a meson is trivial). This new charge then becomes the natural candidate for the interaction responsible for the bound state and the strong force observed in nuclear physics. Indeed, such *non-Abelian* theories – here the gauge group has an SU(3) symmetry – had already been conceived and studied theoretically by Yang and Mills.

A particular property of the interaction is that quarks (and gluons, which turn out to be colour charged too, unlike the electrically neutral photon) are so strongly bound inside hadrons that they can never be detected as free states—however, this property, known as *confinement*, has yet to be proven from first principles; the solution to the theory is unfortunately just too complex. It can, however, be modelled by attributing an effective string-like component to the force field at large distances that grows linearly with separation. That is, in addition to the elementary Coulomb-like contribution to the quark–antiquark potential, we

assume an *effective* linear part:

$$V_{\mathrm{QCD}} \simeq \frac{\alpha_{\mathrm{S}}}{r} + br, \tag{8.1.3}$$

where α_{S} (or α_{QCD}) is the fine-structure constant of the strong interaction while b is a phenomenological so-called *string tension*, experimentally deduced to be of order 1 GeV/fm (*cf.* the baryon mass and size).

A further property, common to all quantum field theories, but of particular importance in QCD, is the so-called *running* of the coupling constant α_{S}. In quantum field theory one can show that all coupling constants are subject to quantum corrections (due to the vacuum polarisation generated by virtual fermion–antifermion fluctuations), which depend on the energy scale of the process under consideration. In the case of non-Abelian theories such as QCD and also the weak part of the unified electroweak interaction (see later) the behaviour is such that α decreases at high energies (or short distances). The reason is an antiscreening effect due to the spontaneous vacuum production of gluon–gluon pairs. As already noted, the discovery of this effect merited a Nobel prize for Gross, Politzer and Wilczek.

The implication is that, while the effective interaction in QCD is too strong for perturbation theory to be applied at low energies (*i.e.* to the bound-state problem), we are still able to perform useful calculations for high-energy processes (just as in QED, for example), as studied in modern-day accelerator experiments. Theoretically, to lowest meaningful order in perturbation theory, we find

$$\alpha_{\mathrm{S}} = \frac{12\pi}{\beta_0 \ln(Q^2/\Lambda^2_{\mathrm{QCD}})}, \tag{8.1.4}$$

where Λ_{QCD} is a new parameter that must be measured experimentally (the current value is of order 100−200 MeV) and β_0 is a first-order perturbative coefficient calculable in quantum field theory ($\beta_0 = 33 - 2N_f$, where N_f is the number of energetically accessible quark flavours in the physical process under study). This behaviour is fully confirmed by a multitude of experimental results. Note that Λ_{QCD} is not technically a new *free* parameter as it merely replaces the value of the coupling constant for some given reference Q^2.

The weak nuclear force

The weak nuclear force is well described at low energies by Fermi theory, as already discussed; it is seen as an interaction at a single space–time point involving four "matter" fields directly. All fermions may participate in such processes and it is the only interaction to which the neutrinos are subject. However, if we consider

the high-energy behaviour of the theoretical cross-sections as calculated by Fermi, we see that they grow as $E_{\rm CM}^2$ and so would eventually diverge. That this must be so follows from a simple dimensional analysis: the Fermi coupling constant $G_{\rm F}$ has dimensions MeV^{-2} and therefore at high enough energies such that all masses may be neglected, the form of any cross-section must be

$$\sigma \propto (\hbar c)^2 G_{\rm F}^2 E_{\rm CM}^2 \times f(\theta). \tag{8.1.5}$$

The implied eventual violation of unitarity is a signal that Fermi theory cannot be considered as complete and is at best an (albeit very good) approximation to the true interaction at low energies.

Such a problem does not occur in, *e.g.*, QED. There the good behaviour may be traced to a dimensionless coupling constant and the photon propagator, which carries the interaction between two separate space–time points (the two interacting currents). In such a theory the high-energy form of any cross-section must be

$$\sigma \propto \frac{(\hbar c)^2 \alpha^2}{E_{\rm CM}^2}. \tag{8.1.6}$$

The solution in the weak theory is then to introduce a similar boson to carry the interaction between weak currents. It turns out that three such fields are necessary: $W^{\pm}$ and Z^0 with masses around 80 GeV and 90 GeV respectively.

Indeed, it is just the large mass in the $W^{\pm}$ and Z^0 propagators that renders this interaction apparently weak while the coupling constant itself is of the same order as (and slightly larger than) the electromagnetic coupling α. The modification is depicted diagrammatically in Fig. 8.1. In mathematical terms, the effect is to

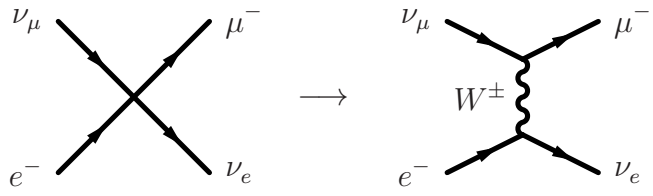


Figure 8.1: The cure for the divergent behaviour of Fermi theory via substitution of the four-point coupling with a boson propagator *à la* QED.

replace the single Fermi constant with two local vertices, separated in space–time, each representing the same weak coupling constant g say, and a propagator for the virtual transmission of the weak boson between the two points:

$$G_{\rm F} \to g \frac{1}{q^2 - M_{\rm W}^2} g, \tag{8.1.7}$$

where q^2 is the squared four-momentum carried by the exchange boson.* Thus, at low energies ($q^2 \ll M_{\mathrm{W}}^2$) the large mass kills all energy dependence in the propagator and leaves an apparently constant four-point interaction as hypothesised by Fermi while the high-energy the behaviour ($\sim 1/q^2$) of the propagator cures the divergences of the Fermi theory.

However, it is not quite so simple: the gauge symmetry of electromagnetism turns out to be fundamental in guaranteeing the general good behaviour of the theory and so a mechanism is needed whereby the effective masses of the new bosons are generated via an interaction energy (much as one might imagine happens to a photon inside a diamagnetic medium). This is the so-called *Higgs*† mechanism, a spontaneous breaking of certain symmetries in the initial theory, which requires the introduction of a new scalar field—actually two doublets, of which the visible remnant is the much sought-after and recently discovered *Higgs boson*. The construction of such a theory also requires a partial unification of the weak and electromagnetic forces. The work leading to the formulation of the full electroweak theory is due to Glashow, Salam and Weinberg.‡ The model thus constructed has a richer gauge symmetry than merely the U(1) of QED: we now have something like $\mathrm{SU}(2)_{\mathrm{W}} \times \mathrm{U}(1)_Y$.§

One further interesting indication coming from the weak interaction regards the number of generations. Among the decay channels of the Z^0 boson there is the possibility to produce a neutrino–antineutrino pair. Each generation (having a light neutrino $2m_\nu < m_{Z^0}$) thus adds an important contribution to the decay rate for the Z^0. The measured value is compatible with exactly three light neutrinos.

Electromagnetism

We have already discussed the electromagnetic interaction: it is the best understood; many quantities are measured experimentally very precisely and can also be

* Note that for the $e^- \nu_\mu \to \mu^- \nu_e$ scattering process depicted, simple kinematics forces $q^2 < 0$. The high-energy limit is then q^2 large but negative and the so singularity $q^2 = M_{\mathrm{W}}^2$ is always avoided.

† The 2013 Nobel Prize in 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."

‡ The 1979 Nobel Prize in Physics was awarded jointly to Sheldon Lee 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 $\mathrm{U}(1)_Y$ refer to a quantum number known as hypercharge (Y), which is actually a combination of electric charge and strangeness.

calculated to very high precision using standard quantum field-theoretic perturbative techniques. All the quarks, together with the charged leptons are subject to electromagnetic forces. In this case the coupling constant is known to *increase* with growing energy scale. Such behaviour has been confirmed by the LEP experiments, which extracted a value of $\alpha_{\rm QED}$ at an energy corresponding to the Z^0 mass. It is found, as predicted, that it changes from the standard low-energy value of close to 1/137 to about 1/128.

8.1.3 Running coupling constants and grand unification

As noted earlier, in quantum field theory all coupling constants *run*; that is, they vary with the energy scale of the specific process under consideration—for both QCD and QED this phenomenon is well verified experimentally. Now remarkably, the largest of these decreases (rapidly) while the smallest increases, with the result that all three tend towards a roughly common value for energies around $10^{15}\,\mathrm{GeV}$ (see Fig. 8.2). This suggests that there may be a single fundamental interaction with some large gauge group, SU(5) for example, which somehow spontaneously breaks down into the sub-groups $\mathrm{SU}(3)_{\rm QCD}\times\mathrm{SU}(2)_{\rm W}\times\mathrm{U}(1)_Y$. Note that what we have described earlier for the electroweak theory does not constitute a true unification of the two forces as there remain two separate and unrelated coupling constants; a fully (grand) unified theory (or GUT), as just described, would have just one unique coupling constant at the higher energy scale, which would then run differently in the different sectors to arrive at the three (different) known couplings of QCD and the electroweak theory at the energy scales relevant to present-day physics.

There are many problems for the construction of such a grand unified theory. In particular, the new exchange bosons it would simply lead to interactions that can change quarks into leptons and *vice versa*. A particular consequence of this is the possibility for the proton to decay (via *e.g.*, $p\to e^+\pi^0$). In the simplest extensions of the SM the predicted lifetimes are typically of order $10^{30-31}\,\mathrm{yr}$ while the current experimental upper limits are of order $10^{34}\,\mathrm{yr}$.

As can clearly be seen in Fig. 8.2, the three couplings do not actually meet at a common point in the SM. However, if there were other (new and as yet unknown) fields, the energy dependence would change. In particular, the presence of (very heavy) so-called supersymmetric partners to the known particles (so-called squarks, sleptons, photinos, gluinos *etc.*) could alter the running of the coupling constants in such a way they might indeed all coincide at one particular energy. The new unification point turns out to be of order $10^{16}\,\mathrm{GeV}$ (see Fig. 8.2). This is achieved, however, if and only if the *sparticle* masses are around $1-10\,\mathrm{TeV}$. Such a limit suggests that they might well be within reach of the LHC.

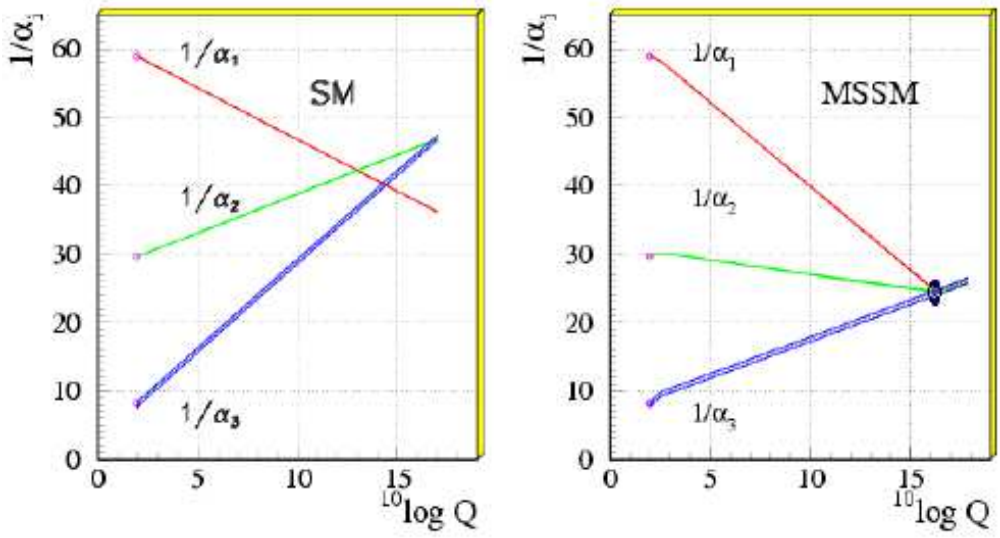


Figure 8.2: The running of the three coupling constants for the SM on the left and the so-called *minimal supersymmetric* SM on the right. From top to bottom the curves are respectively for the QED, weak and strong coupling constants. The figure is taken from Kazakov (2001).

8.2 Bibliography

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.

Appendix A

Background Notes

A.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 would correspond 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 (see App. A.2) 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 is necessary to examine the strength of its interaction to

* 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."

really understand the nature of the muon. Its decay is seen to be weak ($\tau_\mu \sim 2\,\mu$s), 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 the 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 any energy imbalance may be easily 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

$$R^B = \frac{Z}{m\,\alpha}, \tag{A.1.1}$$

which gives $R^B_e \simeq Z{\times}0.6{\times}10^{-10}$ m for an electron. And since then

$$R^B_\mu = \frac{m_e}{m_\mu}\,R^B_e, \tag{A.1.2}$$

for the given mass ratio of approximately 200, this leads to $R^B_\mu \simeq Z{\times}3{\times}10^{-13}$ 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 $R_{\rm nucl} = R_0 A^{1/3}$, where A is just the atomic mass and $R_0 \simeq 1.2\,{\rm fm}$. Using this and Eq. (A.1.2) we obtain

$$f \;=\; \left(\frac{R_{\rm nucl}}{R_\mu^B}\right)^3 \;=\; 0.27\,A\left(\frac{Z}{137}\right)^3. \tag{A.1.3}$$

For aluminium $Z = 13$ and $A = 27$, giving

$$f \;\simeq\; 6\times 10^{-3}. \tag{A.1.4}$$

This will be roughly the fraction of its lifetime that a muon spends inside the 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

$$\Lambda \;=\; v\,f\,\tau, \tag{A.1.5}$$

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

$$\frac{1}{\tau} \;=\; \frac{1}{\tau_d} + \frac{1}{\tau_c}, \tag{A.1.6}$$

where τ_d and τ_c stand for (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

$$\tau_c \;\sim\; 1.5\;\mu{\rm s}. \tag{A.1.7}$$

Inserting this into the formula for the mean free path, Eq. (A.1.5), leads to

$$\Lambda \;\sim\; 20 - 30\;{\rm cm}. \tag{A.1.8}$$

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 Nobel Prize winner Isidor Rabi to comment, in astonishment:

> "Who ordered that?"

A.2 Isospin and SU(2)

In nuclear and particle physics a number of symmetries are apparent. Most simple 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. One such example is

$$^{11}_{5}\mathrm{B} \leftrightharpoons {}^{11}_{6}\mathrm{C}. \tag{A.2.1}$$

While the chemical properties of ^{11}B and ^{11}C atoms are obviously rather different, the nuclei are indeed very similar. When one takes into account the variation due to the effects of Coulomb repulsion, one might even say 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 mass 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 *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 neutral 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 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: I, the total isospin, and I_3, the spin projection along the quantisation axis. 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 of the form $|I,I_3\rangle$, for the nucleon pair, we thus write

$$|p\rangle \;=\; |\tfrac{1}{2},+\tfrac{1}{2}\rangle \quad \text{and} \quad |n\rangle \;=\; |\tfrac{1}{2},-\tfrac{1}{2}\rangle \tag{A.2.2}$$

while the pion triplet becomes

$$|\pi^+\rangle \;=\; |1,+1\rangle, \quad |\pi^0\rangle \;=\; |1,\;0\rangle \quad \text{and} \quad |\pi^-\rangle \;=\; |1,-1\rangle. \tag{A.2.3}$$

The pairs of quantum numbers above then have the same *mathematical* significance as the l,m pairs for angular-momentum states.

For a nucleus the situation is a little more complicated. The I_3 assignment is straightforward: a nucleus containing Z protons and N neutrons clearly has $I_3 = (Z-N)/2$. However, the question of the overall isospin is ambiguous: the composition of A isospin-$\frac{1}{2}$ objects can lead to any value from the minimum of 0 ($\frac{1}{2}$) for A even (odd) up to the maximum possible $A/2$. It is found empirically that light nuclei in their ground state usually have the smallest value of I that can accommodate the I_3. For example, the deuteron ^{4}He and ^{12}C ground-state nuclei all have $I = 0$.

An important consequence of isospin symmetry and its mathematical structure is the possibility to apply Clebsch–Gordan coefficients to combinations of particles. A simple example is the decay of the so-called spin-$\frac{3}{2}$, isospin-$\frac{3}{2}$ Δ resonances, 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^+, which may be 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:

$$|\tfrac{3}{2},\tfrac{1}{2}\rangle \;=\; \sqrt{\tfrac{2}{3}}\;|\tfrac{1}{2},+\tfrac{1}{2}\rangle|1,0\rangle + \sqrt{\tfrac{1}{3}}\;|\tfrac{1}{2},-\tfrac{1}{2}\rangle|1,1\rangle. \qquad \text{(A.2.4)}$$

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 non-trivial fractions are experimentally well verified, thus confirming the SU(2)–isospin picture.

The fact that the lighter nuclei in the ground state usually have the lowest possible value of I consistent with their I_3 may be partially understood by considering the antisymmetrisation requirements for fermions (Fermi–Dirac statistics). The set of six-nucleon states with, say, $I = 3$ contains the state $I_3 = 3$, with all six nucleons protons. For such a state, the wave function must be antisymmetric under nucleon interchange of *any* pair. In contrast, the $I = 0$, $I_3 = 0$ state need only satisfy antisymmetrisation of the three protons and three neutrons separately, but not between the protons and neutrons; antisymmetrisation is thus less of a constraint here.

A.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; *see also*, Sigurgeirsson, T. and Yamakawa, K.A., *ibid.* 319.

Fermi, E., Teller, E. and Weisskopf, V.F. (1947), *Phys. Rev.* **71**, 314.

Heisenberg, W. (1932), *Z. Phys.* **77**, 1.

Street, J.C. and Stevenson, E.C. (1937), *Phys. Rev.* **51**, 1005; **52**, 1003.

Yukawa, H. (1935), *Proc. Phys. Math. Soc. Jap.* **17**, 48.