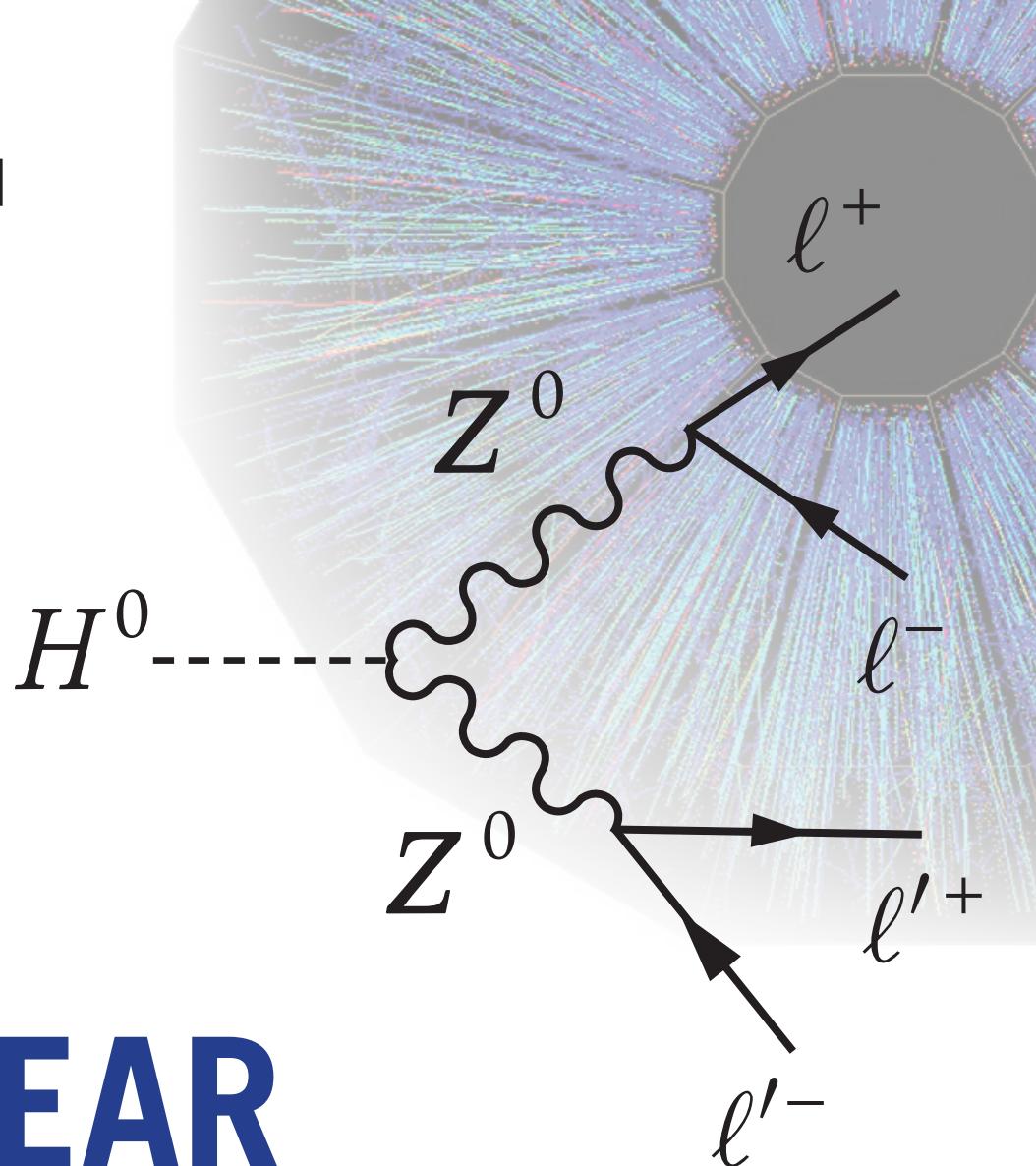


BRIAN R. MARTIN
GRAHAM SHAW

3RD EDITION

NUCLEAR AND PARTICLE PHYSICS

AN INTRODUCTION



WILEY

Nuclear and Particle Physics

Nuclear and Particle Physics

An Introduction

Third Edition

B.R. MARTIN

Department of Physics and Astronomy
University College London, UK

G. SHAW

School of Physics and Astronomy
Manchester University, Manchester, UK

WILEY

This edition first published 2019
© 2019 John Wiley & Sons Ltd

Edition History

Nuclear and Particle Physics - An introduction Wiley - 2006, Nuclear and Particle Physics - Second Edition Wiley 2009.
All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at <http://www.wiley.com/go/permissions>.

The right of B.R. Martin and G. Shaw to be identified as the authors of this work has been asserted in accordance with law.

Registered Offices

John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Office

The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Limit of Liability/Disclaimer of Warranty

In view of ongoing research, equipment modifications, changes in governmental regulations, and the constant flow of information relating to the use of experimental reagents, equipment, and devices, the reader is urged to review and evaluate the information provided in the package insert or instructions for each chemical, piece of equipment, reagent, or device for, among other things, any changes in the instructions or indication of usage and for added warnings and precautions. While the publisher and authors have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials or promotional statements for this work. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Library of Congress Cataloging-in-Publication Data

Names: Martin, B. R. (Brian Robert), author. | Shaw, G. (Graham), 1942– author.
Title: Nuclear and particle physics : an introduction / Brian R. Martin (Department of Physics and Astronomy, University College London, UK), Graham Shaw (School of Physics and Astronomy, University of Manchester, Manchester, UK).
Description: Third edition. | Hoboken, NJ : Wiley, 2019. | Includes index. |
Identifiers: LCCN 2018043095 (print) | LCCN 2018058046 (ebook) | ISBN 9781119344629 (Adobe PDF) | ISBN 9781119344636 (ePub) | ISBN 9781119344612 (pbk.)
Subjects: LCSH: Nuclear physics—Textbooks. | Particles (Nuclear physics)—Textbooks.
Classification: LCC QC776 (ebook) | LCC QC776 .M34 2019 (print) | DDC 539.7/2—dc23
LC record available at <https://lccn.loc.gov/2018043095>

Cover Design: Wiley

Cover Images: Courtesy of Brookhaven National Laboratory

Set in 11/13pt Computer Modern by Aptara Inc., New Delhi, India

Contents

<i>Preface</i>	xi
<i>Notes</i>	xiii
1 Basic concepts	1
1.1 History	1
1.1.1 The origins of nuclear physics	1
1.1.2 The emergence of particle physics: hadrons and quarks	6
1.1.3 The standard model of particle physics	9
1.2 Relativity and antiparticles	11
1.3 Space-time symmetries and conservation laws	13
1.3.1 Parity	14
1.3.2 Charge conjugation	16
1.3.3 Time reversal	17
1.4 Interactions and Feynman diagrams	20
1.4.1 Interactions	20
1.4.2 Feynman diagrams	21
1.5 Particle exchange: forces and potentials	24
1.5.1 Range of forces	24
1.5.2 The Yukawa potential	25
1.6 Observable quantities: cross-sections and decay rates	26
1.6.1 Amplitudes	27
1.6.2 Cross-sections	29
1.6.3 The basic scattering formulas	31
1.6.4 Unstable states	33
1.7 Units	36
Problems 1	37
2 Nuclear phenomenology	41
2.1 Mass spectroscopy	43
2.1.1 Deflection spectrometers	43
2.1.2 Kinematic analysis	45
2.1.3 Penning trap measurements	46
2.2 Nuclear shapes and sizes	51
2.2.1 Charge distribution	52
2.2.2 Matter distribution	56
2.3 Semi-empirical mass formula: the liquid drop model	59
2.3.1 Binding energies	59
2.3.2 Semi-empirical mass formula	60
2.4 Nuclear instability	64

2.5	Decay chains	67
2.6	β decay phenomenology	69
2.6.1	Odd-mass nuclei	70
2.6.2	Even-mass nuclei	71
2.7	Fission	72
2.8	γ decays	76
2.9	Nuclear reactions	76
	Problems 2	81
3	Particle phenomenology	83
3.1	Leptons	83
3.1.1	Lepton multiplets and lepton numbers	83
3.1.2	Universal lepton interactions; the number of neutrinos	86
3.1.3	Neutrinos	88
3.1.4	Neutrino mixing and oscillations	90
3.1.5	Oscillation experiments	93
3.1.6	Neutrino masses and mixing angles	101
3.1.7	Lepton numbers revisited	103
3.2	Quarks	104
3.2.1	Evidence for quarks	104
3.2.2	Quark generations and quark numbers	106
3.3	Hadrons	109
3.3.1	Flavour independence and charge multiplets	109
3.3.2	The simple quark model	113
3.3.3	Hadron decays and lifetimes	117
3.3.4	Hadron magnetic moments and masses	119
3.3.5	Heavy quarkonia	126
3.3.6	Allowed and exotic quantum numbers	133
	Problems 3	135
4	Experimental methods	139
4.1	Overview	139
4.2	Accelerators and beams	142
4.2.1	DC accelerators	142
4.2.2	AC accelerators	143
4.2.3	Neutral and unstable particle beams	150
4.3	Particle interactions with matter	152
4.3.1	Short-range interactions with nuclei	153
4.3.2	Ionisation energy losses	154
4.3.3	Radiation energy losses	157
4.3.4	Interactions of photons in matter	158
4.3.5	Ranges and interaction lengths	159
4.4	Particle detectors	160
4.4.1	Gaseous ionisation detectors	162
4.4.2	Scintillation counters	167
4.4.3	Semiconductor detectors	169

4.4.4	Čerenkov counters and transition radiation	170
4.4.5	Calorimeters	173
4.5	Detector Systems	176
	Problems 4	182
5	Quark dynamics: the strong interaction	185
5.1	Colour	185
5.2	Quantum chromodynamics (QCD)	187
5.2.1	The strong coupling constant	190
5.2.2	Screening, antiscreening and asymptotic freedom	193
5.3	New forms of matter	194
5.3.1	Exotic hadrons	194
5.3.2	The quark–gluon plasma	201
5.4	Jets and gluons	204
5.4.1	Colour counting	205
5.5	Deep inelastic scattering and nucleon structure	207
5.5.1	Scaling	207
5.5.2	The quark–parton model	210
5.5.3	Scaling violations and parton distributions	211
5.5.4	Inelastic neutrino scattering	215
5.6	Other processes	217
5.6.1	Jets	219
5.6.2	Lepton pair production	221
5.7	Current and constituent quarks	224
	Problems 5	226
6	Weak interactions and electroweak unification	229
6.1	Charged and neutral currents	229
6.2	Charged current reactions	231
6.2.1	W^\pm –lepton interactions	232
6.2.2	Lepton–quark symmetry and mixing	234
6.2.3	W -boson decays	238
6.2.4	Charged current selection rules	239
6.3	The third generation	242
6.3.1	More quark mixing	243
6.3.2	Properties of the top quark	246
6.4	Neutral currents and the unified theory	247
6.4.1	Electroweak unification	247
6.4.2	The Z^0 vertices and electroweak reactions	250
6.5	Gauge invariance and the Higgs boson	252
6.5.1	Unification and the gauge principle	253
6.5.2	Particle masses and the Higgs field	255
6.5.3	Properties of the Higgs boson	257
6.5.4	Discovery of the Higgs boson	259
	Problems 6	266

7 Symmetry breaking in the weak interaction	271
7.1 P violation, C violation, and CP conservation	271
7.1.1 Muon decay symmetries	273
7.1.2 Parity violation in electroweak processes	275
7.2 Spin structure of the weak interactions	277
7.2.1 Left-handed neutrinos and right-handed antineutrinos	277
7.2.2 Particles with mass: chirality	279
7.3 Neutral kaons: particle–antiparticle mixing and CP violation	281
7.3.1 CP invariance and neutral kaons	281
7.3.2 CP violation in K_L^0 decay	283
7.3.3 Flavour oscillations and CPT invariance	285
7.4 CP violation and flavour oscillations in B decays	289
7.4.1 Direct CP violation in decay rates	290
7.4.2 $B^0 - \bar{B}^0$ mixing	291
7.4.3 CP violation in interference	295
7.5 CP violation in the standard model	299
Problems 7	302
8 Models and theories of nuclear physics	305
8.1 The nucleon–nucleon potential	305
8.2 Fermi gas model	308
8.3 Shell model	310
8.3.1 Shell structure of atoms	310
8.3.2 Nuclear shell structure and magic numbers	312
8.3.3 Spins, parities, and magnetic dipole moments	315
8.3.4 Excited states	318
8.4 Nonspherical nuclei	319
8.4.1 Electric quadrupole moments	319
8.4.2 Collective model	322
8.5 Summary of nuclear structure models	323
8.6 α decay	324
8.7 β decay	327
8.7.1 $V - A$ theory	327
8.7.2 Electron and positron momentum distributions	329
8.7.3 Selection rules	330
8.7.4 Applications of Fermi theory	332
8.8 γ decay	337
8.8.1 Selection rules	337
8.8.2 Transition rates	339
Problems 8	340
9 Applications of nuclear and particle physics	343
9.1 Fission	343
9.1.1 Induced fission and chain reactions	344
9.1.2 Thermal fission reactors	348
9.1.3 Radioactive waste	352
9.1.4 Power from ADS systems	354

9.2	Fusion	357
9.2.1	Coulomb barrier	357
9.2.2	Fusion reaction rates	358
9.2.3	Nucleosynthesis and stellar evolution	361
9.2.4	Fusion reactors	366
9.3	Nuclear weapons	371
9.3.1	Fission devices	371
9.3.2	Fission/fusion devices	374
9.4	Biomedical applications	377
9.4.1	Radiation and living matter	377
9.4.2	Radiation therapy	380
9.4.3	Medical imaging using ionising radiation	385
9.4.4	Magnetic resonance imaging	390
9.5	Further applications	395
9.5.1	Computing and data analysis	395
9.5.2	Archaeology and geophysics	396
9.5.3	Accelerators and detectors	397
9.5.4	Industrial applications	398
	Problems 9	398
10	Some outstanding questions and future prospects	401
10.1	Overview	401
10.2	Hadrons and nuclei	402
10.2.1	Hadron structure and the nuclear environment	402
10.2.2	Nuclear structure	405
10.3	Unification schemes	407
10.3.1	Grand unification	407
10.3.2	Supersymmetry	412
10.3.3	Strings and things	417
10.4	The nature of the neutrino	418
10.4.1	Neutrinoless double beta decay	420
10.5	Particle astrophysics	426
10.5.1	Neutrino astrophysics	427
10.5.2	Cosmology and dark matter	432
10.5.3	Matter–antimatter asymmetry	438
10.5.4	Axions and the strong <i>CP</i> problem	441
A	Some results in quantum mechanics	445
A.1	Barrier penetration	445
A.2	Density of states	447
A.3	Perturbation theory and the Second Golden Rule	449
A.4	Isospin formalism	452
A.4.1	Isospin operators and quark states	452
A.4.2	Hadron states	454
	Problems A	456

B Relativistic kinematics	457
B.1 Lorentz transformations and four-vectors	457
B.2 Frames of reference	459
B.3 Invariants	461
Problems B	463
C Rutherford scattering	465
C.1 Classical physics	465
C.2 Quantum mechanics	467
Problems C	469
D Gauge theories	471
D.1 Gauge invariance and the standard model	471
D.1.1 Electromagnetism and the gauge principle	471
D.1.2 The standard model	474
D.2 Particle masses and the Higgs field	478
Problems D	481
E Short answers to selected problems	483
<i>References</i>	487
<i>Index</i>	491

Inside Rear Cover: Table of constants and conversion factors

Preface

It is common practice to teach nuclear physics and particle physics together in an introductory undergraduate course, and it is for such a course that this book has been written. The material is presented so that different selections can be made for a short course of about 25–30 lectures depending on the lecturer’s preferences and the students’ backgrounds. On the latter, students should have taken a first course in quantum physics, covering the traditional topics in nonrelativistic quantum mechanics and atomic physics. No prior knowledge of nuclear and particle physics is assumed. A few lectures on relativistic kinematics would also be useful, but this is not essential, as the necessary background is given in an appendix and is only used in a few places in the book.

We have not presented proofs or derivations of all the statements in the text. Rather, we have taken the view that it is more important that students see an overview of the subject, which for many, probably the majority, will be the only time they study nuclear and particle physics. For future specialists, the details will form part of more advanced courses. We have tried to take a direct approach throughout, focusing on the interpretation of experimental data in terms of current models and theories. Space restrictions have still meant that it has been necessary to make a choice of topics, and doubtless other equally valid choices could have been made. This is particularly true in Chapter 9, which deals with applications of nuclear and particle physics.

Since publication of the Second Edition of this book, there have been many important developments in both nuclear and particle physics. These include: the long-awaited discovery of the Higgs boson; substantial progress in neutrino physics and symmetry breaking in the weak interaction; a better understanding of stellar evolution and cosmology; high-precision nuclear mass measurements; increased developments in applying nuclear and particle physics techniques to clinical science; and tighter constraints on difficult-to-measure quantities, such as possible electric dipole moments and the masses of hypothetical particles, which are important for testing new theories of particle physics. Our aim in producing this Third Edition is again to bring the book up-to-date throughout, while leaving its basic philosophy unchanged. In doing this we are grateful to John Wiley and Sons for permission to use material from other books that we have published with them.

Finally, a word about footnotes: readers often have strong views about these (‘Notes are often necessary, but they are necessary evils’ – Samuel Johnson), so, as in previous editions, in this book they are designed to provide ‘non-essential’ information only. For those readers who prefer not to have the flow disrupted, ignoring the footnotes should not detract from understanding the text. Nuclear and particle physics have been, and still are, very important parts of the entire subject of physics and its practitioners have won an impressive number of Nobel Prizes. For historical interest, the footnotes also record many of these awards.

Brian Martin and Graham Shaw
July 2018

Notes

References

References are referred to in the text in the form of a name and date, for example Jones (1997). A list of references with full publication details is given at the end of the book.

Data

Tabulations of nuclear and particle physics data, such as masses, quantum numbers, decay modes, etc., are now readily available at the ‘click of a mouse’ from a number of sites and it is useful for students to get some familiarity with such sources. They are also needed to solve some end-of-chapter problems in the book. Many physical quantities are also readily found by a simple Internet search.

For particle physics, a comprehensive compilation of data, plus brief critical reviews of a number of current topics, may be found in the biannual publications of the Particle Data Group (PDG). The 2018 edition of their definitive *Review of Particle Properties* is referred to in Tanabashi et al. (Particle Data Group) (2018). *Physical Review* D98, 030001 in the references, and also as Particle Data Group (2018). The PDG Review is available online at <http://pdg.lbl.gov> and this site also contains links to other sites where compilations of specific particle data may be found.

Nuclear physics does not have the equivalent of the PDG review, but extensive compilations of nuclear data are available from a number of sources. Examples are: the Berkeley Laboratory Isotopes Project (<http://ie.lbl.gov/education/isotopes.htm>); the National Nuclear Data Center (NNDC), based at Brookhaven National Laboratory, USA (<http://www.nndc.bnl.gov>); the Nuclear Data Centre of the Japan Atomic Energy Research Institute (<http://wwwnndc.tokai-sc.jaea.go.jp/NuC>); and the Nuclear Data Evaluation Laboratory of the Korea Atomic Energy Research Institute (<http://atom.kaeri.re.kr>). All four sites have links to other data compilations.

Problems

Problems are provided for Chapters 1 to 9 and Appendices A to D; they are an integral part of the text. The problems are sometimes numerical and require values of physical constants that are given on the inside rear cover. Some also require data that may be found in the reference sites listed above. Short answers to selected problems are given at the end of the book in Appendix E. Readers may access the full solutions to odd-numbered problems on the book’s website given below, and instructors can access there the full solutions for all problems.

Illustrations

Some illustrations in the text have been adapted from, or are based on, diagrams that have been published elsewhere. We acknowledge, with thanks, permission to use such illustrations from the relevant copyright holders, as stated in the captions.

Website

www.wiley.com/go/martin/nuclear3

Instructors may access PowerPoint slides of all the illustrations from the text on the accompanying website. As indicated above, solutions for all the problems are also available to Instructors, with odd-numbered solutions available to all Readers. Any misprints or other necessary corrections brought to the author's notice will be listed. We would also be grateful for any other comments about the book, which should initially be sent to the Publishers (jcossham@wiley.com).

1

Basic concepts

1.1 History

Although this book will not follow a strictly historical development, to ‘set the scene’ this first chapter will start with a brief review of the most important discoveries that led to the separation of nuclear physics from atomic physics as a subject in its own right, and later work that in its turn led to the emergence of particle physics from nuclear physics.¹

1.1.1 The origins of nuclear physics

In 1896 Becquerel observed that photographic plates were being fogged by an unknown radiation emanating from uranium ores. He had accidentally discovered *radioactivity*, the fact that some chemical elements spontaneously emit radiation. The name was coined by Marie Curie two years later to distinguish this phenomenon from induced forms of radiation. In the years that followed, radioactivity was extensively investigated, notably by the husband and wife team of Pierre and Marie Curie, and by Rutherford and his collaborators.² Other radioactive sources were quickly found, including the hitherto unknown chemical elements polonium and radium,

¹For a readable and lavishly illustrated account, see Close, Marten, and Sutton (1987). An interesting account of the early period, with descriptions of the personalities involved, is given in Segrè (1980), while a very detailed and scholarly account may be found in Pais (1986).

²The 1903 Nobel Prize in Physics was awarded jointly to Henri Becquerel for his discovery and to Pierre and Marie Curie for their subsequent research into radioactivity. Ernest Rutherford had to wait until 1908, when he was awarded the Nobel Prize in Chemistry for his ‘investigations into the disintegration of the elements and the chemistry of radioactive substances’.

discovered by the Curies in 1897.³ It was soon established that there were two distinct types of radiation involved, named by Rutherford α and β rays. We know now that β rays are electrons (the name ‘electron’ had been coined in 1894 by Stoney) and α rays are doubly ionised helium atoms. In 1900 a third type of decay was discovered by Villard that involved the emission of photons, the quanta of electromagnetic radiation, referred to in this context as γ rays. These historical names are still commonly used.

The revolutionary implications of these experimental discoveries did not become fully apparent until 1902. Prior to this, atoms were still believed to be immutable – indestructible and unchanging – an idea with its origin in Greek philosophy and, for example, embodied in Dalton’s atomic theory of chemistry in 1804. This causes a big problem: if the atoms in a radioactive source remain unchanged, where does the energy carried away by the radiation come from? Typically, early attempts to explain the phenomena of radioactivity assumed that the energy was absorbed from the atmosphere or, when that failed, that energy conservation was violated in radioactive processes. However, Rutherford had shown in 1900 that the intensity of the radiation emitted from a radioactive source was not constant, but reduced by a factor of two in a fixed time that was characteristic of the source, but independent of its amount. This is called its *half-life*. In 1902, together with Soddy, he put forward the correct explanation, called the *transformation theory*, according to which the atoms of any radioactive element decay with a characteristic half-life, emitting radiation, and in so doing are transformed into the atoms of a different chemical element. The centuries old belief in the immutability of atoms was shattered forever.

An important question not answered by the transformation theory is: which elements are radioactive and which are stable? An early attempt to solve this problem was made by J.J. Thomson, who was extending the work of Perrin and others on the radiation that had been observed to occur when an electric field was established between electrodes in an evacuated glass tube. In 1897 he was the first to definitively establish the nature of these ‘cathode rays’. We now know they consist of free electrons, denoted e^- (the superscript denotes the electric charge) and Thomson measured their mass and charge.⁴ This gave rise to the speculation that atoms contained electrons in some way, and in 1903 Thomson suggested a model where the electrons were embedded and free to move in a region of positive charge filling the entire volume of the atom – the so-called *plum pudding model*. This model could account for the stability of atoms, but gave no explanation for the discrete wavelengths observed in the spectra of light emitted from excited atoms.

³For these discoveries, Marie Curie won a second Nobel Prize in 1911, this time in Chemistry. The honour would presumably have been shared with her husband had he not been killed in a road accident in 1906.

⁴J.J. Thomson received the 1906 Nobel Prize in Physics for his discovery. A year earlier, Philipp von Lenard had received the 1905 Physics Prize for his work on cathode rays.

The plum pudding model was finally ruled out by a classic series of experiments suggested by Rutherford and carried out by his collaborators Geiger and Marsden starting in 1909. This consisted of scattering α particles from very thin gold foils. In the Thomson model, most of the α particles would pass through the foil, with only a few suffering deflections through small angles. However, Geiger and Marsden found that some particles were scattered through very large angles, even greater than 90° . As Rutherford later recalled, ‘It was almost as incredible as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you’.⁵ He then showed that this behaviour was not due to multiple small-angle deflections, but could only be the result of the α particles encountering a very small, very heavy, positively charged central *nucleus*. (The reason for these two different behaviours is discussed in Appendix C.)

To explain these results, Rutherford in 1911 proposed the *nuclear model* of the atom. In this model, the atom was likened to a planetary system, with the light electrons (the ‘planets’) orbiting about a tiny but heavy central positively charged nucleus (the ‘sun’). The size of the atom is thus determined by the radii of the electrons’ orbits, with the mass of the atom arising almost entirely from the mass of the nucleus. In the simplest case of hydrogen, a single electron orbits a nucleus, now called the *proton* (*p*), with electric charge $+e$, where *e* is the magnitude of the charge on the electron, to ensure that hydrogen atoms are electrically neutral. Alpha particles are just the nuclei of helium, while heavier atoms were considered to have more electrons orbiting heavier nuclei. At about the same time, Soddy showed that a given chemical element often contained atoms with different atomic masses but identical chemical properties. He called this *isotopism* and the members of such families *isotopes*. Their discovery led to a revival of interest in *Prout’s Law* of 1815, which claimed that all the elements had integer atomic mass in units of the mass of the hydrogen atom, called *atomic weights*. This holds to a good approximation for many elements, like carbon and nitrogen, with atomic weights of approximately 12.0 and 14.0 in these units, but does not hold for other elements, like chlorine, which has an atomic weight of approximately 35.5. However, such fractional values could be explained if the naturally occurring elements consisted of mixtures of isotopes. Chlorine, for example, is now known to consist of a mixture of isotopes with atomic weights of approximately 35.0 and 37.0, giving an average value of 35.5 overall.⁶

Although the planetary model explained the α particle scattering experiments, there remained the problem of reconciling it with the observation of stable atoms. In classical physics, the electrons in the planetary model would be continuously accelerating and would therefore lose energy by radiation, leading to the collapse of the atom. This problem was solved by Bohr in 1913, who revolutionised the study of atomic physics by

⁵Quoted on p. 111 of da C. Andrade (1964).

⁶Frederick Soddy was awarded the 1921 Nobel Prize in Chemistry for his work on isotopes.

applying the newly emerging quantum theory. The result was the Bohr–Rutherford model of the atom, in which the motion of the electrons is confined to a set of discrete orbits. Because photons of a definite energy would be emitted when electrons moved from one orbit to another, this model could explain the discrete nature of the observed electromagnetic spectra when excited atoms decayed. In the same year, Moseley extended these ideas to a study of X-ray spectra and conclusively demonstrated that the charge on the nucleus is $+Ze$, where the integer Z was the atomic number of the element concerned, and implying Z orbiting electrons for electrical neutrality. In this way he laid the foundation of a physical explanation of Mendeleev’s periodic table and in the process predicted the existence of no less than seven unknown chemical elements, which were all later discovered.⁷

The phenomena of atomic physics are controlled by the behaviour of the orbiting electrons and are explained in detail by refined modern versions of the Bohr–Rutherford model, including relativistic effects described by the Dirac equation, the relativistic analogue of the Schrödinger equation that applies to electrons, which is discussed in Section 1.2. However, following the work of Bohr and Moseley it was quickly realised that radioactivity was a nuclear phenomenon. In the Bohr–Rutherford and later models, different isotopes of a given element have different nuclei with different nuclear masses, but their orbiting electrons have virtually identical chemical properties because these nuclei all carry the same charge $+Ze$. The fact that such isotopes often have dramatically different radioactive decay properties is therefore a clear indication that these decays are nuclear in origin. In addition, since electrons were emitted in β decays, it seemed natural to assume that nuclei contained electrons as well as protons, and the first model of nuclear structure, which emerged in 1914, assumed that the nucleus of an isotope of an element with atomic number Z and mass number A was itself a tightly bound compound of A protons and $A - Z$ electrons. This provided an explanation of the existence of isotopes and of the approximate validity of Prout’s law when applied to isotopes, because the electron mass is negligible compared to that of the proton. However, although this model persisted for some time, it was subsequently ruled out by detailed measurements of the spins of nuclei (cf. Problem 1.1).

The correct explanation of isotopes and nuclear structure had to wait almost twenty years, until a classic discovery by Chadwick, in 1932. His work followed earlier experiments by Irène Curie (the daughter of Pierre and Marie Curie) and her husband Frédéric Joliot. They had observed that neutral radiation was emitted when α particles bombarded beryllium, and later work had studied the energy of protons emitted when paraffin

⁷Niels Bohr received the 1922 Nobel Prize in Physics for his theoretical work on the structure of atoms. Moseley was nominated for the 1915 Nobel Prizes in both Physics and Chemistry for his pioneering use of X-rays, but was tragically killed in World War I in August 1915 at the age of 27, before a decision was made.

was exposed to this neutral radiation. Chadwick refined and extended these experiments and demonstrated that they implied the existence of an electrically neutral particle of approximately the same mass as the proton, called the *neutron* (n).⁸ The discovery of the neutron led immediately to the correct formulation of nuclear structure, in which an isotope of atomic number Z and mass number A is a bound state of Z protons and $A - Z$ neutrons. There are no electrons bound inside nuclei.

Finally, to complete this historical account, we must go back to another major result: the discovery of the continuous β -decay spectrum by Chadwick in 1914. At that time, nuclear decays were all viewed as a parent nucleus decaying via α , β , or γ decay to give a daughter nucleus plus either an alpha particle, an electron or a photon, respectively. As each possibility would be a two-body decay, energy and momentum conservation implies that the emitted particle would have a unique energy, depending on the masses of the parent and daughter nucleons, which would be the same for all observed decays of a given type. This behaviour is precisely what is observed for α decays and γ decays and the earliest experiments erroneously suggested the same held for β decays. However, when Chadwick measured the energies of the electrons from samples of nuclei he found that the electrons emitted in a given β -decay process had a continuous energy distribution, as shown in Figure 1.1.

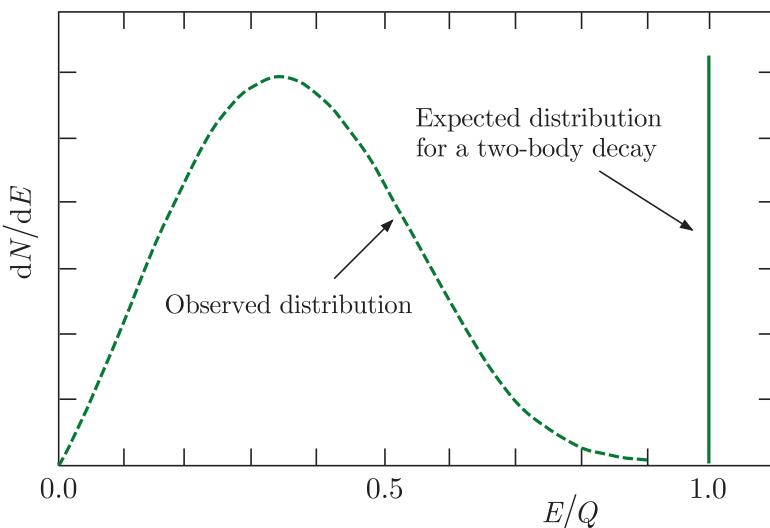


Figure 1.1 The observed electron energy distribution dN/dE in β decay (dashed line) as a function of E/Q , where E is the kinetic energy of the electron and Q is the total energy released. Also shown is the expected energy distribution if β decay were a two-body process (solid line).

After a hiatus due to the first world war, various ideas were suggested to explain this unexpected result, including a remarkable proposal by Bohr in 1929 that energy conservation was violated in β decays, but later abandoned by him in favour of the correct hypothesis proposed by Pauli in

⁸James Chadwick received the 1935 Nobel Prize in Physics for his discovery of the neutron. The discovery was not unexpected, because Rutherford had already deduced that the nucleus must include uncharged constituents with masses similar to that of the proton, and had even coined the name ‘neutron’. Irène Curie and Frédéric Joliot received the 1935 Nobel Prize in Chemistry for ‘synthesising new radioactive elements’.

1930. Pauli proposed that an additional, and hitherto unknown, neutral particle was emitted in β decays and shared the energy released with the electron. This particle had to be very light, since the most energetic electrons in the observed continuous distribution carried off almost all the energy released in the decay, as can be seen in Figure 1.1; it had also to interact so weakly with matter that it invariably escaped detection. Despite this, its existence was rapidly accepted, largely because of its crucial role in the highly successful theory of β decay proposed in 1932 by Fermi, who used the name *neutrino* (meaning ‘little neutral one’) for the new particle after his close friend and colleague Amaldi jokingly suggested it to distinguish Pauli’s particle from Chadwick’s ‘big neutral one’, the neutron.⁹

In conclusion, by 1932 physicists had arrived at a model of the nucleus in which an isotope of atomic number Z and mass number A is a bound state of Z protons and $A - Z$ neutrons. Later workers, including Heisenberg, another of the founders of quantum theory, applied quantum mechanics to the nucleus, now viewed as a collection of neutrons and protons, collectively called *nucleons*. In this case, however, the force binding the nucleus is not the electromagnetic force that holds electrons in their orbits, but a much stronger force that does not depend on the charge of the nucleon (i.e. is charge-independent) and with a very short effective range. This binding interaction is called the *strong nuclear force*. In addition, there is a third force, much weaker than the electromagnetic force, called the *weak interaction*, responsible for β decays, where neutrinos as well as electrons are emitted. These ideas form the essential framework of our understanding of the nucleus today. Nevertheless, there is still no single theory that is capable of explaining all the data of nuclear physics and we shall see that different models are used to interpret different classes of phenomena.

1.1.2 The emergence of particle physics: hadrons and quarks

By the early 1930s, the nineteenth century view of atoms as indivisible elementary particles had been replaced and a smaller group of subatomic particles now enjoyed this status: electrons, protons and neutrons. To these we must add two electrically neutral particles: the photon (γ) and the neutrino (ν). However, this simple picture was not to last, because of the discovery of many new subatomic particles, initially in cosmic rays and later in experiments using particle accelerators.

We start with cosmic rays, which may be conveniently divided into two types: *primaries*, which are high-energy particles, mostly protons,

⁹The neutrino was eventually detected, but not until very much later, by Reines and Cowan in 1956. A brief description of their experiment is given in Section 2.1.1 of Martin and Shaw (2017) and in more detail in Chapter 12 of Trigg (1975). Frederick Reines shared the 1995 Nobel Prize in Physics for his work in neutrino physics, particularly for the detection of the neutrino.

incident on the Earth's atmosphere from all directions in space; and *secondaries*, which are produced when the primaries collide with nuclei in the Earth's atmosphere, with some penetrating to sea level. It was among these secondaries that the new particles were discovered, mainly using a detector devised by C.T.R. Wilson, called the *cloud chamber*. It consisted of a vessel filled with air almost saturated with water vapour and fitted with an expansion piston. When the vessel was suddenly expanded, the air was cooled and became supersaturated. Droplets were then formed preferentially along the trails of ions left by charged particles passing through the chamber. Immediately after the expansion, the chamber was illuminated by a flash of light and the tracks of droplets so revealed were photographed before they had time to disperse. The use of these chambers in cosmic ray studies led to many important discoveries, including, in 1932, the detection of *antiparticles*, to be discussed in Section 1.2.¹⁰ However, the birth of particle physics as a new subject, distinct from atomic and nuclear physics, dates from 1947 with the discovery of *pions* and of *strange particles* by cosmic ray groups at Bristol and Manchester Universities, respectively. We will consider these in turn.

The discovery of pions was not unexpected, since Yukawa had famously predicted their existence in a theory of the strong nuclear forces proposed in 1934. We will return to this in Section 1.5. Here we will simply note that the range of the nuclear force required the pions to have a mass of around one seventh of the proton mass, while the charge independence of the nuclear force required there to be three charge states, denoted π^+ , $\pi^- \pi^0$, with charges $+e$, $-e$ and zero, respectively. This gave rise to a search for these particles in cosmic ray secondaries, and in 1936 Anderson and Neddermeier discovered new subatomic particles that were initially thought to be pions, but are now known to be particles called *muons*. As we shall see in Chapter 3, muons are rather like heavy electrons and, like both electrons and neutrinos, do not interact via the strong force that holds the nucleus together. Charged pions with suitable properties were finally detected in 1947 using photographic emulsions containing a silver halide. The ionisation energy deposited by a charged particle passing through the emulsion causes the formation of a latent image, and the silver grains resulting from subsequent development form a visual record of the path of the particle. The neutral pion was detected somewhat later in 1950.¹¹ Pions interact with each other and with nucleons via forces comparable in strength to the strong nuclear interaction between nucleons and in future we will refer to all such forces as *strong interactions*, reserving the term

¹⁰Wilson built the first cloud chamber in 1911 and shared the 1927 Nobel Prize in Physics. Victor Hess discovered cosmic rays in 1912, by making a series of balloon flights and showing that the intensity of radiation increased at high altitudes, indicating an extraterrestrial origin. He shared the 1936 Nobel Prize in Physics.

¹¹The 1949 Nobel Prize in Physics was awarded to Hideki Yukawa for his prediction of the pion and in 1950 the Nobel Prize in Physics was awarded to Cecil Powell for his leading role in its discovery.

strong nuclear interaction to the special case of nucleon–nucleon interactions. Particles that interact by the strong force are now called *hadrons*. Thus pions and nucleons are examples of hadrons, while electrons, muons and neutrinos are not.

Further work using cloud chambers to detect cosmic ray secondaries led to the discovery in 1947 by Rochester and Butler of new particles, named *kaons*, which, in contrast to the discovery of pions, was totally unexpected. Kaons were almost immediately recognised as a completely new form of matter, because they had supposedly ‘strange’ properties, which will be discussed further in Section 3.3. Other *strange particles* with similar properties were discovered, and in 1953 it was realised that these properties were precisely what would be expected if they were hadrons with nonzero values of an hitherto unknown quantum number, given the name *strangeness* by Gell-Mann, which was conserved in strong and electromagnetic interactions, but not necessarily conserved in the so-called *weak interactions* responsible for β decay. Non-strange particles like the pions and nucleons have zero values of strangeness. This led Gell-Mann, and independently Zweig, to suggest that hadrons were composed of more fundamental particles called *quarks* (q), together with their antiparticles. Three quarks were required at the time, denoted u , d , and s , with fractional electric charges $+2e/3$, $-e/3$, and $-e/3$, respectively. In particular, ordinary matter, i.e. protons and nucleons are composed of u and d quarks only, while the strange particles also contain s quarks. The latter is called the *strange quark* and the strangeness quantum number merely reflects the number of strange quarks and/or antiquarks present.

The 1950s also saw technological developments that enabled high-energy beams of particles to be produced in laboratories, and these rapidly replaced cosmic rays as the source of the high-energy particles required to create new particles in collisions. At the same time, cloud chambers were largely superseded by bubble chambers, a more efficient device in which charged particles were detected by the trail of bubbles left along their tracks through a superheated liquid, rather than droplets in a super-cooled gas.¹² By the mid-1960s this had resulted in the discovery of many more unstable particles and the above *quark model* had considerable success in understanding the properties of the observed hadrons, as we shall see in Section 3.3,¹³ but because no free quarks were detected experimentally, there was initially considerable scepticism for this interpretation. We now know that there is a fundamental reason why quarks cannot

¹²Many beautiful pictures of events observed in both cloud and bubble chambers may be found in Close Marten, and Sutton (1987). Donald Glaser was awarded the 1960 Nobel Prize in Physics for his invention of the bubble chamber and Luis Alvarez received the 1968 prize for its further development and use in discovering new subatomic particles.

¹³Murray Gell-Mann received the 1969 Nobel Prize in Physics for ‘contributions and discoveries concerning the classification of elementary particles and their interactions’. For the origin of the word ‘quark’, he cited the now famous quotation ‘Three quarks for Muster Mark’ from James Joyce’s book *Finnegans Wake*. George Zweig had suggested the name ‘aces’. Subsequently, more than three quarks were discovered, as we shall see.

be observed as free particles (it is discussed in Section 5.1), but at the time many physicists looked upon quarks as a convenient mathematical description, rather than physical particles. However, evidence for the existence of quarks as real particles began to emerge in 1969 from a series of experiments analogous to those of Rutherford and his co-workers, where high-energy beams of electrons and neutrinos were scattered from nucleons. (These experiments are discussed in Section 5.5.) Analysis of the angular distributions of the scattered particles confirmed that the nucleons were themselves bound states of point-like charged entities, with properties consistent with those hypothesised in the quark model, including their fractional electric charges. This is essentially the picture today, where elementary particles are considered to be a small number of fundamental physical entities, including quarks, the electron, neutrinos, the photon and a few others we shall meet, but no longer nucleons.

1.1.3 The standard model of particle physics

Following the discovery of quarks, an ‘in principle’ complete theory of elementary particles gradually emerged, called, rather prosaically, the *standard model*. This aims to explain all the phenomena of particle physics, except those due to gravity, in terms of the properties and interactions of a small number of *elementary* (or *fundamental*) *particles*, which are now defined as being point-like, without internal structure or excited states. Particle physics thus differs from nuclear physics in having a single theory to interpret its data. Here we restrict ourselves to a brief outline of the standard model, which will be developed in more detail later in Chapters 3, 5, 6, and 7.

An elementary particle is characterised by, amongst other things, its mass, its electric charge and its *spin*. The latter is a permanent angular momentum possessed by all particles in quantum theory, even when they are at rest. Spin has no classical analogue and is not to be confused with the use of the same word in classical physics, where it usually refers to the angular momentum of extended objects. The maximum value of the spin angular momentum about any axis is $S\hbar$ ($\hbar \equiv h/2\pi$), where h is Planck’s constant and S is the *spin quantum number*, or *spin* for short. It has a fixed value for particles of any given type (for example $S = 1/2$ for electrons) and general quantum mechanical principles restrict the possible values of S to be $0, 1/2, 1, 3/2, \dots$. Particles with half-integer spin are called *fermions* and those with integer spin are called *bosons*. There are two families of elementary fermions in the standard model: the quarks, which interact via strong forces, and the *leptons*, including electrons, muons, and neutrinos, which do not. In addition, there is a family of spin-1 bosons, which act as force carriers in the theory, and a spin-0 particle, called the *Higgs boson*, which plays a key role in understanding the origin of elementary particle masses within the theory.

The above particles interact via four forces of nature. In decreasing order of strength, these are the strong interaction, which binds the

quarks together into hadrons; the electromagnetic interaction between the charged leptons and quarks; the weak interaction responsible for β decay; and gravity. Although an understanding of all four forces will ultimately be essential in a complete theory, gravity is so weak that it can be neglected in nuclear and particle physics at presently accessible energies. Because of this, we will often refer in practice to the three forces of nature. The standard model specifies the origin of these three forces. In classical physics the electromagnetic interaction is propagated by electromagnetic waves, which are continuously emitted and absorbed. While this is an adequate description at long distances, at short distances the quantum nature of the interaction must be taken into account. In quantum theory, the interaction is transmitted discontinuously by the exchange of photons, which are members of the family of fundamental spin-1 bosons of the standard model. Photons are referred to as the *gauge bosons*, or ‘force carriers’, of the electromagnetic interaction. The use of the word ‘gauge’ originates from the fact that the electromagnetic interaction possesses a fundamental symmetry called *gauge invariance*. For example, Maxwell’s equations of classical electromagnetism are invariant under a specific transformation of the electromagnetic fields, called a gauge transformation. This property is common to all the three interactions of nature we will be discussing and has profound consequences, but we will not need its details in this book.¹⁴ The weak and strong interactions are also mediated by the exchange of spin-1 gauge bosons. For the weak interaction these are the W^+ , W^- , and Z^0 bosons (again the superscripts denote the electric charges) with masses about 80–90 times the mass of the proton. For the strong interaction, the force carriers are called *gluons*. There are eight gluons, all of which have zero mass and are electrically neutral.

In addition to the elementary particles of the standard model, there are other important particles we will be studying. These are the *hadrons*, the bound states of quarks. Nucleons are examples of hadrons, but there are several hundred more, not including nuclei, most of which are unstable and decay by one of the three interactions. For example, the charged pions π^\pm decay via the weak interaction with a lifetime of about 10^{-8} s, while the neutral pion π^0 decays via the electromagnetic interaction with a lifetime of about 10^{-17} s. The existence of quarks was first inferred from the properties of hadrons, as we have seen, and they remain particularly important because free quarks are unobservable in nature. Hence to deduce properties of quarks we are forced to study hadrons. An analogy would be if we had to deduce the properties of nucleons by exclusively studying the properties of nuclei.

Since nucleons are bound states of quarks and nuclei are bound states of nucleons, the properties of nuclei should in principle be deducible from

¹⁴A brief description of gauge invariance and some of its consequences is given, for the interested reader, in Appendix D.

the properties of quarks and their interactions, that is, from the standard model. Although there has been some progress in this direction, in practice this is still beyond present calculational techniques and often nuclear and particle physics are treated as two almost separate subjects. However, there remain some connections between them and in introductory treatments it is still useful to present both subjects together.

The remaining sections of this chapter are devoted to introducing some of the basic theoretical tools needed to describe the phenomena of both nuclear and particle physics, starting with a key concept in the latter: antiparticles.

1.2 Relativity and antiparticles

Elementary particle physics is also called high-energy physics. One reason for this is that if we wish to produce new particles in a collision between two other particles, then because of the relativistic mass–energy relation $E = mc^2$, energies are needed at least as great as the rest masses of the particles produced. The second reason is that to explore the structure of a particle requires a probe whose wavelength λ is smaller than the structure to be explored. By the de Broglie relation $\lambda = h/p$, this implies that the momentum p of the probing particle, and hence its energy, must be large. For example, to explore the internal structure of the proton using electrons requires wavelengths that are much smaller than the radius of the proton, which is roughly 10^{-15} m. This in turn requires electron energies that are greater than 10^3 times the rest energy of the electron, implying electron velocities very close to the speed of light. Hence any explanation of the phenomena of elementary particle physics must take account of the requirements of the theory of special relativity, in addition to those of quantum theory. There are very few places in particle physics where a nonrelativistic treatment is adequate, whereas the need for a relativistic treatment is much less in nuclear physics.

Constructing a quantum theory that is consistent with special relativity leads to the conclusion that for every charged particle of nature, there must exist an associated particle, called an *antiparticle*, with the same mass as the corresponding particle. This important prediction was first made by Dirac and follows from the solutions of the equation he postulated to describe relativistic electrons.¹⁵ The *Dirac equation* for a particle of mass m and momentum \mathbf{p} moving in free space is of the form

$$i\hbar \frac{\partial \Psi(\mathbf{r}, t)}{\partial t} = H(\mathbf{r}, \hat{\mathbf{p}})\Psi(\mathbf{r}, t), \quad (1.1)$$

¹⁵Paul Dirac shared the 1933 Nobel Prize in Physics with Erwin Schrödinger. The somewhat cryptic citation stated ‘for the discovery of new productive forms of atomic theory’.

where we use the notation $\mathbf{r} = (x_1, x_2, x_3) = (x, y, z)$, $\hat{\mathbf{p}} = -i\hbar\nabla$ is the usual quantum mechanical momentum operator and the Hamiltonian was postulated by Dirac to be

$$H = c \boldsymbol{\alpha} \cdot \hat{\mathbf{p}} + \beta mc^2. \quad (1.2)$$

The coefficients $\boldsymbol{\alpha}$ and β are determined by the requirement that the solutions of (1.1) are also solutions of the free-particle *Klein-Gordon equation*

$$-\hbar^2 \frac{\partial^2 \Psi(\mathbf{r}, t)}{\partial t^2} = -\hbar^2 c^2 \nabla^2 \Psi(\mathbf{r}, t) + m^2 c^4 \Psi(\mathbf{r}, t), \quad (1.3)$$

which follows from making the usual quantum mechanical substitutions $\mathbf{p} \rightarrow -i\hbar\nabla$ and $E \rightarrow i\hbar\partial/\partial t$ in the relativistic mass–energy relation $E^2 = p^2 c^2 + m^2 c^4$. This leads to the conclusion that $\boldsymbol{\alpha}$ and β cannot be ordinary numbers; their simplest forms are 4×4 matrices. Thus the solutions of the Dirac equation are four-component wavefunctions (called *spinors*) with the form¹⁶

$$\Psi(\mathbf{r}, t) = \begin{pmatrix} \psi_1(\mathbf{r}, t) \\ \psi_2(\mathbf{r}, t) \\ \psi_3(\mathbf{r}, t) \\ \psi_4(\mathbf{r}, t) \end{pmatrix}. \quad (1.4)$$

The interpretation of (1.4) is that the four components describe the two spin states of a negatively charged electron with positive energy and the two spin states of a corresponding particle having the same mass, but with negative energy. Two spin states arise because in quantum mechanics the projection in any direction of the spin vector of a spin-1/2 particle can only result in one of the two values $\pm 1/2$, referred to as ‘spin up’ and ‘spin down’, respectively. The two energy solutions arise from the two solutions of the relativistic mass–energy relation $E = \pm(p^2 c^2 + m^2 c^4)^{1/2}$. The negative-energy states can be shown to behave in all respects as positively charged electrons called *positrons*, but with *positive* energy.¹⁷ The positron is referred to as the *antiparticle* of the electron. The discovery of the positron by Anderson in 1933, with all the predicted properties, was a spectacular verification of Dirac’s prediction, as was the much later discovery of the antiproton in 1955.¹⁸

Although Dirac originally made his prediction for electrons, the result holds for all charged particles and is true whether the particle is an elementary particle or a hadron. If we denote a particle by P , then the antiparticle is in general written with a bar over it, i.e. \bar{P} . For example,

¹⁶The details may be found in many quantum mechanics books, e.g. pp. 475–477 of Schiff (1968).

¹⁷See, for example, Chapter 1 of Martin and Shaw (2017).

¹⁸Carl Anderson shared the 1936 Nobel Prize in Physics for the discovery of the positron and Emilio Segrè and Owen Chamberlain were awarded the 1959 prize for their discovery of the antiproton.

the antiparticle of the proton p is the antiproton \bar{p} , with negative electric charge, and associated with every quark, q , is an antiquark, \bar{q} . However, for some very common particles the bar is usually omitted. Thus, for example, in the case of the positron e^+ , the superscript denoting the charge makes explicit the fact that the antiparticle has the opposite electric charge to that of its associated particle. The argument does not extend to neutral particles in general and while some have distinct antiparticles, others do not. For example, the neutron has a non-zero magnetic moment, as we shall see below, and there is a distinct antiparticle, the *antineutron* \bar{n} , which has a magnetic moment equal in magnitude to that of the neutron, but opposite in sign. On the other hand, neither the photon γ nor the neutral pion π^0 has a distinct antiparticle.

Electric charge is just one example of a quantum number that has equal and opposite values for particles and antiparticles. We will meet others later. When brought together, particle–antiparticle pairs, each of mass m , can annihilate, releasing their combined rest energy $2mc^2$ as photons or other particles. There is a symmetry between particles and antiparticles, and it is a convention to call the electron the particle and the positron its antiparticle. This reflects the fact that normal matter contains electrons rather than positrons.

Finally, we note that among the many successful predictions of the Dirac equation is that for magnetic moments. A charged particle with spin necessarily has an intrinsic magnetic moment μ , and it can be shown from the Dirac equation that a point-like spin-1/2 particle of charge q and mass m has a magnetic moment $\mu = (q/m) \mathbf{S}$, where \mathbf{S} is its spin vector.¹⁹ Magnetic moment is a vector, and the value μ tabulated is the z component of μ when the z component of spin has its maximum value, i.e. $\mu = q\hbar/2m$. This is a test of the elementarity of a spin-1/2 particle and the measured magnetic moment of the electron is compatible with this assumption. However, the experimental values for the proton and neutron are

$$\mu_p = 2.79e \mathbf{S}/m_p \quad \text{and} \quad \mu_n = 1.91e \mathbf{S}/m_n,$$

which do not obey the Dirac prediction, reflecting the fact that the proton and neutron are not point-like, elementary particles.²⁰

1.3 Space-time symmetries and conservation laws

Symmetries and the invariance properties of the underlying interactions play an important role in physics. Some lead to conservation laws that are universal. Familiar examples are translational invariance, leading to

¹⁹There is a small correction to this predicted value, of order one part in a thousand, which we ignore in this simple account. See, for example, Section 9.6 of Mandl and Shaw (2010).

²⁰The proton magnetic moment was first measured by Otto Stern in 1933 using a molecular beam method that he developed and for this he received the 1943 Nobel Prize in Physics.

the conservation of linear momentum; and rotational invariance, leading to conservation of angular momentum. The latter plays an important role in nuclear and particle physics as it leads to a scheme for the classification of states based, among other quantum numbers, on their spins. This is similar to the scheme used to classify states in atomic physics.²¹ Another very important invariance that we have briefly mentioned is gauge invariance. This fundamental property of all three interactions restricts their forms in a profound way. In its simplest form, it predicts zero masses for all elementary particles. However, there are theoretical solutions to this problem whose experimental verification is described in Section 6.5.

In nuclear and particle physics we need to consider additional symmetries of the interactions and the conservation laws that follow. In the remainder of this section we discuss three space–time symmetries that we will need in later chapters – *parity*, *charge conjugation*, and *time-reversal*.

1.3.1 Parity

Parity was first introduced in the context of atomic physics by Wigner in 1927.²² It refers to the behaviour of a state under a spatial reflection, i.e. $\mathbf{r} \rightarrow -\mathbf{r}$. If we consider a single-particle state, represented for simplicity by a nonrelativistic wavefunction $\Psi(\mathbf{r}, t)$, then under the parity operator \hat{P} ,

$$\hat{P}\Psi(\mathbf{r}, t) \equiv P\Psi(-\mathbf{r}, t). \quad (1.5)$$

Applying the operator again gives

$$\hat{P}^2\Psi(\mathbf{r}, t) = P\hat{P}\Psi(-\mathbf{r}, t) = P^2\Psi(\mathbf{r}, t), \quad (1.6)$$

implying $P = \pm 1$. If the particle is an eigenfunction of linear momentum \mathbf{p} , i.e.

$$\Psi(\mathbf{r}, t) \equiv \Psi_{\mathbf{p}}(\mathbf{r}, t) = \exp[i(\mathbf{p} \cdot \mathbf{r} - Et)/\hbar], \quad (1.7)$$

then

$$\hat{P}\Psi_{\mathbf{p}}(\mathbf{r}, t) = P\Psi_{\mathbf{p}}(-\mathbf{r}, t) = P\Psi_{-\mathbf{p}}(\mathbf{r}, t) \quad (1.8)$$

and so a particle at rest, with $\mathbf{p} = \mathbf{0}$, is an eigenstate of parity. The eigenvalue $P = \pm 1$ for a particle at rest is called the *intrinsic parity*, or just the *parity*, of the particle. Parity is a multiplicative quantum number, and thus for many-particle systems the appropriate generalisation of (1.5) is

$$\hat{P}\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, t) \equiv P_1 P_2 \cdots \Psi(-\mathbf{r}_1, -\mathbf{r}_2, \dots, t),$$

with one intrinsic parity factor P_1, P_2, \dots occurring for each particle present.

²¹These points are explored in more detail in, for example, Chapter 5 of Martin and Shaw (2017).

²²Eugene Wigner shared the 1963 Nobel Prize in Physics, principally for his work on symmetries.

The strong and electromagnetic interactions, but not the weak interactions, are invariant under parity, that is, the interaction Hamiltonian of the system, and hence the equation of motion, remains unchanged under a parity transformation on the position vectors of all particles in the system. Parity is therefore conserved, by which we mean that the total parity quantum number remains unchanged in the interaction. Compelling evidence for parity conservation in the strong and electromagnetic interactions comes from the suppression of transitions between nuclear states that would violate parity conservation. Such decays are not absolutely forbidden, because the Hamiltonian responsible for the transition will always have a small admixture due to the weak interactions between nucleons. However, the observed rates are extremely small compared to analogous decays that do not violate parity and are entirely consistent with the transitions being due to this very small weak interaction component. The evidence for nonconservation of parity in the weak interaction will be discussed in detail in Section 7.1.

In addition to intrinsic parity, there is a contribution to the total parity if the particle has an orbital angular momentum l . In this case its wavefunction is a product of a radial part R_{nl} and an angular part $Y_l^m(\theta, \phi)$:

$$\Psi_{lmn}(\mathbf{r}) = R_{nl} Y_l^m(\theta, \phi), \quad (1.9)$$

where n and m are the principal and magnetic quantum numbers and $Y_l^m(\theta, \phi)$ is a spherical harmonic. It is straightforward to show from the relations between Cartesian (x, y, z) and spherical polar coordinates (r, θ, ϕ) , that is,

$$x = r \sin \theta \cos \phi, \quad y = r \sin \theta \sin \phi, \quad z = r \cos \theta, \quad (1.10)$$

where the parity transformation $\mathbf{r} \rightarrow -\mathbf{r}$ implies

$$r \rightarrow r, \quad \theta \rightarrow \pi - \theta, \quad \phi \rightarrow \pi + \phi, \quad (1.11)$$

and from this it can be shown that

$$Y_l^m(\theta, \phi) \rightarrow Y_l^m(\pi - \theta, \pi + \phi) = (-)^l Y_l^m(\theta, \phi). \quad (1.12)$$

Equation (1.12) may easily be verified directly for specific cases; for example, for the first three spherical harmonics,

$$Y_0^0 = \left(\frac{1}{4\pi} \right)^{1/2}, \quad Y_1^0 = \left(\frac{3}{4\pi} \right)^{1/2} \cos \theta, \quad Y_1^{\pm 1} = \left(\frac{3}{8\pi} \right)^{1/2} \sin \theta e^{\pm i\phi}. \quad (1.13)$$

Hence

$$\hat{P} \Psi_{lmn}(\mathbf{r}) = P \Psi_{lmn}(-\mathbf{r}) = P(-)^l \Psi_{lmn}(\mathbf{r}), \quad (1.14)$$

that is, $\Psi_{lmn}(\mathbf{r})$ is an eigenstate of parity with eigenvalue $P(-1)^l$.

An analysis of the Dirac equation (1.1) for relativistic electrons shows that it is invariant under a parity transformation only if $P(e^+)P(e^-) = -1$. This is a general result for all fermion–antifermion pairs, so it is a convention to assign $P = +1$ to all leptons and $P = -1$ to their antiparticles. We will see in Chapter 3, that in strong and electromagnetic interactions quarks can only be created as part of a quark–antiquark pair, so the intrinsic parity of a single quark cannot be measured. For this reason, it is also a convention to assign $P = +1$ to quarks. Since quarks are fermions, it follows from the Dirac result that $P = -1$ for antiquarks. The intrinsic parities of hadrons then follow from their structure in terms of quarks and the orbital angular momentum between the constituent quarks, using (1.14). This will be explored in Chapter 3 as part of the discussion of the quark model.

1.3.2 Charge conjugation

Charge conjugation is the operation of changing a particle into its antiparticle. Like parity, it gives rise to a multiplicative quantum number that is conserved in strong and electromagnetic interactions, but violated in the weak interaction. In strong interactions this can be tested experimentally, by, for example, measuring the rates of production of positive and negative mesons in $p\bar{p}$ annihilations, and is found to hold.

In discussing charge conjugation, we will need to distinguish between states such as the photon γ and the neutral pion π^0 that do not have distinct antiparticles and those such as the π^+ and the neutron that do. Particles in the former class we will collectively denote by a , and those of the latter type will be denoted by b . It is also convenient at this point to extend our notation for states. Thus we will represent a state of type a having a wavefunction Ψ_a by $|a, \Psi_a\rangle$ and similarly for a state of type b .²³ Then under the charge conjugation operator \hat{C} ,

$$\hat{C}|a, \Psi_a\rangle = C_a|a, \Psi_a\rangle \quad \text{and} \quad \hat{C}|b, \Psi_b\rangle = |\bar{b}, \Psi_{\bar{b}}\rangle, \quad (1.15)$$

where C_a is a phase factor analogous to the phase factor in (1.5).²⁴ Applying the operator twice, in the same way as for parity, leads to $C_a = \pm 1$. From the first equation in (1.15), we see that states of type a are eigenstates of \hat{C} with eigenvalues ± 1 , called their *C parities*. As an example, consider the π^0 . This decays via the electromagnetic interaction to two photons: $\pi^0 \rightarrow \gamma\gamma$. The *C* parity of the photon follows directly from the invariance of Maxwell’s equations under charge conjugation and is

²³This is part of the so-called ‘Dirac notation’ in quantum mechanics. However, we will only need the notation and not the associated mathematics.

²⁴A phase factor C_b could also have been inserted in the second equation of (1.15), but it is straightforward to show that the relative phase of the two states b and \bar{b} cannot be measured, and so a phase introduced in this way would have no physical consequences. (See Problem 1.4.)

$C_\gamma = -1^{25}$ and hence $C_{\pi^0} = C_\gamma^2 = 1$. It follows that the decay $\pi^0 \rightarrow \gamma\gamma\gamma$ is forbidden by C invariance. The experimental limit for the ratio of rates $\pi^0 \rightarrow 3\gamma/\pi^0 \rightarrow 2\gamma$ is less than 3×10^{-8} , which is strong evidence for C invariance in electromagnetic interactions. The evidence for the violation of C invariance in the weak interaction is discussed in detail in Chapter 7.

If a state contains particles that have distinct antiparticles, it can only be an eigenstate of \hat{C} if they are present as particle–antiparticle pairs. As an example of this, consider a $\pi^+\pi^-$ pair with orbital angular momentum L between them. We then have

$$\hat{C}|\pi^+\pi^-; L\rangle = (-1)^L |\pi^+\pi^-; L\rangle, \quad (1.16)$$

because interchanging the pions reverses their relative positions in the spatial wavefunction. The same factor occurs for spin-1/2 fermion pairs $f\bar{f}$, but in addition there are two other factors. The first is $(-1)^{S+1}$, where S is the total spin of the pair. This follows directly from the structure of the spin wavefunctions:

$$\left. \begin{array}{ll} \frac{1}{\sqrt{2}} (\uparrow_1 \uparrow_2 + \downarrow_1 \downarrow_2) & S_z = 1 \\ & S_z = 0 \\ \frac{1}{\sqrt{2}} (\downarrow_1 \downarrow_2) & S_z = -1 \end{array} \right\} \quad S = 1 \quad (1.17a)$$

and

$$\frac{1}{\sqrt{2}} (\uparrow_1 \downarrow_2 - \downarrow_1 \uparrow_2) \quad S_z = 0 \quad S = 0, \quad (1.17b)$$

where $\uparrow_i(\downarrow_i)$ represents particle i having spin ‘up’ (‘down’) in the z direction. A second factor (-1) arises whenever fermions and antifermions are interchanged. This has its origins in quantum field theory.²⁶ Combining these factors, finally we have

$$\hat{C}|f\bar{f}; J, L, S\rangle = (-1)^{L+S} |f\bar{f}; J, L, S\rangle, \quad (1.18)$$

for fermion–antifermion pairs having total, orbital, and spin angular momentum quantum numbers J , L , and S , respectively.

1.3.3 Time reversal

Time-reversal invariance is defined as invariance under the transformation

$$t \rightarrow t' = -t, \quad (1.19)$$

leaving all position vectors unchanged. Like parity and charge conjugation invariance, it is a symmetry of the strong and electromagnetic

²⁵A proof of this is given in Section 5.4.1 of Martin and Shaw (2017). An alternative argument is that electromagnetic fields are produced by moving electric charges, which change sign under charge conjugation, and hence $C_\gamma = -1$.

²⁶See, for example, pp. 249–250 of Gottfried and Weisskopf (1986).

interactions, but is violated by the weak interactions. However, unlike parity and charge conjugation, there is no associated quantum number that is conserved when weak interactions are neglected. To understand this we consider the transformation of a single-particle wavefunction, which must satisfy

$$|\Psi(\mathbf{r}, t)|^2 \xrightarrow{T} |\Psi'(\mathbf{r}, t)|^2 = |\Psi(\mathbf{r}, -t)|^2 \quad (1.20)$$

if the system is T invariant, so that the probability of finding the particle at position \mathbf{r} at time $-t$ becomes the probability of finding it at position \mathbf{r} at time t in the transformed system. In addition, since in classical mechanics linear and angular momentum change sign under (1.19), we would expect the same result

$$\mathbf{p} \xrightarrow{T} \mathbf{p}' = -\mathbf{p}; \quad \mathbf{J} \xrightarrow{T} \mathbf{J}' = -\mathbf{J} \quad (1.21)$$

to hold in quantum mechanics by the correspondence principle. Hence a free-particle wavefunction

$$\Psi_{\mathbf{p}}(\mathbf{r}, t) = \exp[i(\mathbf{p} \cdot \mathbf{r} - Et)/\hbar],$$

corresponding to momentum \mathbf{p} and energy $E = p^2/2m$, must transform into a wavefunction corresponding to momentum $-\mathbf{p}$ and energy E , i.e.

$$\Psi_{\mathbf{p}}(\mathbf{r}, t) \xrightarrow{T} \Psi'_{\mathbf{p}}(\mathbf{r}, t) = \Psi_{-\mathbf{p}}(\mathbf{r}, t) = \exp[-i(\mathbf{p} \cdot \mathbf{r} + Et)/\hbar]. \quad (1.22)$$

A suitable transformation that satisfies both (1.20) and (1.22) is

$$\Psi(\mathbf{r}, t) \xrightarrow{T} \Psi'(\mathbf{r}, t) = \Psi^*(\mathbf{r}, -t) \equiv \hat{T} \Psi(\mathbf{r}, t), \quad (1.23)$$

where we have introduced the time reversal operator \hat{T} by analogy with the parity operator \hat{P} introduced in (1.5). However, quantum mechanical operators \hat{O} that correspond to physical observables must be both linear

$$\hat{O}(\alpha_1 \Psi_1 + \alpha_2 \Psi_2) = \alpha_1 (\hat{O} \Psi_1) + \alpha_2 (\hat{O} \Psi_2) \quad (1.24a)$$

(to ensure that the superposition principle holds) and Hermitian

$$\int (\hat{O} \Psi_1)^* \Psi_2 \, dx = \int \Psi_1^* (\hat{O} \Psi_2) \, dx \quad (1.24b)$$

(to ensure that the eigenvalues of \hat{O} , that is, observable quantities, are real), where $\Psi_{1,2}$ are arbitrary wavefunctions and $\alpha_{1,2}$ are arbitrary complex numbers. In contrast, the definition (1.23) implies

$$\hat{T}(\alpha_1 \Psi_1 + \alpha_2 \Psi_2) = \alpha_1^* (\hat{T} \Psi_1) + \alpha_2^* (\hat{T} \Psi_2) \neq \alpha_1 (\hat{T} \Psi_1) + \alpha_2 (\hat{T} \Psi_2)$$

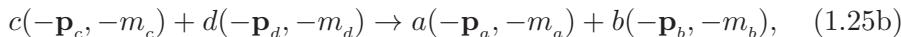
for complex α_1 and α_2 , and one easily verifies that (1.24b) is also not satisfied by \hat{T} . Thus the time reversal operator does not correspond to a

physical observable and there is no observable analogous to parity that is conserved as a consequence of T invariance.

Although T invariance does not give rise to a conservation law, it does lead to a relation between any reaction and the ‘time-reversed’ process related to it by (1.19). Thus reactions like



and their time-reversed counterparts

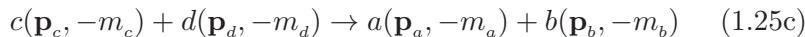


in which the initial and final states are interchanged and the particle momenta (\mathbf{p}_a etc.) and z components of their spins (m_a etc.) are reversed in accordance with (1.21), are related. In particular, if weak interactions are neglected, the rates for reactions (1.25a) and (1.25b) must be equal.

A more useful relation between reaction rates can be obtained if we combine time reversal with parity invariance. Under the parity transformation (1.5), momenta \mathbf{p} change sign while orbital angular momenta $\mathbf{L} = \mathbf{r} \times \mathbf{p}$ do not. If we assume the same behaviour holds for spin angular momenta, then

$$\mathbf{p} \xrightarrow{P} \mathbf{p}' = -\mathbf{p}; \quad \mathbf{J} \xrightarrow{P} \mathbf{J}' = \mathbf{J} \quad (1.26)$$

under parity. The parity-transformed reaction corresponding to (1.25b) is



so that if both P and T invariance holds, all three reactions (1.25a, 1.25b, and 1.25c) must have the same rate. If we average over all spin projections

$$m_i = -S_i, -S_i + 1, \dots S_i \quad (i = a, b, c, d),$$

where S_i is the spin of particle i , then reactions (1.25a) and (1.25c) differ only by the interchange of initial and final states. Consequently, the rates for the reactions

$$i \equiv a(\mathbf{p}_a) + b(\mathbf{p}_b) \leftrightarrow c(\mathbf{p}_c) + d(\mathbf{p}_d) \equiv f \quad (1.27)$$

should be equal, provided that we average over all possible spin states. This relation is called the *principle of detailed balance* and has been accurately confirmed experimentally in a variety of strong and electromagnetic reactions.

Finally, although the weak interaction is not invariant under the above transformations, there is a general result, called the *CPT theorem*, which states that under very general conditions *any* relativistic field theory is invariant under the combined operation of *CPT*, taken in any order. Among other things, *CPT* invariance predicts that the masses and lifetimes of a particle and its antiparticle must be exactly equal. These predictions are consistent with all known data. The measured mass of the

positron, for example, is equal to the mass of the electron within an experimental uncertainty of better than one part in 10^8 .

1.4 Interactions and Feynman diagrams

We now turn to a discussion of particle interactions and how they can be described by the very useful pictorial methods of Feynman diagrams.

1.4.1 Interactions

Reactions involving elementary particles and/or hadrons are conveniently summarised by ‘equations’ in analogy to chemical reactions, in which the different particles are represented by symbols, which usually, but not always, have a superscript to denote their electric charge. For example, in the interaction

$$\nu_e + n \rightarrow e^- + p, \quad (1.28)$$

a neutrino ν_e (the subscript is explained in Section (3.1.3)) collides with a neutron n to give an electron e^- and a proton p , while the equation

$$e^- + p \rightarrow e^- + p \quad (1.29)$$

represents an electron and proton interacting to give the same particles in the final state, but in general travelling in different directions. In such equations, conserved quantum numbers must have the same total values in initial and final states.

Particles may be transferred from initial to final states and vice versa, when they become antiparticles. Thus starting from the process

$$\pi^- + p \rightarrow \pi^- + p, \quad (1.30a)$$

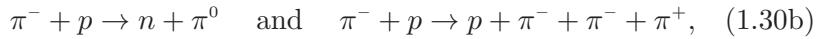
and taking the proton from the initial state to an antiproton in the final state and the negatively charged pion in the final state to a positively charged pion in the initial state, we obtain

$$\pi^+ + \pi^- \rightarrow p + \bar{p}. \quad (1.31)$$

It follows that if (1.30a) does not violate any relevant quantum numbers, then neither does reaction (1.31) and so is also in principle an allowed reaction. The qualification is needed because although (1.31) does not violate any quantum numbers, energy conservation leads to a minimum total energy below which it cannot proceed, because the two pions must have enough energy to create the heavier (i.e. having greater mass) $p\bar{p}$ pair.

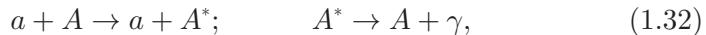
The interactions (1.29) and (1.30a), in which the particles remain unchanged, are examples of *elastic scattering*, in contrast to reactions (1.28) and (1.31), where the final-state particles differ from those in the initial state. Collisions between a given pair of initial particles do not always lead to the same final state, but can lead to different final states

with different probabilities. For example, the collision of a negatively charged pion and a proton can give rise to elastic scattering (1.30a) and a variety of other reactions, such as



depending on the initial energy. In particle physics it is common to also refer (rather imprecisely) to such interactions as ‘inelastic’ scattering.

Similar considerations apply to nuclear physics, but the term *inelastic scattering* is reserved for the case where the final state is an excited state of the parent nucleus A , which subsequently decays, for example via photon emission, i.e.

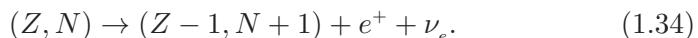


where a is a projectile and A^* is an excited state of A . A useful shorthand notation used in nuclear physics for the general reaction $a + A \rightarrow b + B$ is $A(a, b)B$. It is usual in nuclear physics to further subdivide types of interactions according to the underlying mechanism that produced them. We will return to this in Section 2.9, as part of a more general discussion of nuclear reactions.

Finally, many particles are unstable and spontaneously decay to other, lighter (i.e. having less mass) particles. An example of this is the free neutron (i.e. one not bound in a nucleus), which decays by the β -decay reaction



with a mean lifetime of about 900 seconds.²⁷ The same notation can also be used in nuclear physics. For example, many nuclei decay via the β -decay mechanism. Thus, denoting a nucleus with Z protons and N neutrons as (Z, N) , we have



This is also a weak interaction. This reaction is effectively the decay of a proton bound in a nucleus. Although a *free* proton cannot decay by the beta decay $p \rightarrow n + e^+ + \nu_e$ because it violates energy conservation (the final-state particles have greater total mass than the proton), a proton bound in a nucleus can decay because of its binding energy. The explanation for this is given in Chapter 2.

1.4.2 Feynman diagrams

The forces producing all of the above interactions are due to the exchange of particles and a convenient way of illustrating this is to use *Feynman diagrams*. There are mathematical rules (the *Feynman rules*) and techniques associated with these that enable them to be used to calculate the

²⁷The reason why this decay involves an antineutrino rather than a neutrino will become clear in Chapter 3.

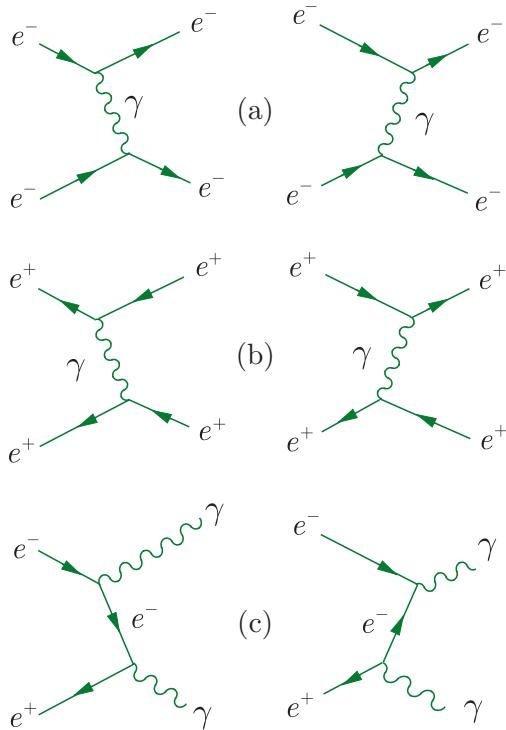


Figure 1.2 Some examples of electromagnetic processes:
 (a) single-photon exchange in $e^- + e^- \rightarrow e^- + e^-$;
 (b) single-photon exchange in $e^+ + e^+ \rightarrow e^+ + e^+$;
 (c) electron–positron annihilation producing two photons. Time runs from left to right and in each case the two diagrams are related by ‘time ordering’, as explained in the text.

quantum mechanical probabilities for given reactions to occur, but in this book Feynman diagrams will only be used as a convenient very useful pictorial description of reaction mechanisms.

We first illustrate them at the level of elementary particles for the case of electromagnetic interactions, which arise from the emission and/or absorption of photons. For example, the dominant interaction between two electrons is due to single photon exchange processes in which the photon is emitted by one of the electrons and absorbed by the other. This mechanism, which gives rise to the familiar Coulomb interaction at large distances, is illustrated in the Feynman diagrams of Figure 1.2a.

In such diagrams, we will use the convention that particles in the initial state are shown on the left and particles in the final state are shown on the right, i.e. the time axis runs from left to right. Spin-1/2 fermions (such as the electron) are drawn as solid lines and photons are drawn as wiggly lines. Arrows pointing to the right indicate that the solid lines represent electrons. In the case of photon exchange between two positrons, which is shown in Figure 1.1b, the arrows on the antiparticle (in this case the positron) lines are conventionally shown as pointing to the left. In interpreting these diagrams, it is important to remember that the direction of the arrows on fermion lines does not indicate the particle’s direction of motion, but merely whether the fermions are particles or antiparticles, and that particles in the initial state are always to the left of the vertex and particles in the final state are always to the right. Pairs of

electrons and positrons can also annihilate to produce photons in the final state and an example of this is shown in Figure 1.2c, which corresponds to the process $e^+ + e^- \rightarrow \gamma + \gamma$.

The two diagrams shown in Figure 1.2c are closely related. If the lines of the first diagram were made of rubber, we could imagine deforming them so that the top vertex occurred after, instead of before, the bottom vertex, and it became the second diagram. The pairs of diagrams shown in each of Figure 1.2a and b are related in the same way. For any given process, diagrams related in this way are called different ‘time orderings’. In practice, it is usual to draw only one time ordering (e.g. the left-hand diagram in Figure 1.2a) leaving the other(s) implied, and we shall usually follow this practice in what follows. In contrast, for any given process, pairs of diagrams that are not related in this way are referred to as ‘topologically distinct’ and must both be retained.

A feature of the above diagrams is that they are constructed from combinations of simple three-line vertices. This is characteristic of electromagnetic processes. Each vertex has a line corresponding to a single photon being emitted or absorbed, while one fermion line has the arrow pointing towards the vertex and the other away from the vertex, guaranteeing charge conservation at the vertex, which is one of the rules of Feynman diagrams.²⁸ For example, a vertex like Figure 1.3 would correspond to a process in which an electron emitted a photon and turned into a positron. This would violate charge conservation and is therefore forbidden.

Feynman diagrams can also be used to describe the fundamental weak and strong interactions which, as mentioned in Section 1.1.3, are mediated by the massive W^+ , W^- and Z^0 bosons and the massless gluons, respectively. This is illustrated by Figure 1.4a and b, which show contributions to the elastic weak scattering reaction $e^- + \nu_e \rightarrow e^- + \nu_e$ due to the exchange of single Z^0 and W bosons, and by Figure 1.4c, which shows the exchange of a gluon g (represented by a coiled line) between two quarks q , which is a strong interaction.

Feynman diagrams that involve hadrons can also be drawn. As illustrations, Figure 1.5a shows the decay of a neutron via an intermediate W boson and Figure 1.5b and c denote the exchange of a neutral and a charged pion, respectively (shown as a dashed line), between a proton and a neutron. Pion exchange diagrams of this type form the basis of Yukawa’s theory of nuclear forces mentioned earlier. He was then able to predict the pion mass by using a fundamental relation between the mass of the exchanged particle and the range of the resulting force discussed in the following two subsections.

We turn now to consider in more detail the relation between exchanged particles and forces.

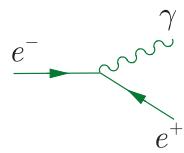


Figure 1.3 The forbidden vertex $e^- \rightarrow e^+ + \gamma$.

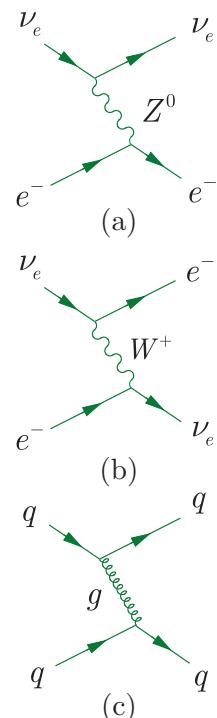


Figure 1.4 Contributions to the elastic weak scattering reaction $e^- + \nu_e \rightarrow e^- + \nu_e$ by (a) the exchange of a Z^0 boson and (b) the exchange of a W boson; (c) gluon exchange contribution to the strong interaction $q + q \rightarrow q + q$.

²⁸Compare Kirchhoff’s laws in electromagnetism.

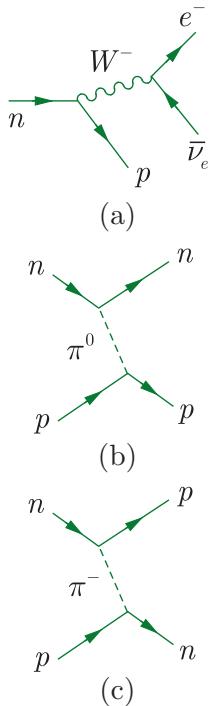


Figure 1.5 (a) The decay of a neutron via an intermediate W boson; (b) single π^0 exchange in the reaction $p + n \rightarrow p + n$; (c) single π^- exchange in the reaction $p + n \rightarrow n + p$.

1.5 Particle exchange: forces and potentials

This section starts with a discussion of the important relationship between forces and particle exchanges and then relates this to potentials. Although the idea of a potential has its greatest use in nonrelativistic physics, nevertheless it is useful to illustrate concepts and is used in later sections as an intermediate step in relating theoretical Feynman diagrams to measurable quantities. The results can be extended to more general situations.

1.5.1 Range of forces

At each vertex of a Feynman diagram, charge is conserved by construction. We will see later that depending on the nature of the interaction (strong, weak or electromagnetic), other quantum numbers are also conserved. However, it is easy to show that energy and momentum cannot in general be conserved simultaneously.

Consider the general case of a reaction $A + B \rightarrow A + B$ mediated by the exchange of a particle X , as shown in Figure 1.6. In the rest frame of the incident particle A , the lower vertex represents the *virtual* process ('virtual' because X does not appear as a real particle in the final state),

$$A(M_A c^2, \mathbf{0}) \rightarrow A(E_A, \mathbf{p}_A c) + X(E_X, -\mathbf{p}_A c), \quad (1.35)$$

where E_A is the *total* energy of the final particle A and \mathbf{p}_A is its 3-momentum.²⁹ Thus, if we denote by P_A the 4-momentum for particle A ,

$$P_A = (E_A/c, \mathbf{p}_A) \quad (1.36)$$

and

$$P_A^2 = E_A^2/c^2 - \mathbf{p}_A^2 = M_A^2 c^2. \quad (1.37)$$

Applying this to the diagram and imposing momentum conservation gives

$$E_A = (p^2 c^2 + M_A^2 c^4)^{1/2} \text{ and } E_X = (p^2 c^2 + M_X^2 c^4)^{1/2}, \quad (1.38)$$

where $p = |\mathbf{p}_A|$. The energy difference between the final and initial states is given by

$$\begin{aligned} \Delta E &= E_X + E_A - M_A c^2 \rightarrow 2pc, & p \rightarrow \infty \\ &\rightarrow M_X c^2, & p \rightarrow 0 \end{aligned} \quad (1.39)$$

and thus $\Delta E \geq M_X c^2$ for all p , i.e. energy is not conserved. However, by the energy-time uncertainty principle, such an energy violation is allowed, but only for a time $\tau \approx \hbar/\Delta E$, so we immediately obtain

$$R \equiv \hbar/M_X c \quad (1.40)$$

as the approximate value of the maximum distance over which X can propagate before being absorbed by particle B . The distance R is called

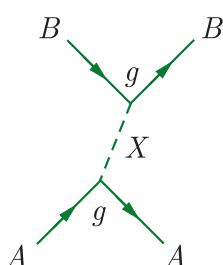


Figure 1.6 Exchange of a particle X in the reaction $A + B \rightarrow A + B$.

²⁹A résumé of relativistic kinematics is given in Appendix B.

the *range* of the interaction, which was the sense of the word used in Section 1.1.1.

The electromagnetic interaction has an infinite range because the exchanged particle is a massless photon. In contrast, the weak interaction is associated with the exchange of very heavy particles – the W and Z bosons. These lead to ranges that from (1.40) are approximately $R_{W, Z} \approx 2 \times 10^{-18}$ m. The fundamental strong interaction between quarks has infinite range because, like the photon, gluons have zero mass. On the other hand, the strong nuclear force has a much shorter range of approximately $(1 - 2) \times 10^{-15}$ m, corresponding to the exchange of pions with the mass predicted by Yukawa, as illustrated in Figure 1.5a and b. We will return briefly to the relation between these two different manifestations of the strong interaction in Section 8.1.

1.5.2 The Yukawa potential

In the limit that M_A becomes large, we can regard B as being scattered by a static potential of which A is the source. This potential will in general be spin dependent, but its main features can be obtained by neglecting spin and considering X to be a spin-0 boson, in which case it will obey the Klein–Gordon equation,

$$-\hbar^2 \frac{\partial^2 \phi(\mathbf{r}, t)}{\partial t^2} = -\hbar^2 c^2 \nabla^2 \phi(\mathbf{r}, t) + M_X^2 c^4 \phi(\mathbf{r}, t). \quad (1.41)$$

The static solution of this equation satisfies

$$\nabla^2 \phi(\mathbf{r}) = \frac{M_X^2 c^2}{\hbar^2} \phi(\mathbf{r}), \quad (1.42)$$

where $\phi(\mathbf{r})$ is interpreted as a static potential. For $M_X = 0$ this equation is the same as that obeyed by the electrostatic potential, and for a point charge $-e$ interacting with a point charge $+e$ at the origin, the appropriate solution is the Coulomb potential

$$V(r) = -e \phi(r) = -\frac{e^2}{4\pi\epsilon_0} \frac{1}{r}, \quad (1.43)$$

where $r = |\mathbf{r}|$ and ϵ_0 is the dielectric constant. The corresponding solution in the case where $M_X^2 \neq 0$ is easily verified by substitution to be

$$V(r) = -\frac{g^2}{4\pi} \frac{e^{-r/R}}{r}, \quad (1.44)$$

where R is the range defined earlier and g , the so-called *coupling constant*, is a parameter associated with each vertex of a Feynman diagram and represents the basic strength of the interaction. For simplicity, we have assumed equal strengths for the coupling of particle X to the particles A and B .

The form of $V(r)$ in (1.44) is called a *Yukawa potential*, after the physicist who in 1935 first introduced the idea of forces due to the exchange of massive particles. As $M_x \rightarrow 0$, $R \rightarrow \infty$ and the Coulomb potential is recovered from the Yukawa potential, while for very large masses the interaction is approximately point-like (zero-range). It is conventional to introduce a dimensionless parameter α_x by

$$\alpha_x = g^2 / 4\pi\hbar c, \quad (1.45)$$

which characterises the strength of the interaction. For the electromagnetic interaction this is denoted

$$\alpha \equiv e^2 / 4\pi\varepsilon_0\hbar c \approx 1/137 \quad (1.46)$$

and is called the *fine structure constant* because it determines the magnitude of the fine structure seen in atomic spectral lines.

In deriving the above potential we implicitly assumed point-like particles, so that (1.41) can be used for all $r \neq 0$. For composite particles, at distances smaller or of the same order as the size of the particles, it breaks down and the interaction must be modified to take account of the particle's size and structure. This is important for the strong forces between hadrons, whose range is not much bigger than the size of the hadrons themselves, so that the strong nuclear force, for example, is only dominated by pion exchange over a limited range and is significantly modified at distances less than 10^{-15}m . The strong nuclear interaction is actually a complicated effect that has its origins in the fundamental strong interactions between the quark distributions within the two hadrons. Similarly, two neutral atoms also experience an electromagnetic interaction (the van der Waals force), which has its origins in the fundamental Coulomb forces, but is of a much shorter range. Although an analogous mechanism is not in fact responsible for the nuclear strong interaction, it is a useful reminder that the force between two *distributions* of particles can be much more complicated than the forces between the individual components. We will return to this point when we discuss the nature of the nuclear potential in more detail in Section 8.1.

1.6 Observable quantities: cross-sections and decay rates

We have mentioned earlier that Feynman diagrams can be translated into probabilities for a process by using a set of mathematical rules (the *Feynman Rules*) that can be derived from the quantum theory of the underlying interaction. In the case of the electromagnetic interaction, the theory is called Quantum Electrodynamics (QED) and is spectacularly successful in explaining experimental results.³⁰ We will not pursue this

³⁰Richard Feynman, Sin-Itiro Tomonaga, and Julian Schwinger shared the 1965 Nobel Prize in Physics for their work on formulating quantum electrodynamics. The Feynman rules are introduced in an accessible way in Griffiths (1987) or more rigorously in Mandl and Shaw (2010).

in detail in this book, but rather will show in principle their relation to *observables*, i.e. quantities that can be measured, concentrating on the cases of two-body scattering reactions and decays of unstable states.

1.6.1 Amplitudes

The intermediate step is the *amplitude* \mathcal{M} , the modulus squared of which is directly related to the probability of the process occurring. To get a qualitative idea of the structure of \mathcal{M} , we will use nonrelativistic quantum mechanics and assume that the coupling constant g^2 is small compared to $4\pi\hbar c$, so that the interaction is a small perturbation on the free particle solution, which will be taken as plane waves.

In lowest-order perturbation theory, the probability amplitude for a particle with initial momentum \mathbf{q}_i to be scattered to a final state with momentum \mathbf{q}_f by a potential $V(\mathbf{r})$ is proportional to³¹

$$\mathcal{M}(\mathbf{q}) = \int V(\mathbf{r}) \exp(i\mathbf{q} \cdot \mathbf{r}/\hbar) d^3\mathbf{r}, \quad (1.47)$$

where $\mathbf{q} \equiv \mathbf{q}_f - \mathbf{q}_i$ is the momentum transfer. The integration may be done using polar coordinates. Taking \mathbf{q} in the z direction gives

$$\mathbf{q} \cdot \mathbf{r} = |\mathbf{q}| r \cos \theta \quad (1.48)$$

and

$$d^3\mathbf{r} = r^2 \sin \theta d\theta dr d\phi, \quad (1.49)$$

where $r \equiv |\mathbf{r}|$. For the Yukawa potential, the integral (1.47) gives

$$\mathcal{M}(\mathbf{q}^2) = \frac{-g^2 \hbar^2}{|\mathbf{q}|^2 + M_X^2 c^2}. \quad (1.50)$$

In deriving (1.50) for the scattering amplitude, we have used potential theory, treating the particle A as a static source. The particle B then scatters through some angle without loss of energy, so that $|\mathbf{q}_i| = |\mathbf{q}_f|$ and the initial and final energies of particle B are equal, $E_i = E_f$. While this is a good approximation at low energies, at higher energies the recoil energy of the target particle cannot be neglected, so that the initial and final energies of B are no longer equal. A full relativistic calculation taking account of this is beyond the scope of this book, but the result is surprisingly simple. Specifically, in lowest-order perturbation theory, one obtains

$$\mathcal{M}(q^2) = \frac{g^2 \hbar^2}{q^2 - M_X^2 c^2}, \quad (1.51)$$

³¹This is called the Born approximation. For a discussion, see, for example, Section 10.2.2 of Mandl (1992) or pp. 397–399 of Gasiorowicz (1974).

where

$$q^2 \equiv (E_f - E_i)^2/c^2 - (\mathbf{q}_f - \mathbf{q}_i)^2 \quad (1.52)$$

is the squared four-momentum transfer. In the low-energy limit, $E_i = E_f$ and (1.51) reduces to (1.50). However, in contrast to (1.50), which was derived in the rest frame of particle A , the form (1.51) is explicitly Lorentz invariant and holds in all inertial frames of reference. It is thus also called the *invariant amplitude*.

In the zero-range approximation, (1.51) reduces to a constant. To see this, we note that this approximation is valid when the range $R = \hbar/M_X c$ is very small compared to the de Broglie wavelengths of all the particles involved. In particular, this implies $q^2 \ll M_X^2 c^2$, and neglecting q^2 in (1.51) gives

$$\mathcal{M}(q^2) = -G, \quad (1.53a)$$

where the constant G is given by

$$\frac{G}{(\hbar c)^3} = \frac{1}{\hbar c} \left(\frac{g}{M_X c^2} \right)^2 = \frac{4\pi\alpha_X}{(M_X c^2)^2} \quad (1.53b)$$

and the right-hand side has the dimensions of inverse energy squared. Thus we see that in the zero-range approximation, the resulting point interaction between A and B is characterised by a single dimensioned coupling constant G and not g and M_X separately.

In the above discussion, we have assumed for simplicity that the exchanged particle is a spin-zero meson. However, by far the most important application of the zero-range approximation is to the weak interaction, in which the exchanged particle is a heavy spin-1 boson. In particular, most observed low-energy weak processes, including β -decay, are dominated by W -exchange; if the spin structure of the interaction is taken into account, the effective low-energy coupling corresponding to (1.53b) is called the *Fermi coupling constant*, and is given by

$$\frac{G_F}{(\hbar c)^3} = \frac{\sqrt{2}}{\hbar c} \left(\frac{g_W}{M_W c^2} \right)^2 = \frac{4\sqrt{2}\pi\alpha_W}{(M_W c^2)^2}, \quad (1.54)$$

where g_W is the weak coupling constant, $\alpha_W = g_W^2/4\pi\hbar c$ and $M_W \approx 80 \text{ GeV}/c^2$ is the mass of the W -boson. As we shall see, this approximation is extremely useful in weak interactions, and the Fermi coupling constant, measured in various processes, is given by

$$G_F/(\hbar c)^3 = 1.166 \times 10^{-5} \text{ GeV}^{-2}. \quad (1.55)$$

The amplitude (1.50) corresponds to the exchange of a single particle, as shown, for example, in Figure 1.6. It is also possible to draw more complicated Feynman diagrams that correspond to the exchange of more than one particle. An example of such a diagram for elastic $e^- e^-$ scattering, where two photons are exchanged, is shown in Figure 1.7. Multiparticle

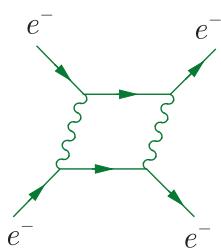


Figure 1.7 Two-photon exchange in the reaction $e^- + e^- \rightarrow e^- + e^-$.

exchange corresponds to higher orders in perturbation theory and higher powers of the appropriate coupling constant g^2 .

The number of vertices in any diagram is called the *order* n , and for electromagnetic processes, when the amplitude associated with any given Feynman diagram is calculated, it always contains a factor of $\alpha^{n/2}$. Since the probability is proportional to the square of the modulus of the amplitude, the former will contain a factor α^n . The probability associated with the single-photon exchange diagrams of Figure 1.2 thus contain a factor of α^2 and the contribution from two-photon exchange is of order α^4 . As $\alpha \approx 1/137$, the latter is usually very small compared to the contribution from a single-photon exchange. This is a general feature of electromagnetic interactions: because the fine structure constant is very small, in many cases only the lowest-order diagrams that contribute to a given process need be taken into account, and more complicated higher-order diagrams with more vertices can to a good approximation be ignored.

1.6.2 Cross-sections

For scattering reactions, the simplest observable is the *cross-section*. In *fixed-target experiments*, a beam of mono-energetic particles is directed on to a stationary target and the rates of production of various particles are measured. It is clear that these rates will be proportional to: (i) the number N of particles in the target illuminated by the beam and (ii) the rate per unit area at which beam particles cross a small surface placed in the beam at rest with respect to the target and perpendicular to the beam direction. This rate is called the *flux* and is given by

$$J = n_b v_i, \quad (1.56)$$

where n_b is the number density of particles in the beam and v_i is the magnitude of their velocity in the rest frame of the target. Hence the rate W_r at which a specific reaction r occurs in a particular experiment can be written in the form

$$W_r = JN\sigma_r, \quad (1.57a)$$

where σ_r , the constant of proportionality, is called the *cross-section* for reaction r . If the beam has a cross-sectional area S , its intensity is $I = JS$ and so an alternative expression for the rate is

$$W_r = N\sigma_r I/S = I\sigma_r n_t t, \quad (1.57b)$$

where n_t is the number of target particles per unit volume and t is the thickness of the target. If the target consists of an isotopic species of atomic mass M_A in atomic mass units (these are defined in Section 1.7 below), then $n_t = \rho N_A / M_A$, where ρ is the density of the target and N_A is Avogadro's constant. Thus, (1.57b) may be written

$$W_r = I\sigma_r (\rho t) N_A / M_A, \quad (1.57c)$$

where (ρt) is a measure of the amount of material in the target, expressed in units of mass per unit area. The form (1.57c) is particularly useful for the case of thin targets commonly used in experiments (such as those of Rutherford and his collaborators) to reduce the probability of multiple scattering.

The basic equation (1.57a) is frequently written in the form

$$W_r = L\sigma_r, \quad (1.58a)$$

where the product

$$L \equiv J N \quad (1.58b)$$

is called the *luminosity*. The luminosity has dimensions $[\text{length}]^{-2}[\text{time}]^{-1}$ and contains all the dependencies on the densities and geometries of the beam and target. The cross-section is independent of these factors, but characteristic of the particular reaction r . It follows from (1.58a) that σ_r has the dimensions of an area and the rate per target particle $J\sigma_r$, at which the reaction occurs is equal to the rate at which beam particles would hit a surface of area σ_r placed in the beam at rest with respect to the target and perpendicular to the beam direction. Since the area of such a surface is unchanged by a Lorentz transformation in the beam direction, the cross-section is the same in all reference frames related by such a transformation. In particular, the cross-sections in the *laboratory frame*, in which the target particles are at rest, and the *centre-of-mass frame*, in which the colliding particles have equal but opposite momenta, are identical.

In scattering experiments, the target is not always stationary and in the case of a *colliding beam experiment*, the ‘target’ is itself another beam. Equations (1.58a) and (1.58b) still hold, but it is more convenient to express the luminosity in a different form. For example, consider an experiment with beams of particles of types 1 and 2, which may or may not be identical, travelling in opposite directions and made to collide at a given point. The beams are not usually continuous, but composed of bunches, and we define N_1 and N_2 to be the numbers of particles per bunch in the two beams and f to be the frequency of collisions between bunches. Since the beams do not have sharp cutoffs at their edges, we will assume Gaussian shaped beams with transverse dimensions s_x and s_y , which yields a cross-sectional area $A = 4\pi s_x s_y$. In one collision between bunches, a particle in beam 1 crosses N_2/A particles in beam 2, regarded as the target, and because it is in a bunch of N_1 particles, the luminosity is

$$L = fN_1N_2/A. \quad (1.58c)$$

A related quantity that is also used is the *integrated luminosity*, defined as the integral of L over the time during which the experiment is performed. Hence the product of the integrated luminosity and the cross-section σ_r gives the total number of events corresponding to the reaction

r that would be observed in the experiment, assuming 100% detection efficiency.

The quantity σ_r is better named the *partial cross-section*, because it is the cross-section for a particular reaction r . Other types of cross-section are also of interest. One is the *total cross-section* σ_{tot} , defined by

$$\sigma_{\text{tot}} \equiv \sum_r \sigma_r, \quad (1.59)$$

where the summation is over all allowed reactions. Another is of particular interest for describing the angular distributions observed in reactions with only two particles in the final state, like the elastic scattering processes (1.29) and (1.30a). In such reactions, if one particle, referred to as the scattered particle, is emitted in a particular direction, then the direction of the other particle is determined by energy-momentum conservation (see Problem 1.14), so that it is sufficient to describe the angular distribution of the scattered particle alone. This is done by specifying another useful quantity, the *differential cross-section* $d\sigma_r(\theta, \phi)/d\Omega$, defined by

$$dW_r \equiv JN \frac{d\sigma_r(\theta, \phi)}{d\Omega} d\Omega, \quad (1.60)$$

where dW_r is the measured rate for the scattered particles to be emitted into an element of solid angle $d\Omega = d\cos\theta d\phi$ in the direction (θ, ϕ) , as shown in Figure 1.8. The partial cross-section σ_r is obtained on integrating the differential cross-section over all angles, i.e.

$$\sigma_r = \int_0^{2\pi} d\phi \int_{-1}^1 \frac{d\sigma_r(\theta, \phi)}{d\Omega} d\cos\theta. \quad (1.61)$$

1.6.3 The basic scattering formulas

The next step is to write the cross-sections in terms of the scattering amplitude $\mathcal{M}(\mathbf{q}^2)$ appropriate for describing the scattering of a nonrelativistic spinless particle from a potential. To do this it is convenient to consider a single beam particle interacting with a single target particle and to confine the whole system in an arbitrary large volume V (which cancels in the final result). The incident flux is then given by

$$J = n_b v_i = v_i/V \quad (1.62)$$

and since the number of target particles is $N = 1$, the differential rate is

$$dW_r = \frac{v_i}{V} \frac{d\sigma_r(\theta, \phi)}{d\Omega} d\Omega. \quad (1.63)$$

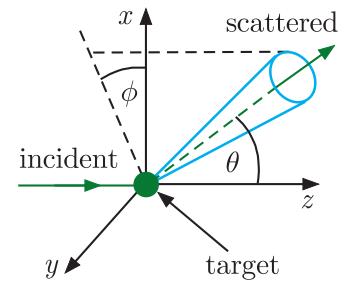


Figure 1.8 Geometry of the differential cross-section. A beam of particles, shown in green, is incident along the z axis and collides with a stationary target at the origin. The differential cross-section is proportional to the rate for particles to be scattered into a small solid angle $d\Omega$ in the direction (θ, ϕ) defining the cone shown in blue.

In quantum mechanics, provided the interaction is not too strong, the transition rate for any process is given in perturbation theory by the Born approximation³²

$$dW_r = \frac{2\pi}{\hbar} \left| \int \psi_r^* V(\mathbf{r}) \psi_i d^3 \mathbf{r} \right|^2 \rho(E_f). \quad (1.64)$$

The term $\rho(E_f)$ is the *density-of-states factor* (see below) and we take the initial and final state wavefunctions to be plane waves:

$$\psi_i = \frac{1}{\sqrt{V}} \exp(i\mathbf{q}_i \cdot \mathbf{r}/\hbar), \quad \psi_f = \frac{1}{\sqrt{V}} \exp(i\mathbf{q}_f \cdot \mathbf{r}/\hbar), \quad (1.65)$$

where the final momentum \mathbf{q}_f lies within a small solid angle $d\Omega$ located in the direction (θ, ϕ) (see Figure 1.8). Then, by direct integration,

$$dW_r = \frac{2\pi}{\hbar V^2} |\mathcal{M}(\mathbf{q}^2)|^2 \rho(E_f), \quad (1.66)$$

where $\mathcal{M}(\mathbf{q}^2)$ is the scattering amplitude defined in (1.47).

The density of states $\rho(E_f)$ that appears in (1.64) is defined so that the number of possible final states with energy lying between E_f and $E_f + dE_f$ is $\rho(E_f)dE_f$. It is given by³³

$$\rho(E_f) = \frac{V}{(2\pi\hbar)^3} q_f^2 \frac{dq_f}{dE_f} d\Omega, \quad (1.67)$$

where, nonrelativistically,

$$dq_f/dE_f = 1/v_f. \quad (1.68)$$

If we use (1.66), (1.67) and (1.68) in (1.63), we have

$$\frac{d\sigma}{d\Omega} = \frac{1}{4\pi^2\hbar^4} \frac{q_f^2}{v_i v_f} |\mathcal{M}(\mathbf{q}^2)|^2. \quad (1.69)$$

Although this result has been derived in the laboratory system, it is also valid in the centre-of-mass system.

The only place where nonrelativistic kinematics have been explicitly used in obtaining (1.69) is in the derivation of the density-of-states factor, so to have a formula that is also true for the general two-body relativistic scattering process $a + b \rightarrow c + d$, we have to re-examine the derivative (1.68) using relativistic kinematics. In this case we can use

$$E_f = E_c + E_d = (q_f^2 c^2 + m_c^2 c^4)^{1/2} + (q_f^2 c^2 + m_d^2 c^4)^{1/2} \quad (1.70)$$

to give

$$\frac{dE_f}{dq_f} = q_f c^2 \left(\frac{1}{E_c} + \frac{1}{E_d} \right), \quad (1.71)$$

³²This equation is a form of the *Second Golden Rule* in quantum mechanics. It is discussed in Section A.3.

³³The derivation is given in detail in Section A.2.

which, using the relativistic relation $\mathbf{v} = \mathbf{p}c^2/E$ (see Eq. (B.9) of Appendix B) and noting that in the centre-of-mass system $\mathbf{p}_c = -\mathbf{p}_d$, yields

$$\frac{dq_f}{dE_f} = \frac{1}{v_f}, \quad (1.72)$$

where v_f is the modulus of the relative velocity of particles c and d . Thus the general interpretation of (1.69) is that $q_f = |\mathbf{q}_c| = |\mathbf{q}_d|$ is the centre-of-mass momentum of the final-state particles and $v_{i,f}$ are the relative velocities in the centre-of-mass of particles a and b , and c and d , respectively.

All the above is for spinless particles, so finally we have to generalise (1.69) to include the effects of spin. Suppose the initial-state particles a and b have spins S_a and S_b and the final-state particles c and d have spins S_c and S_d . The total numbers of spin substates available to the initial and final states are g_i and g_f , respectively, given by

$$g_i = (2S_a + 1)(2S_b + 1) \quad \text{and} \quad g_f = (2S_c + 1)(2S_d + 1). \quad (1.73)$$

If the initial particles are unpolarised (which is the most common case in practice), then we must average over all possible initial spin configurations (because each is equally likely) and sum over the final configurations. Thus, (1.69) becomes

$$\frac{d\sigma}{d\Omega} = \frac{g_f}{4\pi^2\hbar^4} \frac{q_f^2}{v_i v_f} |\bar{\mathcal{M}}_f|^2, \quad (1.74)$$

where

$$|\bar{\mathcal{M}}_f|^2 \equiv |\bar{\mathcal{M}}(\mathbf{q}^2)|^2 \quad (1.75)$$

and the bar over the amplitude denotes a spin-average of the squared matrix element.

1.6.4 Unstable states

In the case of an unstable state, the observable of interest is its *lifetime at rest* τ . However, decays are quantum mechanical processes, governed by statistics, so that individual particles of a given type do not always have the same lifetime. In particle physics, it is usual to specify the *mean lifetime*, averaged over a large number of observed decays, whereas in nuclear physics it is usual to specify the *half-life* $t_{1/2}$, defined as the time for half the nuclei in a sample containing many nuclei to decay. In either case, the fact that these are independent of the number N of nuclei or particles in the sample implies that the average decay rate is proportional to N , i.e. the *activity*

$$\mathcal{A} \equiv -dN/dt = \lambda N, \quad (1.76)$$

where λ is called the *decay constant*. Integrating this equation leads immediately to the *exponential decay law*

$$N(t) = N_0 \exp(-\lambda t), \quad (1.77)$$

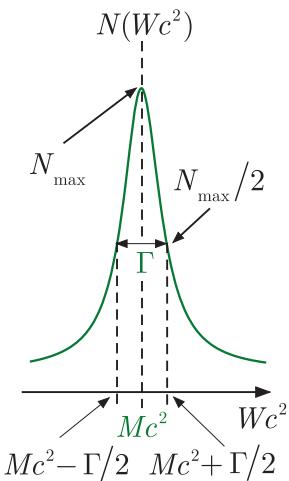


Figure 1.9 The Breit–Wigner formula (1.82).

where N_0 is the initial number of nuclei, i.e. the number at $t = 0$. The mean lifetime is then given by

$$\tau \equiv \frac{\int t N(t) dt}{\int N(t) dt} = \frac{\int_0^\infty t \exp(-\lambda t) dt}{\int_0^\infty \exp(-\lambda t) dt} = \frac{1}{\lambda} \quad (1.78)$$

The half-life also follows directly from (1.77) and is given by

$$t_{1/2} = \ln 2 / \lambda = \tau \ln 2. \quad (1.79)$$

In this book, the term *lifetime* will stand for the mean lifetime in the rest frame of the decaying particle, both for radioactive nuclei and unstable hadrons, unless explicitly stated otherwise. An equivalent quantity is the *natural decay width*, given by $\Gamma = \hbar/\tau$, which is also a measure of the rate of the decay reaction. In general, an initial unstable state will decay to several final states and in this case we define Γ_f as the *partial width* for a specific final state f and

$$\Gamma = \sum_f \Gamma_f \quad (1.80)$$

as the *total decay width*, while

$$B_f \equiv \Gamma_f / \Gamma \quad (1.81)$$

is defined as the *branching ratio* for decay to the state f .

The energy distribution of an isolated unstable state decaying to a final state f has the *Breit–Wigner* form

$$N_f(W) \propto \frac{\Gamma_f}{(W - M)^2 c^4 + \Gamma^2/4}, \quad (1.82)$$

where M is the mass of the decaying state and $N_f(W)dW$ is the number of events in which the invariant mass of the decay products lies between W and $W+dW$.³⁴ The Breit–Wigner formula is shown in Figure 1.9 and is the same formula that describes the widths of atomic and nuclear spectral lines. It is a symmetrical bell-shaped curve with a maximum at $W = M$ and a full width Γ at half the maximum height of the curve, whose height is proportional to the number of events with invariant mass W .

If an unstable state is produced in a scattering reaction by the mechanism $1 + 2 \rightarrow R \rightarrow f$ illustrated in Figure 1.10, then the scattering cross-section for that reaction will show an enhancement described by the same Breit–Wigner formula. In this case, the unstable state is referred to as a *resonance* and the mechanism of Figure 1.10 is called a *resonance*

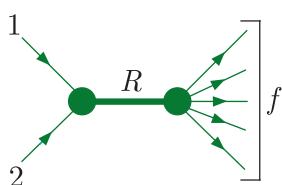


Figure 1.10 Formation and decay of a resonance R in the reaction $1 + 2 \rightarrow f$.

³⁴Proofs of the various Breit–Wigner formulas quoted in this section are quite lengthy. See, for example, Appendix B of Martin and Shaw (2017).

formation. In the vicinity of a resonance of mass M and width Γ , the cross-section for the reaction $i \rightarrow f$ has the form

$$\sigma_{fi} \propto \frac{\Gamma_i \Gamma_f}{(E - Mc^2)^2 + \Gamma^2/4}, \quad (1.83a)$$

where $E = Wc^2$ is the total energy of the system in the centre-of-mass frame. In this case the overall constant can be calculated and if the resonance particle has spin j and the spins of the initial particles are S_1 and S_2 , then

$$\sigma_{fi} = \frac{\pi \hbar^2}{q_i^2} \frac{2j+1}{(2S_1+1)(2S_2+1)} \frac{\Gamma_i \Gamma_f}{(E - Mc^2)^2 + \Gamma^2/4}, \quad (1.83b)$$

where q_i is the magnitude of the initial particle momenta in the centre-of-mass frame.

Finally, in the common situation where $\Gamma \ll Mc^2$, the *narrow width approximation* is often used to simplify calculations. There are different forms of this, but one is to make the replacement

$$\frac{1}{(E - Mc^2)^2 + \Gamma^2/4} \rightarrow \frac{2\pi}{\Gamma} \delta(E - Mc^2) \quad (1.84a)$$

in (1.83a), where δ is the Dirac delta function,³⁵ and the constant $2\pi/\Gamma$ multiplying it is chosen to ensure that both sides of (1.84a) give the same result, to a very good approximation, when integrated over all energies. Alternatively, if we instead express the Breit–Wigner formula as a function of E^2 rather than E , (1.84a) becomes

$$\frac{4(Mc^2)}{(E^2 - (Mc^2)^2)^2 + (Mc^2)^2 \Gamma^2} \rightarrow \frac{4\pi(Mc^2)}{\Gamma} \delta(E^2 - (Mc^2)^2), \quad (1.84b)$$

where we have used the standard result

$$|a|\delta[a(x - a)] = \delta(x - a) \quad (1.85)$$

for any constant a to write $\delta(E - Mc^2)$ in terms of E^2 .

The above formulas assume one is dealing with an isolated resonance. In practice, there may be several overlapping resonances and nonresonant processes contributing to the same reaction that must be taken into account. An example of resonance formation in $\pi^- p$ interactions is given in Figure 1.11, which shows the $\pi^- p$ total cross-section in the centre-of-mass energy range (1.4–2.4 GeV). (The units used in the plots will become clear after the next section.) Two enhancements can be seen that

³⁵The delta function is defined by the conditions $\int f(x)\delta(x-a)dx \equiv f(a)$, where a is a real constant, implying $\delta(x-a) = 0$ if $x \neq a$, and $\int_b^c f(x)\delta(x-a)dx \equiv 1$, $b < a < c$. It is discussed, for example, in Section 13.3.2 of Martin and Shaw (2015), where derivations of its important properties, including (1.85), may be found.

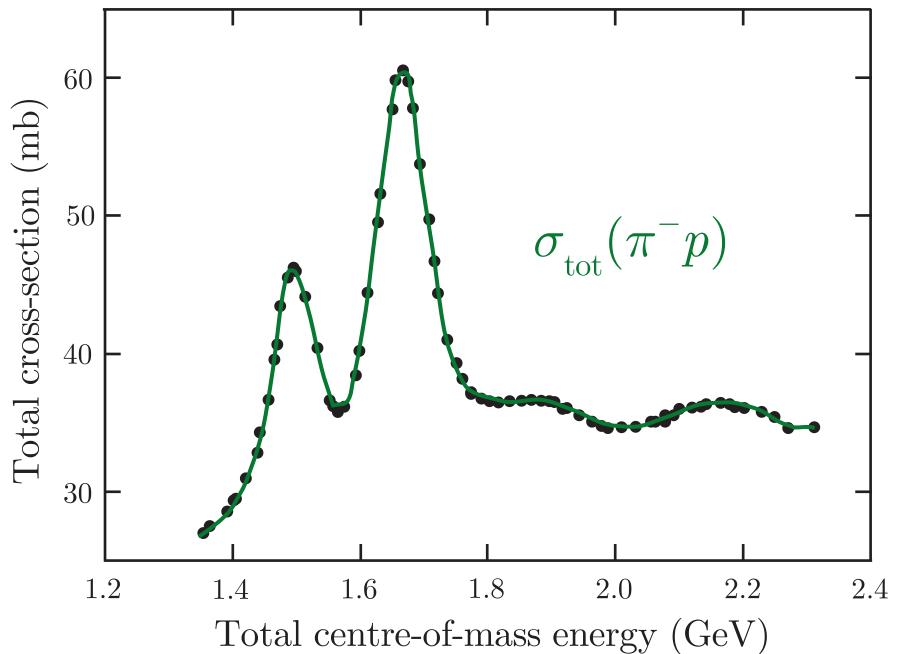


Figure 1.11 Total cross-sections for $\pi^- p$ interactions. Source: Data from Carter *et al.* (1968).

are of the approximate Breit–Wigner resonance form and there are two other maxima at higher energies. In principle, the mass and width of a resonance may be obtained by using a Breit–Wigner formula and varying M and Γ to fit the cross-section in the region of the enhancement. In practice more sophisticated methods are used that simultaneously fit a wide range of data, including differential cross-sections, and also take account of nonresonant contributions to the scattering. The widths obtained from such analyses are of the order of 100 MeV, with corresponding interaction times of order 10^{-23} s, which are typical of hadrons, which decay by strong interactions, as we shall see in Chapter 3. Resonances are also a prominent feature of interactions in nuclear physics and we will return to this in Section 2.9 when we discuss nuclear reaction mechanisms.

1.7 Units

Many branches of science introduce special units that are convenient for their own purposes. Nuclear and particle physics are no exceptions. Distances tend to be measured in femtometres or, equivalently, *fermis*, with $1 \text{ fm} \equiv 10^{-15} \text{ m}$. In these units, the radius of the proton is about 0.8 fm. The range of the strong nuclear force between protons and neutrons is of order 1–2 fm, while the range of the weak force is of order 10^{-3} fm. For comparison, the radii of atoms are of order 10^5 fm. A common unit for

area is the *barn*, defined by $1 \text{ b} = 10^{-28} \text{ m}^2$. For example, the total cross-section for pp scattering (a strong interaction) is a few tens of millibarns (mb) (compare also the $\pi^- p$ total cross-section in Figure 1.11), whereas the same quantity for νp scattering (a weak interaction) is a few tens of femtobarns (fb), depending on the energies involved. Nuclear cross-sections are very much larger and increase approximately like $A^{2/3}$, where A is the total number of nucleons in the nucleus.

Energies are invariably specified in terms of the electron volt, eV, defined as the energy required to raise the electric potential of an electron or proton by one volt. In S.I. units, $1 \text{ eV} = 1.6 \times 10^{-19} \text{ joules}$. The units $1 \text{ keV} = 10^3 \text{ eV}$, $1 \text{ MeV} = 10^6 \text{ eV}$, $1 \text{ GeV} = 10^9 \text{ eV}$ and $1 \text{ TeV} = 10^{12} \text{ eV}$ are also in general use. In terms of these units, atomic ionisation energies are typically a few eV, the energies needed to bind nucleons in heavy nuclei are typically 7–8 MeV per particle and the highest particle energies produced in the laboratory are of order of a few TeV for protons. Momenta are specified in eV/c , MeV/c , etc.

In order to create a new particle of mass M , an energy at least as great as its rest energy Mc^2 must be supplied. The rest energies of the electron and proton are 0.51 MeV and 0.94 GeV respectively, whereas the W and Z^0 bosons have rest energies of 80 GeV and 91 GeV, respectively. Correspondingly their masses are conveniently measured in MeV/c^2 or GeV/c^2 , so that, for example,

$$M_e = 0.51 \text{ MeV}/c^2, \quad M_p = 0.94 \text{ GeV}/c^2, \\ M_W = 80.4 \text{ GeV}/c^2, \quad M_Z = 91.2 \text{ GeV}/c^2.$$

In S.I. units, $1 \text{ MeV}/c^2 = 1.78 \times 10^{-30} \text{ kg}$. In nuclear physics it is also common to express masses in *atomic mass units* (u), defined as 1/12th the mass of the commonest isotope of carbon: $1 \text{ u} = 1.661 \times 10^{-27} \text{ kg} = 931.5 \text{ MeV}/c^2$.

Although experimental results are expressed in the above units, it is usual in particle physics to make theoretical calculations in units chosen such that $\hbar \equiv h/2\pi = 1$ and $c = 1$ (called *natural units*) and many books do this. The corresponding equations in practical units can then be regained, if required, by using the method of dimensions to restore the suppressed factors of \hbar and c . However, as this book is about both nuclear and particle physics, practical units will be used throughout, the sole exception being in Appendix D. A table giving numerical values of fundamental and derived constants, together with some useful conversion factors, is given inside the rear cover of the book.

Problems 1

- 1.1** (a) For what combinations of atomic number Z and mass number A is the associated neutral atom a boson, and for what combinations is it a fermion?

- (b) How would these results be modified if the nucleus were a bound state of protons and electrons, as initially proposed prior to the discovery of the neutron?

1.2 Verify that the spherical harmonic $Y_1^1 = \sqrt{\frac{3}{8\pi}} \sin \theta e^{i\phi}$ is an eigenfunction of parity with eigenvalue $P = -1$.

1.3 A proton and antiproton at rest in an S-state annihilate to produce $\pi^0\pi^0$ pairs. Show that this reaction cannot be a strong interaction.

1.4 Suppose that an intrinsic C -parity factor is introduced into the second equation of (1.15), which then becomes

$$\hat{C}|b, \psi_b\rangle = C_b|\bar{b}, \psi_{\bar{b}}\rangle.$$

Show that the eigenvalue corresponding to any eigenstate of \hat{C} is independent of C_b , so that C_b cannot be measured.

1.5 (a) In the absence of free electric charges, the scalar potential ϕ may be set to zero and the electromagnetic field described solely by the vector potential $\mathbf{A}(\mathbf{r}, t)$. The latter is related to the electric field \mathbf{E} by $\mathbf{E} = -\partial\mathbf{A}/\partial t$ and transforms under charge conjugation as $\hat{C}\mathbf{A}(\mathbf{r}, t) = C_\gamma \mathbf{A}(\mathbf{r}, t)$, where C_γ is the C -parity of the photon. If the electromagnetic interaction is invariant under charge conjugation, deduce the value of C_γ .

(b) Use Maxwell's equations to deduce how the magnetic field $\mathbf{B}(\mathbf{r}, t)$ transforms under P and T ?

1.6 Show that a particle of spin \mathbf{J} can only have a non-zero electric dipole moment if time-reversal invariance is violated.

1.7 Use the principle of detailed balance applied to the reactions $pp \rightleftharpoons \pi^+ d$ to deduce that the spin of the π^+ may be found from the expression

$$S_\pi = \frac{1}{2} \left[\frac{4R}{3} \left(\frac{p_p}{p_\pi} \right)^2 - 1 \right],$$

where $p_{p,\pi}$ are the magnitudes of the proton and pion momenta and

$$R = \frac{d\sigma(pp \rightarrow \pi^+ d)/d\Omega}{d\sigma(\pi^+ d \rightarrow pp)/d\Omega},$$

where the differential cross-sections are at the same total centre-of-mass energy and both beams and projectiles are unpolarised.

1.8 Consider the strong interaction $\pi^- d \rightarrow nn$, where d is a spin-1 S-wave bound state of a proton and a neutron called the deuteron and the initial pion is at rest. Deduce the intrinsic parity of the negative pion.

1.9 Write down equations in symbol form that describe the following interactions:

(a) elastic scattering of an electron antineutrino and a positron;

(b) inelastic production of a pair of neutral pions in proton-proton interactions;

(c) the annihilation of an antiproton with a neutron to produce three pions.

1.10 Draw two topologically distinct Feynman diagrams that can contribute to each of the following processes in lowest order: (a) $\gamma + e^- \rightarrow \gamma + e^-$; (b) $e^+ + e^- \rightarrow e^+ + e^-$; (c) $\nu_e \bar{\nu}_e$ elastic scattering.

- 1.11** Draw a fourth-order Feynman diagram for the reaction (a) $\gamma + \gamma \rightarrow e^+ + e^-$ and (b) $e^+ + e^- \rightarrow e^+ + e^-$.
- 1.12** Show that the Yukawa potential of Eq. (1.44) is the only spherically symmetric solution of the static Klein–Gordon equation (1.41) that vanishes as r goes to infinity.
- 1.13** Verify by explicit integration that

$$\mathcal{M}(q^2) = -g^2 \hbar^2 (|\mathbf{q}|^2 + m^2 c^2)^{-1}$$

is the scattering amplitude (1.47) corresponding to the Yukawa potential (1.44).

- 1.14** A high-energy electron of momentum k is scattered through 90° by a stationary proton and the recoil proton is emitted at an angle θ' to the initial electron direction. Derive an expression for θ' in terms of k and the proton mass m . You may assume the energies are sufficiently large for the electron masses to be neglected throughout.
- 1.15** A thin ('density' of 1 mg cm^{-2}) target of ^{24}Mg ($M_A = 24.3$ atomic mass units) is bombarded with a 10 nA beam of alpha particles. A detector subtending a solid angle of $2 \times 10^{-3} \text{ sr}$ records 20 protons per second. If the scattering is isotropic, determine the cross-section for the $^{24}\text{Mg}(\alpha, p)^{27}\text{Al}$ reaction.
- 1.16** The cross-section for photon scattering from free electrons when $E_\gamma \ll m_e c^2$ is given *in natural units* by

$$\sigma = \frac{8\pi\alpha^2}{3m_e^2}.$$

What is the value of σ in mb?

2

Nuclear phenomenology

In this chapter we start to examine some of the general properties of nuclei that can be learned from experiments, beginning with their masses and what can be deduced about their shapes and sizes. Then we discuss the important topic of nuclear stability and the phenomenology of the various ways that unstable nuclei decay to stable states. Finally, we briefly review the classification of reactions in nuclear physics. However, before any of this, we will recap some notations previously used in Chapter 1 and introduce some new terms, as follows.

Nuclei are specified by:

Z – *atomic number* = the number of protons,

N – *neutron number* = the number of neutrons,

A – *mass number* = the number of nucleons, so that $A = Z + N$.

We will also refer to Z as the *proton number* and A as the *nucleon number*. The charge on the nucleus is $+Ze$, where e is the absolute value of the electric charge on the electron. Nuclei with combinations of these three numbers are also called *nuclides* and are written ${}^A_Z Y$ or A_Y , where Y is the chemical symbol for the element. Some other common nomenclature is:

nuclides with the same atomic number are called *isotopes*,

nuclides with the same neutron number are called *isotones*,

nuclides with the same mass number are called *isobars*.

The concept of isotopes was introduced in Chapter 1. For example, stable isotopes of carbon are ${}^{12}\text{C}$ and ${}^{13}\text{C}$, and the unstable isotope used in dating ancient objects (see Section 9.5.2) is ${}^{14}\text{C}$. All three have $Z = 6$.

The known elements are shown in the Mendeleev periodic table of Table 2.1. The first row of each box shows the atomic number and the chemical symbol of the element and the second row gives its name.

Table 2.1 The Mendelev periodic table of the elements (see text for details)

1 H	Hydrogen	2 He	Helium
1.008		4.003	
3 Li	4 Be	5 B	6 C
Lithium	Beryllium	Boron	Carbon
6.94	9.012	10.81	12.01
11 Na	12 Mg	13 Al	14 Si
Sodium	Magnesium	Aluminum	Silicon
22.99	24.31	26.98	28.09
19 K	20 Ca	21 Sc	22 Ti
Potassium	Calcium	Scandium	Titanium
39.10	40.08	44.96	50.94
37 Rb	38 Sr	39 Y	40 Zr
Rubidium	Strontium	Zirconium	Niobium
85.47	87.62	88.91	91.22
55 Cs	56 Ba	57-71	72 Hf
Cesium	Barium	Hafnium	Tantalum
132.9	137.3	Lanthanides	Tungsten
87 Fr	88 Ra	89-103	104 Rf
Francium	Radium	Dubnium	Bohrium
(223)	(226)	(267)	(268)
Actinides			
57 La	58 Ce	59 Pr	60 Nd
Lanthanide series	Cerium	Praseodym.	Neodym.
138.9	140.1	144.2	(145)
89 A	90 Th	91 Pa	92 U
Actinium series	Thorium	Protactin.	Uranium
(227)	232.0	231.0	238.0
5 B	6 C	7 N	8 O
Boron	Carbon	Nitrogen	Oxygen
10.81	12.01	14.00	16.00
13 Al	14 Si	15 P	16 S
Aluminum	Silicon	Phosphorus	Sulphur
26.98	28.09	30.97	32.06
31 Ga	32 Ge	33 As	34 Se
Gallium	German.	Arsenic	Selenium
58.69	63.55	65.38	69.72
46 Pd	47 Ag	48 Cd	49 In
Palladium	Silver	Cadmium	Indium
106.4	107.9	102.9	101.1
78 Pt	79 Au	76 Os	75 Re
Platinum	Gold	Osmium	Rhenium
197.0	195.1	192.2	186.2
204.4	207.2	190.2	183.8
209.0	207.2	190.2	180.9
209.0	(209.0)	(281)	(278)
111 Rg	112 Ch	113 Nh	114 Fl
Darmstadt.	Rutherford.	Copernic.	Flerovium.
(285)	(286)	(285)	(289)
116 Lv	115 Mc	113 Nh	114 Fl
Moscovium	Livermor.	Nihonium	Flerovium.
(289)	(289)	(285)	(281)
117 Ts	118 Og	116 Lv	114 Fl
Tennessee	Oganesson	Livermor.	Flerovium.
(294)	(294)	(289)	(281)
10 Ne	19 F	17 Cl	18 Ar
Neon	Fluorine	Chlorine	Argon
20.18	19.00	17.45	39.95
35 Br	34 Se	33 As	32 Ge
Bromine	Selenium	Arsenic	Gallium
79.90	78.97	79.90	74.92
52 Te	53 I	51 Sb	50 Sn
Tellurium	Iodine	Antimony	Nickel
121.8	127.6	118.7	114.8
126.9	127.6	118.7	107.9
84 Bi	85 At	82 Pb	81 Tl
Bismuth	Astatine	Lead	Thallium
(210.0)	(222.0)	(209.0)	(207.2)
86 Rn	85 At	83 Bi	81 Tl
Radon	Astatine	Bismuth	Thallium
(222.0)	(210.0)	(209.0)	(207.2)
118 Og	117 Ts	116 Lv	114 Fl
Tennessee	Tennessee	Livermor.	Flerovium.
(294)	(294)	(289)	(281)

Obvious abbreviations are used for long names. The third row shows the atomic mass, including the mass of the orbital electrons, relative to 1/12 of the mass of the ^{12}C atom, the latter being defined as 1 *atomic mass unit* (u). The relative atomic masses are given for convenience to four significant figures where this is known, although many are known to greater accuracy. Where an atomic mass is in brackets, this indicates that it refers to the longest-lived isotope, as no stable isotope exists. The atomic masses shown are weighted by the isotopic abundances found at the Earth's surface. The latter often vary significantly between samples, and in the cases of some very rare unstable elements, such as polonium and francium, the abundances are uncertain because they are too low to be measured accurately.

2.1 Mass spectroscopy

The mass of a nucleus is a fundamental quantity that uniquely defines the nuclide. An accurate knowledge of its value is very important. For example, knowledge of masses can be used to test nuclear models and those of short-lived nuclei can help test the standard model of particle physics in the context of the weak interaction, as well as astrophysical models. We will return briefly to some of these applications in later chapters. A great deal of effort has been devoted to measuring masses, both of stable and unstable nuclei. This is the topic of *mass spectroscopy* that we now discuss.

2.1.1 Deflection spectrometers

A relatively simple way of measuring masses is by passing ion beams through crossed magnetic and electric fields. This technique has a long history dating from the time of the pioneering research of J.J. Thomson, who in 1912 found the isotopes of Ne with masses 20 and 22, research that was continued by his one-time student Aston.¹ The principle of the method is shown in Figure 2.1.

A source of ions produces a beam of particles containing various isotopes with charge q , which passes through a region where there are uniform electric and magnetic fields at right angles, with magnitudes E and B_1 , respectively. The electric field exerts a force qE in one direction and the magnetic field exerts a force qvB_1 in the opposite direction, where v is the speed of the ions. By balancing these forces, ions of a specific speed $v = E/B_1$ can be selected and allowed to pass through a collimating slit. Ions with other velocities (shown as dashed lines in Figure 2.1) are deflected. The beam is then allowed to continue through a second

¹Francis Aston was awarded the 1922 Nobel Prize in Chemistry for the discovery of a large number of nonradioactive isotopes using electromagnetic field spectroscopy.

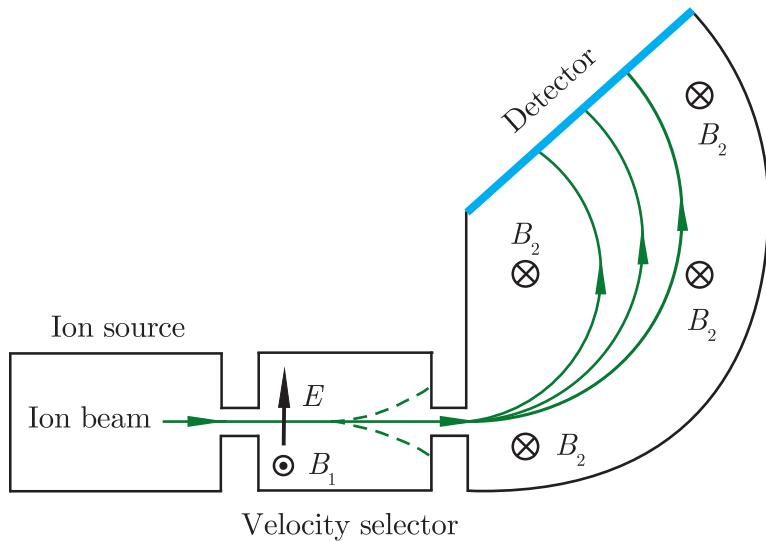


Figure 2.1 Schematic diagram of a deflection mass spectrometer. Source: Adapted from Krane (1988). Copyright (1988) John Wiley & Sons Inc., reprinted with permission.

uniform magnetic field of magnitude B_2 where it bends in a circular path of radius ρ , given by

$$mv = qB_2\rho \quad (2.1)$$

and since q , B_2 , and v are fixed, particles with a fixed ratio q/m will bend in a path with a unique radius. Hence isotopes may be separated and focused onto a detector (historically a photographic plate). In the common case where $B_1 = B_2 = B$,

$$m = qB^2\rho/E. \quad (2.2)$$

To test nuclear models, $\Delta m/m$ must be determined with a precision of at least 10^{-6} . It is clear from (2.2) that this would require an unrealistically precise knowledge of several physical parameters. However, this problem can be overcome by making a series of measurements comparing pairs of ions with similar masses and combining these to give the mass of the ion under study. This is known as the *mass doublet* method and measurements with precisions as good as 10^{-8} or 10^{-9} can be achieved using it. Consider the example² of making an accurate measurement of the mass of the nucleus ^{14}N . Setting the spectrometer for molecules with total mass number 128, the mass difference

$$\Delta_1 = m(\text{C}_9\text{H}_{20}) - m(\text{C}_{10}\text{H}_8) = 12m(^1\text{H}) - m(^{12}\text{C})$$

can be measured, where the very small difference (of order 10^{-9}u) in the molecular binding energies of the two molecules has been neglected. Then, from the above relation,

$$m(^1\text{H}) = \frac{1}{12}[m(^{12}\text{C}) + \Delta_1].$$

²See p. 61 of Krane (1988).

Next, setting the spectrometer for mass 28, the mass difference

$$\Delta_2 = m(\text{C}_2\text{H}_4) - m(\text{N}_2) = 2m(^{12}\text{C}) + 4m(^1\text{H}) - 2m(^{14}\text{N})$$

can similarly be measured, from which

$$m(^{14}\text{N}) = m(^{12}\text{C}) + 2m(^1\text{H}) - \frac{1}{2}\Delta_2 = \frac{7}{6}m(^{12}\text{C}) + \frac{1}{6}\Delta_1 - \frac{1}{2}\Delta_2.$$

Finally, using the measured values

$$\begin{aligned}\Delta_1 &= (0.09390032 \pm 0.00000012) \text{ u} \\ \Delta_2 &= (0.025152196 \pm 0.000000030) \text{ u}\end{aligned}$$

gives the accurate value

$$m(^{14}\text{N}) = (14.00307396 \pm 0.00000002) \text{ u}.$$

2.1.2 Kinematic analysis

Mass spectrometers of the deflection type cannot be used to find the masses of very short-lived nuclei, but in these cases the masses can in principle be determined from kinematic analysis of nuclear reactions as follows. Consider the inelastic reaction $A(a, b)A^*$, where A^* is the short-lived nucleus whose mass is to be determined. The particles a and b could be different, but initially we will set $b = a$. The kinematics of this reaction are then

$$a(E_i, \mathbf{p}_i) + A(m_A c^2, \mathbf{0}) \rightarrow a(E_f, \mathbf{p}_f) + A^*(E^*, \mathbf{p}^*), \quad (2.3)$$

where (E_i, \mathbf{p}_i) and (E_f, \mathbf{p}_f) are the total energies and three-momenta of the initial and final particles a , and we use a star to denote kinematical quantities related to A^* . Equating the total energy before the collision,

$$E_i^{\text{tot}} = T_i + m_a c^2 + m_A c^2, \quad (2.4)$$

to the total energy after the collision,

$$E_f^{\text{tot}} = T_f + T^* + m_a c^2 + m^* c^2, \quad (2.5)$$

where T are the kinetic energies, gives the following expression for the mass difference $(m^* - m_A)$:

$$(m^* - m_A)c^2 = T_i - T_f - T^* = \frac{p_i^2}{2m_a} - \frac{p_f^2}{2m_a} - \frac{p^{*2}}{2m^*}, \quad (2.6)$$

where we have assumed nonrelativistic kinematics. If the initial momentum of the projectile is along the z direction, the scattering angle is θ and we choose the y - z plane to be the scattering plane, then from momentum conservation,

$$p_x^* = 0, \quad p_y^* = p_f \sin \theta, \quad p_z^* = p_i - p_f \cos \theta \quad (2.7)$$

and using these in (2.6) gives

$$m^*c^2 = m_A c^2 + T_i \left(1 - \frac{m_a}{m^*}\right) - T_f \left(1 + \frac{m_a}{m^*}\right) + \frac{2m_a}{m^*} (T_i T_f)^{1/2} \cos \theta. \quad (2.8)$$

This formula can be used iteratively to deduce m^* , and hence the mass of the excited nucleus A^* , from measurements of the initial and final energies of the projectile by initially setting $m^* = m_A$ on the right-hand side, because $(m^* - m_A)$ is small in comparison with m_A . A similar formula to (2.8) may be derived for the general reaction $A(a,b)B$ and is

$$\begin{aligned} m_B c^2 &= m_A c^2 + T_i \left(1 - \frac{m_a}{m_B}\right) - T_f \left(1 + \frac{m_b}{m_B}\right) \\ &\quad + \frac{2}{m_B} (m_a m_b T_i T_f)^{1/2} \cos \theta + Q, \end{aligned} \quad (2.9)$$

where Q is the kinetic energy released in the reaction.

2.1.3 Penning trap measurements

The most precise mass measurements, for both stable and unstable nuclei, come from storage devices that confine ions in three dimensions by the use of well-controlled electromagnetic fields. Current measurements for the masses m of some stable nuclei yield values with accuracies $\delta m/m \sim 10^{-10} - 10^{-11}$, although this degree of precision is not needed for most applications. The extreme precision results from the extended time of observation of the ions, which is limited in principle only by the lifetime of radioisotopes (i.e. isotopes that decay), together with the fact that experiments with single ions are possible and that ions can be stored in ideal conditions.

There are two principal types of ion storage devices: *ion traps*, which use a combination of magnetic and electric fields to effect confinement; and *storage rings*, which use a combination of dipole magnets and quadrupole magnetic lenses to produce the same effect.³ Ion traps are small devices with dimensions of order centimetres, whereas storage rings have dimensions of several metres. Since the accuracy of mass measurements for unstable nuclei using ion traps is at present about an order-of-magnitude better than that obtained in storage rings, we will restrict our discussion to ion traps. To obtain spatial confinement requires a potential minimum in all three dimensions and in principle the best configuration is one that results in the confined particle executing harmonic motion about the centre of confinement. Here we will outline how this is achieved in just one type of trap, the *Penning trap*.

³A detailed account of both types of ion storage devices may be found in Blaum (2006).

2.1.3(a) Confinement field configuration

A particle with charge q and mass m moving with velocity \mathbf{v} in a pure homogeneous magnetic field \mathbf{B} , taken for convenience to be in the z direction, will experience a Lorentz force $\mathbf{F} = q\mathbf{v} \times \mathbf{B}$ and undergo so-called *cyclotron oscillations* at the angular frequency (cf. (2.1))

$$\omega_c = qB/m, \quad (2.10)$$

where $B = |\mathbf{B}|$. Then the mass ratio for two ion species (1) and (2) in the same magnetic field may be obtained from the ratio

$$\frac{\omega_c^{(1)}}{\omega_c^{(2)}} = \frac{q^{(1)} m^{(2)}}{q^{(2)} m^{(1)}}. \quad (2.11)$$

Thus if the ^{12}C is taken as a reference, the other mass is obtained in atomic mass units.

The circular motion of the particle around the magnetic field lines confines it in the radial plane, but does not prevent it spiralling out of the trap along the field lines. To prevent this, a weak axially symmetric restoring electrostatic potential is superimposed along the magnetic field lines to produce a saddle point at the centre, and hence the desired three-dimensional confinement. For positively charged ions, the potential has a minimum along the magnetic field axis, and the simplest form that satisfies this requirement is the quadrupole potential. In the Penning trap this is created by three electrodes of hyperboloidal shapes, two endcaps at a positive potential, and a ring electrode at a lower potential inserted between them. This is illustrated in Figure 2.2a. Alternatively, cylindrical electrodes may be used, illustrated in Figure 2.2b, which are easier to manufacture and align, and near the centre of such a trap the potential closely approximates a quadrupole potential. A related device called a *Paul trap* uses a radio frequency (rf) quadrupole electric field to achieve the same end.⁴

Referring to Figure 2.2a, the coordinates of the electrodes are given by

$$(r/r_0)^2 - (z/z_0)^2 = \pm 1, \quad (2.12)$$

where the positive sign refers to the ring electrode and the negative sign refers to the endcaps. In addition $r_0 = \sqrt{2}z_0$. This structure has rotational symmetry around the z axis and the quadrupole potential inside the electrode configuration is given by

$$\Phi(z, r) = \frac{V}{4d^2}(2z^2 - r^2), \quad (2.13)$$

⁴The Penning trap was invented by Dehmelt, who named it after the pioneering work of Penning in the late 1930s on increasing the efficiency of ionisation vacuum gauges. For the development of ion trap techniques Hans Dehmelt and Wolfgang Paul shared the Nobel Prize in Physics for 1989.

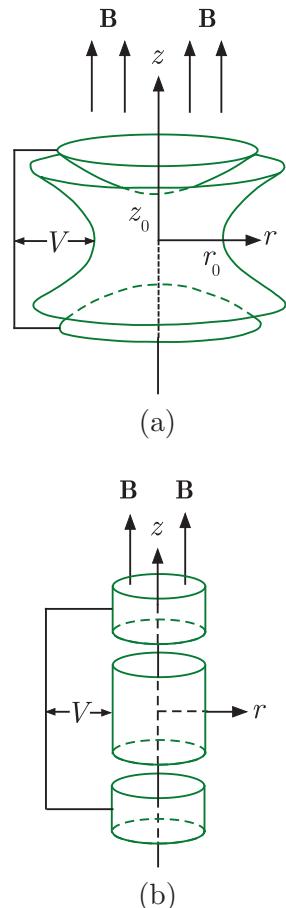


Figure 2.2 Schematic diagram of the arrangement of the electrodes in (a) hyperbolical and (b) cylindrical Penning traps. The magnetic field \mathbf{B} is typically a few tesla and V is a few volts. Source: Blaum et al. (2009). © IOP Publishing. Reproduced with permission.

where V is the voltage applied to the electrodes (with appropriate polarity) and

$$d = \frac{1}{2} (2z_0^2 + r_0^2)^{1/2} \quad (2.14)$$

is a characteristic dimension of the trap.

2.1.3(b) Ion trajectories

Solving the equations of motions for all three coordinates in the above field⁵ results in three independent motional modes: an harmonic trapping motion along the trap axis with *axial oscillation frequency* ω_z , a circular cyclotron motion with *reduced cyclotron frequency* ω_+ , and a circular motion with *magneton frequency* ω_- , where

$$\omega_z = \sqrt{(qV/m d^2)} \quad \text{and} \quad \omega_{\pm} = \frac{1}{2}\omega_c \pm \sqrt{\frac{1}{4}\omega_c^2 - \frac{1}{2}\omega_z^2}. \quad (2.15)$$

From (2.15) we see that the three frequencies are related:

$$\omega_c = \omega_+ + \omega_-, \quad 2\omega_+\omega_- = \omega_z^2, \quad \omega_- < \omega_z < \omega_+. \quad (2.16)$$

The magnetron motion is a slow drift in the $\mathbf{E} \times \mathbf{B}$ field. Collisions with background molecules will increase this drift and a high vacuum is required to prevent early loss of ions. The amplitudes and phases of the harmonic and circular modes depend on the initial conditions, i.e. on the position and velocity of the ion at the moment of creation within the trap volume or the circumstances of its injection from an external source. Numerical values for the trap parameters vary from trap to trap, and also depend on the ion used. For example, for a trap with r_0 and z_0 both of order 5 mm, V of order 10 V and B of order 7 T, the motional frequencies $f \equiv \omega/2\pi$ for an ion with $A/q = 100$ are approximately $f_+ \approx 1$ MHz, $f_z \approx 1$ kHz, and $f_- \approx 5$ kHz. These modes of motion are shown schematically in Figure 2.3. The combined motion is complicated and the particle trajectory traces out a curve called an epitrochoidin.⁶

For bounded motion, the square roots in (2.15) must be positive, leading to the trapping condition $\omega_c^2 - 2\omega_z^2 > 0$ and, hence, from (2.10) and (2.15), the condition on the magnetic field needed to balance the radial component of the applied electric field is

$$B^2 > \frac{2m|V|}{|q|d^2} \quad \text{with} \quad qV > 0. \quad (2.17)$$

For trapping conditions such that $\omega_z \ll \omega_c$, series expansions of (2.15) give the first approximations

$$\omega_- \approx \frac{V}{2d^2 B}, \quad (2.18)$$

⁵The interested reader can find the details in Brown and Gabrielse (1986).

⁶The definition of an epitrochoidin in two dimensions is the locus of a fixed point on a circle that is rotating around the perimeter of another circle.

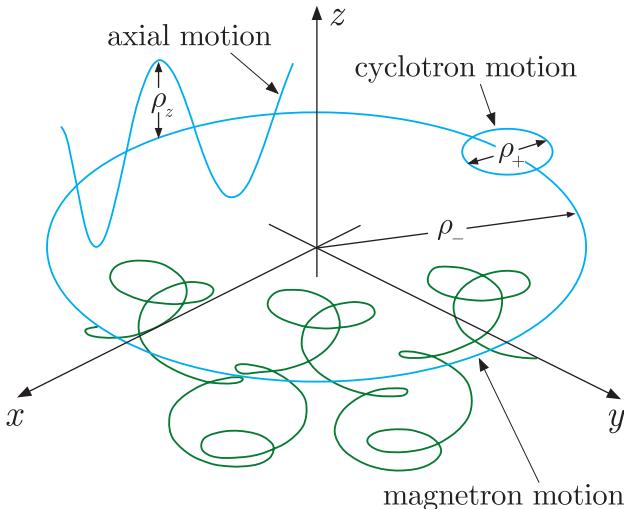


Figure 2.3 Motion of a single charged particle in a Penning trap. The quantities ρ_+ , ρ_- , and ρ_z are the radii of the cyclotron, axial, and magnetron motions, respectively. The green line is the resulting motion from the addition of the three motional modes. The frequencies and amplitudes are not to scale. Source: Blaum et al. (2009). © IOP Publishing. Reproduced with permission.

which is independent of the mass, and

$$\omega_+ = \omega_c - \frac{V}{2d^2 B}. \quad (2.19)$$

Therefore, to determine the mass one can either directly measure the cyclotron frequency or the individual radial frequencies. All of the above is for an ideal Penning trap. In a real trap there will be imperfections of various types (field inhomogeneities, misalignment of the trap and magnetic field axes, etc.) that will lead to systematic uncertainties in the final mass determination unless they are corrected for.⁷

2.1.3(c) Production and trapping of ions

Before any measurements can be made, the ions have to be produced and contained in the trap, and then manipulated in such a way that a frequency measurement can be made. We will discuss these topics only briefly.⁸ The ions are most easily confined if they are produced within the trap and this is the method generally used for stable ions. One simple method is to pass an atomic beam through the trap where it collides with electrons from a filament placed near one of the endcap electrodes. The trap can be filled very rapidly using this method, in a fraction of a second. Short-lived ions, however, are usually delivered from external sources outside the trap but located close to it, and with beam energies that range from several tens of keV to several GeV.⁹ The more exotic species, of considerable interest to astrophysicists, are often available only at very low rates of 100 ions per second or less. Thus highly efficient methods of slowing down and bunching the beams are required to move the incoming ions from their

⁷ Real Penning traps are discussed in Ghosh (1995), Major, Gheorghe, and Werth (2005), and Blaum (2006).

⁸ More details may be found in Footnote 7 and Blaum et al. (2009).

⁹ The production of beams of unstable nuclei is briefly discussed in Section 4.2.3.

initial trajectories to an orbit bound in the trap. One way this can be done is to close the trap around the ion as it passes through. In the Penning trap, the ions pass the first endcap electrode, which is held at ground potential. Ions of sufficient low energy are reflected at the high potential of the second endcap and bounce back towards the entry endcap. Before they have time to exit the trap, the first endcap is returned to its normal high potential, closing the trap and capturing the ions. Other methods are also used.

2.1.3(d) Frequency measurements

There are two techniques used to make the frequency measurement on the trapped ions: destructive and nondestructive. In the former, the ions are lost after the measurement and the trap must be reloaded; in the latter, the ions remain in the trap and may be reused for additional experiments. Nondestructive methods are preferable for stable ions or long-lived radionuclides with low production rates, whereas the destructive method is acceptable for the very short-lived radionuclides, because they will rapidly decay anyway. The most commonly used method, which is routinely used for measurements on short-lived radionuclides, is to manipulate the ion motion by rf fields and measure the time-of-flight (TOF) of the ions after ejection from the trap to an ion detector placed outside the magnetic field. Other methods include the observation of the oscillating image currents induced in the trapping electrodes by motion of the ions. We will just describe the former.

In TOF detection, the ions are passed into the trap as a pulsed low-energy beam and are captured in the centre of the trap. The trapped ions are excited to a finite magnetron radius by a dipole excitation, and this magnetron motion is then converted to modified cyclotron motion by applying an azimuthal rf quadrupole excitation close to the cyclotron frequency for time T_{obs} . The ions are then ejected from the trap by lowering the trapping potential of the downstream endcap electrode and they drift along the field lines to an ion detector. As the ions leave the trap they pass through the magnetic field gradient and are accelerated towards the detector (i.e. their cyclotron motion is converted into longitudinal motion). Resonantly excited ions arrive earlier at the detector than those ions that have been excited nonresonantly, so the experiment is repeated for a range of excitation frequencies spanning the cyclotron frequency. An example of a resulting plot of time-of-flight versus excitation frequency is given in Figure 2.4 and shows a clear resonance.

As can be seen, the line shape agrees well with detailed theoretical calculations, shown as the solid line, and the half-width of the resonance can be shown to be

$$\Delta f_c \equiv \Delta\omega_c/2\pi \approx 1/T_{\text{obs}}. \quad (2.20)$$

Thus the resolving power is

$$R = m/\Delta m = f_c/\Delta f_c \approx f_c T_{\text{obs}} \approx 4 \times 10^6. \quad (2.21)$$

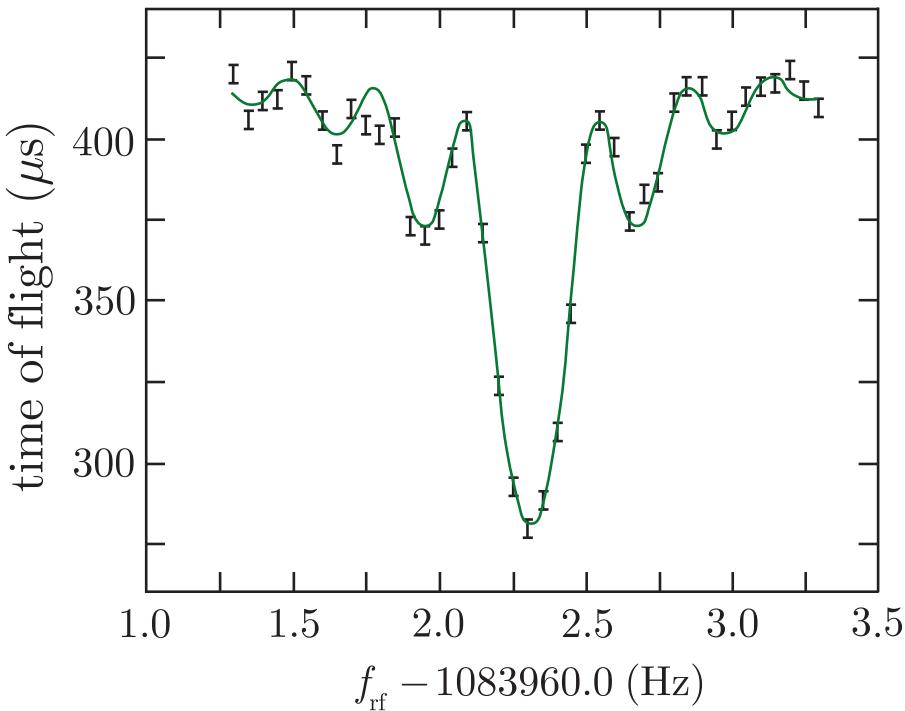


Figure 2.4 Cyclotron resonance for ^{85}Rb obtained for an excitation time $T_{\text{obs}} = 3.6\text{ s}$, obtained using the multi-trap ISOLTRAP system at ISOLDE (CERN). Source: König et al. (1995). Copyright (1995), Elsevier, reprinted with permission.

This is an important parameter because the statistical uncertainty in the mass measurement is given by

$$\left(\frac{\delta m}{m}\right)_{\text{stat}} \approx \frac{1}{R\sqrt{N_{\text{tot}}}} = \frac{2\pi m}{T_{\text{obs}} qB\sqrt{N_{\text{tot}}}}, \quad (2.22)$$

where N_{tot} is the total number of ions recorded in a single resonance. In practice at least about 100 ions have to be recorded within one resonance to get a proper fit. (In the example shown in Figure 2.4, this would give $\delta m/m \sim 3 \times 10^{-8}$.) From these results, and the fact that radionuclides will decay exponentially (see Section 2.5), one can compute that the statistical precision is highest for an observation time that is about 2.9 times the half-life of the short-lived nuclide (see Problem 2.1). It follows from (2.22) that the resolving power, and hence the precision, is increased for highly charged ions. Alternatively, for a given precision, much shorter observation times can be used compared to those for singly-charged ions, thus opening up the possibility of measurements on very short-lived nuclides. An extreme example of a measurement of a very short-lived nuclide is that for ^{11}Li , which has a half-life of $t_{1/2} = 8.8\text{ ms}$ (Smith et al. 2008).

2.2 Nuclear shapes and sizes

The shape and size of a nucleus may be found from scattering experiments; i.e. a projectile is scattered from the nucleus and the angular

distribution of the scattered particles examined, as was done by Rutherford and his collaborators when they deduced the existence of the nucleus. The interpretation is simplest in those cases where the projectile itself has no internal structure, i.e. is an elementary particle, and electrons are often used. In this case the relevant force is electromagnetic and we learn about the *charge distribution* in the nucleus. The first experiments of this type were performed by Hofstadter and his collaborators in the 1950s.¹⁰ If instead of an electron a hadron is used as the projectile, the nuclear strong interaction has also to be taken into account and we can find information about the *matter density*. Neutrons are commonly used so that Coulomb effects are absent. We discuss these two cases in turn.

2.2.1 Charge distribution

We start by considering the determination of the electric charge distribution within the nucleus from the measured cross-sections for elastic electron scattering. At low electron energies, nuclear recoil can be neglected so that the electron energy E remains unchanged after scattering and the nucleus can be treated as a static charge distribution $f(\mathbf{r})$, subject to the normalisation condition

$$\int f(\mathbf{r}) d^3\mathbf{r} = Ze. \quad (2.23)$$

At high energies, nuclear recoil and contributions from magnetic moments both become important, but in this section we will confine ourselves to energies where these are small corrections that can, to a good approximation, be neglected.

The next step is to relate the charge distribution $f(\mathbf{r})$ to the measured cross-sections. In Appendix C we derive a simple formula (C.23) for Coulomb scattering of a particle with charge ze from a static point charge $+Ze$ in the Born approximation, which assumes $Z\alpha \ll 1$. For incident electrons, $z = 1$ and the resulting cross-section is

$$\left(\frac{d\sigma}{d\Omega} \right)_0 = \frac{Z^2 \alpha^2 (\hbar c)^2}{4\beta^4 E^2 \sin^4(\theta/2)}. \quad (2.24)$$

Here E is the total energy of the electron, θ is the electron scattering angle, and $\beta = v/c$, where v is the speed of the initial electron. Note that (2.24) is of order α^2 because it corresponds to the exchange of a single photon. Although this formula has a limited range of applicability, it is useful to discuss the general features of electron scattering. However, in deriving it we have ignored the electron spin, and taking this into account leads to the so-called *Mott cross-section*¹¹

$$\left(\frac{d\sigma}{d\Omega} \right)_{\text{Mott}} = \left(\frac{d\sigma}{d\Omega} \right)_0 [1 - \beta^2 \sin^2(\theta/2)]. \quad (2.25)$$

¹⁰Robert Hofstadter shared the 1961 Nobel Prize in Physics for his pioneering electron scattering experiments.

¹¹A derivation is given in Section 8.7 of Mandl and Shaw (2010).

In the nonrelativistic limit, both (2.24) and (2.25) reduce to the spin-independent *Rutherford cross-section*

$$\left(\frac{d\sigma}{d\Omega} \right)_{\text{Rutherford}} = \frac{(\hbar c)^2 (\alpha Z)^2}{4m^2 v^4 \sin^4(\theta/2)}, \quad (2.26)$$

as used by Rutherford, with appropriate modifications of the projectile charge and mass, to interpret the alpha particle scattering experiments that led to his historic discovery of the nucleus described in Section 1.1.1.

It remains to take account of the spatial extension of the nucleus, described by the charge distribution function $f(\mathbf{r})$, subject to the normalisation condition (2.23). To do this, for a given momentum transfer $\mathbf{q} = \mathbf{p} - \mathbf{p}'$, where \mathbf{p} and \mathbf{p}' are the initial and final electron momenta respectively, we define the *form factor* $F(\mathbf{q}^2)$ as the Fourier transform of the charge distribution:

$$F(\mathbf{q}^2) \equiv \frac{1}{Ze} \int e^{i\mathbf{q}\cdot\mathbf{r}/\hbar} f(\mathbf{r}) d^3\mathbf{r}. \quad (2.27)$$

In the case of a spherically symmetric charge distribution, the angular integrations in (2.27) may be done using spherical polar coordinates to give

$$F(\mathbf{q}^2) = \frac{4\pi\hbar}{Zeq} \int_0^\infty r \rho(r) \sin\left(\frac{qr}{\hbar}\right) dr, \quad (2.28)$$

where $\rho(r)$ is the radial charge distribution, $q = |\mathbf{q}|$, and, since the magnitude of the electron momentum p is unchanged on scattering,

$$\mathbf{q}^2 = 4p^2 \sin^2(\theta/2). \quad (2.29)$$

The final form of the experimental cross-section in this approximation is derived in Appendix C and is given by

$$\left(\frac{d\sigma}{d\Omega} \right)_{\text{expt}} = \left(\frac{d\sigma}{d\Omega} \right)_{\text{Mott}} |F(\mathbf{q}^2)|^2. \quad (2.30)$$

Two examples of measured cross-sections are shown in Figure 2.5. Striking features are the presence of a number of well-defined minima superimposed on a rapid decrease in the cross-section with angle. These features are common to all elastic data, although not all nuclei show so many minima as those shown and their depth and sharpness depends on the nuclear size, as we show below. The minima are due to the form factor and we can make this plausible by taking the simple case where the nuclear charge distribution is represented by a hard sphere such that

$$\begin{aligned} \rho(r) &= \text{constant}, & r \leq a, \\ &= 0, & r > a, \end{aligned} \quad (2.31)$$

where a is a constant. In this case, evaluation of (2.28) gives

$$F(\mathbf{q}^2) = 3[\sin(b) - b \cos(b)]b^{-3}, \quad (2.32)$$

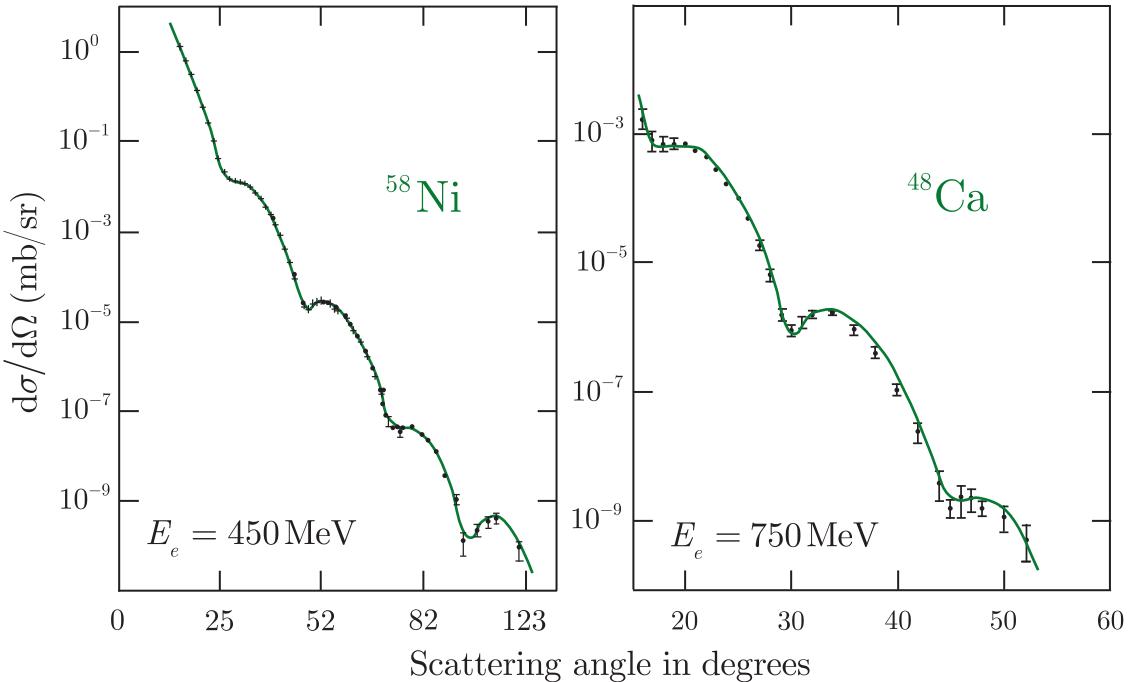


Figure 2.5 Elastic differential cross-sections as a function of the scattering angle for 450 MeV electrons from ^{58}Ni and 750 MeV electrons from ^{48}Ca . The solid lines are fits as described in the text. Source: ^{58}Ni data from Sick et al. (1975) and ^{48}Ca data from Bellicard et al. (1967). Copyright American Physical Society, reprinted with permission.

where $b = qa/\hbar$. Thus $F(\mathbf{q}^2)$ will be zero at values of b for which $b = \tan(b)$. In practice, as we will see below, $\rho(r)$ is not a hard sphere, and although it is approximately constant for much of the nuclear volume, it falls smoothly to zero at the surface. Smoothing the edges of the radial charge distribution (2.31) turns the zeros into minima and modifies their position, but does not alter the argument that the minima in the cross-sections are due to the spatial distribution of the nucleus. The actual positions and depths of the minima result from a combination of the form factor and the point-like amplitude. We shall see below that the minima can give information about the size of the nucleus.

If one measures the cross-section for a fixed energy at various angles (and hence, from (2.29), at various q^2), the form factor can in principle be extracted using (2.30) and one might attempt to find the charge distribution from the inverse Fourier transform

$$f(\mathbf{r}) = \frac{Ze}{(2\pi)^3} \int F(\mathbf{q}^2) e^{-i\mathbf{q}\cdot\mathbf{r}/\hbar} d^3\mathbf{q}. \quad (2.33)$$

However, \mathbf{q}^2 only has a finite range for a fixed initial electron energy, and even within this range the rapid fall in the cross-section means that in practice measurements cannot be made over a sufficiently wide range of angles for the integral in (2.33) to be evaluated accurately. Thus, even

within the approximations used, reliable charge distributions cannot be found from (2.33). Therefore, different strategies must be used to deduce the charge distribution.

In one approach, plausible but very general parameterised forms (for example a sum of Gaussians) are chosen for the charge distribution and are used to modify the point-like electromagnetic interaction. The resulting Schrödinger (or Dirac) equation is solved numerically to produce an amplitude, and hence a cross-section, for electron–nucleus scattering. The parameters of the charge distribution are then varied to give a good fit to the experimental data. The solid curves in Figure 2.5 were obtained in this way.

Some radial charge distributions for various nuclei obtained by these methods are shown in Figure 2.6. They are well represented by the form

$$f(r) = \rho_{\text{ch}}(r) = \frac{\rho_{\text{ch}}^0}{1 + e^{(r-a)/b}}, \quad (2.34)$$

where a and b for medium and heavy nuclei are found to be

$$a \simeq 1.07A^{1/3} \text{ fm}; \quad b \simeq 0.54 \text{ fm}. \quad (2.35)$$

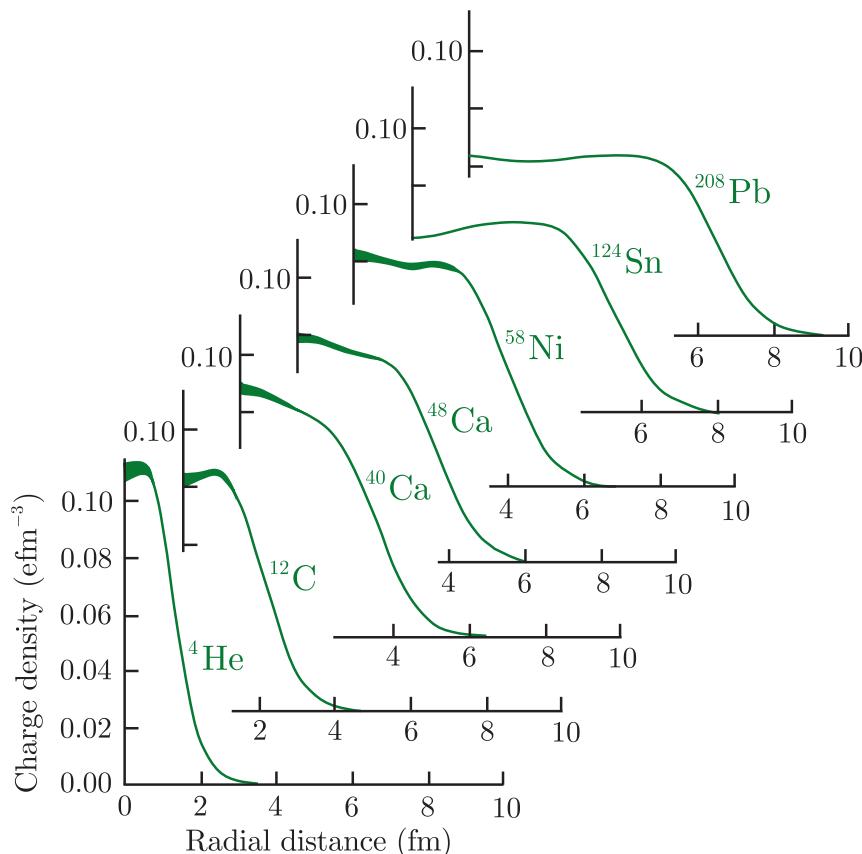


Figure 2.6 Radial charge distributions ρ_{ch} of various nuclei, in units of efm^{-3} . The thickness of the curves near $r = 0$ is a measure of the uncertainty in ρ_{ch} . Source: Adapted from data in Frois (1983).

From this we can deduce that the charge density is approximately constant in the nuclear interior and falls fairly rapidly to zero at the nuclear surface, as anticipated above. The value of ρ_{ch}^0 is in the range $(0.06\text{--}0.08)\text{e fm}^{-3}$ for medium to heavy nuclei and decreases slowly with increasing mass number.

A useful quantity is the *mean square charge radius*,

$$\langle r^2 \rangle \equiv \frac{1}{Ze} \int r^2 f(\mathbf{r}) d^3\mathbf{r}. \quad (2.36)$$

This can be found from the form factor as follows. Expanding the expression (2.27) for $F(\mathbf{q}^2)$ gives

$$F(\mathbf{q}^2) = \frac{1}{Ze} \int f(\mathbf{r}) \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{i|\mathbf{q}|r \cos \theta}{\hbar} \right)^n d^3\mathbf{r} \quad (2.37)$$

and after doing the angular integrations this becomes

$$F(\mathbf{q}^2) = \frac{4\pi}{Ze} \int_0^{\infty} f(r) r^2 dr - \frac{4\pi\mathbf{q}^2}{6Ze\hbar^2} \int_0^{\infty} f(r) r^4 dr + \dots \quad (2.38)$$

From the normalisation of $f(\mathbf{r})$, we finally have

$$F(\mathbf{q}^2) = 1 - \frac{\mathbf{q}^2}{6\hbar^2} \langle r^2 \rangle + \dots \quad (2.39)$$

and thus the mean square charge radius can be found from

$$\langle r^2 \rangle = -6\hbar^2 \frac{dF(\mathbf{q}^2)}{d\mathbf{q}^2} \Big|_{\mathbf{q}^2=0}, \quad (2.40)$$

provided the form factor can be measured at very small values of \mathbf{q}^2 . For medium and heavy nuclei, $\langle r^2 \rangle^{1/2}$ is given approximately by¹²

$$\langle r^2 \rangle^{1/2} = 0.94 A^{1/3} \text{ fm}. \quad (2.41)$$

The nucleus is often approximated by a homogeneous charged sphere. The radius R of this sphere is then quoted as the nuclear radius. The relation of this to the mean square charge radius is $R^2 = \frac{5}{3} \langle r^2 \rangle$, so that

$$R = 1.21 A^{1/3} \text{ fm}. \quad (2.42)$$

2.2.2 Matter distribution

Electrons cannot be used directly to obtain the distribution of neutrons in the nucleus. We could, however, take the presence of neutrons into account

¹²The constant comes from a fit to a range of data, e.g. the compilation for $55 \leq A \leq 209$ given in Barrett and Jackson (1977).

by multiplying $\rho_{\text{ch}}(r)$ by A/Z . Then one finds an almost identical nuclear density in the nuclear interior for all nuclei because the decrease in ρ_{ch}^0 with increasing A is compensated by the increase in A/Z with increasing A . The interior nuclear density is given by

$$\rho_{\text{nuclear}} \simeq 0.17 \text{ nucleons/fm}^3. \quad (2.43)$$

Likewise, the effective nuclear matter radius for medium and heavy nuclei is

$$R_{\text{nuclear}} \simeq 1.2 A^{1/3} \text{ fm}. \quad (2.44)$$

These are important results that will be used extensively later in this chapter and elsewhere in this book.

Although the relations (2.42) and (2.44) are valid for the vast majority of nuclei, there are some very interesting exceptions. These are the so-called *halo nuclei*, which have spatial extents far larger than expectations. (The nucleus of ^{11}Li mentioned at the end of Section 2.1.3 is such a nucleus.) We will return to this briefly in Section 10.2.2.

To probe the matter density of nuclei experimentally, a hadron, typically a neutron or proton, has to be used as the projectile. At high energies, where elastic scattering is only a small part of the total interaction, the nucleus behaves more like an absorbing sphere. In this case, the incident particle of momentum p will have an associated quantum mechanical wave of wavelength $\lambda = h/p$ and will suffer diffraction-like effects, as in optics. To the extent that we are dealing at high energies purely with the nuclear strong interaction (i.e. neglecting the relatively weak Coulomb interaction in the case of a proton for example), the nucleus can be represented by a black disk of radius R and the differential cross-section will have a Fraunhofer-like diffraction form, i.e.

$$\frac{d\sigma}{d\Omega} \propto \left[\frac{J_1(qR)}{qR} \right]^2, \quad (2.45)$$

where $qR \simeq pR\theta$ for small θ and J_1 is a first-order Bessel function. For large qR ,

$$[J_1(qR)]^2 \simeq \left(\frac{2}{\pi qR} \right) \sin^2 \left(qR - \frac{\pi}{4} \right), \quad (2.46)$$

which has zeros at intervals $\Delta\theta = \pi/pR$. The plausibility of this interpretation is borne out by experiment, an example of which is shown in Figure 2.7. The data show a succession of roughly equally spaced minima as suggested by (2.46).

To go further requires solving the equations of motion, but this is far more problematical than in the electron case because hadrons are more likely to be absorbed as they pass through the nucleus and the effective potential is far less well known. However, the analogy with optics can be pursued further in the so-called *optical model*. The essential idea in this model is that a hadron incident on a nucleus may be elastically scattered or it may cause a variety of different reactions. As in the discussion

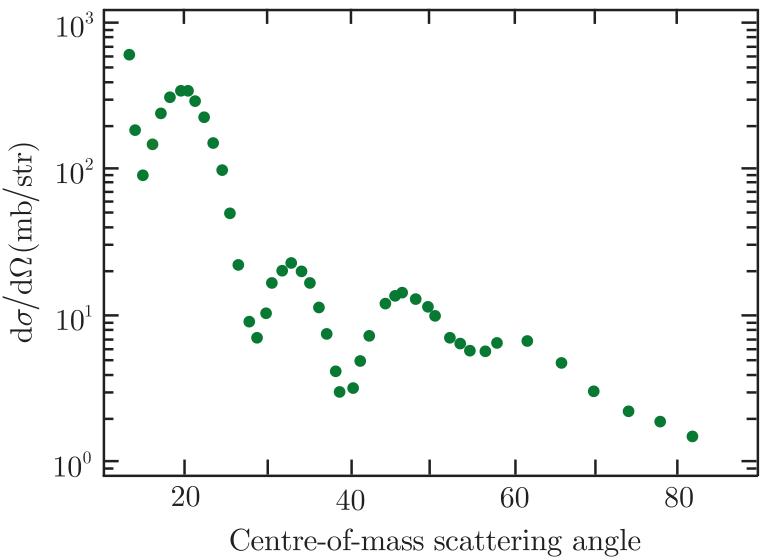


Figure 2.7 Elastic differential cross-sections for 52 MeV deuterons on ^{54}Fe . Source: Hintenberger et al. (1968). Copyright (1968) Elsevier, reprinted with permission.

above, if the incident particle is represented by a wave, then in classical language it may be scattered or it may be absorbed. In optics this is analogous to the refraction and absorption of a light wave by a medium of complex refractive index, and just as the imaginary part of the refractive index takes account of the absorption of the light wave, so in the nuclear case the imaginary part of a complex potential describing the interaction takes account of all the inelastic reactions. It is an essential feature of the model that the properties of nuclei are mainly determined by their size, as this implies that the same form of potential can account for the interaction of particles of different energies with different nuclei. Apart from the theoretical basis provided by analogy with classical optics, the model is essentially phenomenological, in that the values of the parameters of the optical potentials are found by optimising the fit to the experimental data. This type of semi-phenomenological approach is common in both nuclear and particle physics.

In practice, the Schrödinger equation is solved using a parameterised complex potential, where the real part is a sum of the Coulomb potential (for charged projectiles), an attractive nuclear potential and a spin-orbit potential. The imaginary part is assumed to cause the incoming wave of the projectile to be attenuated within the nucleus, thereby allowing for inelastic effects. Originally, mathematical forms like (2.34) were used to parameterise the real and imaginary parts of the potential, but because substantial differences were found between the real and imaginary parts of the potential, different forms are now used for the real and imaginary parts. The free parameters of the total potential are adjusted to fit the data.

The optical model has achieved its greatest success in the scattering of nucleons and a wide range of scattering data can be accounted for to a high degree of precision by the model. Examples of this are shown

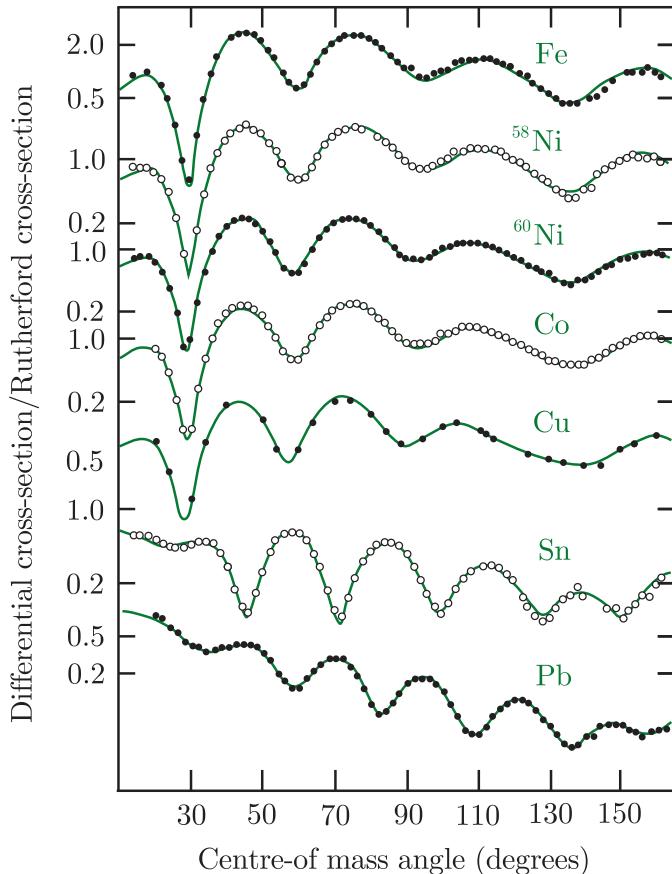


Figure 2.8 Differential cross-sections (normalised to the Rutherford cross-section) for the elastic scattering of 30.3 MeV protons, for a range of nuclei compared with optical model calculations. Source: Satchler (1967). Copyright (1967) Elsevier, reprinted with permission.

in Figure 2.8. The corresponding wavefunctions are extensively used to extract information on nuclear structure. The conclusions are in accord with those above deduced indirectly from electron data.

2.3 Semi-empirical mass formula: the liquid drop model

In this section, we will discuss the experimental data on the forces that bind nucleons in the nucleus and the interpretation of these data by a simple phenomenological formula incorporating the conclusions drawn from the experimental results of earlier sections.

2.3.1 Binding energies

Just as in the case of electrons in atoms, the forces that bind nucleons in nuclei contribute to the total mass of an atom $M(Z, A)$ and in terms of the masses of the proton M_p , neutron M_n , and electron m_e ,

$$M(Z, A) < Z(M_p + m_e) + N M_n. \quad (2.47)$$

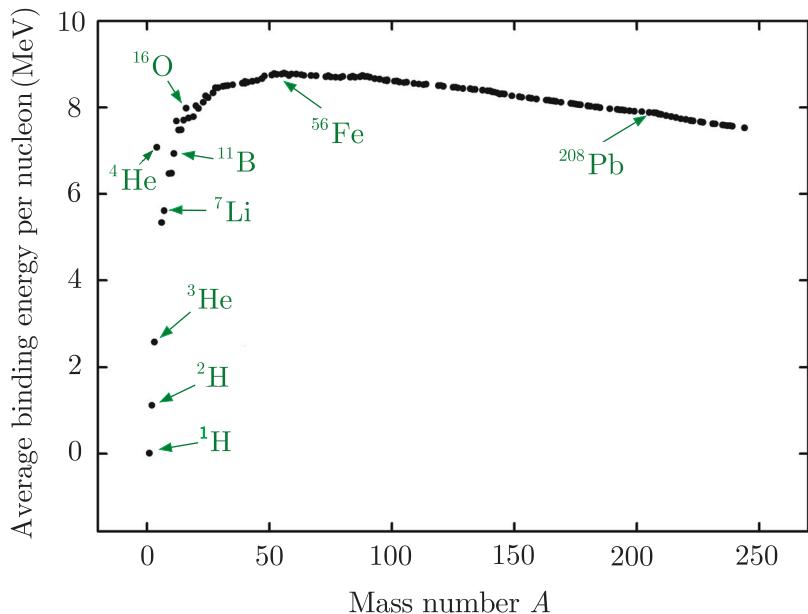


Figure 2.9 Binding energy per nucleon as a function of mass number A for stable and long-lived nuclei.

The *mass deficit* is defined as

$$\Delta M(Z, A) \equiv M(Z, A) - Z(M_p + m_e) - N M_n \quad (2.48)$$

and $-\Delta M c^2$ is called the *binding energy* B .

A commonly used quantity of interest is the *binding energy per nucleon* B/A . This is shown schematically in Figure 2.9 for nuclei that are stable or long lived. This shows that B/A peaks at a value of 8.7 MeV for a mass number of about 56 (close to iron) and thereafter falls very slowly. Excluding very light nuclei, the binding energy per nucleon is between 7 and 9 MeV over a wide range of the periodic table. In the next section we discuss a model that provides an explanation for the shape of this curve.

2.3.2 Semi-empirical mass formula

Apart from the lightest elements and a few special very stable nuclei, the binding energy data of Figure 2.9 can be approximated by a simple formula containing just a few free parameters. This is the *semi-empirical mass formula* (SEMF), first written down in 1935 by Weizsäcker. It is a *semi-empirical* formula because, although it contains a number of constants that have to be found by fitting experimental data, the formula does have a theoretical basis. This arises from the two properties common to most nuclei that we have seen earlier (the exceptions are those with very small A values): (i) the interior mass densities are approximately equal and (ii) their total binding energies are approximately proportional to their masses. There is an analogy here with a classical model of a liquid drop, where for drops of various sizes: (i) interior densities are the same and (ii) heats of vaporisation are proportional to their masses if surface

effects are neglected.¹³ However, the analogy of a nucleus as an incompressible liquid droplet, with the nucleons playing the role of individual molecules within the droplet, cannot be taken too far, because nucleons obey the laws of quantum, not classical, physics.

The semi-empirical mass formula will be taken to apply to *atomic* masses. The atomic mass $M(Z, A)$ may then be written as the sum of six terms $f_i(Z, A)$:

$$M(Z, A) = \sum_{i=0}^5 f_i(Z, A). \quad (2.49)$$

The first of these is the *mass of the constituent nucleons and electrons*,

$$f_0(Z, A) = Z(M_p + m_e) + (A - Z)M_n. \quad (2.50)$$

The remaining terms are various corrections, which we will write in the form of constants a_i multiplied by functions of Z and A with $a_i > 0$.

The most important correction is the *volume* term,

$$f_1(Z, A) = -a_1 A. \quad (2.51)$$

This arises from the fact that the strong nuclear force is short-range and each nucleon therefore feels the effect of only the nucleons immediately surrounding it (the force is said to be ‘saturated’) independent of the size of the nucleus. Recalling the important result deduced in Section 2.2 that the nuclear radius is proportional to $A^{1/3}$, this leads immediately to the binding energy being proportional to the volume, or nuclear mass. The coefficient is negative; i.e. it increases the binding energy, as expected.

The volume term overestimates the effect of the nuclear force because nucleons at the surface are not surrounded by other nucleons. Thus the volume term has to be corrected. This is done by the *surface* term

$$f_2(Z, A) = a_2 A^{2/3}, \quad (2.52)$$

which is proportional to the surface area and decreases the binding energy. In the classical model of a real liquid drop, this term would correspond to the surface tension energy.

The *Coulomb* term accounts for the Coulomb energy of the charged nucleus, i.e. the fact that the protons repel each other. If we had a uniform spherical charge distribution with a radius proportional to $A^{1/3}$, then this term would be proportional to Z^2 with the constant a_3 exactly known (see Problem 2.9). To a very rough approximation, the nucleus can be considered a sphere of uniform charge, but because electrostatic repulsion will only exist for more than one proton, Z^2 must be replaced by $Z(Z-1)$, which represents the Coulomb repulsion of the $Z(Z-1)/2$ pairs of protons in the nucleus. Then

$$f_3(Z, A) = a_3 \frac{Z(Z-1)}{A^{1/3}} \simeq a_3 \frac{Z^2}{A^{1/3}}, \quad (2.53)$$

¹³The heat of vaporisation is the heat absorbed by a given mass of a given material at its boiling point that completely converts the material to a gas at the same temperature.

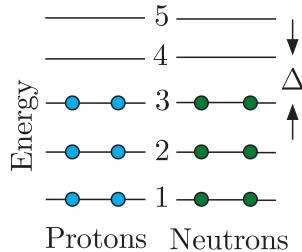


Figure 2.10 Schematic diagram of nuclear energy levels near the highest filled levels.

where the approximation is sufficiently accurate for the values of Z we will be considering.

The next term is the *asymmetry* term

$$f_4(Z, A) = a_4 \frac{(Z - A/2)^2}{A}, \quad (2.54)$$

which accounts for the observed tendency for nuclei to have $Z = N$. (No stable nucleus exists with very large neutron or proton excesses – cf. Figure 2.13 in Section 2.4.) This term is purely quantum mechanical in origin and is due to the Pauli principle. Part of the reason for the form (2.54) can be seen from the diagram of Figure 2.10, which shows the energy levels of a nucleus near the highest filled level in the approximation where all the energy levels are separated by the same energy Δ .¹⁴ Keeping A fixed and removing a proton from level 3 and adding a neutron to level 4 gives $(N - Z) = 2$ and leads to an energy increase of Δ . Repeating this for more protons, we find that the transfer of $(N - Z)/2$ nucleons decreases the binding energy by an amount $\Delta(N - Z)^2/4$. Although we have assumed Δ is a constant, in practice the spacing of levels close to the highest filled level is proportional to A^{-1} ; hence the final form of the asymmetry term (2.54).

The final contribution is the empirical *pairing term* with the form

$$\begin{aligned} f_5(Z, A) &= -f(A), & \text{if } Z \text{ even, } A - Z = N \text{ even,} \\ f_5(Z, A) &= 0, & \text{if } Z \text{ even, } A - Z = N \text{ odd; or, } Z \text{ odd, } A - Z = N \text{ even,} \\ f_5(Z, A) &= +f(A), & \text{if } Z \text{ odd, } A - Z = N \text{ odd.} \end{aligned} \quad (2.55)$$

This arises from the tendency of like nucleons in the same spatial state to couple pairwise to give configurations with spin zero. When coupled like this, the wavefunctions of the two nucleons heavily overlap and so on average they are closer together than when coupled in other configurations, and hence are more tightly bound. When there is an odd number of nucleons, this term does not contribute. Thus, when both Z and N are odd, the binding energy may be increased by converting one of the odd protons to a neutron (or vice versa) so that it can form a pair with its formerly odd partner. The evidence for this is that there are only four stable nuclei with odd N and odd Z , whereas there are 167 with even N and even Z . The form used for f_5 is empirical, but $f(A) = a_5 A^{-1/2}$ represents the trend of the data for the pairing energies and is often used.

To help remember these terms, the notation VSCAP is frequently used, with

$$a_v = a_1, \quad a_s = a_2, \quad a_c = a_3, \quad a_a = a_4, \quad a_p = a_5. \quad (2.56)$$

where the subscripts v , s , c , a , and p stand for *volume*, *surface*, *Coulomb*, *asymmetry*, and *pairing*, respectively. Precise values of the coefficients

¹⁴This is essentially the Fermi gas model, to be discussed in Section 8.2.

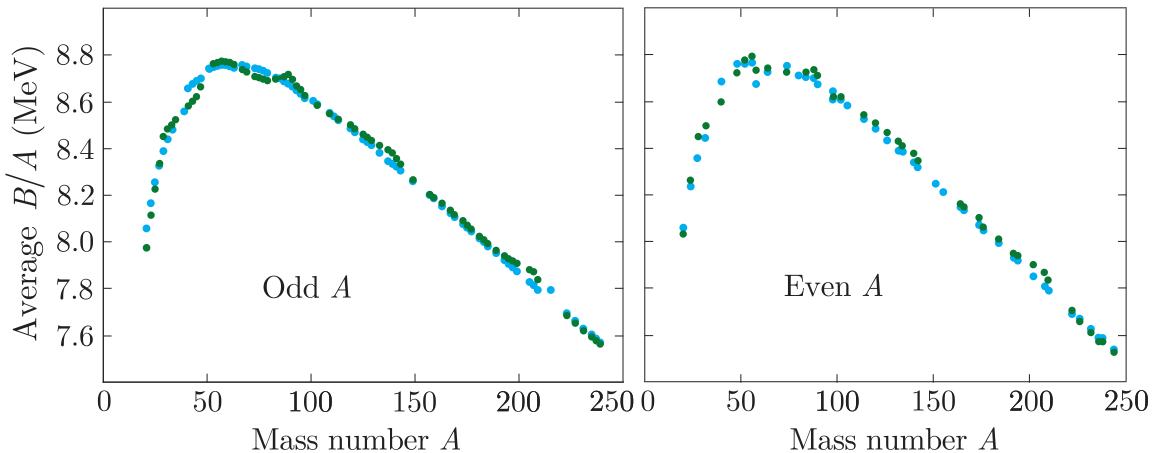


Figure 2.11 Fit to the binding energy data (shown as green circles) for odd- A and even- A nuclei using the SEMF with the coefficients given in the text. The predictions (shown as blue circles) do not lie on smooth curves because Z is not always a smooth function of A .

depend a little on the range of A fitted. One commonly used set is, in units of MeV/c^2 ,¹⁵

$$a_v = 15.56, a_s = 17.23, a_c = 0.697, a_a = 93.14, a_p = 12. \quad (2.57)$$

The fit to the binding energy data for $A > 20$ using these coefficients in the SEMF is shown in Figure 2.11. Overall the fit to the data is remarkably good for such a simple formula, but is not exact of course. For example, there are a small number of regions where the binding energy curves show enhancements that are not reproduced. These enhancements are due to the existence of a ‘shell structure’ of nucleons within the nucleus, similar to the shell structure of electrons in atoms, and will be discussed in Section 8.3. Nevertheless, the SEMF gives accurate values for the binding energies for some 200 stable and many more unstable nuclei. We will use it to analyse the stability of nuclei in what follows.

Using the numerical values of (2.57), the relative sizes of each of the terms in the SEMF may be calculated, and for the case of odd- A are shown in Figure 2.12. For clarity, the curves have been smoothed, because Z is not a smooth function of A . In this diagram, the volume term is shown as positive and the other terms are subtracted from it to give the final SEMF curve.

Finally, from its definition, one might expect the binding energy per nucleon to be equivalent to the energy needed to remove a nucleon from

¹⁵Note that some authors write the asymmetry term proportional to $(Z - N)^2$, which is equivalent to the form used here, but their value for the coefficient a_a will differ by a factor of four from the one in (2.54).

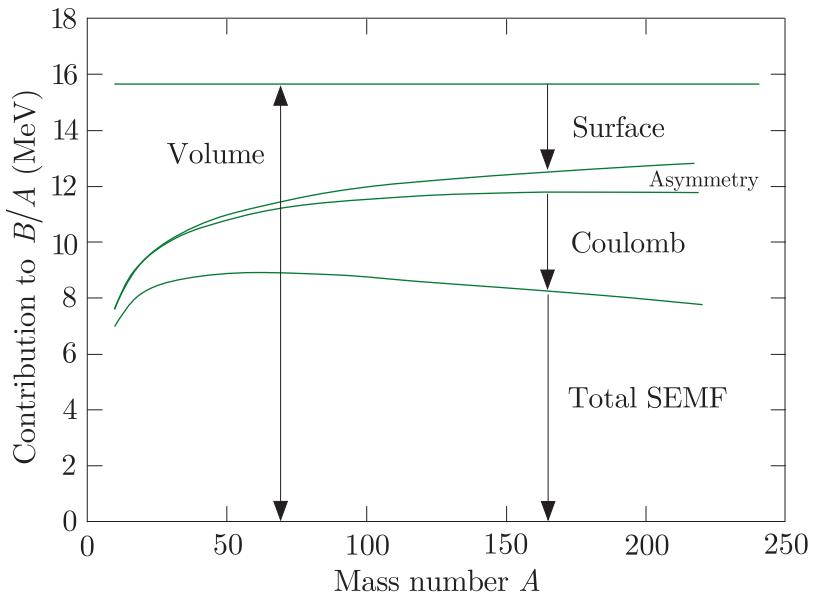


Figure 2.12 Contributions to the binding energy per nucleon as a function of mass number for odd- A from each term in the semi-empirical mass formula. The surface, asymmetry, and Coulomb terms have been plotted so that they subtract from the volume term to give the total SEMF result in the lowest curve.

the nucleus. However, to remove a neutron from a nucleus corresponds to the process



and requires an energy change (the *separation energy*)

$$E_n = [M(Z, A - 1) + M_n - M(Z, A)]c^2 = B(Z, A) - B(Z, A - 1), \quad (2.58b)$$

whereas the removal of a proton corresponds to the process



where X is a different chemical species to Y and requires an energy change

$$\begin{aligned} E_p &= [M(Z - 1, A - 1) + M_p + m_e - M(Z, A)]c^2 \\ &= B(Z, A) - B(Z - 1, A - 1). \end{aligned} \quad (2.59b)$$

Thus, E_p and E_n are only equal to the binding energy per nucleon in an average sense. In practice, measurements show that E_p and E_n can differ substantially from this average and from each other at certain values of (Z, A) . We will see in Chapter 8 that one reason for this is the existence of a shell structure for nucleons within nuclei, mentioned above, which is ignored in the liquid drop model.

2.4 Nuclear instability

A useful way to display the nuclear landscape is the Segrè plot, which is a plot of proton number Z against neutron number N for known nuclear

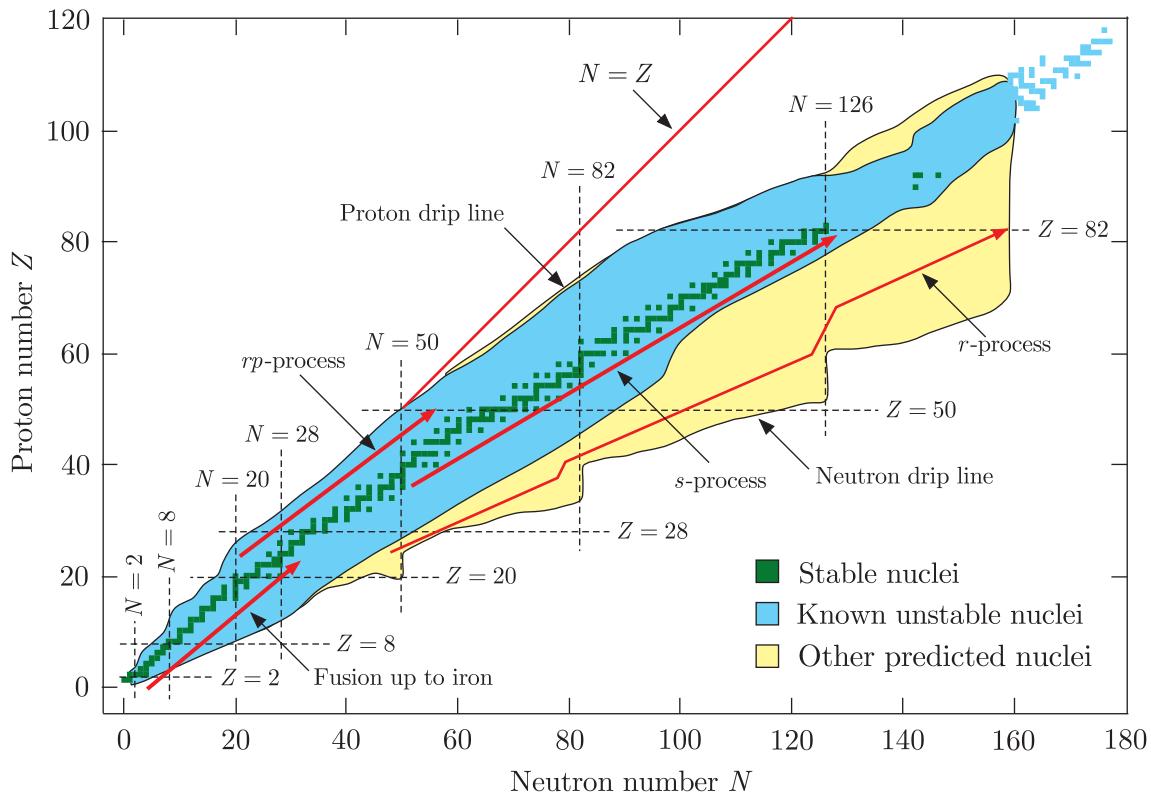


Figure 2.13 The distribution of nuclei. The green squares are the stable nuclei occurring in nature. Known, but unstable, nuclei lie within the blue area, and other predicted nuclei (usually very unstable) lie within the outer yellow area. See text for details. Specific values shown for N and Z are the magic numbers, mentioned below. Source: Bazin (2012), reprinted and adapted by permission from Macmillan Publishers Ltd: copyright 2012.

states. A version is shown in Figure 2.13. Stable nuclei are shown as green squares and lie on a trajectory starting just below the line $Z = N$, but deviating more from it with increasing values of N . Unstable nuclei with known half-lives are located within the blue shaded area, the boundary of which has been simplified as a smooth curve for clarity. In addition, many thousands of other unstable states are predicted to exist, with up to 300 nucleons. Most of these states are expected to be extremely unstable with very small half-lives. They are located within the yellow shaded areas (also drawn with smooth boundaries) on Figure 2.13. The upper and lower boundaries of the yellow area mark the predicted limit of nuclear stability. Adding more neutrons or protons to nuclei along the boundaries leads to structures that are intrinsically unstable and nucleons ‘leak’ out of the ensemble. This is analogous to the situation where droplets of water form from a large water drop, leading to two or more smaller drops. For this reason the boundaries are referred to as the neutron and proton ‘drip lines’.

A very wide range of structures are observed in the Segrè plot. For example, in Section 2.3.2 we mentioned that nucleons arrange themselves in shells and that some configurations of these lead to nuclei that are especially tightly bound. The specific values of N and Z (called *magic numbers*) for $N \leq 126$ at which this occurs are shown in Figure 2.13.¹⁶ A prediction that follows is that there should exist a cluster of tightly bound super-heavy nuclei at even larger values of N and Z , for example at $N = 184$. In principle they could be produced by the collisions of two heavy radioactive nuclei and the study of such reactions is one of the drivers for building nuclear accelerators capable of producing high-energy beams of such particles, as we shall see in Chapter 4.

In the outer regions of the landscape, exotic nuclei may appear, with unusual shapes, including some that have a ‘halo’ of loosely bound neutrons around a central core, and others that are highly non-spherical. Such nuclei may also exhibit unusual decay modes. These outer regions are of great interest because the nuclei there play an important role in the processes that lead to the formation of astrophysical objects, including supernovas, and their subsequent histories. Initially, heavier nuclei are built in stars by a fusion of more neutrons or protons with lighter nuclei, starting from hydrogen, but this process ceases after the nucleus of iron has been formed. Beyond iron, heavier nuclei are produced in several ways, indicated on Figure 2.13 by the red arrows and called the *rp*-process, the *s*-process, and the *r*-process. This is the topic of *nucleosynthesis*, which will be discussed in more detail in Section 9.2.3.

Unstable nuclei decay in various ways, as we have seen in Chapter 1. In β decay, isobars with a surplus of neutrons gain energy by converting a neutron into a proton, via the emission of electrons; conversely, a nucleus with a surplus of protons converts protons to neutrons via positron emission, or by a related process where an atomic electron is captured by the nucleus and the conversion takes place within the nucleus. This is called *electron capture* and, like β decay, is a weak interaction. The electron is usually captured from the innermost atomic shell and the process competes with positron emission in heavy nuclei because the radius of this shell (the K shell) is close to the nuclear radius. The presence of a third particle in the β decay and electron capture processes (called the neutrino by Pauli who first postulated its existence) means that the emitted electrons (or positrons) have a continuous energy spectrum, as mentioned in Section 1.1.1. An extremely rare but related process is double β decay, in which two electrons and two neutrinos are emitted, as we shall see in Section 2.6.2.

The maximum of the curve of binding energy per nucleon is at approximately the position of iron (Fe) and nickel (Ni), which are therefore the most stable nuclides. In heavier nuclei, the binding energy is smaller because of the larger Coulomb repulsion. For still heavier nuclear masses,

¹⁶The origin of magic numbers is discussed in Section 8.3.2.

nuclei can decay spontaneously into two or more lighter nuclei, provided the mass of the parent nucleus is larger than the sum of the masses of the daughter nuclei. Most such nuclei decay via two-body decays and the commonest case is α -decay, where one of the daughter nuclei is a ${}^4\text{He}$ nucleus (i.e. an α particle: ${}^4\text{He} \equiv 2p2n$, with $A = 4$, $Z = N = 2$). The α particle is favoured in such decays because it is a very stable, tightly bound structure. Because this is a two-body decay, the α particle has a unique energy and the total energy released, the *Q value*, is given by

$$Q_\alpha = (M_P - M_D - M_\alpha)c^2 = E_D + E_\alpha, \quad (2.60)$$

where the subscripts refer to parent and daughter nuclei and the α particle, and E is a kinetic energy. The term *fission* is used to describe the rare cases where the two daughter nuclei have similar masses. If the decay occurs without external action, it is called *spontaneous fission* to distinguish it from *induced fission*, where some external stimulus is required to initiate the decay. Spontaneous fission only occurs with a probability greater than that for α emission for nuclei with $Z \geq 110$. The reason for this is discussed in Section 2.7, while α -decay itself is considered in more detail in Section 8.6.

Nuclei may also decay by the emission of photons, with energies in the gamma ray part of the electromagnetic spectrum (*gamma emission*). This occurs when an excited nuclear state decays to a lower state and is a common way whereby excited states lose energy. The lower energy state is often the ground state. A competing process is *internal conversion*, where the nucleus de-excites by ejecting an electron from a low-lying atomic orbit. Both are electromagnetic processes. Electromagnetic decays will be discussed in more detail in Sections 2.8 and 8.8.

2.5 Decay chains

In Section 1.6.4 we saw that the probability per unit time that a given nucleus will decay is specified by its decay constant λ , which is related to the activity \mathcal{A} by (1.76), i.e.

$$\mathcal{A} = -dN/dt = \lambda N, \quad (2.61)$$

where $N(t)$ is the number of radioactive nuclei in the sample at time t . The activity is measured in *becquerels* (Bq), which are defined as one decay per second.¹⁷ If more than one type of decay occurs, then λ is the sum of the decay probabilities for all distinct final states in the same way that the total decay width of an unstable particle is the sum of its partial widths. Integrating (2.61) then gives the exponential decay law (1.77), i.e.

$$N(t) = N_0 \exp(-\lambda t), \quad (2.62)$$

¹⁷An older unit, the curie ($1\text{Ci} = 3.7 \times 10^{10}\text{Bq}$) is also still in common use. A typical laboratory radioactive source has an activity of a few tens of kBq, i.e. μCi .

together with the corresponding result for the activity

$$\mathcal{A}(t) = \lambda N_0 \exp(-\lambda t), \quad (2.63)$$

where N_0 is the initial number of nuclei, i.e. the number at $t = 0$.

In the case of a single decay process $A \rightarrow B$, where B is stable, the number of the final state nuclei $N_B(t)$ grows steadily until the whole sample has decayed, giving

$$N_B(t) = N_A(0) - N_A(t) = N_A(0)[1 - \exp(-\lambda_A t)]. \quad (2.64)$$

However, in many cases the products of radioactive decay are themselves radioactive and so a decay chain results. Consider a decay chain $A \rightarrow B \rightarrow C \rightarrow \dots$ with decay constants $\lambda_A, \lambda_B, \lambda_C, \dots$. The variation of species A with time is given by (2.62), i.e.

$$N_A(t) = N_A(0) \exp(-\lambda_A t), \quad (2.65)$$

but the differential equation for $N_B(t)$ will have an extra term in it to take account of the production of species B from the decay of species A :

$$dN_B(t)/dt = -\lambda_B N_B + \lambda_A N_A. \quad (2.66)$$

The solution of this equation may be verified by substitution to be

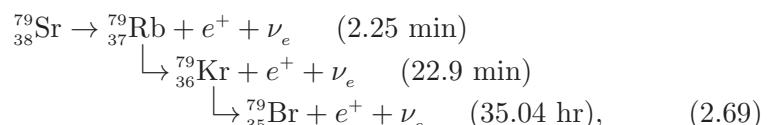
$$N_B(t) = \frac{\lambda_A}{\lambda_B - \lambda_A} N_A(0) [\exp(-\lambda_A t) - \exp(-\lambda_B t)]. \quad (2.67)$$

Similar equations may be found for decay sequences with more than two stages. Thus, for a three-stage sequence

$$N_C(t) = \lambda_A \lambda_B N_A(0) \left[\frac{\exp(-\lambda_A t)}{(\lambda_B - \lambda_A)(\lambda_C - \lambda_A)} + \frac{\exp(-\lambda_B t)}{(\lambda_A - \lambda_B)(\lambda_C - \lambda_B)} + \frac{\exp(-\lambda_C t)}{(\lambda_A - \lambda_C)(\lambda_B - \lambda_C)} \right]. \quad (2.68)$$

The time dependence of the relative numbers of the various components depends of course on the relative sizes of the decay constants.

As an example, the variation of the components as a function of time is shown in Figure 2.14 for the specific case:



with the half-lives shown. Here $\lambda_A > \lambda_B > \lambda_C$, and the final nucleus is stable. This illustrates the general features for this type of sequence, that whereas $N_A(t)$ for the initial species falls monotonically with time and $N_D(t)$ for the final stable species rises monotonically, $N_B(t)$ and $N_C(t)$ for the intermediate species rise to maxima before falling. Note that at any time the sum of the components is a constant, as expected.

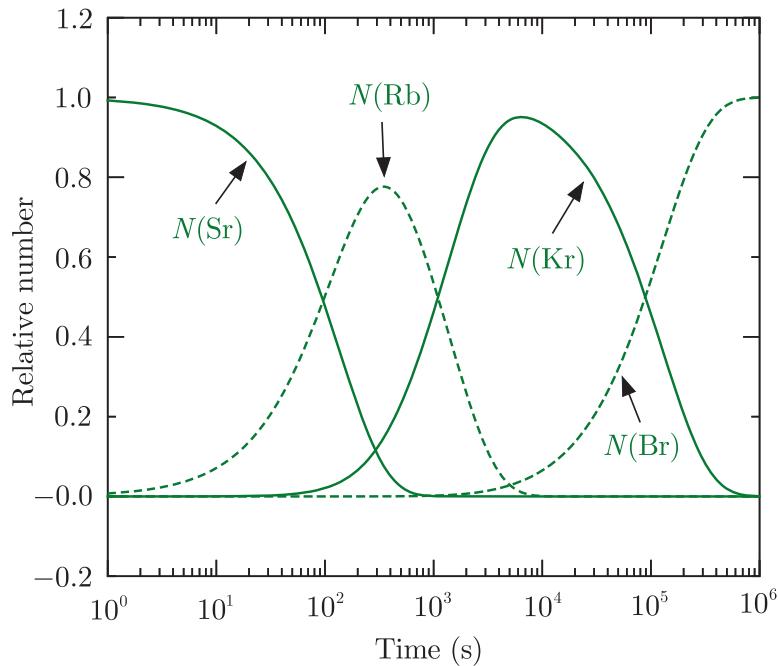


Figure 2.14 Time variation of the relative numbers of nuclei in the decay chain (2.69).

In the following sections we consider the phenomenology of some of the various types of radioactivity in more detail and in Chapter 8 we will return to discuss various models and theories that provide an understanding of these phenomena.

2.6 β decay phenomenology

By rearranging terms, the semi-empirical mass formula (2.49) may be written

$$M(Z, A) = \alpha A - \beta Z + \gamma Z^2 + \frac{\delta}{A^{1/2}}, \quad (2.70)$$

where

$$\begin{aligned} \alpha &= M_n - a_v + \frac{a_s}{A^{1/3}} + \frac{a_a}{4}, & \beta &= a_a + (M_n - M_p - m_e), \\ \gamma &= \frac{a_a}{A} + \frac{a_c}{A^{1/3}}, & \delta &= a_p. \end{aligned} \quad (2.71)$$

$M(Z, A)$ is thus a quadratic in Z at fixed A and has a minimum at $Z = \beta/2\gamma$. For a fixed value of A , a stable nucleus will have an integer value of Z closest to the solution of this equation. For odd A , the SEMF is a single parabola, but for even A the even-even and odd-odd nuclei lie on two distinct vertically shifted parabolas, because of the pairing term. The nucleus with the smallest mass in an isobaric spectrum is stable with respect to β decay. Other nuclei with the same value of A , but values of Z not at the minimum, are usually unstable and will decay. Lifetimes of β emitters vary enormously from milliseconds to 10^{16} yr. They depend very sensitively on the Q value for the decay and on the properties of the

nuclei involved, for example their spins. We will consider the two cases of odd and even A separately using specific values of A to illustrate the main features.

2.6.1 Odd-mass nuclei

Odd-mass nuclei can arise from even- N , odd- Z , or even- Z , odd- N configurations and in practice the number of nuclei that are stable against β decay are roughly equally distributed between these two types. The example we take is the case of the $A = 111$ isobars, which are shown in Figure 2.15a. The circles show the experimental data as *mass excess* values in atomic mass units, where

$$\text{mass excess} \equiv M(Z, A)(\text{in atomic mass units}) - A. \quad (2.72)$$

The curve is the theoretical prediction from the SEMF using the numerical values of the coefficients (2.57). The minimum of the parabola corresponds to the isobar $^{111}_{48}\text{Cd}$ with $Z = 48$.

Isobars with more neutrons, such as $^{111}_{45}\text{Rh}$, $^{111}_{46}\text{Pd}$, and $^{111}_{47}\text{Ag}$, decay by converting a neutron to a proton, i.e.

$$n \rightarrow p + e^- + \bar{\nu}_e, \quad (2.73)$$

so that

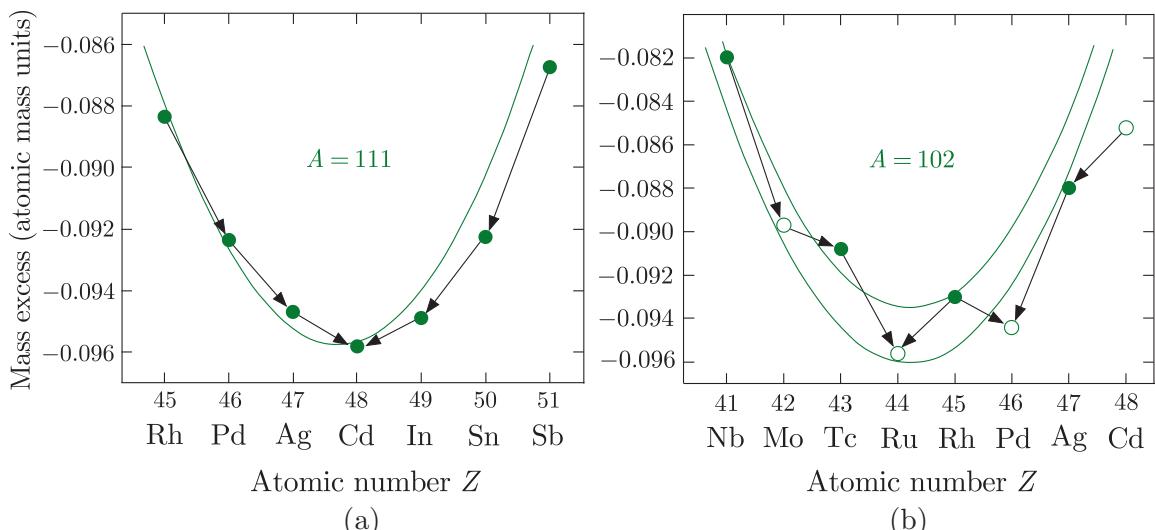


Figure 2.15 (a) Mass parabola of the $A = 111$ isobars. (b) Mass parabolas of the $A = 102$ isobars. The circles are experimental data. The curves are the prediction of the SEMF. Possible β decays are indicated by arrows. In (b) the open circles are even-even nuclei and closed circles are odd-odd nuclei; the upper curve is for odd-odd nuclei and the lower curve is for even-even nuclei.

and



with the half-lives shown. This decay sequence is shown in Figure 2.15a. Electron emission is energetically possible whenever the mass of the daughter atom $M(Z+1, A)$ is smaller than its isobaric neighbour, i.e.

$$M(Z, A) > M(Z+1, A). \quad (2.75)$$

Recall that we are referring here to *atoms*, so that the rest mass of the created electron is automatically taken into account.

Isobars with proton excess decay via



i.e. positron emission, which, although not possible for a free proton, is possible in a nucleus because of the binding energy. Thus, for example, the nuclei ${}_{51}^{111}\text{Sb}$, ${}_{50}^{111}\text{Sn}$, and ${}_{49}^{111}\text{In}$ could in principle decay by positron emission, which is energetically possible if

$$M(Z, A) > M(Z-1, A) + 2m_e, \quad (2.77)$$

and takes account of the creation of a positron and the existence of an excess of electrons in the parent atom.

It is also theoretically possible for this sequence of transitions to occur by electron capture, mentioned in Section 2.4. This process gives rise to a ‘hole’ and causes electrons from higher levels to cascade downwards and in so doing emit characteristic X-rays. Electron capture is energetically allowed if

$$M(Z, A) > M(Z-1, A) + \varepsilon, \quad (2.78)$$

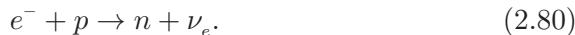
where ε is the excitation energy of the atomic shell of the daughter nucleus. The process competes with positron emission, and in practice for the nuclei above this is what happens. Thus, we have



and



with the half-lives shown, which are manifestations of the primary reaction



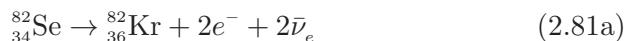
So once again we arrive at the stable isobar.

2.6.2 Even-mass nuclei

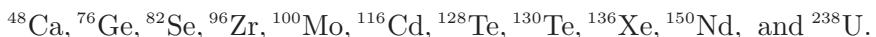
Even-mass nuclei can arise from even- N , even- Z , or odd- Z , odd- N configurations, but, for reasons that are explained below, nearly all even-mass

nuclei that are stable against β decay are of the even–even type, with only a handful of odd–odd types known. Consider as an example the case of $A = 102$ shown in Figure 2.15b. (Recall that the plot is of mass excess, which is a very small fraction of the total mass.) The lowest isobar is $^{102}_{44}\text{Ru}$ and is β stable. The isobar $^{102}_{46}\text{Pd}$ is also stable since its two odd–odd neighbours both lie above it. Thus there are two β -stable isobars. This is a common situation for A even, although no two neighbouring isobars are known to be stable. Odd–odd nuclei always have at least one more strongly bound even–even neighbour nucleus in the isobaric spectrum. They are therefore unstable. The only exceptions to this are a few very light nuclei.

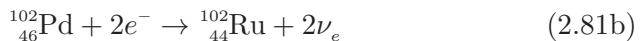
In a small number of even–even nuclei, although β decay is energetically forbidden, the decay $(A, Z) \rightarrow (A, Z+2)$ is energetically allowed and in principle could occur by the emission of two electrons (and of course two antineutrinos). This is referred to as *double beta decay*. It is a second-order weak interaction and is the rarest known type of radioactive decay, with lifetimes of order $10^{19} - 20$ yr. It was first observed in 1987 in the decay



and has subsequently been seen in a total of 11 isotopes:



Related to double beta decay is the possibility of *double electron capture*, i.e. $(A, Z) \rightarrow (A, Z-2)$. For example, referring to Figure 2.15b, in principle the reaction



could occur. There are 35 nuclei where double electron capture is theoretically possible, but to date none has been observed in a direct experiment. Apart from the extremely long lifetimes involved, the experiments are very difficult to do, because the only detectable particles are X-rays in the energy region 1–10 keV, where the backgrounds are usually very high. The only positive evidence is from a geochemical experiment with ^{130}Ba .

2.7 Fission

Spontaneous fission has been defined as the process whereby a parent nucleus breaks into two daughter nuclei of approximately equal masses without external action. The SEMF predicts that the energy release is a maximum when the two fragments have exactly equal masses, but experimentally precisely equal masses are found to be very unlikely. This is illustrated in Figure 2.16 for the case of the isotope ^{254}Fm . Similar results are found for nuclei fissioned by low-energy neutrons, but for fission by very energetic particles the masses are closer to being equal. This is the behaviour that would be expected if its origin were the shell structure

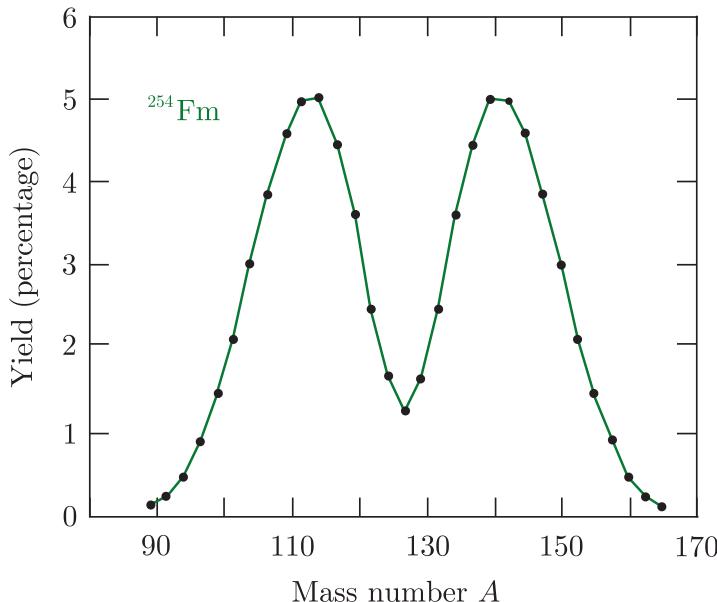
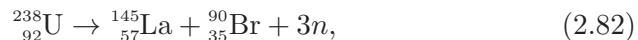


Figure 2.16 Mass distribution of fission fragments from the spontaneous fission of ^{254}Fm . Source: Brandt et al. (1963). Copyright (1963) American Physical Society, reprinted with permission.

of nuclei. The points on Figure 2.16 lie on a smooth curve, but for some nuclei there are often irregularities, which again are due to a nuclear shell structure.

The binding energy curve shows that spontaneous fission is energetically possible for nuclei with $A > 100$.¹⁸ An example is



which illustrates the asymmetrical nature of the fission fragments. This has a release of about 154 MeV of energy, which is carried off as kinetic energy of the fission products. Heavy nuclei are neutron-rich and so necessarily produce neutron-rich decay products, including free neutrons. The fission products are themselves usually some way from the line of β stability and will decay by a series of steps. Thus $^{145}_{57}\text{La}$ decays to the β -stable $^{145}_{60}\text{Nd}$ by three stages, releasing a further 8.5 MeV of energy, which in this case is carried off by the electrons and neutrinos emitted in β decay. Although the probability of fission increases with increasing A , it is still a very rare process. For example, in $^{238}_{92}\text{U}$, the transition rate for spontaneous fission is about $3 \times 10^{-24} \text{ s}^{-1}$ compared with about $5 \times 10^{-18} \text{ s}^{-1}$ for α decay, a branching fraction of 6×10^{-7} , and even for the heavier case of ^{254}Fm shown in Figure 2.16, the branching ratio is 0.06% compared to a branching ratio of 99.94% for α decay. Spontaneous emission only becomes dominant in very heavy elements with $A \geq 270$, as we shall now show.

¹⁸Fission in heavy nuclei was discovered by Otto Hahn and Fritz Strassman and first identified as such by Lise Meitner and Otto Frisch. Hahn was subsequently awarded the 1944 Nobel Prize in Chemistry for his work.

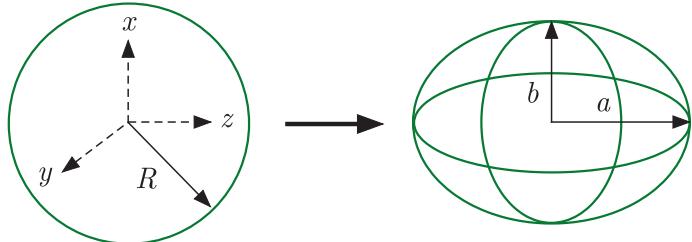


Figure 2.17 Deformation of a heavy nucleus.

To understand spontaneous fission we can again use the liquid drop model. In the SEMF we have assumed that the drop (i.e. the nucleus) is spherical, because this minimises the surface area. However, if the surface is perturbed for some reason from spherical to prolate, the surface term in the SEMF will increase and the Coulomb term will decrease (assuming the volume remains the same because the drop is incompressible) and the relative sizes of these two changes will determine whether the nucleus is stable against spontaneous fission.

For a fixed volume we can parameterise the deformation by the semi-major and semi-minor axes of the ellipsoid a and b , respectively, as shown in Figure 2.17. One possible parameterisation that preserves the volume is

$$a = R(1 + \varepsilon), \quad b = R/(1 + \varepsilon)^{1/2}, \quad (2.83)$$

where ε is a small parameter, so that

$$V = \frac{4}{3}\pi R^3 = \frac{4}{3}\pi ab^2. \quad (2.84)$$

To find the new surface and Coulomb terms one has to find the expression for the surface of the ellipsoid in terms of a and b and expand it in a power series in ε . The algebra is unimportant; the results are

$$E_s = a_s c^2 A^{2/3} (1 + \frac{2}{5}\varepsilon^2 + \dots) \quad (2.85a)$$

and

$$E_c = a_c c^2 Z^2 A^{-1/3} (1 - \frac{1}{5}\varepsilon^2 + \dots). \quad (2.85b)$$

Hence the change in the total energy is

$$\Delta E = (E_s + E_c) - (E_s + E_c)_{\text{SEMF}} = \frac{\varepsilon^2}{5} (2a_s c^2 A^{2/3} - a_c c^2 Z^2 A^{-1/3}). \quad (2.86)$$

If $\Delta E < 0$, then the deformation is energetically favourable and spontaneous fission can occur. From (2.86), this happens if

$$\frac{Z^2}{A} \geq \frac{2a_s}{a_c} \approx 49, \quad (2.87)$$

where we have used the experimental values for the coefficients a_s and a_c given in (2.57).

Spontaneous fission is a potential barrier problem and this is shown in Figure 2.18. The solid line corresponds to the shape of the potential in the parent nucleus. The *activation energy* shown in Figure 2.18 determines the probability of spontaneous fission. In order to fission, the nucleus could in principle tunnel through the barrier, but the fragments are large and the probability for this to happen is extremely small.¹⁹ For heavy nuclei, the activation energy is about 6 MeV, but disappears for very heavy nuclei. For such nuclei, the shape of the potential corresponds closer to the dashed line and the slightest deformation will induce fission.

Another possibility for fission is to supply the energy needed to overcome the Coulomb barrier by a flow of neutrons. Because of the absence of a Coulomb force, a neutron can get very close to the nucleus and be captured by the strong nuclear attraction. The resulting nucleus may then be excited to a state above the fission barrier and therefore splits up. This process is an example of *induced fission*. The energy released when the neutron is captured by a nucleus having an odd neutron number is enhanced by the presence of the pairing energy (2.55). This small extra contribution makes a crucial difference to nuclear fission properties. For example, ultra low-energy ('thermal') neutrons can induce fission in ^{235}U , whereas only higher energy ('fast') neutrons induce fission in ^{238}U . This is because ^{235}U is an even–odd nucleus and ^{238}U is even–even. Therefore, the ground state of ^{235}U will lie higher (less tightly bound) in the potential well of its fragments than that of ^{238}U . Hence to induce fission, a smaller energy will be needed for ^{235}U than for ^{238}U . In principle, fission may be induced in ^{235}U using even zero-energy neutrons.²⁰

We consider this quantitatively as follows. The capture of a neutron by ^{235}U changes an even–odd nucleus to a more tightly bound even–even (compound) nucleus of ^{236}U and releases the binding energy of the last neutron. In ^{235}U this is 6.5 MeV. As the activation energy (the energy needed to induce fission) is about 5 MeV for ^{236}U , neutron capture releases sufficient energy to fission the nucleus. The kinetic energy of the incident neutron is irrelevant and even zero-energy neutrons can induce fission in ^{235}U . In contrast, neutron capture in ^{238}U changes it from an even–even nucleus to an even–odd nucleus, i.e. changes a tightly bound nucleus to a less tightly bound one. The energy released (the binding energy of the last neutron) is about 4.8 MeV in ^{239}U and is less than the 6.5 MeV required for

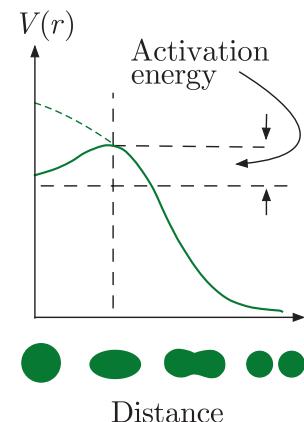


Figure 2.18 Potential energy during different stages of a fission reaction.

¹⁹The special case of α decay is discussed in Section 8.6. There we will show that the lifetime for such decays is expected to have an exponential dependence on the height of the fission barrier and this is observed qualitatively in fission data.

²⁰Enrico Fermi was a pioneer in the field of induced fission and received the 1938 Nobel Prize in Physics for 'demonstrations of the existence of new radioactive elements produced by neutron irradiation, and for his related discovery of nuclear reactions brought about by slow neutrons'. Fermi's citation could equally have been about his experimental discoveries and theoretical work in a wide range of areas from nuclear and particle physics to condensed matter physics and astrophysics. He was probably the last 'universal physicist'.

fission. For this reason, fast neutrons with energy of at least the difference between these two energies are required to fission ^{238}U .

2.8 γ decays

When a heavy nucleus disintegrates by α or β decay, or by fission, the daughter nucleus is often left in an excited state. If this state is below the excitation energy for fission, it will de-excite, usually by emitting a high-energy photon. The energy of these photons is determined by the average energy level spacings in nuclei, and is typically a few MeV in the γ -ray part of the electromagnetic spectrum. In practice, lifetimes are very sensitive to the amount of energy released in the decay and in the nuclear case other factors are also very important, particularly the quantity of angular momentum carried off by the photon. Typical lifetimes of nuclear levels decaying by photon emission are about 10^{-12} s.

The role of angular momentum and parity in γ decays is crucial. If the initial excited state has a total spin \mathbf{J}_i and magnetic quantum number m_i , and the final nucleus has a total spin \mathbf{J}_f and magnetic quantum number m_f , then by angular momentum conservation, the total angular momentum \mathbf{L} of the emitted photon²¹ is given by

$$\mathbf{L} = \mathbf{J}_i - \mathbf{J}_f, \quad (2.88)$$

with

$$J_i + J_f \geq L \geq |J_i - J_f| \quad (2.89)$$

and

$$m_i = M + m_f, \quad (2.90)$$

where M is the magnetic quantum number corresponding to \mathbf{L} . It is also necessary to take account of parity conservation in these electromagnetic processes, as both the initial and final nuclear level and the photon all have intrinsic parities. We will not pursue the consequences of these conservation laws further here, but defer a more detailed discussion until Section 8.8.

2.9 Nuclear reactions

In Chapter 1 and earlier sections of the present chapter we discussed various aspects of reactions, including their classification. In particle physics, because the projectiles and targets have relatively simple structures, this is usually all that is required in classifying reactions. In nuclear physics, the target has a richer structure and it is useful to classify reactions in

²¹As this is the total angular momentum, logically it would be better to employ the symbol \mathbf{J} . However, as \mathbf{J} is invariably used for the nuclear spin, \mathbf{L} will be used in what follows.

more detail. In this section we do this, drawing together our previous work and also anticipating some reactions that will be met in later chapters.²²

Elastic scattering was defined in Chapter 1 as an interaction where the initial and final particles are identical, i.e. $a + A \rightarrow a + A$. We also defined inelastic scattering as the situation where the final particles are the same chemical species, but one or more is in an excited state, e.g. $a + A \rightarrow a + A^*$; and in Section 2.1.2 we showed how the kinematics of such reactions could be used to determine the mass of the excited state. Elastic and inelastic scattering are examples of so-called *direct reactions*. These are defined as ones where the incident particle interacts in a time comparable to the time taken to transit the nucleus. They are more likely when the incident particle has an energy corresponding to a de Broglie wavelength closer to the size of a nucleon rather than that of the nucleus. The collisions are largely peripheral, with only a relatively small fraction of the available energy transferred to the target. Another type of direct reaction is $^{16}\text{O}(p, d)^{15}\text{O}$, i.e.



This is an example of a *pickup reaction*, because one or more nucleons (in this case a neutron) is stripped off the target nucleus and carried away by the projectile. The ‘inverse’ of this reaction is $^{16}\text{O}(d, p)^{17}\text{O}$. This is an example of a *stripping reaction*, because one or more nucleons (in this case again a neutron) is stripped off the projectile and transferred to the target nucleus.

The theoretical interpretation of direct reactions is based on the assumption that the projectile experiences the average potential of the target nucleus. For example, we have seen in the optical model of Section 2.2.2 how this approach can be used to analyse differential cross-sections for elastic scattering and be used to extract information about nuclear shapes and sizes. It also leads to the prediction of resonance structures, with widths typically of order 1 MeV separated by a few MeV, as observed in cross-section as functions of centre-of-mass energy for nucleon scattering from light nuclei. One way of viewing these widths is as a consequence of the reaction time for a direct reaction, typically 10^{-22} s, making use of the uncertainty relation between energy and time, $\Delta E \Delta t \sim \hbar$.

A second important class of interactions is where the projectile becomes loosely bound in the nucleus and shares its energy with all the nuclear constituents. This is called a *compound nucleus reaction*. The time for the system to reach statistical equilibrium depends on the nuclear species, the type of projectile and its energy, but will always be much longer than the transit time and is typically several orders of magnitude longer. An important feature of these reactions is that the properties

²²Hans Bethe received the 1967 Nobel Prize in Physics for his contributions to the theory of nuclear reactions, especially his discoveries concerning the energy production in stars discussed in Section 9.2.3(b).

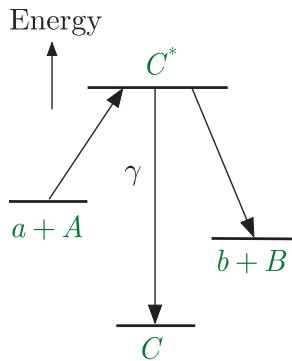


Figure 2.19 Energy-level diagram showing the excitation of a compound nucleus C^* and its subsequent decay.

of the compound nucleus determine its subsequent behaviour and not the mechanism by which it was formed. The compound nucleus is in an excited state and is inherently unstable. Eventually, by a statistical fluctuation, one or more nucleons will acquire sufficient energy to escape and the nucleus either emits particles or de-excites by radiating gamma rays.

If the compound nucleus is created in a region of excitation where its energy levels are well separated, the cross-section will exhibit well-defined resonances described by the Breit–Wigner formulas of Section 1.6.4. These processes are depicted schematically in the energy-level diagram of Figure 2.19, which correspond to $a + A \rightarrow C^* \rightarrow b + B$, where C^* is the compound nucleus, and $a + A \rightarrow C^* \rightarrow C + \gamma$, where C is the ground state corresponding to the excited state C^* . In practice, there could be many final states to which C^* could decay.

Because the time for a compound nucleus to reach statistical equilibrium is much longer than the transit time for a direct reaction, the cross-sections for the compound nucleus process can show variations on much smaller energy scales than those for direct reactions. The density of levels in the compound nucleus is high and so a very small change in the incident energy suffices to alter completely the intermediate states, and hence the cross-section. An example is shown in Figure 2.20, which gives the total cross-section for neutron scattering from ^{12}C at neutron laboratory energies of a few MeV. Peaks corresponding to resonance formation in ^{13}C are clearly identified. Their widths vary from a few tens to a few hundreds of keV, consistent with the characteristic times for compound nucleus formation and decay. The mean widths of compound nucleus excitations depend on the incident energy and the target nucleus, decreasing both with energy and rapidly with nuclear mass. Neutrons, because they are neutral, have a high probability of being captured by nuclei and their cross-sections are rich in compound nucleus effects, particularly at very low energies. This is discussed further below.

The division of reactions into direct and compound nuclei is not exhaustive and situations can occur where particles are ejected from the

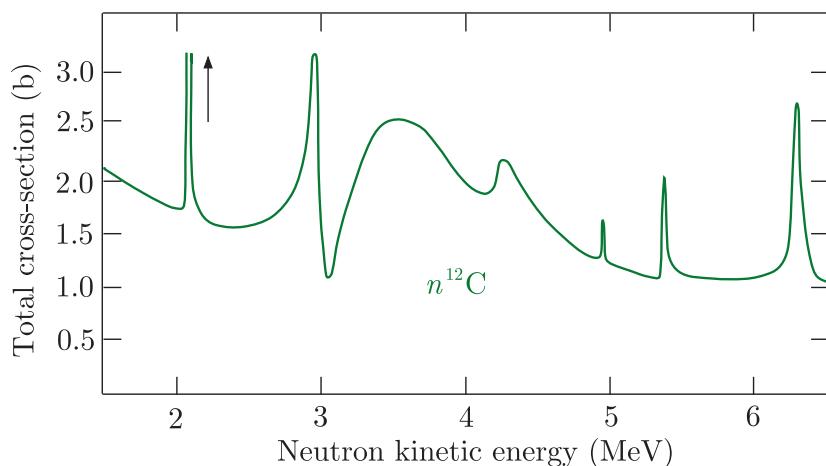


Figure 2.20 Total cross-section for $n^{12}\text{C}$ interactions. Source: Fossan et al. (1961). Copyright (1961) American Physical Society, reprinted with permission.

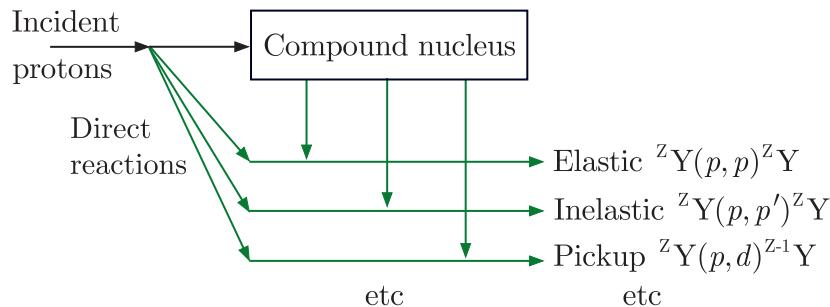


Figure 2.21 Direct and compound nucleus reactions in nuclear reactions initiated by protons.

nucleus before full statistical equilibrium has been reached. Also, in the collisions of complex heavy ions, there is an appreciable probability for an additional reaction mechanism, called *deep inelastic scattering*, that is intermediate between direct and compound nucleus reactions. In this case, the probability for complete fusion of the colliding ions is small, but there can be substantial transfer of the incident kinetic energy to internal excitations of the ions. We will not discuss this or other mechanisms further, but we will meet the concept of deep inelastic scattering again in Section 5.5 in the context of exploring the internal structure of nucleons. In practice, the various mechanisms feed the same final states as direct reactions. This is illustrated schematically in Figure 2.21 for reactions initiated using protons as the projectile.

The general form of the yield $N(E)dE$ of secondary particles at a fixed angle as a function of the outgoing energy E , i.e. the number of particles with energy E between E and $E+dE$, is shown schematically in Figure 2.22 for the case of an incident proton on medium-mass nuclei, where the incident particle has an energy of a few times that of the Coulomb barrier. At the upper end of the plot there are a number of distinct peaks due to elastic, inelastic, and transfer reactions. Then as the excitation energy is reduced, the more closely spaced energy levels in the final nucleus are not fully resolved because of the spread in energy of the incident beam and the uncertainty in the experimental measurements of energy. At the lowest energies there is a broad continuum mainly

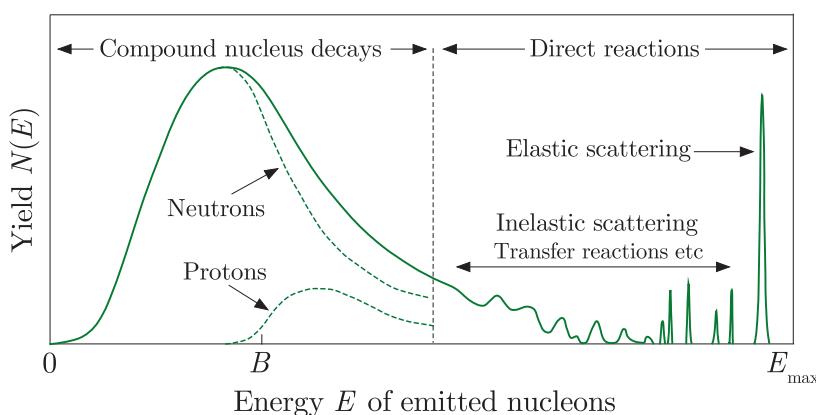


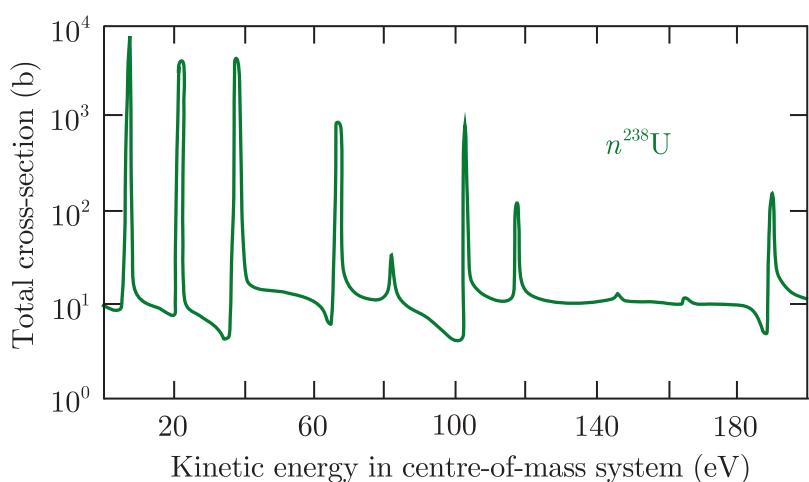
Figure 2.22 Typical spectrum of energies of nucleons emitted at a fixed angle in inelastic proton–nucleus reactions, where the beam energy is a few times that of the Coulomb barrier. The dashed lines show the separate contributions of protons and neutrons in the energy region where compound nucleus reactions dominate. The energy of the proton Coulomb barrier is marked as B . Source: Adapted from Lilley (2001). Copyright John Wiley & Sons, Inc., reprinted with permission.

due to the decays of compound nuclei formed by the absorption of the projectile nucleon by the target nucleus. The nucleons ‘evaporating’ from a compound nucleus will have relatively low energies because the statistical probability of a particle acquiring a large fraction of the available energy is small. The differential cross-sections for the two processes are very different. Direct reactions lead to cross-sections peaked in the forward direction, falling rapidly with angle and with oscillations, as we have seen in the case of elastic scattering in Section 2.2.1 (Figure 2.5). In contrast, differential cross-sections in the region where an isolated compound nucleus is formed are fairly isotropic and approximately symmetric about 90° .

Many medium and large- A nuclei can capture very low-energy ($10 - 100$ eV) neutrons very readily. The neutron separation energy for the final nucleus is approximately 6 MeV and thus capture leads to a compound nucleus with an excitation energy above the ground state by this separation energy. Such excitation often occurs in a region of high density of narrow states that show up as a rich resonance structure in the corresponding neutron total cross-section. An example is shown in Figure 2.23. The value of the cross-section at the resonance peaks can be many orders of magnitude greater than the geometrical cross-section based on the size of the nucleus. This is because the cross-section is determined dominantly by the area associated with the wavelength λ of the projectile, i.e. $\pi\lambda^2$, which is very large because λ is large.

Once formed, the compound nucleus can decay to any final state consistent with the relevant conservation laws. If this includes neutron emission, it will be the preferred decay. However, for production by very slow (thermal) neutrons, with energies of order 0.02 eV, the available decay kinetic energy will reflect the initial energy of the projectile, which is very small. Therefore, in these cases, photon emission is often preferred. We shall see in Section 9.1.2 that the fact that radiative decay is the dominant decay mode of compound nuclei formed by thermal neutrons

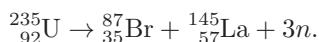
Figure 2.23 Total cross-section for very low energy neutron interactions with ^{238}U , showing many very narrow resonances (with intrinsic widths of order 10^{-2} eV) corresponding to excited states of ^{239}U .
Source: Garber and Kinsey (1976), courtesy of Brookhaven National Laboratory.



is important in the use of nuclear fission to produce power in nuclear reactors.

Problems 2

- 2.1** Show that mass measurements in an ideal Penning trap have maximum precision when the observation time is about 2.9 times the nuclide half-life.
- 2.2** Electrons with momentum $330 \text{ MeV}/c$ are elastically scattered through an angle of 10^0 by a nucleus of ^{56}Fe . If the charge distribution on the nucleus is that of a uniform hard sphere, and assuming the Born approximation is valid, by what factor would you expect the Mott cross-section to be reduced?
- 2.3** Show explicitly that (2.39) follows from (2.37).
- 2.4** A beam of electrons with energies of 250 MeV is scattered through an angle of 10^0 by a heavy nucleus. It is found that the differential cross-section is 65% of that expected from scattering from a point nucleus. Estimate the root mean square radius of the nucleus.
- 2.5** Show that the q^2 dependence of the form factor corresponding to the charge distribution $\rho(r) = \rho_0 \exp(-r/a)/r$, where ρ_0 and a are constants, is of the form $[(a^2 q^2/\hbar^2) + 1]^{-1}$.
- 2.6** A sample of 1 g of a radioactive isotope of atomic mass 208 decays via β emission, and 75 counts are recorded in a 24 hour period. If the detector efficiency is 10%, estimate the mean life of the isotope.
- 2.7** Nuclei of $^{212}_{86}\text{Rn}$ decay by α emission to $^{208}_{84}\text{Po}$ with a mean life of 23.9 min . The $^{208}_{84}\text{Po}$ nuclei in turn decay, also by α -emission, to the stable isotope $^{204}_{82}\text{Pb}$ with a mean life of 2.9 yr . If initially the source is pure $^{212}_{86}\text{Rn}$, how long will it take for the rate of α emission in the final decay to reach a maximum?
- 2.8** Natural lanthanum has an atomic weight of 138.91 and contains 0.09% of the isotope $^{138}_{57}\text{La}$. The latter has two decay modes: $^{138}_{57}\text{La} \rightarrow ^{138}_{58}\text{Ce} + e^- + \bar{\nu}_e$ (β^- decay) and $^{138}_{57}\text{La} + e^- \rightarrow ^{138}_{56}\text{Ba}^* + \nu_e$ (electron capture) followed by the electromagnetic decay of the excited state $^{138}_{56}\text{Ba}^* \rightarrow ^{138}_{56}\text{Ba} + \gamma$ (radiative decay). There are $7.8 \times 10^2 \beta^-$ particles emitted per second per kilogram of natural lanthanum and there are 50 photons emitted per 100 β^- particles. Estimate the mean lifetime of $^{138}_{57}\text{La}$.
- 2.9** Show that the electrostatic self-energy of a uniformly charged sphere of charge Ze and radius $1.2A^{1/3}\text{fm}$ is of the form $aZ^2/A^{1/3}$ and find the value of the constant a . Compare your result to the corresponding term in the semi-empirical mass formula.
- 2.10** Use the SEMF to estimate the energy released in the spontaneous fission reaction



- 2.11** The most stable nucleus with $A = 111$ is $^{111}_{48}\text{Cd}$ (see Figure 2.15). By what percentage would the fine structure constant α have to change if the most stable nucleus with $A = 111$ were to be $^{111}_{47}\text{Ag}$? Assume that altering α does not change particle masses.

- 2.12** The transuranic isotope $^{269}_{108}\text{Hs}$ decays 100% via α emission, i.e. $^{269}_{108}\text{Hs} \rightarrow ^{265}_{106}\text{Sg} + \alpha$, where the kinetic energy of the α particle is $E_\alpha = 9.23$ MeV. Assuming the known masses of $^{265}_{106}\text{Sg}$ and the α particle, calculate the mass of the $^{269}_{108}\text{Hs}$ nucleus in atomic mass units.
- 2.13** The isotope $^{238}_{94}\text{Pu}$ decays via α emission to the essentially stable isotope $^{234}_{92}\text{U}$ with a mean lifetime of 126.7 yr and a release of 5.49 MeV of kinetic energy. This energy is converted to electrical power in a space probe designed to reach planet X in a journey planned to last four years. If the efficiency of power conversion is 5% and on reaching planet X the probe requires at least 200 W of power to perform its landing tasks, how much $^{238}_{94}\text{Pu}$ would be needed at launch?
- 2.14** The reaction $^{45}_{21}\text{Sc}(d, p)^{46}_{21}\text{Sc}$ has a Q -value of 6.54 MeV and a resonance when the incident deuteron laboratory kinetic energy is 2.76 MeV. Would you expect the same resonance to be excited in the reaction $^{43}_{20}\text{Ca}(\alpha, n)^{46}_{22}\text{Ti}$ and if so at what value of the laboratory kinetic energy of the alpha particle? You may use the fact that the β decay $^{46}_{21}\text{Sc} \rightarrow ^{46}_{22}\text{Ti} + e^- + \bar{\nu}_e$ has a Q -value of 2.37 MeV and the mass difference between the neutron and a hydrogen atom is 0.78 MeV/c².
- 2.15** A radioisotope with decay constant λ is produced at a constant rate P starting at time $t = 0$. Show that the number of atoms at time $t > 0$ is $N(t) = P[1 - \exp(-\lambda t)]/\lambda$.
- 2.16** Radioactive ^{36}Cl (half-life 3×10^5 yr) is produced by irradiating 1 g of natural nickel chloride (NiCl_2 , molecular weight 129.6) in a neutron beam of flux $F = 10^{14}$ cm⁻²s⁻¹. If the neutron absorption cross-section $^{35}\text{Cl}(n, \gamma)^{36}\text{Cl}$ is $\sigma = 43.6$ b and 75.8% of natural chlorine is ^{35}Cl , use the result of Problem 2.15 to estimate the time it would take to produce a 3×10^5 Bq source of ^{36}Cl .
- 2.17** Consider the total cross-section data for the $n^{238}\text{U}$ interaction shown in Figure 2.23. There is a resonance R at the centre-of-mass neutron kinetic energy $E_n = 10$ eV with width $\Gamma = 10^{-2}$ eV and the total cross-section there is $\sigma_{\max} = 9 \times 10^3$ b. Use this information to find the partial widths $\Gamma_{n, \gamma}$ for the decays $R \rightarrow n + ^{238}\text{U}$ and $R \rightarrow \gamma + ^{239}\text{U}$ if these are the only two significant decay modes. The spin of the ground state of ^{238}U is zero.

3

Particle phenomenology

In this chapter we look at some of the phenomena of particle physics – the properties of leptons and quarks, and the bound states of the latter, the hadrons. In later chapters we will discuss theories and models that attempt to explain these and other particle data.

3.1 Leptons

We have seen that the spin-1/2 leptons are one of the three classes of elementary particles in the standard model and we start with a discussion of their basic properties, including the concept of ‘lepton universality’ and the question of how many distinct types of neutrino can exist. Then we look in more detail at the neutral leptons, the neutrinos. In particular, we discuss the clear evidence for the existence of the phenomenon of neutrino oscillations, which can only exist if neutrinos have nonzero masses. The section concludes with a review of the experimental evidence for small but nonzero neutrino masses and its implications.

3.1.1 Lepton multiplets and lepton numbers

There are six known leptons, and they occur in pairs, called *generations*, which we write, for reasons that will become clear presently, as

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}, \quad \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}, \quad \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}. \quad (3.1)$$

Each generation comprises a *charged lepton* with electric charge $-e$ and a *neutral neutrino*. The three charged leptons (e^- , μ^- , τ^-) are the familiar electron, together with two heavier particles, the *mu lepton* (usually called the *muon*, or just *mu*) and the *tau lepton* (usually called the *tauon*, or

just *tau*). The associated neutrinos are called the *electron neutrino*, *mu neutrino*, and *tau neutrino*, respectively.¹ The six distinct types of leptons are also referred to as having different ‘flavours’. In addition to the leptons, there are six corresponding antileptons:

$$\begin{pmatrix} e^+ \\ \bar{\nu}_e \end{pmatrix}, \quad \begin{pmatrix} \mu^+ \\ \bar{\nu}_\mu \end{pmatrix}, \quad \begin{pmatrix} \tau^+ \\ \bar{\nu}_\tau \end{pmatrix}. \quad (3.2)$$

Ignoring gravity, the charged leptons interact only via electromagnetic and weak forces, whereas for the neutrinos, only weak interactions have been observed.² Because of this, neutrinos, which are all believed to have extremely small masses, can be detected only with considerable difficulty.

Table 3.1 Properties of leptons. All have spin 1/2. Masses are given in units of MeV/ c^2 , except for the electron neutrino. The antiparticles (not shown) have the same masses as their associated particles, but the electric charges (Q) and lepton numbers (L_ℓ , $\ell = e, \mu, \tau$) are reversed in sign. The neutrinos are shown as stable, but we will return to this in Section 3.1.5 when we discuss neutrino oscillations.

Name and symbol	Mass	Q	L_e	L_μ	L_τ	Lifetime (s)	Major decays
Electron e^-	0.511	-1	1	0	0	Stable	None
Electron neutrino ν_e	$<2\text{eV}/c^2$	0	1	0	0	Stable	None
Muon (mu) μ^-	105.7	-1	0	1	0	2.197×10^{-6}	$e^- \bar{\nu}_e \nu_\mu$ (100%)
Muon neutrino ν_μ	<0.19	0	0	1	0	Stable	None
Tauon (tau) τ^-	1776.9	-1	0	0	1	2.903×10^{-13}	$\mu^- \bar{\nu}_\mu \nu_\tau$ (17.4%) $e^- \bar{\nu}_e \nu_\tau$ (17.8%) $\nu_\tau + \text{hadrons}$ ($\sim 64\%$)
Tauon neutrino ν_τ	<18.2	0	0	0	1	Stable	None

The masses and lifetimes of the leptons are listed in Table 3.1. The electron is stable, for reasons that will become clear shortly. The muons decay by the weak interaction processes

$$\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu; \quad \mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu, \quad (3.3a)$$

with a lifetime of $(2.1969811 \pm 0.0000022) \times 10^{-6}$ s. The tau also decays by the weak interaction, but with a much shorter lifetime of $(2.903 \pm 0.005) \times 10^{-13}$ s. This illustrates what we have already seen in nuclear physics, that lifetimes depend sensitively on the energy released in the decay, i.e. the Q value. Because it is heavier than the muon, the tau has sufficient energy to decay to many different final states, which can

¹Leon Lederman, Melvin Schwartz, and Jack Steinberger shared the 1988 Nobel Prize in Physics for their use of neutrino beams and the discovery of the muon neutrino. Martin Perl shared the 1995 Nobel Prize in Physics for his pioneering work in lepton physics and in particular for the discovery of the tau lepton.

²Although neutrinos have zero electric charge, they could in principle have a charge distribution that would give rise to a magnetic moment (like neutrons) and hence electromagnetic interactions. This is forbidden in the standard model because neutrinos are assumed to be point-like.

include both hadrons and leptons. However, about 35% of decays lead to purely leptonic final states, via reactions that are very similar to muon decay, for example:

$$\tau^+ \rightarrow \mu^+ + \nu_\mu + \bar{\nu}_\tau; \quad \tau^- \rightarrow e^- + \bar{\nu}_e + \nu_\tau. \quad (3.3b)$$

Associated with each generation of leptons is a quantum number called a *lepton number*. The first of these lepton numbers is the *electron number*, defined for any state by

$$L_e \equiv N(e^-) - N(e^+) + N(\nu_e) - N(\bar{\nu}_e), \quad (3.4)$$

where $N(e^-)$ is the number of electrons present, $N(e^+)$ is the number of positrons present, and so on. For single-particle states, $L_e = 1$ for e^- and ν_e ; $L_e = -1$ for e^+ and $\bar{\nu}_e$; and $L_e = 0$ for all other particles. *Muon and tauon numbers* are defined in a similar way and their values for all single-particle states are summarised in Table 3.1, together with the values of their other quantum numbers. For multiparticle states, the lepton numbers of the individual particles are added. For example, the final state in neutron beta decay (i.e. $n \rightarrow p + e^- + \bar{\nu}_e$) has

$$L_e(\text{final}) = L_e(p) + L_e(e^-) + L_e(\bar{\nu}_e) = (0) + (1) + (-1) = 0, \quad (3.5)$$

like the initial state, which has $L_e(\text{initial}) = L_e(n) = 0$.

In the standard model, the value of each lepton number is postulated to be separately conserved in any reaction. The decays of (3.3) illustrate this principle of *lepton number conservation*. In electromagnetic interactions, this reduces to the conservation of $N(e^-) - N(e^+)$, $N(\mu^-) - N(\mu^+)$, and $N(\tau^-) - N(\tau^+)$, since neutrinos are not involved. This implies that the charged leptons can only be created or annihilated in particle–antiparticle pairs. For example, in the electromagnetic reaction

$$e^+ + e^- \rightarrow \mu^+ + \mu^- \quad (3.6)$$

an electron pair is annihilated and a muon pair is created by the mechanism of Figure 3.1.

In weak interactions more general possibilities are allowed that still conserve lepton numbers. For example, in the tau-decay process $\tau^- \rightarrow e^- + \bar{\nu}_e + \nu_\tau$, a tau converts to a tau neutrino and an electron is created together with an antineutrino, rather than a positron. The dominant Feynman graph corresponding to this process is shown in Figure 3.2.

Lepton number conservation, like electric charge conservation, plays an important role in understanding reactions involving leptons. Observed reactions conserve lepton numbers, while reactions that do not conserve lepton numbers are forbidden and are not observed. For example, the neutrino scattering reaction

$$\nu_\mu + n \rightarrow \mu^- + p \quad (3.7)$$

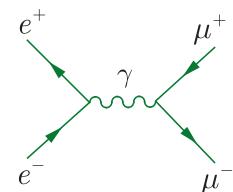


Figure 3.1 Single-photon exchange in the reaction $e^+ e^- \rightarrow \mu^+ \mu^-$.

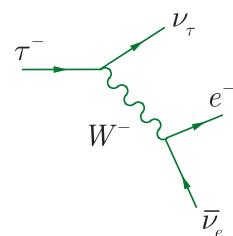


Figure 3.2 Dominant Feynman diagram for the decay $\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$.

is observed experimentally, whereas the apparently similar reaction

$$\nu_\mu + n \rightarrow e^- + p \quad (3.8)$$

that violates both L_e and L_μ conservation, is not. Another example that violates both L_e and L_μ conservation is $\mu^- \rightarrow e^- + \gamma$. If this decay were allowed, the dominant decay of the muon would be electromagnetic and the muon lifetime would be much shorter than its observed value. A number of possible decays that violate one or more lepton numbers have been searched for but without success, and the upper limits for the branching ratios of such decays are between 10^{-7} and 10^{-13} . This is very strong evidence that lepton numbers are conserved to a high degree of accuracy in reactions.

Finally, conservation laws explain the stability of the electron. The electron is stable because electric charge is conserved in all interactions and the electron is the lightest charged particle. Hence decays to lighter particles that satisfy all other conservation laws, like $e^- \rightarrow \nu_e + \gamma$, are necessarily forbidden by electric charge conservation. In the same way, one would expect lepton number conservation to imply that the lightest particles with nonzero values of the three lepton numbers – the three neutrinos – are stable, whether they have zero masses or not, and this is shown in Table 3.1. However, we will return to this question, and that of lepton number conservation, in Sections 3.1.6 and 3.1.7 below.

3.1.2 Universal lepton interactions; the number of neutrinos

The three neutrinos have similar properties, but the three charged leptons are strikingly different. For example, the mass of the muon is roughly 200 times greater than that of the electron and consequently its magnetic moment is 200 times smaller; high-energy electrons are stopped by modest thicknesses of a centimetre or so of lead, while muons are the most penetrating form of radiation known, apart from neutrinos; and the tauon lifetime is many orders of magnitude smaller than the muon lifetime, while the electron is stable. It is therefore a remarkable fact that all experimental data are consistent with the assumption that the interactions of the electron and its associated neutrino are identical with those of the muon and its associated neutrino, and of the tauon and its neutrino, provided the mass differences are taken into account. This property, called *lepton universality*, can be verified with great precision, because we have a precise theory of electromagnetic and weak interactions (to be discussed in Chapter 6), which enables predictions to be made of the mass dependence of all observables.

For example, when we discuss experimental methods in Chapter 4, we will show that the *radiation length*, which is a measure of how far a charged particle travels through matter before losing a certain fraction of its energy by radiation, is proportional to the squared mass of the radiating particle. Hence it is about 4×10^4 times greater for muons than

for electrons, explaining their much greater penetrating power in matter. Another example is provided by the leptonic decays (3.3) of the muon and the tau to electrons and neutrinos. These weak interactions are accurately accounted for by the zero-range approximation discussed in Section 1.6.1, so that the decay widths Γ and decay rates Γ/\hbar are proportional to the square of the Fermi coupling constant G_F whose value is given in (1.55). In addition, the masses of the final state electrons and neutrinos are very small compared to the masses of the decaying muons and tauons, and can to a good approximation be neglected. Hence, in this approximation, apart from \hbar and c , the only relevant physical constants are G_F and the mass $m_\ell (\ell = \tau, \mu)$ of the decaying lepton, so that the decay width must be given by an expression of the form

$$\Gamma = K G_F^2 \hbar^\alpha c^\beta m_\ell^\gamma,$$

where the dimensionless constant K must have the same value for muon and tauon decay if universality holds. The coefficients α, β, γ are then determined by comparing dimensions, and since widths have the dimensions of energy, one easily sees, using (1.55), that

$$\Gamma = K G_F^2 (m_\ell c^2)^5 / (\hbar c)^6,$$

so that the decay widths are proportional to the fifth power of the decaying particle masses.³ Thus, from universality, the ratio of the decay widths is given to a good approximation by

$$\frac{\Gamma(\tau^- \rightarrow e^- + \bar{\nu}_e + \nu_\tau)}{\Gamma(\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu)} \approx \left(\frac{m_\tau}{m_\mu} \right)^5 = 1.345 \times 10^6. \quad (3.9)$$

This prediction is in good agreement with experiment, so that the mass difference between the tau and muon accounts very well for the huge difference between the tau and muon lifetimes. The above are just some of the most striking manifestations of the universality of lepton interactions.

A question that arises naturally is whether there are more generations of leptons, with identical interactions, waiting to be discovered. This question has been answered, under reasonable assumptions, by an experimental study of the decays of the Z^0 boson. This particle, one of the gauge bosons associated with the weak interaction, is copiously produced in e^+e^- collisions if the total energy is tuned to the Z^0 mass and enables its mass and total width Γ_z to be very accurately found. It decays, among other final states, to neutrino pairs, and if we assume universal lepton interactions and neutrino masses that are small compared to the

³The result is closely related to *Sargent's Rule*, which accounts for the enormous variation in the lifetimes of nuclei decaying via β decay, and is illustrated briefly in Section 8.7.4(c).

mass of the Z^0 ,⁴ the decay rates to a given neutrino pair will all be equal. Thus,

$$\Gamma_{\text{neutrinos}} \equiv \Gamma_{\nu_e} + \Gamma_{\nu_\mu} + \Gamma_{\nu_\tau} + \dots = N_\nu \Gamma_\nu, \quad (3.10)$$

where N_ν is the number of neutrino species and Γ_ν is the partial decay width to any given pair of neutrinos. The total decay width of the Z^0 may then be written

$$\Gamma_Z = \Gamma_{\text{hadrons}} + 3\Gamma(Z^0 \rightarrow \ell^+ \ell^-) + N_\nu \Gamma_\nu, \quad (3.11)$$

giving

$$N_\nu \Gamma_\nu = \Gamma_Z - \Gamma_{\text{hadrons}} - 3\Gamma(Z^0 \rightarrow \ell^+ \ell^-), \quad (3.12)$$

where Γ_{hadrons} is the measured decay width to all hadrons and $\Gamma(Z^0 \rightarrow \ell^+ \ell^-)$ is the width for decay to charged leptons, respectively. Although the decay width to a specific neutrino species Γ_ν is not directly measured, it can be calculated in the standard model, and combining this with experimental data for the other decay modes, a value of N_ν may be found. The best value using all available data is $N_\nu = 2.92 \pm 0.05$, which is consistent with the expectation for three neutrino species, but not four. The conclusion is that only three generations of leptons exist, if we assume universal lepton interactions and exclude very large neutrino masses. Why there are just three generations of leptons remains a mystery, particularly as the extra two generations seem to tell us nothing fundamental that cannot be deduced from the interactions of the first generation.

3.1.3 Neutrinos

As we mentioned in Chapter 1, the existence of the *electron neutrino* ν_e was first postulated by Pauli in 1930. He did this in order to understand the observed nuclear β decays

$$(Z, N) \rightarrow (Z + 1, N - 1) + e^- + \bar{\nu}_e \quad (3.13)$$

and

$$(Z, N) \rightarrow (Z - 1, N + 1) + e^+ + \nu_e, \quad (3.14)$$

which were discussed in Section 2.6. The neutrinos and antineutrinos emitted in these decays are not observed experimentally, but are inferred from energy and angular momentum conservation. In the case of energy, if the antineutrinos were not present in the first of the reactions, the energy E_e of the emitted electron would be a unique value equal to the difference in rest energies of the two nuclei, i.e.

$$E_e = \Delta Mc^2 = [M(Z, N) - M(Z + 1, N - 1)]c^2, \quad (3.15)$$

⁴More precisely, we assume $m_\nu \leq M_Z/2$, so that the decays $Z \rightarrow \nu\bar{\nu}$ are not forbidden by energy conservation.

where for simplicity we have neglected the extremely small kinetic energy of the recoiling nucleus. However, if an antineutrino is present, the electron energy would not be unique, but lie in the range

$$m_e c^2 \leq E_e \leq (\Delta M - m_{\bar{\nu}_e})c^2, \quad (3.16)$$

depending on how much of the kinetic energy released in the decay is carried away by the neutrino. Experimentally, the observed energies span the whole of the above range and in principle a measurement of the energy of the electron near its maximum value of $E_e = (\Delta M - m_{\bar{\nu}_e})c^2$ determines the neutrino mass. The most accurate results come from tritium (${}^3\text{H}$) decay. When experimental errors are taken into account, the experimentally allowed range is

$$0 \leq m_{\bar{\nu}_e} < 2 \text{ eV}/c^2 \approx 4 \times 10^{-6} m_e. \quad (3.17)$$

We will discuss this determination of $m_{\bar{\nu}_e}$ in more detail in Section 8.7.4, after we have considered the theory of β decay.

The masses of both ν_μ and ν_τ can similarly be directly inferred from the e^- and μ^- energy spectra in the leptonic decays of muons and tauons, using energy conservation. The results from these and other decays show that the neutrino masses are very small compared with the masses of the associated charged leptons. The present limits are given in Table 3.1.

Small neutrino masses, compatible with the above limits, can be ignored in most circumstances, and there are theoretical attractions in assuming neutrino masses are precisely zero, as was done in the original formulation of the standard model. However, we will show in the following section that there is now strong evidence for physical phenomena that could not occur if the neutrinos had exactly zero mass. The consequences of neutrinos having small masses have therefore to be taken seriously.

Because neutrinos only have weak interactions, they can only be detected with great difficulty. For example, electron neutrinos and antineutrinos of sufficient energy can in principle be detected by observing the *inverse β decay* processes



and



However, the probability for these and other processes to occur is extremely small. In particular, the neutrinos and antineutrinos emitted in β decays, with energies of order 1 MeV, have mean free paths in matter of order 10^6 km.⁵ Nevertheless, if the neutrino flux is intense enough and the detector is large enough, the reactions can be observed. In particular, uranium fission fragments are neutron-rich and decay by electron

⁵The mean free path is roughly the average distance a particle would have to travel in a medium for there to be a significant probability of an interaction. It is defined more precisely in Section 4.3.4.

emission to give an antineutrino flux that can be of order $10^{17} \text{ m}^{-2}\text{s}^{-1}$ in the vicinity of a nuclear reactor, which derives its energy from the decay of nuclei (see Section 9.1.2). These antineutrinos will occasionally interact with protons in a large detector, enabling examples of the inverse β decay reaction to be observed. As mentioned in Chapter 1 (Footnote 9), electron neutrinos were first detected in this way by Reines and Cowan in 1956, and their interactions have been studied in considerable detail since.

The mu neutrino ν_μ has been detected using the reaction $\nu_\mu + n \rightarrow \mu^- + p$ and other reactions. In this case, well-defined high-energy ν_μ beams can be created in the laboratory by exploiting the decay properties of pions, which are particles we have mentioned briefly in Chapter 1 and which we will meet in more detail later in this chapter. The probability of neutrinos interacting with matter increases rapidly with energy (this will be demonstrated at the end of Section 5.5.4) and for large detectors, events initiated by such beams are so copious that they have become an indispensable tool in studying both the fundamental properties of weak interactions and the internal structure of the proton. Finally, in 2000, the first examples of tau neutrino scattering were reported, so that 70 years after Pauli first suggested the existence of a neutrino, all three types had finally been directly detected.

3.1.4 Neutrino mixing and oscillations

When the standard model was first proposed, neutrinos were assumed to have zero mass, but by 2000 this hypothesis was no longer tenable, due to the discovery of *neutrino oscillations*. This is the name given to the situation where a beam of neutrinos of one type, for example ν_μ , develops components of other types, for example ν_e and/or ν_τ , as it travels over long distances. For this to occur, there must in addition be *neutrino mixing*. This is the assumption that the neutrino states that couple to electrons, muons, and tauons, respectively, do not have definite masses, but instead are linear combinations of three other states ν_1 , ν_2 , and ν_3 that do have definite masses m_1 , m_2 , and m_3 , i.e. are eigenstates of mass. As we shall immediately see, neutrino oscillations can then occur, but only if the masses m_1 , m_2 , and m_3 are different and so cannot all be zero.

Instead of considering mixing between all three ‘flavour’ states ν_e , ν_μ , and ν_τ , it is simpler, and often a good approximation, to consider the mixing between just two of them, which we will denote ν_α and ν_β . In order to preserve the orthonormality of the states, we can write

$$\nu_\alpha = \nu_i \cos\theta_{ij} + \nu_j \sin\theta_{ij} \quad (3.20)$$

and

$$\nu_\beta = -\nu_i \sin\theta_{ij} + \nu_j \cos\theta_{ij}, \quad (3.21)$$

where ν_i and ν_j are the two mass eigenstates involved. Here ν_α is shorthand for $|\nu_\alpha, \psi\rangle$, etc. and θ_{ij} is a *mixing angle* that must be determined from experiment.

If $\theta_{ij} \neq 0$ then some interesting predictions follow. For example, when a ν_α neutrino is produced with momentum \mathbf{p} at time $t = 0$, the ν_1 and ν_2 components will have slightly different energies, E_1 and E_2 , due to their slightly different masses. In quantum mechanics, their associated waves will therefore have slightly different frequencies, giving rise to a phenomenon somewhat akin to the ‘beats’ heard when two sound waves of slightly different frequencies are superimposed. As a result of this, one finds that the original beam of ν_α neutrinos develops a ν_β component whose intensity oscillates as it travels through space, while the intensity of the ν_α neutrino beam itself is correspondingly reduced, i.e. ν_α neutrinos will ‘disappear’.

This effect follows from simple quantum mechanics. To illustrate this we will consider a ν_α neutrino produced with momentum \mathbf{p} at time $t = 0$. The initial state is therefore

$$|\nu_\alpha, \mathbf{p}\rangle = |\nu_i, \mathbf{p}\rangle \cos\theta_{ij} + |\nu_j, \mathbf{p}\rangle \sin\theta_{ij}, \quad (3.22)$$

where we use the notation $|P, \mathbf{p}\rangle$ to denote a state of a particle P having momentum \mathbf{p} . After time t this will become

$$a_i(t)|\nu_i, \mathbf{p}\rangle \cos\theta_{ij} + a_j(t)|\nu_j, \mathbf{p}\rangle \sin\theta_{ij}, \quad (3.23)$$

where

$$a_i(t) = e^{-iE_i t/\hbar} \quad \text{and} \quad a_j(t) = e^{-iE_j t/\hbar} \quad (3.24)$$

are the usual oscillating time factors associated with any quantum mechanical stationary state.⁶ For $t \neq 0$, the linear combination (3.22) does not correspond to a pure ν_α neutrino state, but can be written as a linear combination

$$A(t)|\nu_\alpha, \mathbf{p}\rangle + B(t)|\nu_\beta, \mathbf{p}\rangle \quad (3.25)$$

of ν_α and ν_β states, where the latter is

$$|\nu_\beta, \mathbf{p}\rangle = -|\nu_i, \mathbf{p}\rangle \sin\theta_{ij} + |\nu_j, \mathbf{p}\rangle \cos\theta_{ij}. \quad (3.26)$$

The functions $A(t)$ and $B(t)$ are found by solving (3.22) and (3.26) for $|\nu_i, \mathbf{p}\rangle$ and $|\nu_j, \mathbf{p}\rangle$, and then substituting the results into (3.23) and comparing with (3.25). This gives

$$A(t) = a_i(t) \cos^2\theta_{ij} + a_j(t) \sin^2\theta_{ij} \quad (3.27)$$

and

$$B(t) = \sin\theta_{ij} \cos\theta_{ij} [a_j(t) - a_i(t)]. \quad (3.28)$$

The probability of finding a ν_β state is therefore, using (3.24),

$$P(\nu_\alpha \rightarrow \nu_\beta) = |B(t)|^2 = \sin^2(2\theta_{ij}) \sin^2[(E_j - E_i)t/2\hbar] \quad (3.29)$$

and thus oscillates with time, while the probability of finding a ν_α neutrino is reduced by a corresponding oscillating factor. Irrespective of which

⁶See, for example, Chapter 1 of Mandl (1992).

neutrino states ν_α and ν_β are involved, the oscillations vanish if the mixing angle is zero or if the corresponding mass eigenstates $\nu_{i,j}$ have equal masses, and hence equal energies, as can be seen explicitly from (3.29). In particular, such oscillations are not possible if ν_i and ν_j both have zero masses.

In practice, the time t is determined by measuring the distance L that the neutrinos have travelled from their source, since their momenta are always very much greater than their possible masses and they travel extremely close to the speed of light. Hence, to a very good approximation, $t = L/c$ and

$$E_j - E_i = (m_j^2 c^4 + p^2 c^2)^{1/2} - (m_i^2 c^4 + p^2 c^2)^{1/2} \approx \frac{m_j^2 c^4 - m_i^2 c^4}{2pc}, \quad (3.30)$$

where $p = |\mathbf{p}|$, so that (3.29) may be written as

$$P(\nu_\alpha \rightarrow \nu_\beta) \approx \sin^2(2\theta_{ij}) \sin^2[L/L_0], \quad (3.31a)$$

where the ‘oscillation length’ L_0 is given by

$$L_0 = \frac{4E(\hbar c)}{(m_j^2 - m_i^2)c^4}, \quad (3.31b)$$

where $E = pc$ and

$$P(\nu_\alpha \rightarrow \nu_\alpha) = 1 - P(\nu_\alpha \rightarrow \nu_\beta). \quad (3.31c)$$

As we shall see, typical oscillation lengths range from a few hundred metres to a few hundred kilometres, so that neutrino oscillations can be safely neglected under most laboratory conditions. Nevertheless, neutrino oscillations have now been observed in many experiments.

The formulas derived above assume that the neutrinos are propagating in a vacuum. This is usually a very good approximation, because of the enormous mean free paths for neutrinos to interact with matter. However, if the neutrinos travel very long distances through dense matter, their propagation can be significantly modified by their interactions with the dense material they encounter, although the result that neutrino oscillations can only occur if neutrinos have nonzero masses remains unchanged. This is called the MSW effect, after Mikheyev, Smirnov, and Wolfenstein, who first discussed it. It arises from the fact that electron neutrinos interact with electrons in a different way from muon and tauon neutrinos (see Problem 3.3), and hence have different forward scattering amplitudes. When this is taken into account, the factors $\sin(2\theta_{ij})$ and $\Delta m_{ji}^2 \equiv m_j^2 - m_i^2$ that occur in the oscillation formulas (3.31) are modified as follows:⁷

$$\Delta m_{ij}^2 \rightarrow C \Delta m_{ij}^2 \text{ and } \sin(2\theta_{ij}) \rightarrow \sin(2\theta_{ij})/C, \quad (3.32a)$$

⁷For a derivation, see, for example, Chapter 7 of Giunti and Kim (2007).

where

$$C = \sqrt{[\cos(2\theta_{ij}) - A]^2 + \sin^2(2\theta_{ij})} \quad (3.32b)$$

and

$$A = \pm \frac{2\sqrt{2}G_F N_e E}{\Delta m_{ij}^2 c^4}. \quad (3.32c)$$

In these formulas, N_e is the electron density in matter, G_F is the Fermi weak interaction coupling, and the positive (negative) sign refers to neutrinos (antineutrinos). From them, we see that the magnitude of the effect depends upon the sign of A , and hence on whether neutrinos or antineutrinos are observed. The sign of A also depends on the sign of the squared mass difference. This is important, because on substituting (3.31b) into (3.31a) one finds that oscillations in vacuum depend only on the magnitude of the squared mass difference and not its sign. Hence matter effects, if they can be detected, can be used to determine the sign of the mass difference, which is otherwise undetermined. They can also, in special circumstances, be very large, because if A is positive and the electron density N_e and neutrino energy E are such that $\cos(2\theta_{ij}) - A = 0$, a resonant enhancement occurs, giving large effects even for small mixing angles. The MSW effect is an important factor when solar neutrino data are analysed, as we shall see in the following Section 3.1.5(c). However, in the other experiments discussed in Section 3.1.5, it is too small to be important.

3.1.5 Oscillation experiments

The experiments that have established the existence and properties of neutrino oscillations are conveniently classified into four main types, depending on the source of neutrinos: atmospheric neutrinos, neutrino beam experiments, solar neutrinos, and reactor experiments. We will consider each in turn.

3.1.5(a) Atmospheric neutrinos

The first experiment to produce definitive evidence for muon neutrino oscillations was that of a Japanese group in 1998 using the giant SuperKamiokande detector to study *atmospheric neutrinos* produced by the action of cosmic rays.⁸ The SuperKamiokande detector is shown in Figure 3.3. (Detectors will be discussed in detail in Chapter 4, so the description here will be brief.) It consists of a cylindrical stainless steel tank of roughly 40 m diameter and 40 m high, containing about

⁸Takaaki Kajita, from the SuperKamiokande collaboration, was awarded the 2015 Nobel Prize in Physics ‘for the discovery of neutrino oscillation, which shows that neutrinos have mass’. The prize was shared with Arthur McDonald, from the Sudbury Neutrino Observatory, for an experiment that showed that the so-called ‘solar neutrino problem’ was due to oscillations, as discussed in Section 3.1.5(c) below.

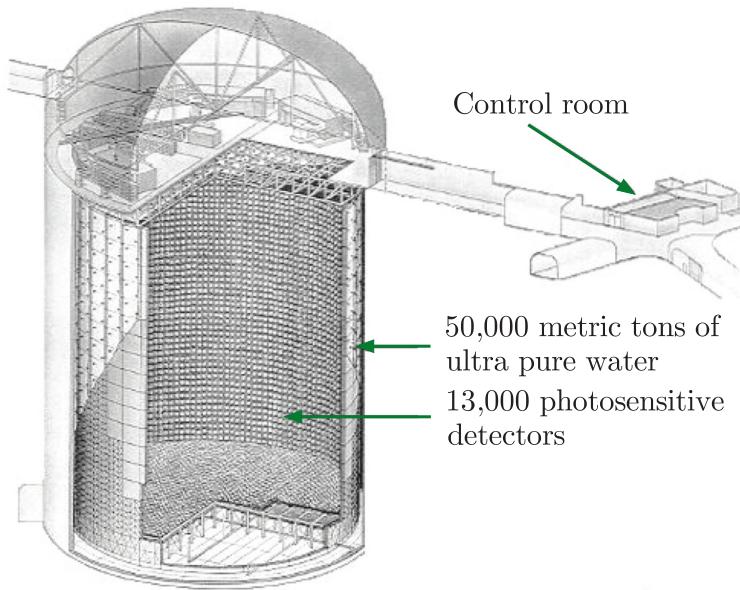


Figure 3.3 A schematic diagram of the SuperKamiokande detector.
Source: Adapted with permission from an original University of Hawaii, Manoa, illustration.

50 000 metric tons of very pure water. The detector is situated deep underground in the Japanese Alps, at a depth equivalent to 2700 m of water. This is to use the rocks above to shield the detector from cosmic ray muons. The volume is separated into inner and outer regions. The walls of the large inner region are lined with 11 200 light-sensitive devices called photomultipliers. When neutrinos with energies above 1 GeV interact with nuclei in the water, the velocities of the electrons and muons produced are greater than the speed of light in water. Because of this, a shock wave of light, called Čerenkov radiation, is emitted. (This is analogous to the shock wave emitted when an aircraft exceeds the speed of sound in air.) This Čerenkov radiation is detected by the photomultipliers and used to infer properties of the particles that produced it.⁹ The outer region of water acts as a shield against low-energy particles entering the detector from outside. An additional 1200 photomultipliers are located there to detect muons that enter or exit the detector.

When cosmic ray protons collide with atoms in the upper atmosphere they create many pions, which in turn create neutrinos mainly by the decay sequences

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu, \quad \pi^+ \rightarrow \mu^+ + \nu_\mu \quad (3.33)$$

and

$$\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu, \quad \mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu. \quad (3.34)$$

From this, one would naively expect to detect two muon neutrinos for every electron neutrino. However, the ratio was observed to be about 1.3 to 1 on average, suggesting that the muon neutrinos produced might

⁹Čerenkov radiation and other aspects of particle detection are discussed in more detail in Chapter 4.

be oscillating into other species. Confirmation for this was found by exploiting the fact that the detector could measure the direction of the detected neutrinos to study the azimuthal dependence of the effect. Since the flux of cosmic rays that led to neutrinos with energies above about 1 GeV is isotropic, the production rate for neutrinos should be the same all around the Earth. In particular, one can compare the measured flux from neutrinos produced in the atmosphere directly above the detector, which have a short flight path before detection, with those incident from directly below, which have travelled a long way through the Earth before detection, and so have had plenty of time to oscillate (perhaps several cycles). Experimentally, it was found that the yield of electron neutrinos from above and below were the same within a margin of error and consistent with expectation for no oscillations. However, while the yield of muon neutrinos from above accorded with the expectation for no significant oscillations, the flux of muon neutrinos from below was a factor of about two lower. This is direct evidence for muon neutrino oscillations.

In a later development of the experiment, the flux of muon neutrinos was measured as a function of L/E by estimating L from the reconstructed neutrino direction. Values of L range from 15 to 13 000 km. The results are shown in Figure 3.4 in the form of the ratio of $P(\nu_\mu \rightarrow \nu_\mu)$ of the observed number of events to the theoretical expectation if there were no oscillations. The data show clear evidence for a deviation of this ratio from unity, particularly at large values of L/E . Other experiments set limits on $P(\nu_\mu \rightarrow \nu_e)$ and taking these into account the most plausible hypothesis is that muon neutrinos are changing into tau neutrinos, which for the neutrino energies concerned could not be detected by

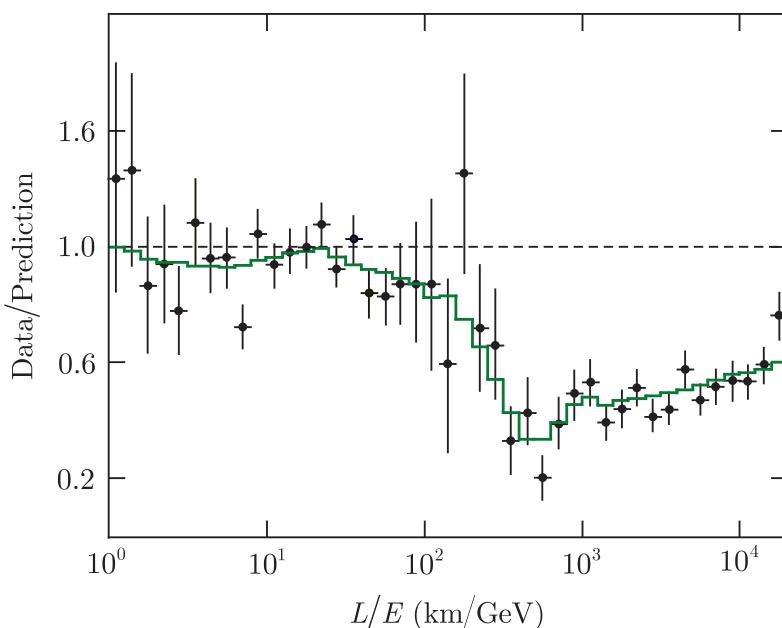


Figure 3.4 Data from the SuperKamiokande detector showing evidence for neutrino oscillations in atmospheric neutrinos. The green line is the fit from the model of two-neutrino mixing with parameters given by (2.32a) and (2.32b). Source: Adapted from Ashie et al. (2004) with permission. Copyright (2004) by the American Physical Society.

SuperKamiokande.¹⁰ As can be seen in Figure 3.4, the data are well-described by the two-state mixing formalism of the previous sections and if we define

$$\Delta(m_{ij}^2) \equiv m_i^2 - m_j^2,$$

then the experiment yields the values¹¹

$$1.9 \times 10^{-3} \lesssim |\Delta(m_{32}^2)| \lesssim 3.0 \times 10^{-3} \text{ (eV/c}^2\text{)}^2 \quad (3.35\text{a})$$

and

$$\sin^2(2\theta_{23}) \gtrsim 0.9, \quad \text{i.e. } 36^\circ \lesssim \theta_{23} \lesssim 54^\circ. \quad (3.35\text{b})$$

3.1.5(b) Neutrino beam experiments

The results obtained using atmospheric neutrinos have been confirmed by two experiments using muon neutrino beams produced at accelerators.¹² Since the mass difference (3.35a) implies oscillation lengths of order 400 km for neutrinos of energy 1 GeV, this requires the beam intensity to be measured by two detectors a considerable distance apart. In the T2K (Tokyo to Kamioka) experiment, a neutrino beam with a mean energy of 1.3 GeV produced at the KEK accelerator laboratory in Japan was directed to the SuperKamiokande detector 250 km away, with the near detector at 300 m from the source; in the MINOS experiment, a neutrino beam with a peak energy of 3 GeV produced at Fermi National Accelerator Laboratory (FNAL) near Chicago was directed through a near detector at the laboratory to a far detector in the Soudan mine in Minnesota about 735 km away. Both experiments confirmed the existence of neutrino oscillations, with parameters consistent with those obtained in SuperKamiokande, as illustrated in Figure 3.5.

3.1.5(c) Solar neutrinos

Further evidence for neutrino oscillations comes from our knowledge of the Sun. We shall see in Section 9.2.3 that the energy of the Sun is due to various nuclear reactions and these produce a huge flux of electron neutrinos that can be detected at the surface of the Earth. Since the astrophysics of the Sun and nuclear production processes are well understood, this flux can be calculated with some confidence by what is known as the ‘standard solar model’.¹³ However, the measured count rate is about a factor

¹⁰One such experiment, called the Opera Collaboration, located at the Gran Sasso National Laboratory in Italy, studies muon neutrino decays using a beam directed from the CERN laboratory in Geneva (see Figure 4.8). Opera is able to directly observe tau neutrinos and has provisionally reported five events of the type $\nu_\mu \rightarrow \nu_\tau$, which is consistent with theoretical expectations (Agafonova et al. 2015).

¹¹We label the two neutrinos involved in this experiment as ν_2 and ν_3 , rather than ν_1 and ν_2 , to conform with the convention used by the Particle Data Group (Tanabashi et al. 2018).

¹²The production of such beams is briefly discussed in Section 4.2.3.

¹³This model is discussed in, for example, Chapter 4 of Phillips (1994).

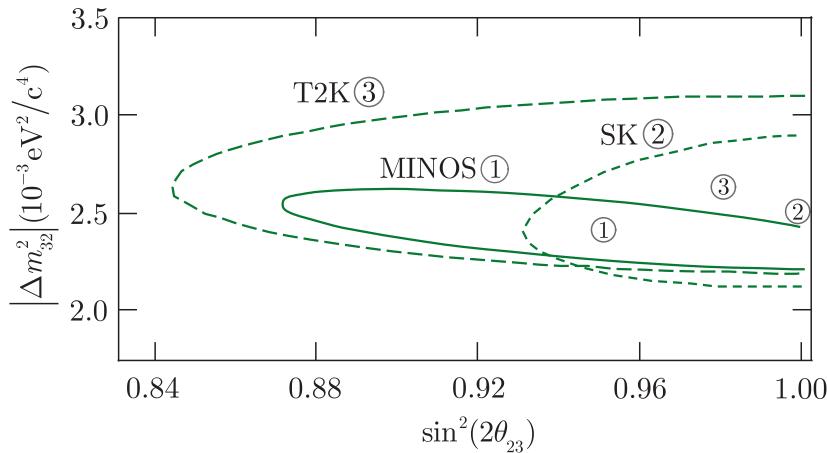


Figure 3.5 The allowed regions for the parameters $|\Delta m_{32}^2|$ and $\sin^2(2\theta_{23})$ at the 90% confidence level obtained by two-neutrino mixing fits to later atmospheric neutrino results from SuperKamiokande (SK), and neutrino data from T2K and MINOS. Source: Adapted from Abe et al. (2013). The smaller circled numbers correspond to the best-fit parameters from the respective experiments, identified by the same numbers.

of two lower than the theoretical expectation. This is the so-called *solar neutrino problem*. It was first investigated by Davis and co-workers in an experiment that was started in the late 1960s and studied the reaction



to detect the neutrinos. The principal component of this experiment is a huge tank containing tetrachloroethylene (C_2Cl_4) in which a single argon-37 atom is produced on average every few days by the reaction (3.36). These atoms are unstable, with a half-life of 35 days. They are extracted by flushing the tank with helium gas every few weeks and counted by observing their decays. The experiment was located deep underground in a gold mine in South Dakota, USA, to reduce the number of background events in which argon-37 atoms are produced by reactions involving cosmic ray muons rather than neutrinos. The inferred neutrino flux is expressed in terms of the ‘solar neutrino unit’ (SNU pronounced ‘snew’), defined as one capture event per second for every 10^{36} target atoms. After more than 20 years of running, the neutrino flux was measured to be $2.55 \pm 0.17 \pm 0.18$ SNU, where the first error is statistical and the second systematic. This measured rate is much smaller than the expected rate of 7.3 ± 2.3 SNU predicted by the standard solar model. The discrepancy between these two values constitutes the solar neutrino problem and was confirmed by a second experiment, called Kamiokande II, which was a smaller precursor of the SuperKamiokande detector described above. Unlike the ${}^{37}\text{Cl}$ experiment, the Kamiokande II experiment was sensitive to the direction of the incoming neutrinos, and verified that they came from the direction of the Sun.¹⁴

¹⁴Raymond Davis Jr and Masatoshi Koshiba (of the Kamiokande collaboration) were jointly awarded the 2002 Nobel Prize in Physics, principally for the detection of solar neutrinos.

Reaction (3.36) has a threshold energy of 0.81 MeV and is therefore only sensitive to relatively high-energy neutrinos from the Sun. Such neutrinos come predominantly from the β decay



where the neutrinos have an average energy ≈ 7 MeV. Although the neutrinos from (3.37) have been extensively studied, this decay contributes only about 10^{-4} of the total solar neutrino flux. It is therefore important to detect neutrinos from other reactions, in particular from the reaction

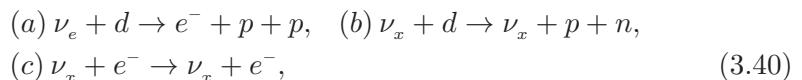


where d is a spin-1 S-wave bound state of a proton and a neutron, called the deuteron. This is the primary reaction that produces the energy of the Sun and contributes approximately 90% of the solar neutrino flux. (It is discussed in more detail in Section 9.2.3.) The neutrinos in this reaction have average energies of ≈ 0.26 MeV and cannot be detected by reaction (3.36). Instead, the reaction



has been used, which has a threshold energy of 0.23 MeV. (The experiments could also detect neutrinos from the solar reaction $e^- + {}^7\text{Be} \rightarrow {}^7\text{Li} + \nu_e$.) Just as for the original experiments of Davis et al., there were formidable problems in identifying the radioactive products from this reaction, which produced only about 1 atom of ${}^{71}\text{Ge}$ per day in a target of 30 tons of gallium. Nevertheless, results from these experiments, called SAGE and GALLEX, confirm the deficit of electron neutrinos and find between 60 and 70% of the flux expected in the absence of oscillations. Since the predicted neutrino flux arising from (3.38) is insensitive to the details of the standard solar model, it is very unlikely that shortcomings in the latter could be the source of this discrepancy.

The proof that neutrino oscillations are indeed the solution to the solar neutrino problem was definitively established by an experiment at the Sudbury Neutrino Observatory (SNO) in Canada in 2002. This experiment used a water Čerenkov detector, like Kamiokande and SuperKamiokande, but instead of normal water it used so-called ‘heavy water’, D_2O , where D is a deuterium atom. SNO was therefore able to study the reactions



where x denotes any lepton (e, μ, τ) and d is the deuteron. The cross-section for (b) is independent of the lepton type (this is a consequence of ‘lepton universality’ discussed in Section 3.1.2) and hence is independent of any possible oscillations. Since the observed flux was consistent with expectations, this confirmed the correctness of the solar model. On the

other hand, the observed flux from (a) was only about 1/3 of expectations, implying that about 2/3 of the electron neutrinos had transformed to μ and/or τ neutrinos before being detected at the surface of Earth. The flux for (c) would then be due to a mixture of approximately 1/3 of electron neutrinos and 2/3 of μ/τ neutrinos. Because the cross-section for $\nu_x e^- \rightarrow \nu_x e^-$ is different for ν_e and $\nu_{\mu,\tau}$ (see Problem 3.3), the expected rate for (c) is below what would be expected if there were no oscillations. The data were consistent with this assumption.

3.1.5(d) Reactor experiments

As we have mentioned in Section 3.1.3, nuclear reactors are a copious source of electron antineutrinos. These have typical energies of a few MeV, and when the solar neutrino results are analysed, taking into account their interactions with matter, they imply that a substantial fraction of the antineutrinos from a nuclear reactor would change to antineutrinos of other flavours after travelling a distance of order 100 km from their source. This prediction has been tested by the KamLAND group using a very large detector filled with liquid scintillator and surrounded by phototubes, which detect antineutrinos with energies above 1.8 MeV indirectly by detecting the scintillation light that is produced by positrons from the inverse beta decay reaction $\bar{\nu}_e + p \rightarrow e^+ + n$. (The KamLAND detector is described in more detail in Section 4.5.) The detector was used to measure the $\bar{\nu}_e$ flux arriving from more than 60 reactors in Japan and South Korea after the neutrinos had travelled distances of between 150 and 200 km to reach the Kamioka mine, where the detector is located. The first results, published in 2003, showed that the $\bar{\nu}_e$ flux was only about 60% of that expected from the known characteristics of the reactors, and subsequent data clearly showed the oscillatory data of the effect, as can be seen in Figure 3.6. A simultaneous analysis of the data shown in this figure together with the solar neutrino data, using two-component mixing, yields the results:

$$\Delta m_{21}^2 = 7.50_{-0.20}^{+0.19} \times 10^{-5} (\text{eV}/c^2)^2 \quad \text{and} \quad \tan^2 \theta_{12} = 0.444_{-0.030}^{+0.036}. \quad (3.41a)$$

In this analysis, it is essential to include matter effects (i.e. the MSW effect discussed earlier) in the description of the solar neutrino data, in order to obtain a consistent account. Because these are far from negligible, both the sign and magnitude of the squared mass difference are determined in this case.

From (3.35), and (3.41a), we see that the splitting between the squared masses of neutrinos ν_1 and ν_2 is very small compared to the corresponding splitting between neutrinos ν_2 and ν_3 . Hence

$$|\Delta m_{13}^2| \approx |\Delta m_{23}^2| \approx 2 \times 10^{-3} (\text{eV}/c^2)^2,$$

so that, for neutrinos of a given energy the oscillation length (3.31) between states 1 and 3 is much shorter than the oscillation length between

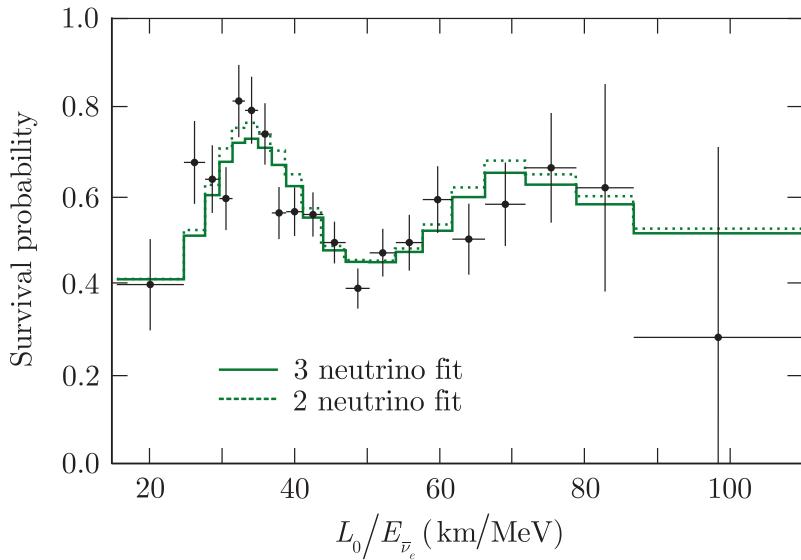


Figure 3.6 The ratio of the observed $\bar{\nu}_e$ spectrum to that predicted without oscillations (the ‘survival probability’) observed in the KamLAND experiment, showing the best fits obtained using two-neutrino and three-neutrino mixing models. Source: From Gando et al. (2011), reproduced with permission of the American Physical Society.

states 1 and 2. For reactor neutrinos, this gives rise to oscillation lengths of order kilometres rather than hundreds of kilometres. To investigate this, the KamLAND and solar neutrino data were re-analysed taking into account the possibility of mixing between all three neutrinos. The resulting fit is shown in Figure 3.6 and yields parameter values

$$\Delta m_{21}^2 = 7.50^{+0.19}_{-0.20} \times 10^{-5} (\text{eV}/c^2) \quad \text{and} \quad \tan^2 \theta_{12} = 0.452^{+0.035}_{-0.033}, \quad (3.41b)$$

which are essentially unchanged from those obtained using two-component mixing, together with a very small value

$$\sin^2 \theta_{13} = 0.020 \pm 0.016$$

for the third mixing angle, which is not convincingly different from zero.

The first clear evidence for nonzero oscillations between ν_1 and ν_3 was obtained in 2012 by three reactor experiments, with the detectors sufficiently close to the reactors for the contribution from the long baseline oscillations due to $\nu_1 - \nu_2$ mixing to be very small. These are: the Daya Bay collaboration, which observed neutrinos from six reactors at the Daya Bay nuclear power plant in China using six detectors distributed between three sites at distances of approximately 470, 580, and 1650 m; the RENO collaboration, which measured the neutrino fluxes from six reactors at the Yonggwang nuclear site in Korea using identical detectors at approximately 300 and 1400 m from the reactor centre; and the Double Chooz experiment in France, which measured the flux from two reactors with a detector approximately 1 km from the cores. The results for $\sin^2(2\theta_{13})$ obtained by the three experiments were

$0.089 \pm 0.010 \pm 0.005$ $0.100 \pm 0.010 \pm 0.012$ $0.109 \pm 0.030 \pm 0.025$	(Daya Bay, An et al. 2013), (RENO, Seo 2013), (Double Chooz, Abe et al. 2012),
---	--

where in each case the two errors are statistical and systematic, respectively. These results are in excellent agreement and clearly demonstrate the existence of oscillations between neutrino species 1 and 3 for the first time.

3.1.6 Neutrino masses and mixing angles

In order to extract reliable information from all the oscillation data it is necessary to extend the discussion of Section 3.1.4 to include mixing between all three neutrino mass states ν_1 , ν_2 , and ν_3 rather than just between two. We will not discuss this in detail,¹⁵ except to say that a global fit to the data from all oscillation experiments in principle yields unique values for the mixing angles and squared masses, although at present this is not possible with the available data.

To see this, we first reiterate that, for the solar neutrino data, the interactions with matter play an important role. As a bonus, this enables the sign of Δm_{21}^2 to be measured, whereas oscillations in free space only determine the magnitude of Δm_{ij}^2 , as can be seen from (3.31). Matter effects can also be used to determine the sign of Δm_{32}^2 . In muon neutrino oscillation experiments using neutrino beams generated at accelerators, like the MINOS and T2K experiments discussed in Section 3.1.5(b) above, the neutrinos travel large distances through the Earth's crust before reaching the detector, and the sign of Δm_{32}^2 can in principle also be determined when neutrino interactions with matter are taken into account. However, the effect is small and the experiments are not yet precise enough to determine the sign in this way.

Because of this, two solutions for the mass hierarchy are compatible with the current data: the so-called ‘normal’ hierarchy, $m_3 > m_2 > m_1$, and the ‘inverted’ hierarchy, $m_2 > m_1 > m_3$. This ambiguity also implies a further ambiguity depending of which quadrant θ_{23} lies in.

The current parameters resulting from a global fit to all the data are given in PDG (2018):

$$\begin{aligned}\sin^2(\theta_{12}) &= 0.307 \pm 0.013; & \Delta m_{21}^2 &= (7.53 \pm 0.18) \times 10^{-5} (eV/c^2)^2, \\ \sin^2(\theta_{13}) &= (2.12 \pm 0.08) \times 10^{-2};\end{aligned}\tag{3.42a}$$

Normal hierarchy:

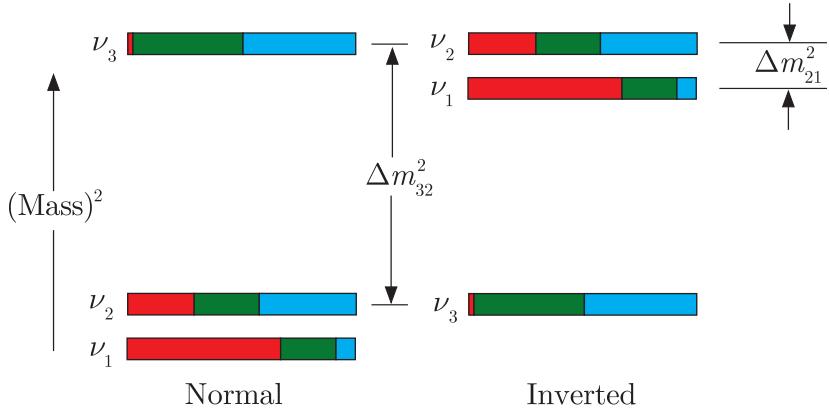
$$\sin^2(\theta_{23}) = \begin{cases} 0.417_{-0.028}^{+0.025} & 0 < \theta_{23} \leq \pi/2 \\ 0.597_{-0.030}^{+0.024} & \pi/2 < \theta_{23} \leq \pi \end{cases}; \quad \begin{aligned}\Delta m_{32}^2 &= (2.51 \pm 0.13) \\ &\times 10^{-3} (eV/c^2)^2,\end{aligned}\tag{3.42b}$$

¹⁵The general form of the mixing will be discussed in Section 7.5. As we shall see there, in addition to the three mixing angles, the mixing matrix contains a phase angle δ that can be nonzero if a symmetry called ‘CP invariance’, which is discussed in Chapter 7, is violated in neutrino oscillations. No such violation has, so far, been detected.

Inverted hierarchy:

$$\sin^2(\theta_{23}) = \begin{cases} 0.421^{+0.033}_{-0.025} & 0 < \theta_{23} \leq \pi/2 \\ 0.592^{+0.023}_{-0.030} & \pi/2 < \theta_{23} \leq \pi \end{cases}; \quad \Delta m_{32}^2 = -(2.56 \pm 0.0) \times 10^{-3} (\text{eV}/c^2)^2. \quad (3.42\text{c})$$

Figure 3.7 Three-neutrino squared-mass spectra in both the ‘normal’ and ‘inverted’ mass hierarchies that are consistent with the values for squared mass differences and mixing angles given in (3.35), (3.41), and (3.42). Also shown is the approximate fractional flavour compositions of ν_1 , ν_2 , and ν_3 : ν_e (red), ν_μ (green), ν_τ (blue).



Both mass hierarchies are illustrated qualitatively in Figure 3.7, where we also show the approximate flavour decomposition of the mass eigenstates resulting from three-component mixing with the parameters above. As can be seen, the lighter of the two states ν_1 and ν_2 that dominate solar neutrino oscillations is predominately an electron neutrino, while ν_3 has large ν_μ and ν_τ components, but a very small electron neutrino component.

We can now return to the interpretation of the bound (3.17) and consider its consequences for the ‘mass’ of the electron neutrino. The point here is that neutrinos with definite flavours, like the electron neutrino, are superpositions of the mass eigenstates ν_1 , ν_2 , and ν_3 and do not themselves have definite masses. Rather, in accord with the standard theory of measurement in quantum mechanics, a measurement of the mass of the electron neutrino can yield any one of the three values m_1 , m_2 , or m_3 . However, if $m_i^2 > m_j^2$, one can easily show that $(m_i - m_j)^2 < m_i^2 - m_j^2$. Hence, (3.35a) and (3.41) set upper limits on the neutrino mass differences that are approximately

$$m_2 - m_1 \lesssim 10^{-2} \text{ eV}/c^2, \quad |m_3 - m_2| \lesssim 5 \times 10^{-2} \text{ eV}/c^2. \quad (3.43)$$

In other words, the neutrino masses are almost equal compared to a mass scale of $1 \text{ eV}/c^2$. Hence it is safe to interpret (3.17) as implying

$$m_i \leq 2 \text{ eV}/c^2 \quad (3.44)$$

for all three neutrinos ν_1 , ν_2 , and ν_3 . This limit is very small compared to the mass of the lightest charged particle, the electron, and is a great improvement on the limits obtained from muon and tauon decay experiments given in Table 3.1.

Finally, it is worth remarking that it is possible to obtain bounds on neutrino masses from cosmology. The most exacting of these comes from using the current standard cosmological model to analyse the large-scale structure of the universe, and the current value latest bound is

$$\sum_{i=1,3} m_i < 0.67 \text{ eV}/c^2. \quad (3.45)$$

This is more stringent than (3.44), but unlike the latter is not a direct measurement of mass. We will return to these cosmological arguments briefly in Section 10.5.2(a).

3.1.7 Lepton numbers revisited

In the previous sections, we have assumed that lepton number conservation holds and can be used to identify the neutrino flavour emitted or absorbed in any weak reaction. However, lepton number violation could in principle be induced in such reactions by the existence of neutrino oscillations. An example is the decay

$$\tau^- \rightarrow \mu^- + \gamma, \quad (3.46)$$

which violates both L_μ and L_τ conservation. This could arise by the mechanism of Figure 3.8, in which the tau neutrino is emitted at the first vertex, then oscillates into a muon neutrino before being reabsorbed at the second vertex. Thus the question arises as to the validity of our assumption of lepton number conservation. To investigate this, we will make a very rough estimate of the branching ratio for this decay and this will show that in practice such effects are totally negligible in the standard model due to the short-range of the weak interaction and can indeed be safely ignored.

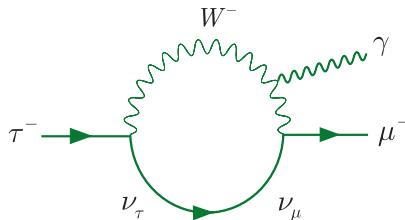


Figure 3.8 A Feynman diagram contributing to the decay $\tau^- \rightarrow \mu^- + \gamma$. There are two other diagrams, where the photon is emitted by either the μ^- meson or the τ^- .

Figure 3.8 involves two weak interaction vertices, the emission of a photon and an oscillation, while the leptonic decay modes (3.3b) are purely weak interactions. Therefore, the branching fraction of the decay (3.46) is expected to be of order-of-magnitude

$$B(\tau^- \rightarrow \mu^- + \gamma) = O[\alpha P(\nu_\tau \rightarrow \nu_\mu) B(\tau^- \rightarrow e^- \nu_\tau \bar{\nu}_e)], \quad (3.47a)$$

where $P(\nu_\tau \rightarrow \nu_\mu)$ is the probability of oscillation and α is the fine structure constant. Thus, using $B(\tau^- \rightarrow e^- \nu_\tau \bar{\nu}_e) \sim O(10^{-1})$, we have

$$B(\tau^- \rightarrow \mu^- + \gamma) = O[10^{-3} P(\nu_\tau \rightarrow \nu_\mu)]. \quad (3.47b)$$

The term $P(\nu_\tau \rightarrow \nu_\mu)$ can be estimated from the two-component mixing model of $\nu_\mu \leftrightarrow \nu_\tau$ oscillations used to describe the atmospheric neutrino data in Section 3.1.5(a). From (3.31a), we then have

$$P(\nu_\tau \rightarrow \nu_\mu) \approx \sin^2(2\theta_{23}) \sin^2(L/L_0) \approx (L/L_0)^2,$$

where we have used $\sin^2(2\theta_{23}) \gtrsim 0.9$ (see Figure 3.5) and assumed $L \ll L_0$. Here L is the typical distance travelled by the neutrino and is of the same order as the range $\hbar/M_W c$ of the weak interaction. The oscillation length is given by

$$L_0 = 4E(\hbar c)/(\Delta m_{32}^2 c^4),$$

where E is the typical energy of the neutrinos in Figure 3.8. Substituting these results into (3.47a) and (3.47b) gives

$$B(\tau^- \rightarrow \mu^- + \gamma) = \mathcal{O}\left[\frac{10^{-3}(m_3^2 - m_2^2)^2 c^4}{16E^2 M_W^2}\right]. \quad (3.47c)$$

For E we make the crude dimensional estimate $m_\mu c^2 \leq E \leq m_\tau c^2$ for a tauon decaying at rest. Finally, using the experimental value (3.43) for Δm_{32}^2 we have the rough estimate

$$B(\tau^- \rightarrow \mu^- + \gamma) \approx 10^{-48} - 10^{-50}. \quad (3.48)$$

This is utterly negligible for all practical purposes and should be compared to the experimental upper limit of $\sim 7 \times 10^{-8}$.

The above conclusion is confirmed by a more detailed treatment of both this and other reactions. Lepton number violation in weak interactions is completely negligible within the standard model, but is predicted to occur in some extensions of the model discussed in Section 10.3.

3.2 Quarks

We turn now to the strongly interacting particles – the quarks and their bound states, the hadrons. These also interact by the weak and electromagnetic interactions, although such effects can often be neglected compared to the strong interactions. To this extent we are entering the realm of ‘strong interaction physics’.

3.2.1 Evidence for quarks

Several hundred hadrons (not including nuclei) have been observed since pions were first produced in the laboratory in the early 1950s and all have zero or integer electric charges: $0, \pm 1$, or ± 2 in units of e . They are all bound states of the fundamental spin-1/2 quarks, whose electric charges are either $+2/3$ or $-1/3$, and/or antiquarks, with charges $-2/3$ or $+1/3$. The quarks themselves have never been directly observed as single, free

particles and, as remarked earlier, this fact initially made it difficult for quarks to be accepted as anything other than convenient mathematical quantities for performing calculations. Only later when the fundamental reason for this was realised (it will be discussed in Chapter 5) were quarks universally accepted as physical entities. However, there is also compelling experimental evidence for their existence. The evidence comes from three main areas: *hadron spectroscopy*, *lepton scattering* and *jet production*.

3.2.1(a) Hadron spectroscopy

This is the study of the static properties of hadrons: their masses, lifetimes, and decay modes, and especially the values of their quantum numbers, including spin, electric charge, and several more that we define in Section 3.2.2 below. As mentioned in Chapter 1, the existence and properties of quarks were first inferred from hadron spectroscopy by Gell-Mann and independently by Zweig in 1964 and the close correspondence between experimentally observed hadrons and those predicted by the quark model, which we will examine in more detail later, remains one of the strongest reasons for our belief in the existence of quarks.

3.2.1(b) Lepton scattering

It was mentioned in earlier chapters that in the early 1960s, experiments were first performed where electrons were scattered from protons and neutrons. These strongly suggested that nucleons were not elementary. By the late 1960s this work had been extended to higher energies and with projectiles that included muons and neutrinos. In much the same way as Rutherford deduced the existence of nuclei in atoms, high-energy lepton scattering, particularly at large momentum transfers, revealed the existence of point-like entities within the nucleons, which we now identify as quarks.

3.2.1(c) Jet production

High-energy collisions can cause the quarks within hadrons, or newly created quark–antiquark pairs, to fly apart from each other with very high energies. Before they can be observed, these quarks are converted into ‘jets’ of hadrons (a process referred to as *fragmentation*) whose production rates and angular distributions reflect those of the quarks from which they originated. They were first clearly identified in experiments at the DESY laboratory in Hamburg in 1979, where electrons and positrons were arranged to collide ‘head-on’ in a magnetic field. An example of a ‘two-jet’ event is shown in Figure 3.9.

The picture is a computer reconstruction of an end view along the beam direction; the solid lines indicate the reconstructed charged particle trajectories taking into account the known magnetic field, which is also parallel to the beam direction; the dotted lines indicate the reconstructed

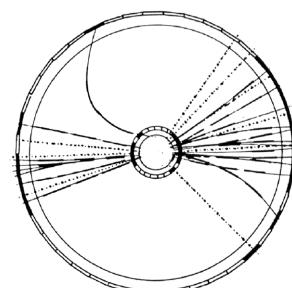


Figure 3.9 Computer reconstruction of a typical ‘two-jet’ event observed in the JADE jet chamber. The figure shows an end view along the beam direction. Source: After Orito (1979), Fermilab, with permission.

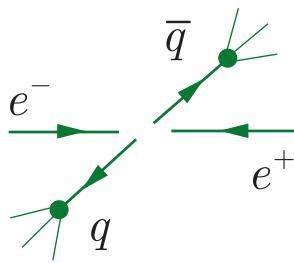


Figure 3.10 Mechanism for two-jet production in e^+e^- annihilation reaction.

trajectories of neutral particles, which were detected outside this device by other means. The production rate and angular distribution of the observed jets closely matches that of quarks produced in the reaction

$$e^+ + e^- \rightarrow q + \bar{q}, \quad (3.49)$$

by the mechanism of Figure 3.10. Such jets have now been observed in many reactions and are strong evidence for the existence of quarks within hadrons.

The failure to detect free quarks is not an experimental problem. Firstly, free quarks would be easily distinguished from other particles by their fractional charges and their resulting ionisation properties.¹⁶ Secondly, electric charge conservation implies that a fractionally charged particle cannot decay to a final state composed entirely of particles with integer electric charges. Hence the lightest fractionally charged particle, i.e. the lightest free quark, would be stable and so presumably easy to observe. Finally, some of the quarks are not very massive (see below) and because they interact by the strong interaction, one would expect free quarks to be copiously produced in, for example, high-energy proton–proton collisions. However, despite careful and exhaustive searches in ordinary matter, in cosmic rays and in high-energy collision products, free quarks have never been observed. The conclusion – that quarks exist solely within hadrons and not as isolated free particles – is called *confinement*. It is for this reason that we are forced to study the properties of hadrons, the bound states of quarks.

The modern theory of strong interactions, called *quantum chromodynamics (QCD)*, which is discussed in Chapter 5, offers at least a qualitative account of confinement, although much of the detail eludes us due to the difficulty of performing accurate calculations. In what follows, we shall assume confinement and use the properties of quarks to interpret the properties of hadrons.

3.2.2 Quark generations and quark numbers

Six distinct types, or *flavours*, of spin-1/2 quarks are now known to exist. Like the leptons, they occur in pairs, or *generations*, denoted

$$\begin{pmatrix} u \\ d \end{pmatrix}, \quad \begin{pmatrix} c \\ s \end{pmatrix}, \quad \begin{pmatrix} t \\ b \end{pmatrix}. \quad (3.50)$$

Each generation consists of a quark with charge +2/3 (*u*, *c*, or *t*), together with a quark of charge -1/3 (*d*, *s*, or *b*), in units of *e*. They are called the *down* (*d*), *up* (*u*), *strange* (*s*), *charmed* (*c*), *bottom* (*b*), and *top* (*t*) quarks. The quantum numbers associated with the *s*, *c*, *b*, and *t* quarks are called

¹⁶We will see in Chapter 4 that energy losses in matter due to ionisation are proportional to the square of the charge and thus would be ‘anomalously’ small for quarks.

Table 3.2 Properties of quarks. All have spin-1/2. The masses shown are the constituent quark masses, in units of GeV/c^2 . The antiparticles (not shown) have the same masses as their associated particles, but the electric charges Q are reversed in sign. In the major decay modes, X denotes other particles allowed by the appropriate conservation laws. Typical lifetimes are given for such decays

Name	Symbol	Mass	Q	Major decays	Lifetime (s)
down	d	$m_d \approx 0.3$	$-1/3$		
up	u	$m_u \approx m_d$	$2/3$		
strange	s	$m_s \approx 0.5$	$-1/3$	$s \rightarrow u + X$	$10^{-8} - 10^{-10}$
charm	c	$m_c \approx 1.5$	$2/3$	$c \rightarrow s + X$ $c \rightarrow d + X$	$10^{-12} - 10^{-13}$
bottom	b	$m_b \approx 4.5$	$-1/3$	$b \rightarrow c + X$	$10^{-12} - 10^{-13}$
top	t	$m_t \approx 173$	$2/3$	$t \rightarrow b + X$	$\sim 10^{-25}$

strangeness, charm, bottom, and top, respectively.¹⁷ The antiquarks are denoted

$$\begin{pmatrix} \bar{d} \\ \bar{u} \end{pmatrix}, \quad \begin{pmatrix} \bar{s} \\ \bar{c} \end{pmatrix}, \quad \begin{pmatrix} \bar{b} \\ \bar{t} \end{pmatrix}, \quad (3.51)$$

with charges $+1/3$ (\bar{d}, \bar{s} , or \bar{b}) and $-2/3$ (\bar{u}, \bar{c} , or \bar{t}). Approximate quark masses are given in Table 3.2. Except for the top quark, these masses are inferred indirectly from the observed masses of their hadron bound states, together with models of quark binding.¹⁸ Because of this, they are also referred to as *constituent* quark masses.

The stability of quarks in hadrons – like the stability of protons and neutrons in nuclei – is influenced by their interaction energies. However, for the s , c , and b quarks these effects are small enough for them to be assigned approximate lifetimes of $10^{-8} - 10^{-10}$ s for the s quark and $10^{-12} - 10^{-13}$ s for both the c and b quarks. The top quark is much heavier than the other quarks and its lifetime is of order 10^{-25} s. This lifetime is so short that, when top quarks are created, they decay too quickly to form observable hadrons. In contrast to the other quarks, our knowledge of the top quark is based entirely on observations of its decay products.

It is worth emphasising that when we talk about ‘the decay of quarks’ we always mean that the decay takes place within a hadron, with the other bound quarks acting as ‘spectators’, i.e. not taking part in the interaction. Thus, for example, in this picture neutron decay at the quark level is given by the Feynman diagram of Figure 3.11 and no free quarks are observed. Note that it is assumed that the exchanged particle interacts with only one constituent quark in the nucleons. This is the essence of the *spectator*

¹⁷The quantum numbers associated with the quark numbers ‘bottom’ and ‘top’ were originally called ‘beauty’ and ‘truth’, respectively, but the former names are now more commonly used.

¹⁸An analogy would be to deduce the mass of nucleons from the masses of nuclei via a model of the nucleus.

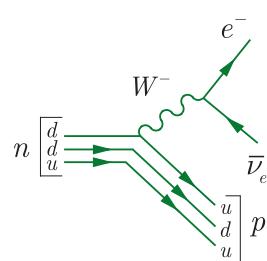


Figure 3.11 Quark Feynman diagram for the decay $n \rightarrow pe^- \bar{\nu}_e$ in the spectator model.

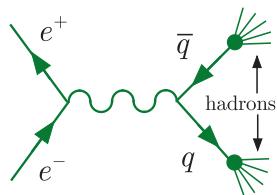


Figure 3.12 Production mechanism for the reaction $e^+e^- \rightarrow q\bar{q}$.

model. (This is similar to the idea of a single nucleon decaying within a radioactive nucleus.)

In strong and electromagnetic interactions, quarks can only be created or destroyed as particle–antiparticle pairs, just like electrons, as we discussed in Section 3.1.1. This implies, for example, that in electromagnetic processes corresponding to the Feynman diagram of Figure 3.12, the reaction $e^+ + e^- \rightarrow c + \bar{c}$, which creates a $c\bar{c}$ pair, is allowed, but the reaction $e^+ + e^- \rightarrow c + \bar{u}$, producing a $c\bar{u}$ pair, is forbidden.¹⁹ More generally, it implies conservation of each of the six *quark numbers*

$$N_f \equiv N(f) - N(\bar{f}) \quad (f = u, d, s, c, b, t), \quad (3.52)$$

where $N(f)$ is the number of quarks of flavour f present and $N(\bar{f})$ is the number of antiquarks of flavour \bar{f} present. For example, for single-particle states, $N_c = 1$ for the c quark, $N_{\bar{c}} = -1$ for the \bar{c} antiquark, and $N_e = 0$ for all other particles. Similar results apply for the other quark numbers N_f , and for multi-particle states the quark numbers of the individual particles are added. Thus a state containing the particles u, u, d has $N_u = 2$, $N_d = 1$, and $N_f = 0$ for the other quark numbers with $f = s, c, b, t$.

In weak interactions, more general possibilities are allowed, and only the total quark number

$$N_q \equiv N(q) - N(\bar{q}) \quad (3.53)$$

is conserved, where $N(q)$ and $N(\bar{q})$ are the total number of quarks and antiquarks present, irrespective of their flavour. This is illustrated by the decay modes of the quarks themselves, some of which are listed in Table 3.2, which are all weak interaction processes, and we have seen it also in the decay of the neutron in Figure 3.11. Another example is the main decay mode of the charmed quark, which is

$$c \rightarrow s + u + \bar{d}, \quad (3.54)$$

in which a c quark is replaced by an s quark and a u quark is created together with a \bar{d} antiquark. This clearly violates conservation of the individual quark numbers N_c, N_s, N_u , and N_d , but the total quark number N_q is conserved.

In practice, it is convenient to replace the total quark number N_q in analyses by the *baryon number*, defined by

$$B \equiv N_q/3 = [N(q) - N(\bar{q})]/3. \quad (3.55)$$

Like the electric charge and the lepton numbers introduced in the last section, the baryon number is conserved in *all known interactions*,

¹⁹Again, these reactions and associated Feynman diagrams do not imply that free quarks are created. Spectator quarks are implicitly present to form hadrons in the final state.

and unlike the lepton number, there are no experiments that suggest otherwise.²⁰

3.3 Hadrons

In principle, the properties of atoms and nuclei can be explained in terms of their proton, neutron, and electron constituents, although in practice many details are too complicated to be accurately calculated. However, the properties of these constituents can be determined without reference to atoms and nuclei by studying them directly as free particles in the laboratory. In this sense atomic and nuclear physics are no longer fundamental, although they are still very interesting and important if we want to understand the world we live in. In the case of hadrons, the situation is more complicated. Their properties are explained in terms of a few fundamental quark constituents, but the properties of the quarks themselves can only be studied experimentally by appropriate measurements on hadrons. Whether we like it or not, studying quarks without hadrons is not an option.

3.3.1 Flavour independence and charge multiplets

One of the fundamental properties of the strong interaction is *flavour independence*. This is the statement that the strong force between two quarks at a fixed distance apart is independent of which quark flavours u, d, s, c, b, t are involved. Thus, for example, the strong forces between us and ds pairs are identical. The same principle applies to quark–antiquark forces, which are, however, *not* identical to quark–quark forces, because in the former case annihilations can occur. Flavour independence does not apply to the electromagnetic interaction, since the quarks have different electric charges, but compared to the strong force between quarks, the electromagnetic force is a small correction. In addition, when applying flavour independence one must take proper account of the quark mass differences, which can be nontrivial. However, there are cases where these corrections are small or easily estimated, and the phenomenon of flavour independence is plain to see.

One consequence of flavour independence is the striking observation that hadrons occur in families of particles with approximately the same masses, called *charge multiplets*. Within a given family, all particles have the same spin-parity and the same baryon number, strangeness, charm, and beauty, but differ in their electric charges. Examples are the triplet of pions (π^+, π^0, π^-) and the nucleon doublet (p, n). This behaviour reflects an approximate symmetry between u and d quarks. It arises because, as

²⁰However, there are *theories* beyond the standard model that predict baryon number nonconservation, although there is no experimental evidence at present to support this prediction. These theories are discussed in Section 10.3.

we shall see in Section 3.3.4(b), these two quarks have only a very small mass difference

$$m_d - m_u = (3 \pm 1)\text{MeV}/c^2, \quad (3.56)$$

so that in this case mass corrections can to a good approximation be neglected. For example, consider the proton and neutron. We shall see in the next section that their quark content is $p(938) = uud$ and $n(940) = udd$. If we neglect the small mass difference between the u and d quarks and also the electromagnetic interactions, which is equivalent to setting all electric charges to zero, so that the forces acting on the u and d quarks are exactly equal, then replacing the u quark by a d quark in the proton would produce a ‘neutron’ that would be essentially identical to the proton. Of course, the symmetry is not exact because of the small mass difference between the u and d quarks and the presence of electromagnetic forces, and it is these that give rise to the small differences in mass within multiplets.

Flavour independence of the strong forces between u and d quarks also leads directly to the *charge independence of nuclear forces*, that is the equality of the force between any pair of nucleons, provided the two particles are in the same spin state. Subsumed in the idea of charge independence is the idea of *charge symmetry*, i.e. the equality of the proton–proton and neutron–neutron forces, again provided the two particles are in the same spin state. Evidence for the latter is found in studies of nuclei with the same value of A , but with the values of N and Z interchanged (called *mirror nuclei*). An example is shown in Figure 3.13. The two nuclei $^{11}_5\text{B}$ and $^{11}_6\text{C}$ have the same number of np pairs, but $^{11}_5\text{B}$ has 10 pp pairs and 15 nn pairs, whereas $^{11}_6\text{C}$ has 15 pp pairs and 10 nn pairs. Thus, allowing for the Coulomb interaction, the approximate equality of the level structures of these two nuclei, as seen in Figure 3.13, means charge symmetry is approximately verified. To test charge independence in a nuclear context we would have to look at the level structure in three related nuclei such as $^{11}_4\text{Be}$, $^{11}_5\text{B}$, and $^{11}_6\text{C}$. Here the test is not so clear-cut because an np pair is not subject to the restrictions of the Pauli principle like pp and nn pairs and there is evidence, as we shall see in Section 8.1, that the np force is stronger in the $S = 1$ state than in the $S = 0$ state. Nevertheless, the measured energy levels in such triplets of nuclei support the idea of approximate charge independence of nuclear forces.

The symmetry between u and d quarks is called *isospin symmetry*²¹ and greatly simplifies the interpretation of hadron physics. It is described by the same mathematics as ordinary spin, and hence the name. For example, the proton and neutron are viewed as the ‘up’ and ‘down’ components of a single particle, the nucleon N , that has an isospin quantum number $I = 1/2$, with I_3 values $1/2$ and $-1/2$, assigned to the proton and neutron, respectively, where I_3 is analogous to the magnetic quantum number in the case of ordinary spin. Likewise, the three pions

²¹Werner Heisenberg received the 1932 Nobel Prize in Physics for his contributions to the creation of quantum mechanics and the idea of isospin symmetry.

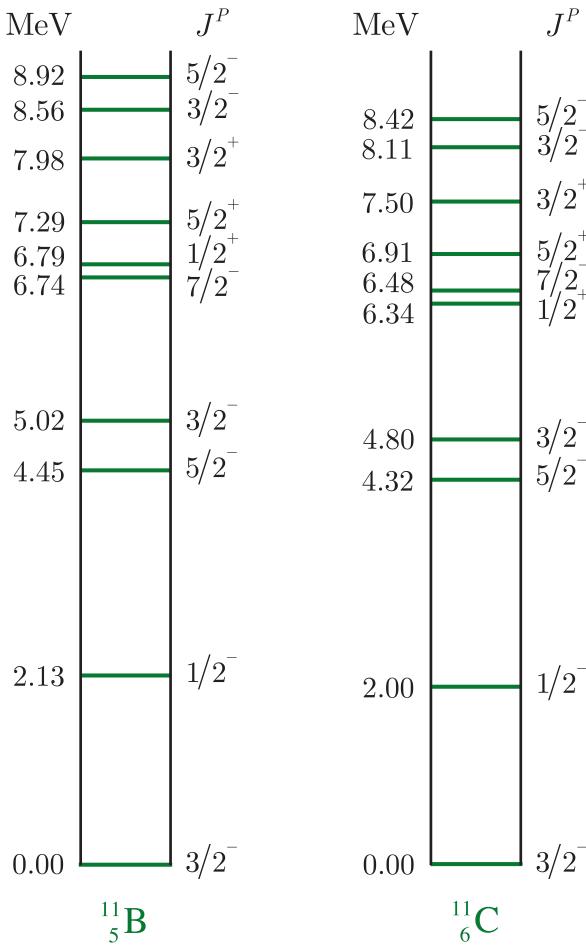


Figure 3.13 Low-lying energy levels with spin-parity J^P of the mirror nuclei $^{11}_5\text{B}$ and $^{11}_6\text{C}$. Source: Data from Ajzenberg-Selove (1990).

π^+ , π^- , and π^0 are part of a triplet π with $I = 1$ corresponding to I_3 values 1, 0, and -1 , respectively. In discussing the strong interactions between pions and nucleons, it is then only necessary to consider the πN interaction with a total isospin of either $1/2$ or $3/2$.

As an example, we will consider some predictions for the hadronic resonance state $\Delta(1232)$. The $\Delta(1232)$ has $I = 3/2$ and four charge states Δ^{++} , Δ^+ , Δ^0 , and Δ^- (see Table 3.3), corresponding to $I_3 = 3/2$, $1/2$, $-1/2$, $-3/2$, respectively. If we use the notation $|\pi N; I, I_3\rangle$ for a πN state, then $|\pi N; 3/2, 3/2\rangle$ is the unique state $\pi^+ p$ and may be written

$$|\pi N; 3/2, 3/2\rangle = |\pi; 1, 1\rangle |N; 1/2, 1/2\rangle. \quad (3.57)$$

The other πN states may then be obtained by applying quantum mechanical shift (ladder) operators to (3.57), as is done when constructing ordinary spin states.²² This gives

$$|\pi N; 3/2, 1/2.\rangle = -\sqrt{\frac{1}{3}}|\pi^+ n\rangle + \sqrt{\frac{2}{3}}|\pi^0 p\rangle \quad (3.58)$$

²²The mathematical formalism of isospin is described, for the interested reader, in Appendix A, Section A4.

Table 3.3 Some examples of baryons and mesons, with their major decay modes.
Masses are in MeV/c^2

Particle	Mass	Lifetime (s)	Major decays
$\pi^+(u\bar{d})$	139.6	2.60×10^{-8}	$\mu^+\nu_\mu$ (~100%)
$\pi^0(u\bar{u}, d\bar{d})$	135.0	8.52×10^{-17}	$\gamma\gamma$ (~100%)
$K^+(u\bar{s})$	493.7	1.24×10^{-8}	$\mu^+\nu_\mu$ (64%) $\pi^+\pi^0$ (21%)
$D^-(d\bar{c})$	1869.7	1.04×10^{-12}	Several seen
$B^-(b\bar{u})$	5279.3	1.64×10^{-12}	Several seen
$p(uud)$	938.3	Stable	None
$n(udd)$	939.6	880.2	$pe^-\bar{\nu}_e$ (100%)
$\Lambda(uds)$	1115.7	2.63×10^{-10}	$p\pi^-$ (64%) $n\pi^0$ (36%)
$\Xi^0(uss)$	1314.9	2.90×10^{-10}	$\Lambda\pi^0$ (99%)
$\Delta^{++}(uuu)$	~1232	$\sim 0.6 \times 10^{-23}$	$p\pi^+$ (100%)
$\Omega^-(sss)$	1672.5	0.82×10^{-10}	ΛK^- (68%) $\Xi^0\pi^-$ (24%)
$\Lambda_c^+(udc)$	2286.5	2.00×10^{-13}	Several seen

and hence isospin invariance predicts

$$\frac{\Gamma(\Delta^+ \rightarrow \pi^+ n)}{\Gamma(\Delta^+ \rightarrow \pi^0 p)} = \frac{1}{2}, \quad (3.59)$$

which is in good agreement with experiment.

Secondly, by constructing all the πN isospin states by analogy with (3.57) and (3.58) it follows that

$$|\pi^- p\rangle = \sqrt{\frac{1}{3}}|\pi N; 3/2, -1/2\rangle - \sqrt{\frac{2}{3}}|\pi N; 1/2, -1/2\rangle \quad (3.60a)$$

and

$$|\pi^0 n\rangle = \sqrt{\frac{2}{3}}|\pi N; 3/2, -1/2\rangle + \sqrt{\frac{1}{3}}|\pi N; 1/2, -1/2\rangle. \quad (3.60b)$$

Then, if \mathcal{M}_I is the amplitude for scattering in a pure isospin state I , isospin invariance gives

$$\mathcal{M}(\pi^- p \rightarrow \pi^- p) = \frac{1}{3}\mathcal{M}_{3/2} + \frac{2}{3}\mathcal{M}_{1/2} \quad (3.61a)$$

and

$$\mathcal{M}(\pi^- p \rightarrow \pi^0 n) = \frac{\sqrt{2}}{3}\mathcal{M}_{3/2} - \frac{\sqrt{2}}{3}\mathcal{M}_{1/2}. \quad (3.61b)$$

At the mass of the $\Delta(1232)$, the resonance dominates and the available energy is such that the total cross-section is dominated by the elastic ($\pi^- p \rightarrow \pi^- p$) and charge-exchange ($\pi^- p \rightarrow \pi^0 n$) reactions. In addition, because the $\Delta(1232)$ has $I = 3/2$, $\mathcal{M}_{3/2} \gg \mathcal{M}_{1/2}$, so

$$\sigma_{\text{total}}(\pi^- p) = \sigma(\pi^- p \rightarrow \pi^- p) + \sigma(\pi^- p \rightarrow \pi^0 n) \propto \frac{1}{3}|\mathcal{M}_{3/2}|^2 \quad (3.62a)$$

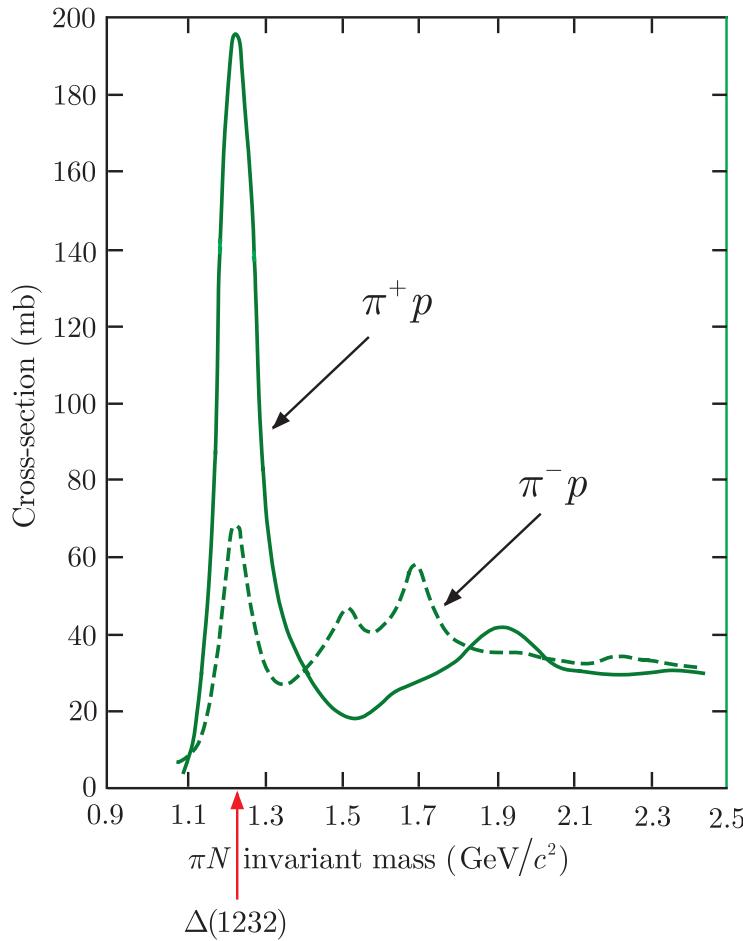


Figure 3.14 Total cross-sections for $\pi^- p$ and $\pi^+ p$ scattering.

and

$$\sigma_{\text{total}}(\pi^+ p) \propto |\mathcal{M}_{3/2}|^2. \quad (3.62b)$$

Thus, neglecting small kinematic corrections due to mass differences, isospin symmetry predicts that in the vicinity of the $\Delta(1232)$ state

$$\frac{\sigma_{\text{total}}(\pi^+ p)}{\sigma_{\text{total}}(\pi^- p)} = 3. \quad (3.63)$$

Figure 3.14 shows the two total cross-sections at low energies. There are clear peaks with Breit–Wigner forms at a mass of 1232 MeV corresponding to the production of the $\Delta(1232)$ and the ratio of the peaks is in good agreement with the prediction (3.63).

3.3.2 The simple quark model

The observed hadrons are of three types: *baryons* and their antiparticles *antibaryons*, which have half-integral spin, and *mesons*, which have integral spin. In the *simple quark model of hadrons* the baryons are assumed

to be bound states of three quarks (qqq), antibaryons are assumed to be bound states of three antiquarks ($\bar{q}\bar{q}\bar{q}$), and mesons are assumed to be bound states of a quark and an antiquark ($q\bar{q}$). The baryons and antibaryons have baryon numbers 1 and -1 respectively, while the mesons have baryon number 0. Hence the baryons and antibaryons can annihilate each other in reactions that conserve baryon numbers to give mesons or, more rarely, photons or lepton–antilepton pairs, in the final state.

The lightest known baryons are the proton and neutron, with the quark compositions:

$$p = uud, \quad n = udd. \quad (3.64)$$

These particles have been known to be constituents of atomic nuclei since the 1930s and their electric charges are assumed to be familiar. For other hadrons, it is conventional to use a superscript to indicate the electric charge in units of e , and we will often give the hadron masses in brackets in units of MeV/c^2 . Thus the pions are $\pi^\pm(140)$ and $\pi^0(135)$. Pions are the lightest known mesons and have the quark compositions

$$\pi^+ = u\bar{d}, \quad \pi^0 = u\bar{u}, d\bar{d}, \quad \pi^- = d\bar{u}. \quad (3.65)$$

While the charged pions have a unique composition, the neutral pion is composed of $u\bar{u}$ or $d\bar{d}$ pairs with equal probability. Pions are copiously produced in high-energy collisions by strong interaction processes such as $p + p \rightarrow p + n + \pi^+$.

The pions and nucleons are composed of u and d quarks only, which are therefore constituents of ‘ordinary’ matter, together with their antiquarks. In contrast, strange particles, as discussed briefly in Section 1.1.2, contain strange quarks s and/or their antiquarks \bar{s} , and are characterised by non-zero values of the strangeness quantum number S , which, apart from a sign, is just the strangeness quark number, i.e.

$$S = -N_s. \quad (3.66)$$

Kaons are the lightest strange mesons, with the quark compositions

$$K^+(494) = u\bar{s}, \quad K^0(498) = d\bar{s}, \quad (3.67)$$

where K^+ and K^0 have $S = +1$ and their antiparticles K^- and \bar{K}^0 have $S = -1$, while the lightest strange baryon is the *lambda*, with the quark composition $\Lambda(1116) = uds$. Subsequently, hadrons containing c and b quarks have also been discovered, with nonzero values of the *charm* and *bottom* quantum numbers defined by

$$C \equiv N_c \equiv N(c) - N(\bar{c}) \quad \text{and} \quad \tilde{B} \equiv -N_b \equiv -[N(b) - N(\bar{b})]. \quad (3.68)$$

The above examples illustrate just some of the many different combinations of quarks that form baryons or mesons. These and some further examples are shown in Table 3.3 and very extensive listings are given in the tables of the Particle Data Group (Tanabashi et al. 2018).

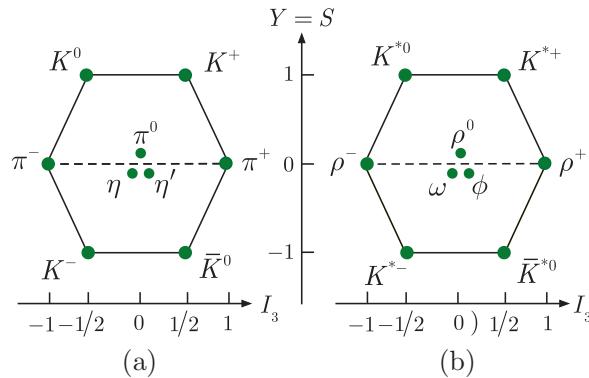


Figure 3.15 The lowest-lying states with (a) $J^P = 0^-$ and (b) $J^P = 1^-$ that are composed of u , d , and s quarks.

To proceed more systematically one could, for example, construct all the meson states of the form $q\bar{q}$, where q can be any of the six quark flavours. Each of these is labelled by its spin and its intrinsic parity P . The lightest such states have the spins of the quark and the antiquark antiparallel with no orbital angular momentum between them and so have spin-parity $J^P = 0^-$. (Recall from Section 1.3.1 that quarks and antiquarks have opposite parities.) If, for simplicity, we consider those states composed of just u , d , and s quarks, there will be nine such mesons with quantum numbers that may be identified with the mesons (K^0 , K^+), (\bar{K}^0 , K^-), (π^\pm , π^0), and two neutral particles, which are called η and η' . This resulting *supermultiplet* is shown in the form of a so-called ‘weight diagram’ in Figure 3.15a as a plot of Y , the *hypercharge*, defined as

$$Y \equiv B + S + C + \tilde{B} + T, \quad (3.69)$$

against I_3 , the third component of isospin. This can be extended to the case where the spins of the quark and the antiquark are parallel, but still with no orbital angular momentum between them, and so have spin-parity $J^P = 1^-$. The weight diagram for the resulting spectrum of states is shown in Figure 3.15b.

These ideas may be extended to baryons qqq , provided we assume that the combined space and spin wavefunctions are *symmetric* under the interchange of like quarks. At first sight this seems impossible, since quarks are spin-1/2 fermions and hence obey the Pauli exclusion principle. The explanation will emerge in Chapter 5, where we introduce the fundamental concept of ‘colour’. Here we shall simply assume that only overall symmetric space–spin wavefunctions are allowed, implying that like quark pairs with zero angular momentum must have parallel spins. When this is imposed, one easily shows that for baryons composed solely of light quarks u, d, s , the lowest-lying supermultiplets with zero orbital angular momentum consist of eight $J^P = 1/2^+$ and ten $J^P = 3/2^+$ baryons, corresponding precisely to the experimentally observed states shown in Figure 3.16a and b, respectively.²³

²³In contrast, if one assumes that like quarks in a state with zero angular momentum must have antiparallel spins a quite different spectrum is obtained (see Problem 3.9).

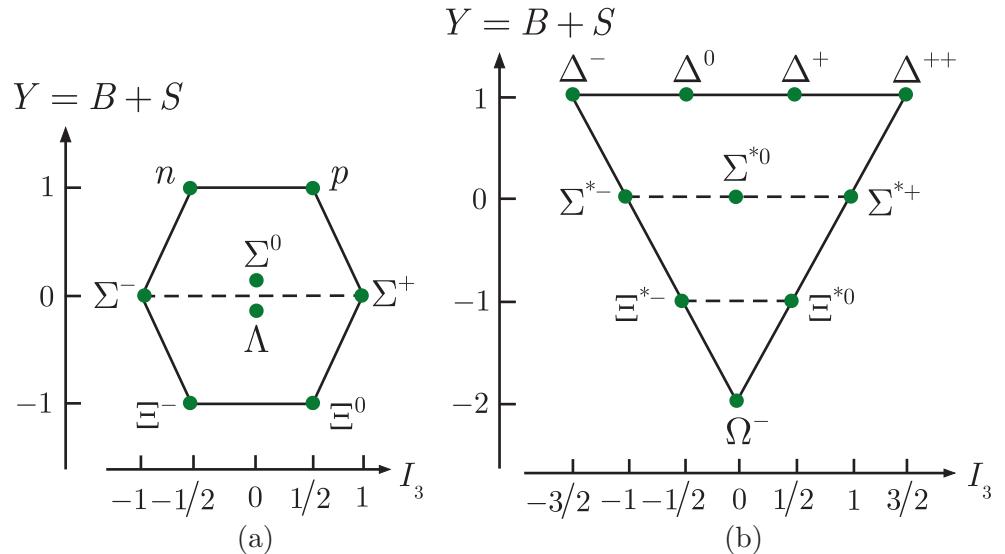


Figure 3.16 The lowest-lying states with (a) $J^P = 1/2^+$ and (b) $J = 3/2^+$ that are composed of u , d , and s quarks.

The scheme may also be extended to include c and b quarks, although the predicted states are more difficult to observe experimentally and have not yet all been confirmed. This is illustrated in Figure 3.17, which shows the predicted $J^P = 3/2^+$ baryon states formed from u , d , s , and c quarks when all three quarks have their spin aligned, but still with zero orbital angular momentum between them. The bottom plane corresponds to states with zero charm and contains the same 10 states shown in Figure 3.16b, while the second plane corresponds to the six predicted states with charm $C = +1$. The latter have also all been observed experimentally and, because these contain a charmed quark, are typically about

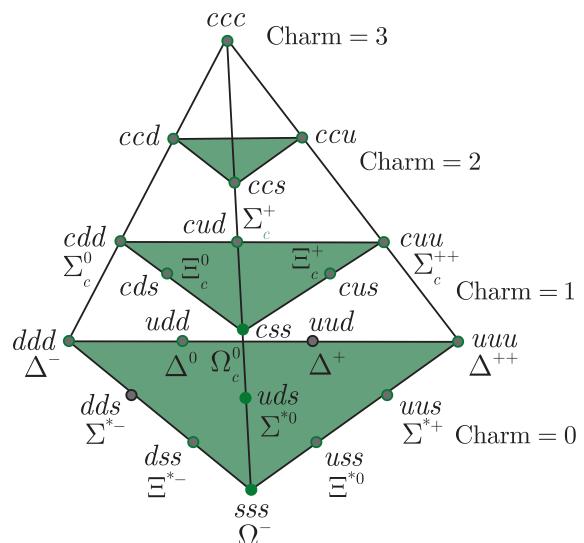


Figure 3.17 The $J = 3/2^+$ baryon states composed of u , d , s , and c quarks.

$1.2 \text{ GeV}/c^2$ heavier than the corresponding $C = 0$ states. The even heavier predicted $J^P = 3/2^+$ states containing two or three charmed quarks are yet to be observed.

So far we have confined ourselves to the lightest hadron states, which are the simplest to discuss and the easiest to observe. However, for any given quark combination, many other heavier states are expected to exist, and many of them have been detected experimentally. For example, if we consider the combination $u\bar{s}$, the lightest hadron is the $K^+(492)$ with $J^P = 0^-$, which can be regarded as the ‘ground state’ of the $u\bar{s}$ system. The next lightest, the ‘first excited state’, is the $K^{*+}(892)$ meson with $J^P = 1^-$, which is the $u\bar{s}$ state shown in Figure 3.15b. The next excited state is the $K^{*+}(1270)$, which is a P -wave state with quark orbital angular momentum $L = 1$, and at least nine other heavier excited states have been detected. These excited states, including the $K^{*+}(890)$, are resonances that decay by the strong interaction, with very short lifetimes of order 10^{-23} s . The mass distribution of their decay products is described by the Breit–Wigner formula discussed in Section 1.6.4, while the spin-parity of a resonance may be found from an analysis of the angular distributions of its decay products, using angular momentum and parity conservation. Similar behaviour is found for other quark systems and it is part of the triumph of the quark model that it successfully accounts for excited states of the various quark systems, as well as their ground states, when the internal motion of the quarks is taken into account, using reasonable models of the interquark potentials.²⁴

3.3.3 Hadron decays and lifetimes

From experiments such as electron scattering we know that hadrons have typical effective radii r of order 1 fm and hence associated timescales r/c of order 10^{-23} s . The vast majority are highly unstable resonances, corresponding to excited states of the various quark systems, which decay to lighter hadrons by the strong interaction, with lifetimes of this order. The $K^{*+}(892) = u\bar{s}$ resonance, mentioned above, is an example. It decays to $K^+\pi^0$ and $K^0\pi^+$ final states with a lifetime of $1.3 \times 10^{-23} \text{ s}$. The quark description of the process $K^{*+} \rightarrow K^0 + \pi^+$ is

$$u\bar{s} \rightarrow d\bar{s} + u\bar{d}.$$

From this we see that the final state contains the same quarks as the initial state, plus an additional $d\bar{d}$ pair, so that the quark numbers N_u and N_d are separately conserved. This is characteristic of strong and electromagnetic processes, which are only allowed if each of the quark numbers N_u , N_d , N_s , N_c , and N_b is separately conserved. The quark diagram for this decay is shown in Figure 3.18a.

²⁴This will be discussed explicitly, for the two particular quark systems $c\bar{c}$ and $b\bar{b}$, in Section 3.3.5 below.

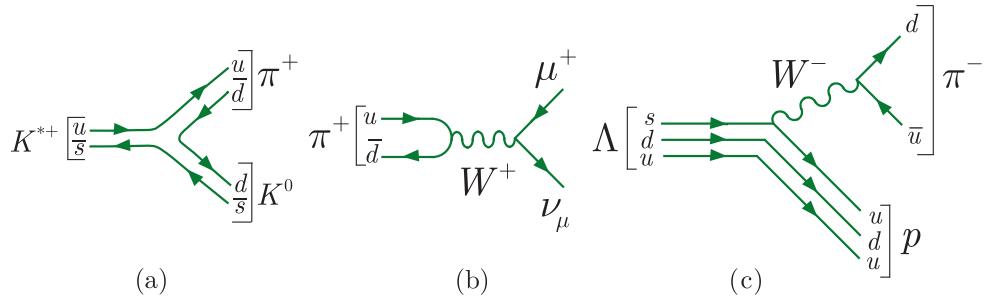


Figure 3.18 Feynman diagrams for the decays (a) $K^{*+} \rightarrow K^0 + \pi^+$, (b) $\pi^+ \rightarrow \mu^+ + \nu_\mu$, and (c) $\Lambda \rightarrow p + \pi^-$.

Since leptons and photons do not have strong interactions, hadrons can only decay by the strong interaction if lighter states composed solely of other hadrons exist with the same quantum numbers. While this is possible for the majority of hadrons, it is not usually possible for the lightest state corresponding to any given quark combination. These hadrons, which cannot decay by strong interactions, are long-lived on a timescale of order 10^{-23} s and are often called *stable particles*. It is more accurate to call them *long-lived particles*, because except for the proton they are not absolutely stable, but decay by either the electromagnetic or weak interaction.

The proton is stable because it is the lightest particle with a nonzero baryon number and the baryon number is conserved in all known interactions. A few of the other long-lived hadrons decay by electromagnetic interactions to final states that include photons. These decays, like the strong interaction, conserve all the individual quark numbers. An example is the neutral pion, which has $N_u = N_d = N_s = N_c = N_b = 0$ and decays by the reaction

$$\pi^0(u\bar{u}, d\bar{d}) \rightarrow \gamma + \gamma, \quad (3.70)$$

with a lifetime of 0.8×10^{-16} s. However, most of the long-lived hadrons have nonzero values for at least one of the quark numbers and can only decay by the weak interaction, in which individual quark numbers do not have to be conserved. For example, the positive pion decays with a lifetime of 2.6×10^{-8} s by the reaction

$$\pi^+ \rightarrow \mu^+ + \nu_\mu, \quad (3.71)$$

while the $\Lambda(1116) = uds$ baryon decays mainly by the reactions

$$\Lambda \rightarrow p + \pi^-, \quad n + \pi^0, \quad (3.72)$$

with a lifetime of 2.6×10^{-10} s. The quark interpretations of these reactions are

$$u\bar{d} \rightarrow \mu^+ + \nu_\mu, \quad (3.73)$$

in which a u -quark annihilates with a \bar{d} -antiquark, violating both N_u and N_d conservation, and for lambda decay to a charged pion,

$$sud \rightarrow uud + d\bar{u}, \quad (3.74)$$

in which an s quark turns into a u quark and a $d\bar{u}$ pair is created, violating N_d and N_s conservation. The lowest-order Feynman diagrams corresponding to these two reactions are shown in Figure 3.18b and c, respectively.

We see from the above that the strong, electromagnetic, or weak nature of a given hadron decay can be determined by inspecting quark numbers. The resulting lifetimes can then be summarised as follows. Strong decays lead to lifetimes that are typically of order 10^{-23} s. Electromagnetic decay rates are suppressed by powers of the fine structure constant α relative to strong decays, leading to observed lifetimes in the range $10^{-16} - 10^{-21}$ s. Weak decays give longer lifetimes, which depend sensitively on the characteristic energy of the decay. A useful measure of the decay energy is the Q value, the kinetic energy released in the decay of the particle at rest, which we have met before in Section 2.4. In the weak interactions of hadrons, Q values of order $(10^2 - 10^3)$ MeV are typical, leading to lifetimes in the range $10^{-7} - 10^{-13}$ s, but there are some exceptions, notably neutron decay, $n \rightarrow p + e^- + \bar{\nu}_e$, for which

$$Q = m_n - m_p - m_e - m_{\bar{\nu}_e} \approx 0.8 \text{ MeV}. \quad (3.75)$$

This is unusually small, leading to a lifetime of about 10^3 s. Thus hadron decay lifetimes are reasonably well understood and span some 27 orders of magnitude, from about 10^{-24} s to about 10^3 s. The typical ranges corresponding to each interaction are summarised in Table 3.4.

3.3.4 Hadron magnetic moments and masses

The quark model can make predictions for hadronic magnetic moments and masses in a way that is analogous to the semi-empirical mass formula for nuclear masses, i.e. the formulas have a theoretical basis, but contain parameters that have to be determined from experiment. We will examine both cases.

Table 3.4 Typical lifetimes of hadrons decaying by the three interactions

Interaction	Lifetimes (s)
Strong	$10^{-22} - 10^{-24}$
Electromagnetic	$10^{-16} - 10^{-21}$
Weak	$10^{-7} - 10^{-13}$

3.3.4(a) Magnetic moments

We will consider the magnetic moments of the $1/2^+$ octet of baryons, which have all been measured except that of the very short-lived $\Sigma^0(1193)$. If the quarks have zero orbital angular momenta, then these magnetic moments are just the sums of contributions from the quark magnetic

moments, which we assume to be of the Dirac form $\mu = (q/m) \mathbf{S}$ given in Section 1.2. That is, we assume

$$\mu_q \equiv \langle q, S_z = 1/2 | \hat{\mu}_z | q, S_z = 1/2 \rangle = e_q e \hbar / 2m_q = (e_q M_p / m_q) \mu_N. \quad (3.76)$$

where e_q is the quark charge in units of e and

$$\mu_N \equiv e \hbar / 2M_p \quad (3.77)$$

is the nuclear magneton.

We start with those $1/2^+$ baryons B with the quark configuration aab . Since they have no orbital angular momentum, the aa pair must be in the symmetric spin-1 state, as discussed in Section 3.3.2, with parallel spins and magnetic moment $2\mu_a$. The spin-up baryon state is given by²⁵

$$\begin{aligned} |B; S = 1/2, S_z = 1/2\rangle &= \sqrt{\frac{2}{3}} |b; S = 1/2, S_z = -1/2\rangle |aa; S = 1, S_z = 1\rangle \\ &\quad - \sqrt{\frac{1}{3}} |b; S = 1/2, S_z = 1/2\rangle |aa; S = 1, S_z = 0\rangle. \end{aligned} \quad (3.78)$$

The first term corresponds to a state with magnetic moment $2\mu_a - \mu_b$, since the b quark has $S_z = -1/2$; the second term corresponds to a state with magnetic moment μ_b , since the aa pair has $S_z = 0$ and does not contribute. Hence the baryon magnetic moment is given by

$$\mu_B = \frac{2}{3}(2\mu_a - \mu_b) + \frac{1}{3}\mu_b = \frac{4}{3}\mu_a - \frac{1}{3}\mu_b. \quad (3.79)$$

In particular, the proton magnetic moment is given by

$$\mu_p = \frac{4}{3}\mu_u - \frac{1}{3}\mu_d = \frac{M_p}{m_{u,d}}\mu_N, \quad (3.80a)$$

where we have neglected the small mass difference between the u and d quarks.

The remaining states in the octet are $\Lambda(1116)$ and $\Sigma^0(1193)$, which both have the quark composition uds . The ud pair in the $\Sigma^0(1193)$ must have parallel spins, because it belongs to the same isospin multiplet as the $\Sigma^+(1193) = uus$ and is obtained from it by replacing a u quark with a d quark, with the same wavefunction. Hence in the $\Lambda(1116)$, the ud pair must be in a spin-0 state and consequently makes no contribution to the Λ spin or magnetic moment, so that

$$\mu_\Lambda = \mu_s = -\frac{1}{3} \frac{M_p}{m_s} \mu_N, \quad (3.80b)$$

where we have used (3.76). The predictions (3.80a) and (3.80b) agree with the measured values of the p and Λ magnetic moments for the constituent quark masses

$$m_s = 510 \text{ MeV}/c^2, \quad m_{u,d} = 336 \text{ MeV}/c^2. \quad (3.81)$$

²⁵See Section A.4 of Appendix A.

Substituting these values into (3.76) and (3.79) and using the quark assignments of Table 3.5 leads to predictions for the other magnetic moments, which are compared with experiment in Table 3.5.

Table 3.5 A comparison of the observed magnetic moments of the $1/2^+$ baryon octet and the predictions of the simple quark model, (3.80a) and (3.80b), for $m_u = m_d = 336 \text{ MeV}/c^2$, $m_s = 510 \text{ MeV}/c^2$. Particle masses are in (MeV/c^2) .

Particle	Mass	Prediction (μ_N)	Experiment (μ_N)
$p(uud)$	938	2.79	2.793^a
$n(udd)$	940	-1.86	-1.913^a
$\Lambda(uds)$	1116	-0.61	-0.613 ± 0.004
$\Sigma^+(uus)$	1189	2.69	2.458 ± 0.010
$\Sigma^-(dds)$	1197	-1.04	-1.160 ± 0.025
$\Xi^0(uss)$	1315	-1.44	-1.250 ± 0.014
$\Xi^-(dss)$	1322	-0.51	-0.6507 ± 0.0025

^aThe errors on the proton and neutron magnetic moments are of the order 6×10^{-8} and 5×10^{-7} , respectively.

The agreement is reasonable, but not exact. This suggests that while the assumption that the observed baryons are three-quark states with zero orbital angular momentum is a good approximation, there could be a small admixture of other states with the same conserved quantum numbers. One possibility is suggested by considering the magnetic moment of the deuteron. Since it is the ground state of the np system, it is reasonable to assume that $L = 0$, so that the deuteron's magnetic moment arises purely from the spin magnetic moments of its constituent nucleons, and since $J = 1$ the neutron and proton spins must be parallel. The magnetic moment should therefore be given by

$$\mu_d = \mu_p + \mu_n = 2.793 - 1.913 = 0.880\mu_N,$$

whereas the experimental value is $0.857\mu_N$. The difference is accounted for if the deuteron is not a pure 3S_1 state, but contains a small admixture of the 3D_1 state with $L = 2$. This illustrates the fact that orbital angular momentum (as opposed to total angular momentum) is only approximately conserved and the ground states of strongly interacting systems are not in general pure S -waves, but may contain small admixtures of states with nonzero angular momenta. Such admixtures are neglected in the simple quark model, but may well occur. A second possibility is that in addition to the dominant qqq state, there may be small admixtures of states with additional quarks, for example $qqqq\bar{q}$. Such states are not forbidden by any physical principles, but are neglected in the simple quark model on grounds of simplicity. We will return to this in Sections 3.3.6 and 5.3.1.

3.3.4(b) Masses

We now turn to the prediction of hadron masses. The mass differences between members of a given supermultiplet are conveniently separated

into the small mass differences between members of the same isospin multiplet and the much larger mass differences between members of different isospin multiplets. If we neglect the former, then a first approximation would be to assume that the mass differences within a given supermultiplet are due solely to differences in the constituent quark masses. If we concentrate on hadrons with quark structures composed of u , d , and s quarks, since their masses are the best known from experiment, this assumption leads directly to the relations

$$M_{\Xi} - M_{\Sigma} = M_{\Xi} - M_{\Lambda} = M_{\Lambda} - M_N = m_s - m_{u,d} \quad (3.82)$$

for the $1/2^+$ baryon octet and

$$M_{\Omega} - M_{\Xi^*} = M_{\Xi^*} - M_{\Sigma^*} = M_{\Sigma^*} - M_{\Delta} = m_s - m_{u,d} \quad (3.83)$$

for the $3/2^+$ decuplet. These give numerical estimates for $m_s - m_{u,d}$ in the range 120–200 MeV/c², which is consistent with the estimate from magnetic moments above.

These results support the suggestion that baryon mass differences (and by analogy meson mass differences) are dominantly due to the mass differences of their constituent quarks. However, this cannot be the complete explanation, because if it were then the $1/2^+$ nucleon would have the same mass as the $3/2^+$ $\Delta(1232)$, as they have the same quark constituents, and similarly for other related particles in the $1/2^+$ octet and $3/2^+$ decuplet. The absence of orbital angular momentum in these states means that there is nothing equivalent to the ‘fine structure’ of atomic physics. The difference lies in the spin structures of these states.

If we take the case of two spin-1/2 particles with magnetic moments μ_i and μ_j separated by a distance r_{ij} , then the interaction energy is proportional to $\mu_i \cdot \mu_j / r_{ij}^3$. If, in addition, the particles are point-like and have charges e_i and e_j , the moments will be of the Dirac form $\mu_i = (e_i/m_i)\mathbf{S}_i$. Then for two particles in a relative S -state it can be shown that the interaction energy is given by²⁶

$$\Delta E = -\frac{2\pi}{3} \frac{e_i e_j}{m_i m_j} |\psi(0)|^2 \mathbf{S}_i \cdot \mathbf{S}_j, \quad (3.84)$$

where $\psi(0)$ is the wavefunction at the origin, $r_{ij} = 0$. (When averaged over all space, the interaction is zero except at the origin.) In atomic physics this is known as the *hyperfine interaction* and causes very small splittings in atomic energy levels. In the hadron case, the electric charges must be replaced by their strong interaction equivalents, with appropriate changes to the overall numerical factor. The resulting interaction is called (for reasons that will be clear in Chapter 5) the *chromomagnetic interaction*. As we cannot calculate the equivalent quark–quark wavefunction, for the

²⁶See, for example, p. 307 of Bjorken and Drell (1964).

purposes of a phenomenological analysis we will write the contribution to a hadron mass as

$$\Delta M = \left(\frac{a}{\hbar^2} \right) \frac{\mathbf{S}_1 \cdot \mathbf{S}_2}{m_1 m_2}, \quad (3.85)$$

where a is a constant to be determined from experiment. This assumes that $|\psi(0)|^2$ is the same for all states, which will not be exactly true.

Consider, firstly, the case of mesons. By writing the total spin squared as

$$\mathbf{S}^2 \equiv (\mathbf{S}_1 + \mathbf{S}_2)^2 = \mathbf{S}_1^2 + \mathbf{S}_2^2 + 2\mathbf{S}_1 \cdot \mathbf{S}_2, \quad (3.86)$$

we easily find that the expectation values of $\mathbf{S}_1 \cdot \mathbf{S}_2$ are $-3\hbar^2/4$ for the $S=0$ mesons and $\hbar^2/4$ for the $S=1$ mesons. The masses may be written

$$M(\text{meson}) = m_1 + m_2 + \Delta M, \quad (3.87)$$

where $m_{1,2}$ are the masses of the constituent quarks and

$$\Delta M(J^P = 0^- \text{ meson}) = -\frac{3a}{4} \frac{1}{m_1 m_2}, \quad \Delta M(J^P = 1^- \text{ meson}) = \frac{a}{4} \frac{1}{m_1 m_2}. \quad (3.88)$$

The masses of the members of the 0^- and 1^- meson supermultiplets then follow from a knowledge of their quark compositions. For example, the K -mesons have one u or d quark and one s quark and so

$$M_K = m + m_s - \frac{3a}{4mm_s}, \quad (3.89)$$

where we write $m = m_{u,d}$ from now on. The corresponding equations for the other masses are given in Table 3.6. The quark masses can then be inferred from the two relations

$$m_\pi + 3m_\rho = 8m, \quad m_K + 3m_{K^*} = 4(m + m_s),$$

which follow from these equations, and the parameter a from the resulting masses, together with the mass of the rho meson. The results are

$$m = 0.306 \text{ GeV}/c^2, \quad m_s = 0.487 \text{ GeV}/c^2, \quad a = 0.0592 (\text{GeV}/c^2)^3, \quad (3.90)$$

and the predicted masses for other mesons are in good agreement with experiment, as is also shown in Table 3.6. Note that the quark mass values are similar to, but somewhat smaller than, those obtained previously from the baryon magnetic moments, reflecting the different approximations made. In addition, in Table 3.6 we have assumed that the ω -meson corresponds to a mixture of the light quark combinations $u\bar{u}$, $d\bar{d}$, like the neutral pion (3.64), while the ϕ is a pure $s\bar{s}$ state. As can be seen from Table 3.6, this automatically leads to the result $m_\rho = m_\omega$, which is in good agreement with experiment. On the other hand, the analogous assumption for the pseudoscalar nonet leads to the result $m_\pi = m_\eta$, indicating

that the eta and eta-prime are both mixtures of the quark states $u\bar{u}$, $d\bar{d}$ and $s\bar{s}$, and we will not pursue the matter further.

Table 3.6 Predicted meson masses (in GeV/c^2) in the approximation (3.87) compared with experimental values. These have been obtained using $m = 0.306 \text{ GeV}/c^2$, $m_s = 0.487 \text{ GeV}/c^2$, and $a = 0.0592 (\text{GeV}/c^2)^3$

Particle	Mass	Prediction	Experiment
π	$2m - \frac{3a}{4m^2}$	0.14	0.140
K	$m + m_s - \frac{3a}{4mm_s}$	0.50	0.495
ρ	$2m + \frac{a}{4m^2}$	0.77	0.775
ω	$2m + \frac{a}{4m^2}$	0.77	0.783
K^*	$m + m_s + \frac{a}{4mm_s}$	0.89	0.892
ϕ	$2m_s + \frac{a}{4m_s^2}$	1.04	1.019

For baryons we have three quarks and the mass formula becomes

$$M = m_1 + m_2 + m_3 + \left(\frac{b}{\hbar^2} \right) \sum_{i < j} \frac{\mathbf{S}_i \cdot \mathbf{S}_j}{m_i m_j}, \quad i, j = 1, 3, \quad (3.91)$$

where b is a constant to be determined from experiment. (There is no reason for b to be equal to the constant a used in the meson case because the quark wavefunctions and numerical factors in the hadronic equivalents of (3.84) will be different in the two cases.) In the case of the $3/2^+$ decuplet, all three quarks have their spins aligned and every pair therefore combines to make spin-1. Thus, for example,

$$(\mathbf{S}_1 + \mathbf{S}_2)^2 = \mathbf{S}_1^2 + \mathbf{S}_2^2 + 2\mathbf{S}_1 \cdot \mathbf{S}_2 = 2\hbar^2$$

giving $\mathbf{S}_1 \cdot \mathbf{S}_2 = \hbar^2/4$ and in general

$$\mathbf{S}_1 \cdot \mathbf{S}_2 = \mathbf{S}_1 \cdot \mathbf{S}_3 = \mathbf{S}_2 \cdot \mathbf{S}_3 = \hbar^2/4.$$

Using this result, the mass of the $\Sigma^*(1385)$, for example, may be written

$$M_{\Sigma^*} = 2m + m_s + \frac{b}{4} \left(\frac{1}{m^2} + \frac{2}{mm_s} \right).$$

In the case of the $1/2^+$ octet, we have

$$(\mathbf{S}_1 + \mathbf{S}_2 + \mathbf{S}_3)^2 = \mathbf{S}_1^2 + \mathbf{S}_2^2 + \mathbf{S}_3^2 + 2(\mathbf{S}_1 \cdot \mathbf{S}_2 + \mathbf{S}_1 \cdot \mathbf{S}_3 + \mathbf{S}_2 \cdot \mathbf{S}_3) = 3\hbar^2/4$$

and hence

$$\mathbf{S}_1 \cdot \mathbf{S}_2 + \mathbf{S}_1 \cdot \mathbf{S}_3 + \mathbf{S}_2 \cdot \mathbf{S}_3 = -3\hbar^2/4. \quad (3.92)$$

In addition, we have to consider the symmetry of the spin wavefunctions of individual hadrons. For example, the spins of the u and d pair in the Λ must combine to give $\mathbf{S} = \mathbf{0}$, as discussed in Section 3.3 above. Thus, in this case, $(\mathbf{S}_u + \mathbf{S}_d)^2 = 0$, so that $\mathbf{S}_u \cdot \mathbf{S}_d = -3\hbar^2/4$. Then

$$M_\Lambda = m_u + m_d + m_s + b \left[\frac{\mathbf{S}_u \cdot \mathbf{S}_d}{m_u m_d} + \frac{\mathbf{S}_u \cdot \mathbf{S}_s}{m_u m_s} + \frac{\mathbf{S}_d \cdot \mathbf{S}_s}{m_d m_s} \right].$$

Finally, setting $m_u = m_d = m$ gives

$$\begin{aligned} M_\Lambda &= 2m + m_s + b \left[\frac{\mathbf{S}_u \cdot \mathbf{S}_d}{m^2} + \frac{(\mathbf{S}_u \cdot \mathbf{S}_d + \mathbf{S}_u \cdot \mathbf{S}_s + \mathbf{S}_d \cdot \mathbf{S}_s - \mathbf{S}_u \cdot \mathbf{S}_d)}{m m_s} \right] \\ &= 2m + m_s - \frac{3b}{4m^2}, \end{aligned} \quad (3.93)$$

where we have used (3.92). The resulting formulas for all the $1/2^+$ octet and $3/2^+$ decuplet masses are shown in Table 3.7 and imply

$$m_N + m_\Delta = 6m, \quad m_\Delta - m_N = 6b/4m^2, \quad m_N - m_\Lambda = m - m_s.$$

Substituting the experimental masses gives

$$m = 0.362 \text{ GeV}/c^2, \quad m_s = 0.539 \text{ GeV}/c^2, \quad b = 0.0256 (\text{GeV}/c^2)^3, \quad (3.94)$$

and the resulting predictions for the other masses agree well with experiment, as shown in Table 3.7. Again, the agreement between these

Table 3.7 Predicted baryon masses (in GeV/c^2) in the approximation (3.91) compared with experimental values. These have been obtained using $m = 0.362 \text{ GeV}/c^2$, $m_s = 0.539 \text{ GeV}/c^2$, and $b = 0.0256 (\text{GeV}/c^2)^3$

Particle	Mass	Prediction	Experiment
N	$3m - \frac{3b}{4m^2}$	0.94	0.939
Λ	$2m + m_s - \frac{3b}{4m^2}$	1.12	1.116
Σ	$2m + m_s + \frac{b}{4} \left(\frac{1}{m^2} - \frac{4}{mm_s} \right)$	1.18	1.193
Ξ	$m + 2m_s + \frac{b}{4} \left(\frac{1}{m_s^2} - \frac{4}{mm_s} \right)$	1.32	1.318
Δ	$3m + \frac{3b}{4m^2}$	1.23	1.232
Σ^*	$2m + m_s + \frac{b}{4} \left(\frac{1}{m^2} + \frac{2}{mm_s} \right)$	1.38	1.385
Ξ^*	$m + 2m_s + \frac{b}{4} \left(\frac{2}{mm_s} + \frac{1}{m_s^2} \right)$	1.53	1.533
Ω	$3m_s + \frac{3b}{4m_s^2}$	1.68	1.673

quark masses and those previously obtained is reasonable, but not perfect. This is to be expected. Since free quarks are not observed, their masses cannot be measured directly and, in the absence of a precise theory, the constituent masses are best regarded as effective masses inferred indirectly in the context of particular models and environments; while reasonable models of the light hadrons always give approximately the same values, there will inevitably be some variation reflecting the different approximations made.²⁷

Overall, what we learn from the above is that the constituent quark model is capable of giving a reasonably consistent account of hadron masses and magnetic moments, at least for the low-lying states, provided a few parameters are allowed to be found from experiment.

3.3.5 Heavy quarkonia

In previous sections we have mainly discussed states composed of light quarks. In this section we will discuss states formed from pairs of heavy quarks and antiquarks, called charmonium ($c\bar{c}$) and bottomonium ($b\bar{b}$). These states are particularly interesting because the c and b quarks are so heavy that they move slowly enough within the resulting hadrons to be treated nonrelativistically to a first approximation. (This is one of the few places in particle physics where a nonrelativistic calculation is adequate.) There is thus a strong analogy with the e^+e^- bound system positronium, except that the potential energy is different. As in positronium, states can be labelled by their angular momentum quantum numbers, J , L , and S , where the spin S can be 0 or 1, and by their principal quantum number $n = \nu + L + 1$, where ν is the number of radial nodes, not counting nodes at the origin. Alternatively, instead of n , the radial quantum number $n_r = \nu + 1 = n - L = 1, 2, 3, \dots$ is sometimes used to label the states. The resulting states for charmonium and bottomonium with principal quantum number $n = 1$ or $n = 2$ are listed in Table 3.8, together with the values of their parity

$$P = P_q P_{\bar{q}} (-1)^L = (-1)^{L+1} \quad (q = c, b), \quad (3.95)$$

C-parity

$$C = (-1)^{L+S}, \quad (3.96)$$

and the observed mesons that correspond to them. We shall see later that the measured energies of these states can be used to partially determine the potential energies in the two systems, leading to the important conclusion that the forces acting in the $c\bar{c}$ and $b\bar{b}$ systems are the same within errors.

²⁷We will return to this point in Section 10.2.1.

Table 3.8 Predicted $c\bar{c}$ and $b\bar{b}$ states with principal quantum numbers $n = 1$ and 2 and radial quantum number $n_r = n - L$, compared with experimentally observed states. Masses are given in MeV/c²

$2S+1 L_J$	n	n_r	J^{PC}	$c\bar{c}$ state	$b\bar{b}$ state
$^1 S_0$	1	1	0^{-+}	$\eta_c(2984)$	$\eta_b(9399)$
$^3 S_1$	1	1	1^{--}	$J/\psi(3097)$	$\Upsilon(9460)$
$^3 P_0$	2	1	0^{++}	$\chi_{c0}(3415)$	$\chi_{b0}(9859)$
$^3 P_1$	2	1	1^{++}	$\chi_{c1}(3511)$	$\chi_{b1}(9893)$
$^3 P_2$	2	1	2^{++}	$\chi_{c2}(3556)$	$\chi_{b2}(9912)$
$^1 P_1$	2	1	1^{+-}	$h_c(3525)$	$h_b(9899)$
$^1 S_0$	2	2	0^{-+}	$\eta_c(3638)$	$\eta_b(9999)^a$
$^3 S_1$	2	2	1^{--}	$\psi(3687)$	$\Upsilon(10023)$

^aThe existence of this state is not well established.

3.3.5(a) Charmonium

The first charmonium state to be discovered was the $J/\psi(3097)$.²⁸ Its discovery caused considerable excitement because it confirmed the existence of the charm quantum number that had been predicted several years earlier, even though the $J/\psi(3097)$ itself has zero overall charm (referred to as ‘hidden charm’). Hence it was a very important piece of evidence in favour of the standard model. It is most conveniently studied by first forming those states V^0 that have the same J^{PC} values 1^{--} as the photon (cf. Table 3.8) and then observing their decays. The basic formation process is shown in Figure 3.19: the electron–positron pair annihilates to a virtual photon, which converts to the V^0 , which then decays, mainly into hadrons. This is only possible when the total centre-of-mass energy is equal to the V^0 mass, leading to a peak in the total cross-section for electron–positron annihilation into hadrons. This cross-section is usually plotted in terms of the ratio

$$R \equiv \frac{\sigma(e^+e^- \rightarrow \text{hadrons})}{\sigma(e^+e^- \rightarrow \mu^+\mu^-)}, \quad (3.97a)$$

where the measured cross-section for muon pair production,²⁹

$$\sigma(e^+e^- \rightarrow \mu^+\mu^-) = \frac{4\pi\alpha^2}{3E_{\text{CM}}^2}, \quad (3.97b)$$

is a smoothly varying function of the centre-of-mass energy E_{CM} , as shown in Figure 3.20.

²⁸The rather clumsy notation is because it was discovered independently by two groups, led by Burton Richter and Samuel Ting. Richter’s group was studying the reactions $e^+e^- \rightarrow \text{hadrons}$ and named it the ψ particle. Ting’s group discovered it in $p\text{Be}$ scattering reactions and called it the J . It is now known as the J/ψ . Richter and Ting shared the 1976 Nobel Prize in Physics for the discovery.

²⁹This cross-section can be calculated theoretically (see, for example, pp. 135–139 of Mandl and Shaw 2010). However, it is sufficient for our purposes to take this result from experiment.

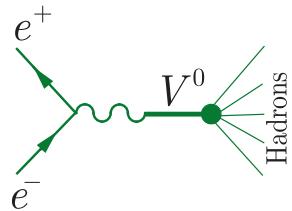
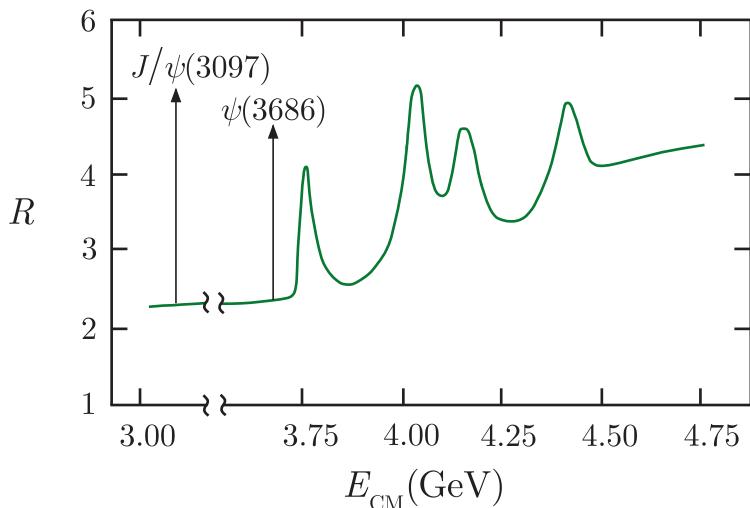


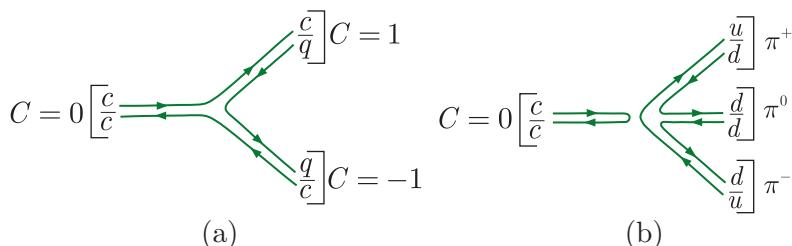
Figure 3.19 Mechanism for the formation of mesons V^0 with quantum numbers $J^{PC} = 1^{--}$ in electron–positron annihilation, and their subsequent decay to hadrons.

Figure 3.20 Measured cross-section ratio (3.97a) in the vicinity of the threshold for producing pairs of charmed particles. The extremely narrow peaks associated with the $J/\psi(3097)$ and $\psi(3686)$ are indicated by arrows because they extend far above the scale shown. In the partially suppressed region between these peaks, the ratio is essentially constant. Source: Data from Brandelik et al. (1978) and Schindler et al. (1980).



As can be seen in Figure 3.20, there are two sharp peaks in the ratio, corresponding to the masses of the $J/\psi(3097)$ and $\psi(3686)$, followed by several relatively low, broad peaks at energies above 3730 MeV. This latter energy is called the *charm threshold* and is twice the mass of the lightest meson having nonzero charm, the $D^0(1865)$. Charmonium states that lie above this threshold are resonances that decay readily into pairs of charmed particles by the mechanism of Figure 3.21a, involving the creation of a light quark–antiquark pair, and the broad peaks in Figure 3.20 are evidence for the existence of several such states. In contrast, charmonium states that lie below this threshold are all very narrow and cannot decay by the mechanism of Figure 3.21a because of energy conservation; instead they decay much more slowly to noncharmed hadrons by mechanisms like that of Figure 3.21b, involving the annihilation of a charmed quark–antiquark pair.

Figure 3.21 Quark diagrams for (a) the decay of a charmonium state to a pair of charmed mesons, and (b) an example of a decay to noncharmed mesons.



In general, strong interaction processes involving the creation or annihilation of $c\bar{c}$ and $b\bar{b}$ pairs are always very heavily suppressed compared to processes involving the creation or annihilation of light quark–antiquark pairs ($u\bar{u}$, $d\bar{d}$, $s\bar{s}$). This characteristic, which does not apply to weak or electromagnetic interactions, is called the OZI rule, after Okubo, Zweig, and Iizuka, who first formulated it.³⁰ In this context it explains why the

³⁰The dynamical origin of the OZI rule will be discussed in Section 5.2.1.

states below the charm threshold are much narrower than those above it. Consequently, as narrow states are much easier to identify experimentally than broad ones, the spectrum is much better known below the charmed threshold than above it. We will therefore confine ourselves to this part of the spectrum in the rest of this section.

The $J/\psi(3097)$ and $\psi(3686)$ are the only well-established charmonium states listed in Table 3.8 that can be produced by the mechanism of Figure 3.19, because they are the only ones with the same quantum numbers $J^{PC} = 1^{--}$ as the photon. Some of the other states shown in Table 3.8, however, can be found by studying the radiative decays of the $J/\psi(3097)$ and $\psi(3686)$. In particular, the $\psi(3686)$ decays mostly into hadrons, but about 25% of the time it decays by the radiative transitions

$$\psi(3686) \rightarrow \chi_{ci} + \gamma \quad (i = 0, 1, 2), \quad (3.98)$$

where $\chi_{c0}(3415)$, $\chi_{c1}(3511)$, and $\chi_{c2}(3556)$ are three new charmonium states, called chi-states. These particles themselves decay, mostly to give hadrons in the final state, and an analysis of these decays shows that they have J^{PC} values 0^{++} , 1^{++} , and 2^{++} , respectively. We therefore identify them with the 3P_0 , 3P_1 , and 3P_2 states of charmonium, which have these predicted J^{PC} values, as can be seen from Table 3.8. In addition, the $n = 1$, 1S_0 ground state $\eta_c(2980)$ has been observed in the radiative transitions

$$\psi(3686) \rightarrow \eta_c(2984) + \gamma \quad (3.99a)$$

and

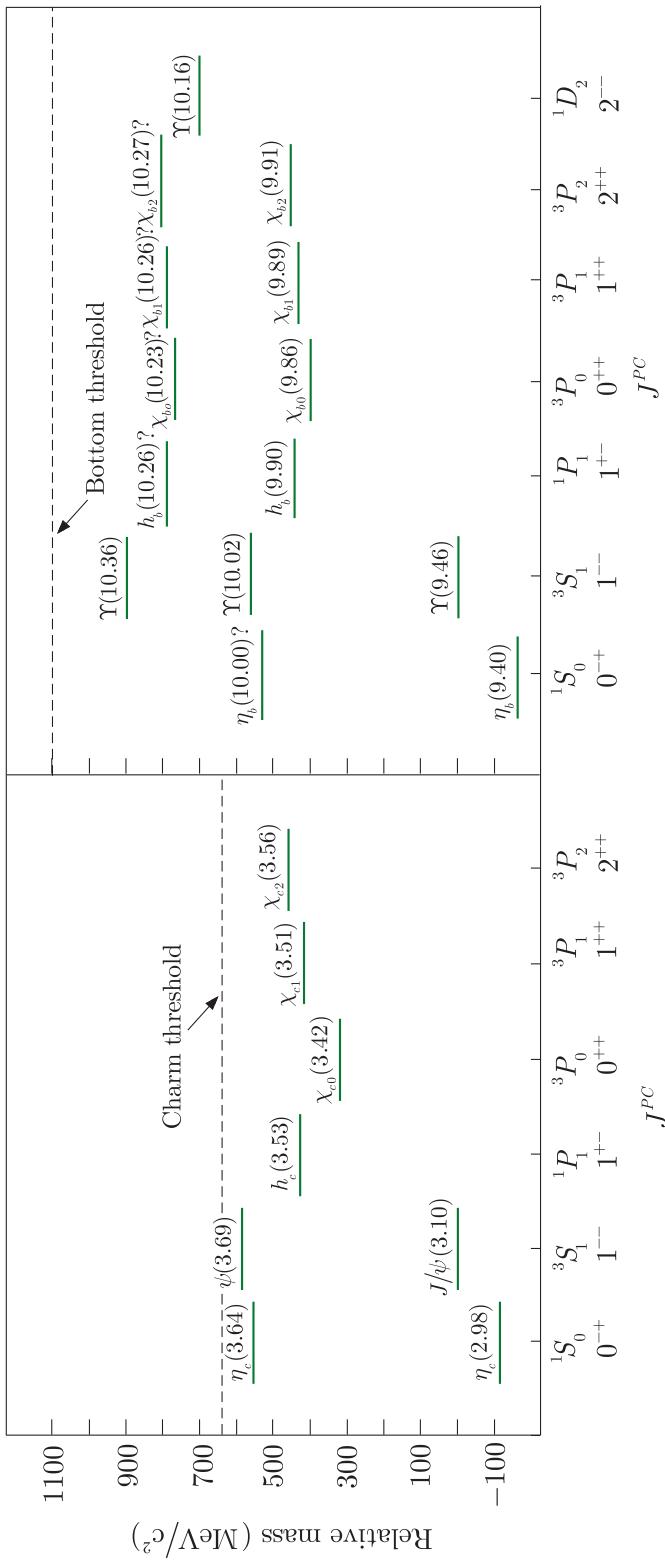
$$J/\psi(3097) \rightarrow \eta_c(2984) + \gamma, \quad (3.99b)$$

despite the small branching ratios of about 3×10^{-3} and 10^{-2} , respectively. The remaining charmonium states in Table 3.8 with $n \leq 2$ are the 1S_0 state $\eta_c(3639)$ and the 1P_1 state $h_c(3526)$. The former has been seen in $\gamma\gamma$ interactions and the latter in $p\bar{p}$ annihilations.

To summarise, all eight predicted states of charmonium with principal quantum numbers $n = 1$ or $n = 2$ have been identified and are the only states observed below the charm threshold. The resulting experimental spectrum is shown in Figure 3.22. Despite the difference in energy scales, it bears a striking resemblance to the spectrum of positronium, except that the $2S$ levels are somewhat higher in energy than the $2P$ levels, instead of being degenerate with them.

3.3.5(b) Bottomonium

The bottomonium spectrum is observed in much the same way as the charmonium spectrum, and we confine ourselves here to stating the results, which are summarised in Figure 3.22. In this case there are three bands of narrow states, corresponding to principal quantum numbers $n = 1$, 2, and 3, lying below the bottom threshold of $10.56 \text{ GeV}/c^2$. This is just twice the mass of the $B^\pm(5279)$ mesons, which are the lightest particles with a nonzero bottom, and bottomonium states above this threshold are broad resonances that decay to pairs of particles with



$\tilde{B} = 0$ by a mechanism analogous to that of Figure 3.21a. As can be seen in Figure 3.22, the ordering of the levels and the magnitudes of their splittings are similar in charmonium and bottomonium, indicating an underlying similarity between the forces acting in the two systems.

3.3.5(c) The quark–antiquark potential

The strong resemblance of the experimental spectra of charmonium and bottomonium to that of positronium, as shown in Figure 3.22, suggests that the spectra can be understood in terms of a simple nonrelativistic approximation. In the centre-of-mass frame of the $q\bar{q} = c\bar{c}$ or $b\bar{b}$ pair, the Schrödinger equation is

$$-\frac{1}{2\mu}\nabla^2\psi(\mathbf{r}) + V(r)\psi(\mathbf{r}) = E\psi(\mathbf{r}), \quad (3.100)$$

where $r = |\mathbf{r}|$ is the distance between the quarks, $\mu = m_q/2$ is their reduced mass, and we have neglected spin-dependent effects. In this approximation, the energies E and the particle masses

$$M(q\bar{q}) = 2m_q + E \quad (3.101)$$

depend only on the principal quantum number n and the orbital angular momentum quantum number L , for a given potential $V(r)$. In the particular case of a Coulomb-like potential

$$V(r) \propto r^{-1}, \quad (3.102)$$

the energies depend on n only, but for other r -dependencies this degeneracy is broken.

This is illustrated in Figure 3.23, where the spectrum of S - and P -wave states resulting from a Coulomb potential (3.102) is compared with that resulting from a simple harmonic oscillator potential

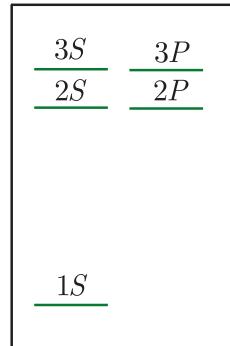
$$V(r) \propto r^2. \quad (3.103)$$

The heavy quarkonia spectra of Figure 3.22 are intermediate between these possibilities, and can be fitted using a potential of the form

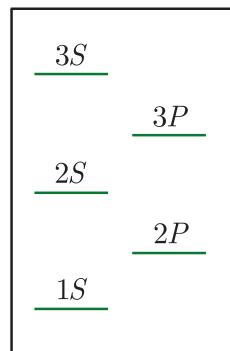
$$V(r) = -\frac{a}{r} + b r, \quad (3.104a)$$

which behaves like a Coulomb potential at small r , but rises linearly at large r values. The constants a and b are determined by solving the Schrödinger equation numerically and adjusting their values to good fits to the experimental spectra.³¹ In this way one finds that the values

$$a = 0.30 \quad \text{and} \quad b = 0.23 \text{ GeV}^2 \quad (3.104b)$$



(a) Coulomb



(b) Oscillator

Figure 3.23 Patterns of S - and P -wave energy levels arising from (a) Coulomb-like (r^{-1}) and (b) simple harmonic oscillator (r^2) potentials for principal quantum numbers $n = 1, 2, 3$. The strengths of the potentials are assumed to be such that the $3s$ states have the same energy, measured relative to the $1s$ state, in each case.

³¹The agreement is not perfect, because relativistic effects are not completely negligible. These include a spin-dependent interaction to account for the fine structure seen in Figure 3.22.

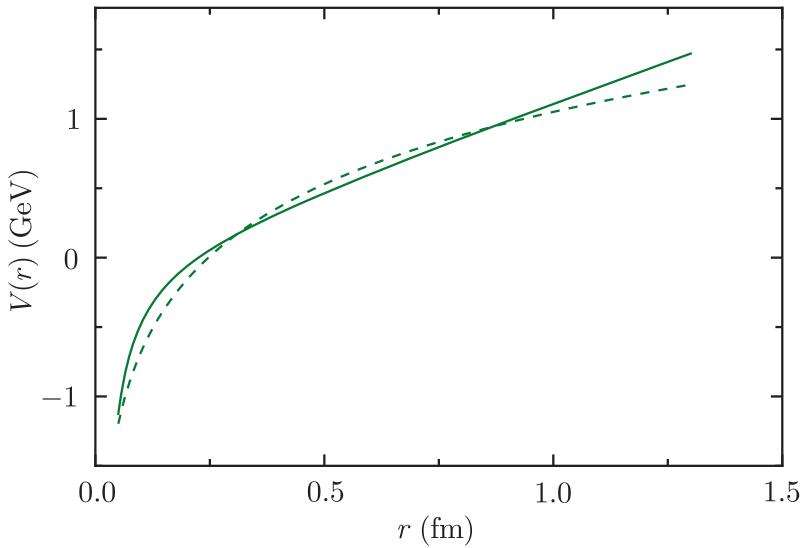


Figure 3.24 Heavy quark-antiquark potentials obtained from fitting the energy levels of charmonium and bottomonium. The solid and dashed lines show the results obtained from the forms (3.105a) and (3.106a), respectively.

can be used to fit both the charmonium and bottomonium spectra, and the resulting potential is shown in Figure 3.24. The potential is not completely determined by this analysis, because of uncertainties in the quark masses and because equally satisfactory results can be obtained using other forms, for example³²

$$V(r) = a \ln(br), \quad (3.105a)$$

where

$$a = 0.75 \text{ GeV} \quad \text{and} \quad b = 0.80 \text{ GeV}. \quad (3.105b)$$

However, all the potentials that successfully explain the spectrum for given quark masses are in good agreement with each other in the range

$$0.2 \text{ fm} \lesssim r \lesssim 1.0 \text{ fm}, \quad (3.106)$$

as illustrated in Figure 3.24 for the cases (3.104b) and (3.105b). We therefore conclude that the potential is well-determined in the range (3.106), and that, within this range at least, the potentials for charmonium ($c\bar{c}$) and bottomonium ($b\bar{b}$) are, within small experimental uncertainties, the same, as expected from flavour independence.³³

³²In this book, logarithms written in the form ‘ \ln ’ denote that the base is e . Logarithms written in the form ‘ \log ’ are to base 10.

³³These analyses also determine the mean square radii

$$r_0^2 \equiv \bar{r}^2 = \int r^2 |\psi(r)|^2 d^3r$$

of the various states, where the wavefunctions are determined by the solutions of the Schrödinger equation for successful potentials like (3.104a) and (3.105a). The resulting values of r_0 for the various states shown in Figures 3.22 span just the range (3.106) in which the potential is well-determined.

3.3.6 Allowed and exotic quantum numbers

The values of the internal quantum numbers

$$Q, B, C, S, \text{ and } \tilde{B} \quad (3.107)$$

that characterise any given hadron are severely restricted in the simple quark model. For example, consider the allowed values of charge and strangeness, starting with hadrons for which³⁴ $C = \tilde{B} = 0$. For baryons, the possible qqq combinations are then

$$sss, ssq_i, sq_iq_j, q_iq_jq_k,$$

where s is the strange quark and $q_i = u, d$. The allowed strangeness values are therefore $S = -3, -2, -1$, and 0, and the allowed charges are obtained by adding the quark charges. Thus a baryon sss with $S = -3$ must have charge $Q = -1$, while a baryon $ssq_j = ssu, ssd$ with $S = -2$ can have $Q = 0$ or $Q = -1$. These values, together with the corresponding results for $S = 0$ and -1 are collected together in Table 3.9a, and the corresponding values for meson combinations $q\bar{q}$, left as an exercise for the reader, are given in Table 3.9b.

Table 3.9 The combinations of charge and strangeness allowed for hadrons in the simple quark model for states with $C = \tilde{B} = 0$

(a) Baryons		(b) Mesons	
S	Q	S	Q
0	2, 1, 0, -1	1	1, 0
-1	1, 0, -1	0	1, 0, -1
-2	0, -1	-1	0, -1
-3	-1		

For hadrons with $C = \tilde{B} = 0$, combinations of S and Q that do not occur in these tables are called *exotic*, since they can arise only from quark combinations like

$$qq\bar{q}\bar{q} \text{ and } qqqq\bar{q} \quad (3.108)$$

that are not included in the simple quark model. Examples of hadrons with exotic quantum numbers³⁵ would be a meson $dd\bar{u}\bar{s}$ with $S = 1$ and $Q = -1$ or a baryon $uuud\bar{s}$ with $S = 1$ and $Q = 2$. Furthermore, the discussion is easily extended to include nonzero charm and bottom. For

³⁴The known hadrons all have top $T = 0$, as noted in Section 3.3.2, and so we ignore top in what follows.

³⁵It is important to distinguish between the terms ‘hadrons with exotic quantum numbers’ and the term ‘exotic hadrons’, which may already be familiar to some readers. The latter term is not introduced in this book until Section 5.1.3 and in this chapter the word ‘exotic’ is only used with reference to quantum numbers.

example, if we consider hadrons in which both C and \tilde{B} are nonzero, the only possible meson configurations are $c\bar{b}$ with $Q = C = \tilde{B} = 1$ and $S = 0$, and their antiparticles $b\bar{c}$, with $Q = C = \tilde{B} = -1$ and $S = 0$. All other combinations are exotic.

From this we see that, in the simple quark model, the internal quantum numbers of hadrons are restricted to relatively few allowed combinations, while all other combinations – the vast majority – are exotic and should not be observed. This is a highly nontrivial prediction, since, when a new hadron is identified, its quark content is unknown, but despite this, its internal quantum numbers can easily be determined by applying the appropriate conservation laws to its observed production and decay processes.

At present, well over two hundred hadrons with internal quantum numbers compatible with the expectations of the simple quark model have been clearly established. In addition, from the 1960s onwards there have been numerous searches for hadrons with exotic internal quantum numbers, i.e. internal quantum numbers that are incompatible with the simple quark model and could only arise from more complicated quark structures, like the *tetraquark* and *pentaquark* quark states given in (3.108) for mesons and baryons, respectively. From time to time, experiments claimed evidence for states with exotic internal quantum numbers, but as more experiments were performed and far greater quantities of data were produced, the evidence for their existence receded, and by 2015 the consensus was that there was no convincing evidence for their existence. However, in 2016 a new candidate emerged. Specifically, the D0 collaboration measured the invariant mass distribution of $B_s^0\pi^\pm$ pairs produced in high-energy proton–antiproton collisions, where $B_s^0(5367) = s\bar{b}$ is a long-lived pseudoscalar meson with a lifetime of about 1.5×10^{-12} s. The results (Abasov et al. 2016) show a clear peak at an invariant mass of $5568 \text{ MeV}/c^2$, with a width of about 20 MeV, which they interpreted as resulting from the strong interaction decay

$$X^\pm \rightarrow B_s^0 + \pi^\pm \quad (3.109a)$$

of an unknown meson $X^\pm(5568)$, with the quantum numbers

$$Q = \pm 1, B = 0, S = -1, \tilde{B} = +1. \quad (3.109b)$$

These quantum numbers are exotic, and the simplest possible quark assignments are $s\bar{b}u\bar{d}$ and $s\bar{b}d\bar{u}$ for the positively and negatively charged states, respectively. On the other hand, the LHCb collaboration (LHCb Collaboration 2016) has subsequently studied the invariant mass distribution of $B_s^0\pi^\pm$ pairs produced in high-energy proton–proton collisions, with a view to confirming the above result, but failed to see any $X^\pm(5568)$ signal, despite very high statistics. While these results do not directly contradict each other, since they are obtained using different particle beams at different energies, the contrast is surprising, and it is clear that further experimental work is required before the $X^\pm(5568)$ can be regarded as a well-established state.

To summarise, it is clear that the quantum numbers of the overwhelming majority of known hadrons conform to the values required by the simple quark model; and while the proposed $X^\pm(5568)$ state discussed above remains a possibility, not a single example of a hadron with exotic internal quantum numbers can yet be regarded as firmly established. Nonetheless, multiquark states like (3.108) do not necessarily give rise to exotic quantum numbers, and, despite the considerable successes of the simple quark model, $q\bar{q}$ and $3q$ states are not the whole story, as we shall see in Section 5.3.1.

Problems 3

- 3.1** Which of the following reactions are allowed and which are forbidden by the conservation laws appropriate to weak interactions?
- (a) $\nu_\mu + p \rightarrow \mu^+ + n$
 - (b) $\nu_e + p \rightarrow n + e^- + \pi^+$
 - (c) $\Lambda \rightarrow \pi^+ + e^- + \bar{\nu}_e$
 - (d) $K^+ \rightarrow \pi^0 + \mu^+ + \nu_\mu$
 - (e) $\nu_e + p \rightarrow e^- + \pi^+ + p$
 - (f) $\tau^+ \rightarrow \mu^+ + \bar{\nu}_\mu + \nu_\tau$
- 3.2** Draw a fourth-order Feynman diagram for the weak reaction $e^- + \mu^+ \rightarrow \nu_e + \bar{\nu}_\mu$.
- 3.3** In Section 3.1.3 it is stated that electron neutrinos interact with electrons in a different way from muon and tauon neutrinos. Justify this remark by considering the lowest-order Feynman diagrams for $\nu_e + e^- \rightarrow \nu_e + e^-$ and $\nu_\mu + e^- \rightarrow \nu_\mu + e^-$.
- 3.4** Show that the oscillation length in (3.31b) may be written $L_0 = E/(1.27\Delta m_{ij}^2)$, where L_0 is expressed in km, E in GeV, and Δm_{ij}^2 in $(\text{eV}/c^2)^2$.
- 3.5** A KamLAND-type experiment detects $\bar{\nu}_e$ neutrinos at a distance of 200 m from a nuclear reactor and finds that the flux is $(90 \pm 10)\%$ of that expected if there were no oscillations. Assuming a two-component model with maximal mixing ($\theta = 45^\circ$) and a mean neutrino energy of 3 MeV, use this result to estimate the squared mass difference of the $\bar{\nu}_e$ and its oscillating partner.
- 3.6** If the Sun is assumed to be a uniform spherical plasma consisting of nucleons, with radius 7×10^5 km and total mass 2×10^{30} kg, calculate the mean free path $\lambda = 1/n\sigma$ of solar neutrinos from the dominant reaction (3.38) where n is the number of nucleons per unit volume and σ , the neutrino-nucleon cross-section, may be written $\sigma = 0.7E_L \times 10^{-42} \text{ m}^2$, where E_L is the neutrino laboratory energy in GeV.
- 3.7** Use the results of Section 3.3.1 to deduce a relation between the total cross-sections for the reactions $\pi^- p \rightarrow K^0 \Sigma^0$, $\pi^- p \rightarrow K^+ \Sigma^-$, and $\pi^+ p \rightarrow K^+ \Sigma^+$ at a fixed energy.
- 3.8** At a certain energy $\sigma(\pi^+ n) \approx \sigma(\pi^- p)$, whereas $\sigma(K^+ n) \neq \sigma(K^- p)$. Comment on this.
- 3.9** Consider a scenario where overall hadronic wavefunctions Ψ consist of just spin and space parts, i.e. $\Psi = \psi_{\text{space}} \psi_{\text{spin}}$, and the Pauli principle is obeyed. What would be the resulting multiplet structure of the lowest-lying baryon states composed of u, d , and s quarks?

- 3.10** Six observed hadrons have the quantum numbers $(Q, B, S, C, \tilde{B}) = (2, 1, 0, 1, 0), (0, 1, -2, 1, 0), (0, 0, 1, 0, -1), (0, -1, 1, 0, 0), (0, 1, -1, 1, 0)$, and $(-1, 1, -3, 0, 0)$, respectively. Identify their quark constituents.
- 3.11** Find the values of the parity P and, where appropriate, the charge conjugation C for the ground-state ($J = 0$) mesons π^\pm and π^0 , and their first excited ($J = 1$) states ρ^\pm and ρ^0 , where the latter have a mass of about $770 \text{ MeV}/c^2$. Why does the charged pion have a longer lifetime than the charged ρ ? Explain also why the decay $\rho^0 \rightarrow \pi^+ \pi^-$ has been observed, but not the decay $\rho^0 \rightarrow \pi^0 \pi^0$.
- 3.12** Classify the following as strong, electromagnetic, weak, or forbidden reactions:
- | | |
|---|--|
| (a) $p + \bar{p} \rightarrow \pi^+ + \pi^-$ | (b) $p \rightarrow e^+ + \gamma$ |
| (c) $\Sigma^0 \rightarrow \Lambda + \gamma$ | (d) $p + p \rightarrow \Sigma^+ + n + K^0 + \pi^+$ |
| (e) $\Xi^- \rightarrow \Lambda + \pi^-$ | (f) $\Delta^+ \rightarrow p + \pi^0$ |
- 3.13** The particle Y^- can be produced in the strong interaction process $K^- + p \rightarrow K^+ + Y^-$. Deduce its baryon number, strangeness, charm, and bottom, and, using these, its quark content. The Y^- (1311) decays by the reaction $Y^- \rightarrow \Lambda + \pi^-$. Give a rough estimate of its lifetime.
- 3.14** Consider the reaction
- $$K^- + p \rightarrow \Omega^- + K^+ + K^0,$$
- followed by the sequence of decays
- $$\begin{array}{ccc} \Omega^- \rightarrow \Xi^0 + \pi^- & & K^+ \rightarrow \pi^+ + \pi^0 \\ \downarrow \pi^0 + \Lambda & & \downarrow \gamma + \gamma, \\ \downarrow \mu^+ + \nu_\mu & \text{and} & K^0 \rightarrow \pi^+ + \pi^- + \pi^0. \end{array}$$
- The quark decompositions of the Ω^- and Ξ^0 states are given in Table 3.3. Classify each process as strong, weak, or electromagnetic and give your reasons.
- 3.15** Draw the lowest-order Feynman diagram for the decay $K^+ \rightarrow \mu^+ + \nu_\mu + \gamma$ and hence deduce the form of the overall effective coupling.
- 3.16** Draw the lowest-order Feynman diagrams at the quark level for the following decays:
- $D^- \rightarrow K^0 + \pi^-$,
 - $\Lambda \rightarrow p + e^- + \bar{\nu}_e$,
- where the quark compositions of the D^- and Λ are given in Table 3.3.
- 3.17** Calculate the quark model predictions for the magnetic moment of the Σ^0 (1193) and the $L = 0, J^P = 3/2^+$ baryon Ω^- (1672), which is the ground state of the sss system.
- 3.18** Verify the expression in Table 3.7 for the mass of the $1/2^+$ Σ baryon.
- 3.19** Which of the following possible hadronic states with quantum numbers $(Q, B, S, C, \tilde{B}) = (1, 0, 0, 1, 1), (-1, 1, -2, 0, -1), (0, 0, 1, 0, 1)$, and $(-1, 1, 0, 1, -1)$ are compatible with the simple quark model, and which are not?
- 3.20** A hypothetical spin-1 bottomonium resonance Υ with mass $9 \text{ GeV}/c^2$ is observed in the reaction $e^+ e^- \rightarrow \mu^+ \mu^-$ and in the total electron–positron annihilation cross-section. The resonance is too narrow for its width to be

resolved directly, but the cross-sections integrated over the vicinity of the resonance are

$$\int \sigma_\mu(E) dE = 10 \text{ nb MeV} \text{ and } \int \sigma_t(E) dE = 300 \text{ nb MeV},$$

where E is the total centre-of-mass energy, $\sigma_\mu = \sigma(e^+e^- \rightarrow \mu\mu)$, and σ_t is the total cross-section. Assuming the cross-sections are dominated by a Breit–Wigner resonance of the form discussed in Section 1.6.4, evaluate the partial widths $\Gamma(\Upsilon \rightarrow e^+e^-)$, $\Gamma(\Upsilon \rightarrow \mu^+\mu^-)$, and the total decay width.

- 3.21** Derive the allowed combinations of charm C and electric charge Q for mesons and baryons in the simple quark model.

4

Experimental methods

In earlier chapters we discussed the results of a number of experiments, but said very little about how experiments are done. In this chapter we consider experimental methods. This is a very extensive subject. Consider, for example, the measurement of nuclear lifetimes. For lifetimes of a few minutes to several hours, the simplest method is to directly observe the activity \mathcal{A} as a function of time. Then from (2.61), the decay rate, and hence the lifetime, is obtainable directly from the slope of a plot of $\ln \mathcal{A}$ versus t . However, this method is impractical for lifetimes outside this range, and since nuclear lifetimes span an enormous range of values, from 10^{15} y to 10^{-15} s, a very wide variety of techniques has to be employed. The same is true for particle physics, where lifetimes also span a very wide range of values, some as small as 10^{-23} s. The aim of this chapter will not be to give a comprehensive review of specific experimental methods, but rather to emphasise the physical principles behind them. More details may be found in specialised texts.¹

4.1 Overview

To explore the structure of nuclei (nuclear physics) or hadrons (particle physics) requires projectiles whose wavelengths are at least as small as the effective radii of the nuclei or hadrons. This determines the minimum value of the momentum $p = h/\lambda$, and hence the energy, required. The majority of experiments are conducted using beams of particles produced by machines called *accelerators*. This has the great advantage that

¹See, for example, Ferbel (1992), Leo (1994), Poenaru and Greiner (1997), Kleinknecht (1998), Knoll (1999), Green (2000), and Grupen and Schwartz (2011). There is also a series of reviews in Tanabashi et al. (Particle Data Group) (2018).

the projectiles are of a single type and have energies that may be controlled by the experimenter.² Beams that are essentially mono-energetic may be prepared and can be used to study the energy dependence of interactions. In a *fixed-target* experiment, the beam, once established, is directed on to a target that is stationary in the laboratory. Many experiments in both particle and nuclear physics, particularly the latter, are of this type. In particle physics, high energies are also required to produce new and unstable particles and this reveals a disadvantage of fixed-target experiments when large centre-of-mass energies are required. The centre-of-mass energy is important because it is a measure of the energy available to create new particles. In the laboratory frame at least some of the final-state particles must be in motion to conserve momentum. Consequently, at least some of the energy of the initial beam must reappear as kinetic energy of the final-state particles, and is therefore unavailable for particle production. In contrast, in the centre-of-mass frame the total momentum is zero, and in principle all the energy is available for particle production.

To find the centre-of-mass energy we use the expression

$$E_{\text{CM}}^2 = (P_t + P_b)^2 c^2, \quad (4.1)$$

where P is the particle's 4-momentum and the subscripts t and b refer to the target and beam, respectively.³ For a fixed-target experiment in the laboratory we have

$$P_t = (m_t c, \mathbf{0}), \quad P_b = (E_{\text{Lab}}/c, \mathbf{p}_b). \quad (4.2)$$

Expanding (4.1) gives

$$E_{\text{CM}}^2 = (P_t^2 + P_b^2 + 2P_t P_b) c^2 \quad (4.3)$$

and using $P_t^2 = m_t^2 c^2$, etc., together with the general result

$$P_i P_j = E_i E_j / c^2 - \mathbf{p}_i \cdot \mathbf{p}_j, \quad (4.4)$$

we have

$$E_{\text{CM}} = (m_b^2 c^4 + m_t^2 c^4 + 2m_t c^2 E_{\text{Lab}})^{1/2}. \quad (4.5)$$

At high energies this increases only like $E_{\text{Lab}}^{1/2}$ and so a decreasing fraction of the beam energy is available for particle production, the rest going to impart kinetic energy to the target.

In a *colliding-beam* accelerator, two beams of particles travelling in almost opposite directions are made to collide at a small or zero crossing angle. If for simplicity we assume the particles in the two beams have

²Nevertheless, even in particle physics, some important experiments are performed without using accelerators. For example, some of those described in Chapter 3 on neutrino oscillations used cosmic rays and nuclear reactors as their source of neutrinos. We will return to this in Section 4.5.

³A brief summary of relativistic kinematics is given in Appendix B.

the same mass and laboratory energy E_{Lab} and collide at a zero crossing angle, then the total centre-of-mass energy is

$$E_{\text{CM}} = 2E_{\text{Lab}}. \quad (4.6)$$

This increases linearly with the energy of the accelerated particles, and hence is a significant improvement on the fixed-target result. Colliding beam experiments are not, however, without their own disadvantages. The colliding particles have to be stable, which limits the interactions that can be studied and the collision rate in the intersection region is smaller than that achieved in fixed-target experiments, because the particle densities in the beams are low compared to a solid or liquid target.

In addition to its energy, the performance of an accelerator is characterised by its luminosity. This was defined in Section 1.6.2 and gives the rate for a reaction in terms of its cross-section. Luminosity is of particular importance for characterising the performance of a collider and it is worth remembering that the general formula for luminosity given in (1.58b) reduces in the case of a collider to the useful form (1.58c), i.e.

$$L = f N_1 N_2 / A, \quad (4.7a)$$

where N_i ($i = 1, 2$) are the numbers of particles in the colliding bunches, A is the cross-sectional area of the beam, and f is the frequency of collisions between bunches. An important related quantity is the *integrated luminosity* L_I , which is the integral of L over time,

$$L_I = \int L dt. \quad (4.7b)$$

Finally, details of the particles produced in the collision, for example their types and energies, are deduced by observing their interactions with the material of *detectors*, which are placed in the vicinity of the interaction region. A wide range of detectors is available. Some have very specific characteristics; others serve more than one purpose. Modern experiments, particularly in particle physics, typically use several types in a single experiment.

In this chapter we start by describing some of the different types of accelerator that have been built, the beams that they can produce and also how beams of neutral and unstable particles can be prepared. Then we discuss the ways in which particles interact with matter and review how these mechanisms are exploited in the construction of a range of particle detectors and how these individual detectors are combined to form complex detector systems. Uses of the latter are illustrated by examples from both accelerator and nonaccelerator experiments. The latter have been very important historically and have become increasingly so again, particularly in the field of neutrino physics, as was illustrated in Chapter 3.

4.2 Accelerators and beams

Accelerators use electromagnetic forces to boost the energy of stable charged particles. These are injected into the machine from a device that provides a high-intensity source of low-energy particles, for example an electron gun (a hot filament) or a proton ion source. The accelerators used for nuclear structure studies may be classified into those that develop a steady accelerating field (DC machines) and those in which radio frequency (rf) electric fields are used (AC machines). All accelerators for particle physics are of the latter type. We start with a brief description of DC machines.

4.2.1 DC accelerators

The earliest practical type of DC accelerator was the *Cockcroft–Walton machine*, in which ions passed through sets of aligned electrodes operated at successively higher potentials. These machines were limited to energies of about 1 MeV, but are still sometimes used as injectors as part of a multistage process of accelerating particles to higher energies.⁴

The most important DC machine in current use is the *Van de Graaff accelerator* and an ingenious version of this, known as the tandem Van de Graaff, which doubles the energy of the simple machine, is shown schematically in Figure 4.1. The key to this type of device is to establish a very high voltage. The Van de Graaff accelerator achieves this by using the fact that the charge on a conductor resides on its outer surface, and hence if a conductor carrying the charge touches another conductor it will transfer its charge to the outer surface of the second conductor.

In Figure 4.1, a high-voltage source at I passes positive ions to a belt via a comb arrangement at C. The belt is motor driven via the pulleys at P and the ions are carried on the belt to a second pulley where they are collected by another comb located within a metal vessel T. The charges are then transferred to the outer surface of the vessel, which acts as an extended terminal. In this way a high voltage is established on T. Singly-charged negative ions are injected from a source and accelerated along a vacuum tube towards T. Within T there is a stripper S (for example a thin carbon foil) that removes two or more electrons from the projectiles to produce positive ions. The latter then continue to accelerate through the second half of the accelerator, increasing their energy still further, and finally may be bent and collimated to produce a beam of positive ions. This brief account ignores many technical details. For example, an inert gas at high pressure is used to minimise electrical breakdown by the high voltage. The highest energy Van de Graaff accelerator can achieve a

⁴Sir John Cockcroft and Ernest Walton received the 1951 Nobel Prize in Physics for the development of their accelerator and the subsequent nuclear physics experiments they did using it.

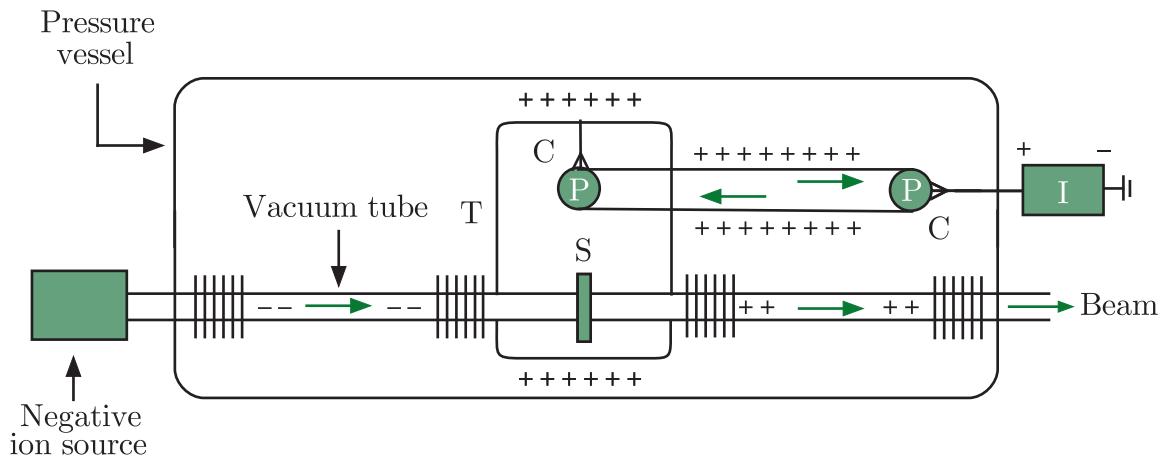


Figure 4.1 Principle of the tandem Van de Graaff accelerator (see text for a detailed description).

potential of about 30–40 MeV for singly-charged ions and greater if more than two electrons are removed by the stripper. It was an important tool for nuclear research for many years, but is now more used in fields such as nuclear medicine and condensed matter physics.

4.2.2 AC accelerators

Accelerators using rf electric fields may conveniently be divided into *linear* and *cyclic* varieties.

4.2.2(a) Linear accelerators

In a linear accelerator (or *linac*) for accelerating ions, particles pass through a series of metal pipes called *drift tubes* that are located in a vacuum vessel and connected successively to alternate terminals of an rf oscillator, as shown in Figure 4.2. Positive ions accelerated by the field move towards the first drift tube. If the alternator can change its direction as the ions pass through that tube, then they will be accelerated again

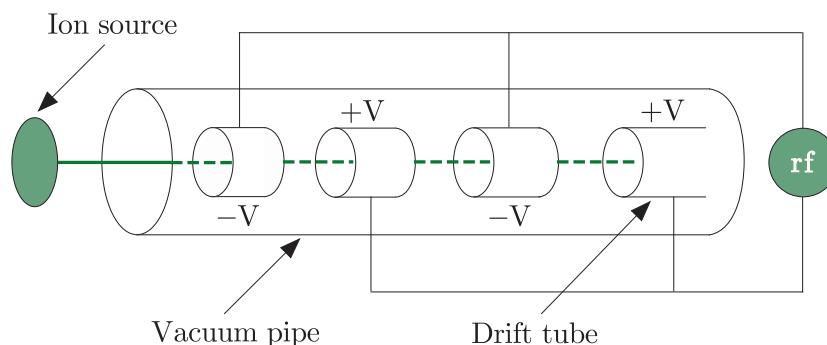


Figure 4.2 Acceleration in a linear ion accelerator.

on their way between the exit of the first and entry to the second tube, and so on. Thus the particles will form bunches and their final energy will be the sum of the energies received at the gaps between the drift tubes. Because the particles are accelerating, their speed is increasing and hence the lengths of the drift tubes have to increase to ensure continuous acceleration. To produce a useful beam the particles must keep in phase with the rf field and remain focused.

Proton linacs of this type are often used in medical physics applications and in particle physics as *injectors*; that is they produce proton beams of moderate energy that are injected into a more powerful machine, usually a synchrotron (described below), where they are accelerated to much higher energies. An example is the linac at J-PARC (Japan Proton Accelerator Research Complex), which accelerates protons to an energy of 3 GeV before they are injected into the main J-PARC ring, where they are further accelerated to 50 GeV. Many linear accelerators for ions also exist worldwide. An example is the Argonne Tandem Linear Accelerator System (ATLAS) at the Argonne National Laboratory, USA. ATLAS is a leading facility for nuclear structure research and was the world's first superconducting linear accelerator for ions. It can produce ion beams of any elements from hydrogen to uranium with energies as high as 17 MeV per nucleon and also secondary radioactive beams that are used, for example, to study reactions of interest to astrophysics.

For electrons, whose speed very rapidly approaches the speed of light, a variation of this method is used. In this case, the accelerator consists of a straight tube in the form of a series of cylindrical metal cavities. Power is fed to the accelerator from a series of devices called klystrons, which produce electromagnetic radiation in the form of microwave pulses that are transported via waveguides to the accelerator. There they generate an oscillating electric field pointing along the direction of the metal tube and a magnetic field in a circle around the interior of the accelerating tube. The magnetic field helps to keep the beam focused, and the frequency of the microwaves is adjusted so that the electrons arrive at each cavity of the accelerator at the optimal time to receive the maximum energy boost from the electric field. As long as this phase relationship can be maintained, the particles will be continuously accelerated. Many electron linacs exist worldwide. The largest ever built was the 3 km long SLC machine at the SLAC laboratory in Stanford, USA, which was operational until 2008. It had a maximum energy of 50 GeV and consisted of 80 000 copper cavities separated by copper discs with a small hole at the centre to transmit the beam.

An ingenious way of reducing the enormous lengths of high-energy linacs has been developed at the Continuous Electron Beam Accelerator Facility (CEBAF) at the Jefferson Laboratory in the USA. This utilises the fact that above about 50 MeV electron velocities are very close to the speed of light and thus electrons of very different energies can be accelerated in the same drift tube. Instead of a single long linac, the CEBAF machine consists of two much shorter linacs and the beam from

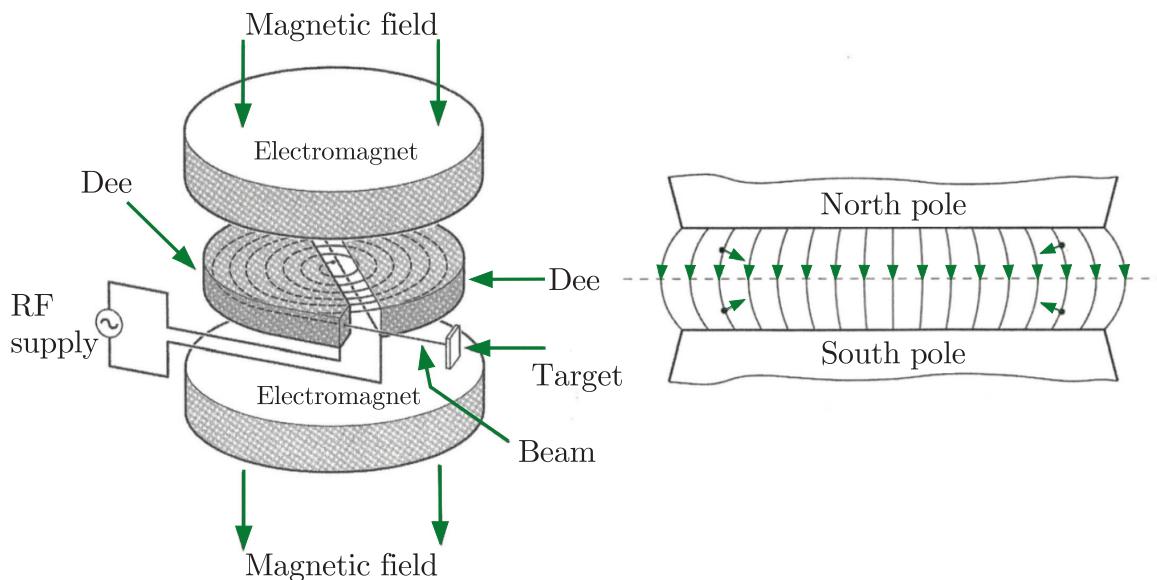


Figure 4.3 Schematic diagram of a cyclotron. Source: Adapted from Krane (1988). Copyright (1988) John Wiley & Sons, reprinted with permission.

one is bent and passed through the other. This can be repeated for up to four cycles. Even with the radiation losses inherent in bending the beams, very intense beams can be produced with energies up to 12.0 GeV. CEBAF has provided much important information in the energy region where nuclear physics and particle physics descriptions overlap.

4.2.2(b) Cyclic accelerators

Cyclic accelerators used for low-energy nuclear physics experiments are of a type called *cyclotrons*.⁵ They are also used to produce beams of particles for medical applications, including proton beams for radiation therapy.⁶ There are several types of cyclotron; to illustrate the principles, we will describe just one. This is shown schematically in Figure 4.3. The accelerator consists of two ‘dee’-shaped sections across which an rf electric field is established. Charged ions are injected into the machine near its centre and are constrained to traverse outward spiral trajectories by a magnetic field. The ions are accelerated each time they pass across the gap between the dees. At the maximum radius, which corresponds to the maximum energy, the beam is extracted. The shape of the magnetic field, which is also shown in Figure 4.3, ensures that forces act on particles not orbiting in the median plane to move them closer to this plane. This brief

⁵The cyclotron was invented by Ernest Lawrence, who received the 1939 Nobel Prize in Physics for this and the experimental work he did using it.

⁶This is discussed in Section 9.4.2.

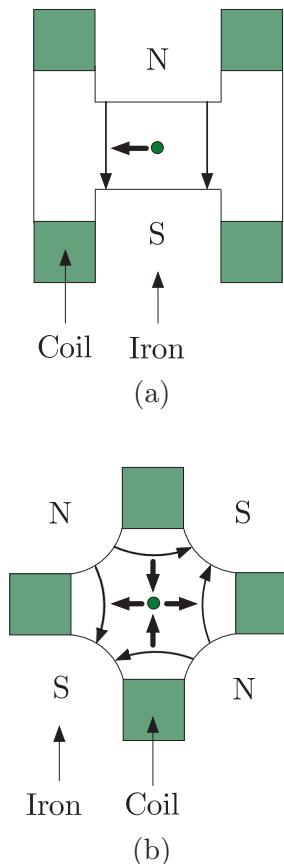


Figure 4.4 Cross-section of: (a) a typical bending (dipole) magnet; (b) a focusing (quadrupole) magnet. The thin arrows indicate field directions; the thick arrows indicate the force on a negative particle travelling into the paper.

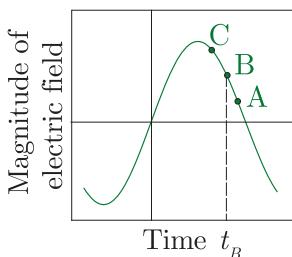


Figure 4.5 Magnitude of the electric field as a function of time at a fixed point in the rf cavity (see text for details).

description ignores the considerable problems that have to be overcome to ensure that the beam remains focused during the acceleration.

Cyclic accelerators used in particle physics are called *synchrotrons*. They operate in a somewhat different way to cyclotrons. The principle of a *synchrotron* is analogous to that of a linear accelerator, but where the acceleration takes place in a closed orbit rather than in a straight line. The beam of particles travels in an evacuated tube called the *beam pipe* and is constrained in a circular, or near circular, path by an array of dipole magnets called bending magnets (see Figure 4.4a). Acceleration is achieved as the beam repeatedly traverses one or more cavities placed in the ring where energy is given to the particles. Since the particles travel in a circular, or near circular, orbit they continuously emit radiation, called in this context *synchrotron radiation*. For a relativistic particle of mass m and a given energy, the energy loss⁷ is proportional to $1/m^4$. For electrons the loss is thus very severe, although it can be reduced by increasing the radius of the ring. The need to compensate for these losses by the input of large amounts of rf power limits the energies of electron synchrotrons. The highest energy was achieved by the LEP accelerator at the CERN Laboratory, Geneva, which was operational until 2000. It had a radius of over 4 km and accelerated electrons and positrons to an energy of approximately 100 GeV.

The momentum in GeV/c of an orbiting particle with unit charge is given by $p = 0.3 B \rho$, where B is the magnetic field in Tesla and ρ , the radius of curvature, is measured in metres. Because p is increased during acceleration, B must also be steadily increased if ρ is to remain constant, and the final momentum is limited both by the maximum field available and by the size of the ring. With conventional electromagnets, the largest field attainable over an adequate region is about 1.5 T, and even with superconducting coils it is only of order 10 T. Hence the radius of the ring must be very large to achieve very high energies. For example, the Tevatron accelerator, which operated at the Fermi National Laboratory, Chicago, until 2011, accelerated protons and antiprotons to energies of 1 TeV and had a radius of 1 km.

In the course of its acceleration, a beam may make typically 10^7 traversals of its orbit before reaching its maximum energy. Consequently, stability of the orbit is vital, both to ensure that the particles continue to be accelerated and that they do not strike the sides of the vacuum tube. In practice the particles are accelerated in bunches, each being synchronised with the rf field. In equilibrium, a particle increases its momentum just enough to keep the radius of curvature constant as the field B is increased during one rotation and the circulation frequency of the particle is in step with the rf of the field. This is illustrated in Figure 4.5. Particle B is assumed to be in the equilibrium orbit, synchronous with the rf field and arrives at time t_B . Particle A (C) is behind (ahead of)

⁷See, for example, p.661 of Jackson (1975).

the rf phase and receives a lower (higher) momentum increase from the field than particle B. This will reduce (increase) the radius of its orbit and, since its velocity $v \approx c$, increase (decrease) its rotational frequency relative to particle B. In practice, the particles remain in the bunch, but their trajectories oscillate about the stable orbits. These oscillations are controlled by a series of focusing magnets, usually of the quadrupole type, which are placed at intervals around the beam and act like optical lenses. A schematic diagram of one of these is shown in Figure 4.4b. Each focuses the beam in one direction and de-focuses it in the orthogonal direction, so alternate magnets have their field directions reversed to keep the particles in a stable orbit. With obvious changes, a similar principle is used in linear accelerators.

In addition to the energy of the beam, one is also concerned to produce a beam of high intensity, so that interactions will be plentiful. The intensity is ultimately limited by defocusing effects, such as the mutual repulsion of the particles in the beam, and a number of technical problems have to be overcome that are outside the scope of this brief account.

4.2.2(c) Fixed target machines and colliders

Both linear and cyclic accelerators can be divided into *fixed-target* and *colliding beam* machines. The latter are also known as *colliders*, or sometimes, in the case of cyclic machines, *storage rings*.⁸ In fixed-target machines, as mentioned previously, particles are accelerated to the highest operating energy and then the beam is extracted from the machine and directed on to a stationary target, which is usually a solid or a liquid. Much higher energies have been achieved for protons than electrons, because of the large radiation losses inherent in electron machines mentioned earlier. The intensity of the beam is such that large numbers of interactions can be produced, which can either be studied in their own right or used to produce secondary beams.

An interesting proton synchrotron (PS) for studies of nuclear physics and low-energy hadron physics is the COSY facility shown in Figure 4.6 located at the Research Centre Jülich, Germany. Low-energy protons are pre-accelerated in a cyclotron and then injected into a synchrotron, where they are further accelerated to momenta in the range 600–3700 MeV/c, corresponding to energies in the range 175–2880 MeV. The protons can be stored in the ring for appreciable times and are used in experiments, either by extracting the beam in the usual way or by allowing the circulating beam to interact with a very thin internal target. Thus we have a mixture of storage rings and fixed targets.

⁸The use of the terms *storage rings* and *colliders* as synonymous is not strictly correct, because the former can also describe machines that store a single beam for use on both internal and external fixed targets. An example is the COSY accelerator shown in Figure 4.6 and described in the text.



Figure 4.6 The COSY accelerator for studying nuclear and low-energy hadron physics. Source: Reproduced with permission of Forschungszentrum Jülich.

The main disadvantage of fixed-target machines for particle physics has been mentioned earlier: the need to achieve large centre-of-mass energies to produce new particles. Almost all new machines for particle physics are therefore colliders. The largest collider that has been built is the Large Hadron Collider (LHC), at CERN, Geneva, which became operational in the Summer of 2008. This is a massive pp accelerator of circumference 27 km and luminosity $L = 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, with each beam having an energy of 7 TeV. A section of the beam line in the LHC tunnel is shown in Figure 4.7 and a schematic diagram of the CERN site showing the LHC and some of the other accelerators there is shown in Figure 4.8. The acceleration process for the LHC starts with a linac whose beam is boosted in

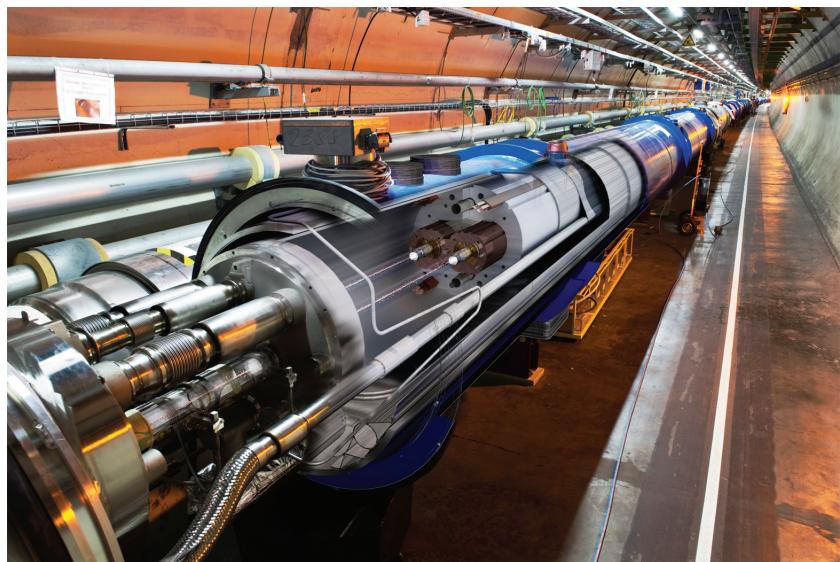


Figure 4.7 LHC tunnel and beam lines. The blue cylinders contain dipole magnets and the liquid helium system required to cool them to superconducting temperatures. The cut-away section is an artist's impression and shows the internal structure with the magnets and the two beam lines. Source: Reproduced by permission of CERN.

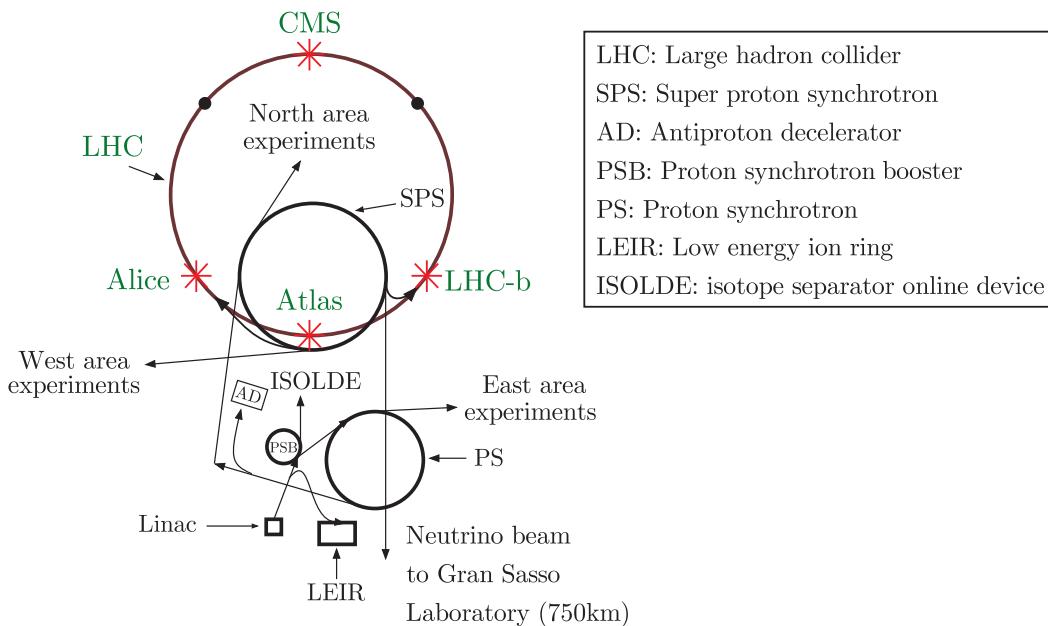


Figure 4.8 A schematic diagram of the CERN site showing the LHC and some of its other accelerators. Source: Reproduced by permission of CERN.

energy in the PSB (Proton Synchrotron Booster) and passed to the PS, a machine that is still the source of beams for lower-energy experiments. The beam energy is increased still further in the SPS (Super Proton Synchrotron) that also provides beams for a range of experiments as well as the injection beams for the LHC itself. As shown in Figure 4.8, there are four beam intersection points around the LHC ring, with a large detector (ALICE, CMS, LHC-b, and ATLAS) located at each. The extracted neutrino beam shown at the bottom of the diagram is sent to the Gran Sasso Laboratory in Italy 730 km away and is used for experiments on neutrino experiments, including oscillation experiments of the type discussed in Chapter 3.

Another important large collider currently operational is the Relativistic Heavy Ion Collider (RHIC), located at Brookhaven National Laboratory (BNL), USA. This machine, which began operation in 2000, was the first collider in the world capable of accelerating heavy ions. Like the LHC, there are several stages, involving a linac, a tandem Van de Graaff, and a synchrotron, before the ions are injected into the main machine. There they form two counter-circulating beams controlled by two 4 km rings of superconducting magnets and are accelerated to an energy of 100 GeV/nucleon. Thus the total centre-of-mass energy is 200 GeV/nucleon. Collisions occur at six intersection points, where major experiments are sited. RHIC primarily accelerates ions of gold and is used to study matter at extreme energy-densities, where a new state of matter called a ‘quark-gluon plasma’ is predicted to occur. This is discussed in Section 5.3.2.

4.2.2(d) Future accelerators

Research and development work for new accelerators is actively ongoing. In particle physics, an international collaboration is working on a proposal to build an enormous electron–positron collider using superconducting rf technology with the aim of complementing and extending the capabilities of the LHC. This machine, the International Linear Collider (ILC), would consist of two linear accelerators, each 15–25 km long, initially producing beams of 250 GeV electrons and positrons that would collide head-on 14 000 times per second. A later stage is envisaged that would double the beam energies. Work on designing the ILC has been ongoing for several years, as have discussions about finding a suitable site for it, but until the design is agreed and funding approved, there can be no definite timescale for the building work to begin and its eventual completion. An alternative design, exploiting new accelerator techniques, called the Compact Linear International Collider (CLIC), is being pursued at CERN. It would be built in stages and would eventually deliver energies up to 3 TeV, with a machine about the same length as the ILC and be built close to the CERN site. As for the ILC, there is much work to be done before any construction could possibly start. An alternative proposal to produce high energies using electron–positron collisions is based in China, where there are plans to build a very large Circular Electron Positron Collider (CEPC). This would have a circumference of about 54 km and produce electrons and positrons with energies of 240 GeV. The projected start date for construction is 2022, subject to funding being approved.

Another notable example is the FAIR (Facility for Antiproton and Ion Research in Europe) accelerator complex to be built near Darmstadt in Germany by an international consortium of European countries plus India. The heart of FAIR comprises two large accelerator rings (a double synchrotron) that can accelerate a wide range of ions from hydrogen ions (protons) to those of uranium. The primary high-energy beams will be used to create secondary beams of antiprotons and other stable and unstable (radioactive) ions. The secondary ion beams may either be used directly or diverted to a complex arrangement of further rings for additional experiments. Some areas that will be studied are: exotic states and the nature of quark binding in hadrons; highly compressed matter, as existed in the early universe and is found in the cores of very dense stellar objects such as neutron stars; and fundamental questions in nuclear structure. Very heavy ions such as uranium will be used to generate rare very short-lived nuclei that played a key role in the formation of elements in stars. Construction on the FAIR site started in 2017, with experiments expected to start in 2025. FAIR will be a leading centre for research in hadron and nuclear physics for many years to come.

4.2.3 Neutral and unstable particle beams

The primary particles used in accelerators must in general be stable and charged, but one is also interested in the interaction of neutral particles,

e.g. photons and neutrons, as well as those of unstable particles, such as charged pions and radioactive nuclei. Beams appropriate for performing such experiments are produced in a number of ways.

We have seen that neutrons are the natural product of many radioactive decays and we will see in Section 9.1.2 that a large flux of neutrons is present in a nuclear fission reactor. Typically these will have a spectrum concentrated at low energies of 1–2 MeV, but extending as high as 5–6 MeV. Purpose-built reactors exist for research purposes, such as the ILL reactor at the Institut Laue-Langevin, France. Another source of neutrons is via the spallation process. This usually starts with a beam of negative hydrogen ions that is prepared in a two-stage process involving initial acceleration in a linac followed by further acceleration in a circular machine, during which the electrons are stripped from the orbiting ions, leaving a bunched proton beam with an energy of several hundred MeV. This is then directed on to a heavy target and the resulting interaction drives out neutrons from the target and provides an intense pulsed source of neutrons.

Several spallation sources using different initial beam energies and different target materials are now available. An early example, built in 1995, is ISIS, located at the Rutherford Appleton Laboratory. The final proton beam at ISIS has an energy of 800 MeV and the target is the heavy metal tantalum. The most recent spallation source to be built is the China Spallation Neutron Source (CSNS), which, using 1.6 GeV protons and a solid tungsten target, plans to produce a 100 kW beam of neutrons. Currently, the most intense source of pulsed neutrons is provided by the Spallation Neutron Source (SNS) constructed in 2006 at Oak Ridge National Laboratory in the USA. This uses a target of liquid mercury. If beams of lower-energy neutrons are required these are produced by slowing down faster neutrons in *moderators*, which are materials with a large cross-section for elastic scattering but a small cross-section for absorption. In Section 9.1.2 we will see that moderators are vital for the successful safe extraction of power from fission nuclear reactors. The Spallation Neutron Source (SINQ) at the Paul Scherrer Institute in Switzerland produces extremely low energy neutrons that are used, for example, in attempts to detect the neutron electric dipole moment, which will be discussed in Section 10.5.3. SINQ produces continuous beams of neutrons, in contrast to the other sources above, which all produce pulsed beams.

Beams of unstable particles can be formed provided their constituents live long enough to travel appreciable distances in the laboratory. In particle physics one way of doing this is to direct an extracted primary beam on to a heavy target. In the resulting interactions with the target nuclei, many new particles are produced which, using electromagnetic fields, may then be analysed into secondary beams of well-defined momentum. Such beams will ideally consist predominantly of particles of one type, but if this cannot be achieved, then the wanted species may have to be identified by other means. Beams of radioactive ions are produced by two complementary methods. In one, the radioactive nuclei are first made in a target/ion source, then extracted in the form of ions, and, after selection

of a particular species by mass, using electromagnetic fields, they are accelerated in a second accelerator to the energy required for experiments. This method can produce high-quality beams comparable to those of stable particles, but the intensity of the beams is poor for short-lived nuclei with half-lives below milliseconds. A second method uses energetic beams of heavy ions impinging on a thin target. Interactions with the target nuclei results in fission or fragmentation, and the resulting nuclei leave the target with velocities close to those of the projectiles, so do not need further acceleration to transport them to the secondary target. However, because many different species are produced, reaction products have to be identified and tagged by mass, charge, and momentum and the secondary reactions have to be studied on an event-by-event basis. Both these methods will be incorporated in the FAIR complex mentioned above.

If the secondary beams are composed of unstable particles, they can themselves be used to produce further beams formed from their decay products. For example, if a high-energy beam of protons interacts with a heavy target, secondary particles will be produced, most of which will be pions. (Other possible particles produced are kaons, which have to be produced with another particle with the opposite strangeness quantum number to conserve strangeness. This is an example of so-called *associated production*.) A collimator can be used to select particles in a particular direction and the π^- component can subsequently be removed and focused into a mono-energetic beam by selective use of electrostatic fields and bending and focusing magnets. This beam of charged pions can be used to produce further secondary beams. For example, the π^- is unstable and decays almost entirely by the reaction $\pi^- \rightarrow \mu^- + \bar{\nu}_\mu$. Therefore, if the pions are passed down a long vacuum pipe, many will decay in flight to give muons and antineutrinos, which will mostly travel in essentially the same direction as the initial beam. The muons and any remaining pions can then be removed by passing the beam through a very long absorber or by deflection in a magnetic field, leaving the neutrinos. In this case the final neutrino beam will have a momentum spectrum reflecting the initial momentum spectrum of the pions, and since neutrinos are electrically neutral, no further momentum selection using magnets is possible.

4.3 Particle interactions with matter

In order to be detected, a particle must undergo an interaction with the material of a detector. In this section we discuss these interactions, but only in sufficient detail to be able to understand the detectors themselves.

The first possibility is that the particle interacts with an atomic nucleus. For example, this could be via the strong nuclear interaction if it is a hadron or by the weak interaction if it is a neutrino. We know from the work of Chapter 1 that both are *short-range interactions*. If the energy is sufficiently high, new particles may be produced, and such reactions are often the first step in the detection process. In addition

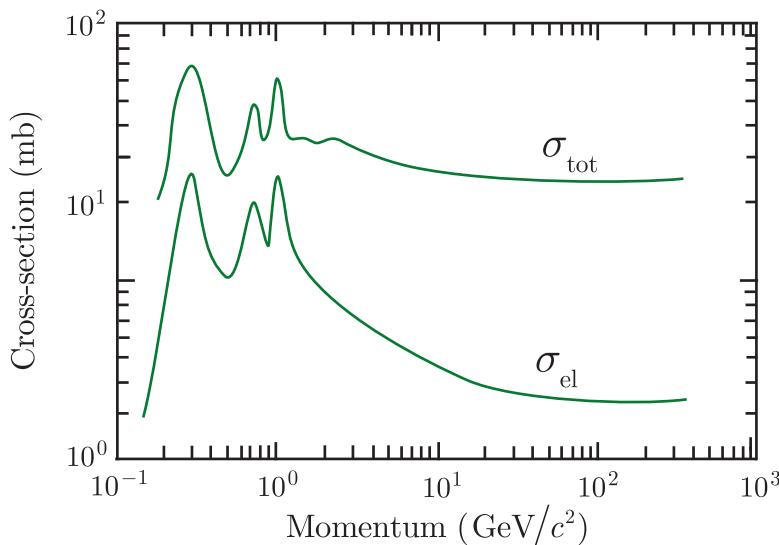


Figure 4.9 Total and elastic cross-sections for $\pi^- p$ scattering as functions of the pion laboratory momentum.

to these short-range interactions, a charged particle will also excite and ionise atoms along its path, giving rise to *ionisation energy losses*, and emit radiation, leading to *radiation energy losses*. Both of these processes are due to the long-range electromagnetic interaction. They are important because they form the basis of most detectors for charged particles. Photons are also directly detected by electromagnetic interactions, and at high energies their interactions with matter lead predominantly to the production of $e^+ e^-$ pairs via the *pair production* process $\gamma \rightarrow e^+ + e^-$ in the Coulomb fields of the nucleus and the atomic electrons.⁹ All these types of interactions are described in the following sections.

4.3.1 Short-range interactions with nuclei

For hadrons, the most important short-range interactions with nuclei are due to the strong nuclear force, which, unlike the electromagnetic interaction, is as important for neutral particles as for charged ones. Both elastic scattering and inelastic reactions may result. At high energies, many inelastic reactions are possible, most of them involving the production of several particles in the final state.

Many hadronic cross-sections show considerable structure at low energies due to the production of hadronic resonances, but at energies above about 3 GeV, total cross-sections are usually slowly varying in the range 10 – 50 mb and are much larger than the elastic cross-section. (The example of $\pi^- p$ scattering is shown in Figure 4.9.) This is of the same order-of-magnitude as the ‘geometrical’ cross-section $\pi r^2 \approx 30$ mb, where

⁹As noted at the beginning of Section 1.5.1, most primary processes, including $\gamma \rightarrow e^+ + e^-$, violate energy momentum conservation in free space, so that a further interaction with an electron or nucleus is needed to restore energy momentum conservation overall.

$r \approx 1 \text{ fm}$ is the approximate range of the strong interaction between hadrons. Total cross-sections on nuclei are much larger (see, for example, Figure 2.20), increasing roughly like the square of the nuclear radius, i.e. like $A^{2/3}$.

A special case is the detection of *thermal* neutrons (defined as those with kinetic energies below about 0.02 eV). We have seen in Chapter 2 that neutrons in this region have very large cross-sections for being absorbed, leading to the production of a compound nucleus that decays by delayed emission of a γ ray. Examples of these so-called ‘neutron activation reactions’ are $^{63}\text{Cu}(n, \gamma)^{64}\text{Cu}$ and $^{55}\text{Mn}(n, \gamma)^{56}\text{Mn}$.

The probability of a hadron–nucleus interaction occurring as the hadron traverses a small thickness dx of material is given by $n\sigma_{\text{tot}} dx$, where n is the number of nuclei per unit volume in the material. Consequently, the mean distance travelled before an interaction occurs is given by

$$l_c = 1/n\sigma_{\text{tot}}, \quad (4.8)$$

and is called the *collision length* or *mean free path*. An analogous quantity is the *absorption length*, defined by

$$l_a = 1/n\sigma_{\text{inel}}, \quad (4.9)$$

which governs the probability of an inelastic collision. As examples, the interaction lengths are between 10 and 40 cm for nucleons of energy in the range 100–300 GeV interacting with metals in the range aluminium to lead.

Neutrinos and antineutrinos can also be absorbed by nuclei, leading to reactions of the type

$$\bar{\nu}_\ell + p \rightarrow \ell^+ + X, \quad (4.10)$$

where ℓ is a lepton and X denotes any hadron or set of hadrons allowed by the conservation laws. Such processes are weak interactions (because they involve neutrinos) and so the associated cross-sections are extremely small compared to the cross-sections for strong interaction processes. The corresponding interaction lengths are therefore enormous. Nonetheless, in the absence of other possibilities such reactions are the basis for detecting neutrinos. Finally, photons can be absorbed by nuclei, giving *photoproduction* reactions such as $\gamma + p \rightarrow X$. However, these electromagnetic interactions are only used to detect photons at low energies, because at higher energies there is a far larger probability for e^+e^- pair production in the Coulomb field of the nucleus. We will return to this in Section 4.3.4 below.

4.3.2 Ionisation energy losses

Ionisation energy losses are important for all charged particles, and for particles other than electrons and positrons they dominate over radiation energy losses at all but the highest attainable energies. The theory of such losses, which are due predominantly to Coulomb scattering from the

atomic electrons, was worked out by Bethe, Bloch and others in the 1930s. The result is called the Bethe–Bloch formula, and for spin-0 bosons with charge $\pm q$ (in units of e), mass M , and velocity v , it takes the approximate form

$$-\frac{dE}{dx} = \frac{D q^2 n_e}{\beta^2} \left[\ln \left(\frac{2m_e c^2 \beta^2 \gamma^2}{I} \right) - \beta^2 - \frac{\delta(\gamma)}{2} \right], \quad (4.11)$$

where x is the distance travelled through the medium,

$$D = 4\pi \alpha^2 \hbar^2 / m_e = 5.1 \times 10^{-25} \text{ MeV cm}^2, \quad (4.12)$$

m_e is the electron mass, $\beta = v/c$, and $\gamma = (1 - \beta^2)^{-1/2}$. The other constants refer to the properties of the medium: n_e is the electron density; I is the mean ionisation potential of the atoms averaged over all electrons, and is given approximately by $I = 10 Z \text{ eV}$ for Z greater than 20; and δ is a dielectric screening correction, which is important only for highly relativistic particles. The corresponding formula for spin-1/2 particles differs from this, but in practice the differences are small and may be neglected in discussing the main features of ionisation energy losses.

It is common practice to absorb the density ρ of the medium by dividing (4.11) by ρ so that

$$-\frac{1}{\rho} \frac{dE}{dx} \rightarrow -\frac{dE}{dx} \quad (4.13)$$

and expressing dE/dx in terms of an equivalent thickness in g cm^{-2} . Examples of the behaviour of $-dE/dx$ for muons, pions, and protons traversing a range of materials are shown in Figure 4.10, using this convention. As can be seen, $-dE/dx$ falls rapidly as the velocity increases from zero because of the $1/\beta^2$ factor in the Bethe–Bloch equation. All particles have a region of ‘minimum ionisation’ for $\beta\gamma$ in the range 3–4. Beyond this, β tends to unity and the logarithmic factor in the Bethe–Bloch formula gives a ‘relativistic rise’ in $-dE/dx$.

The magnitude of the energy loss depends on the medium. The electron density is given by $n_e = \rho N_A Z/A$, where N_A is Avogadro’s number, and ρ and A are the mass density and atomic weight of the medium, so the mean energy loss is proportional to the density of the medium. The remaining dependence on the medium is relatively weak because $Z/A \approx 0.5$ for all atoms except the very light and the very heavy elements, and because the ionisation energy I only enters the Bethe–Bloch formula logarithmically. In the ‘minimum ionisation’ region where $\beta\gamma \approx 3 – 4$, the minimum value of $-dE/dx$ can be calculated from (4.11) and for a particle with unit charge is given approximately by

$$\frac{1}{\rho} \left(-\frac{dE}{dx} \right)_{\min} \approx 3.5 \frac{Z}{A} \text{ MeV g}^{-1} \text{cm}^2. \quad (4.14)$$

Ionisation losses are proportional to the squared charge of the particle, so that a fractionally charged particle with $\beta\gamma \geq 3$ would have a much lower rate of energy loss than the minimum energy loss of any integrally

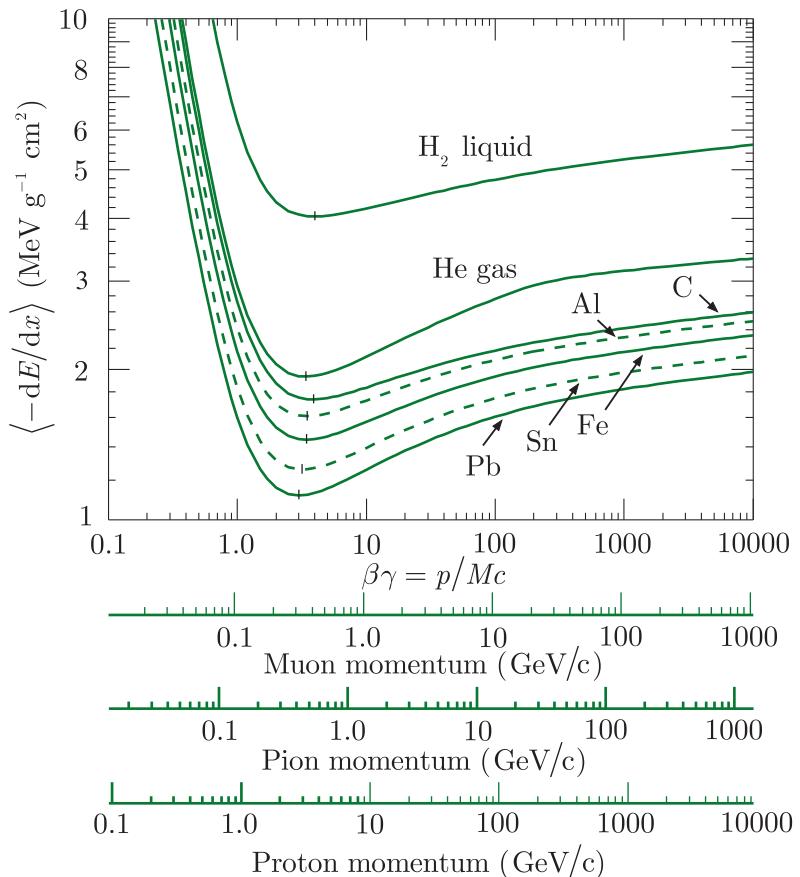


Figure 4.10 Ionisation energy loss for muons, pions, and protons on a variety of materials. See the text for an explanation of the units of dE/dx . Source: Adapted from Tanabashi et al. (Particle Data Group) (2018).

charged particle. This has been used as a means of identifying possible free quarks, but without success.

From the knowledge of the rate of energy loss, we can calculate the energy attenuation as a function of distance travelled in the medium. This is called the *Bragg curve*. The maximum ionisation loss occurs near the end of the path where the speed is smallest and the curve has a pronounced peak (the *Bragg peak*) close to the endpoint before falling rapidly to zero at the end of the particle's path length. For particles whose energy loss is dominated by ionisation, the *range* R , i.e. the mean distance a particle travels before it comes to rest, is given by

$$R = \int_0^R dx = \int_0^{\beta_{\text{initial}}} \left[-\frac{dE}{dx} \right]^{-1} \frac{dE}{d\beta} d\beta = \frac{M}{q^2 n_e} F(\beta_{\text{initial}}), \quad (4.15)$$

where F is a function of the initial ‘scaled’ velocity β_{initial} and we have used the relation $E = \gamma Mc^2$ to show the dependence on the projectile mass M . For nonrelativistic particles, $\beta_{\text{initial}} \ll 1$, the rate of energy loss is very high and the particle comes to rest very quickly.

The range as given by (4.15) is actually an average value because scattering is a statistical process and there will therefore be a spread of

values for individual particles. The spread is greater for light particles and smaller for heavier particles such as α particles. These properties have implications for the use of radiation in therapeutic situations, where it may be necessary to deposit energy within a small region at a specific depth of tissue, for example to precisely target a cancer. This and other biological effects of radiation are discussed in Section 9.4.

Because neutrons are uncharged, direct detection is not possible by ionisation methods. However, they can be detected via the action of the charged products of induced direct nuclear reactions. Commonly used reactions are ${}^6\text{Li}(n, \alpha){}^3\text{H}$, ${}^{10}\text{B}(n, \alpha){}^7\text{Li}$, and ${}^3\text{He}(n, p){}^3\text{H}$. All these reactions are exothermic and so are very suitable for detecting neutrons with energies below about 20 MeV. Moreover, as nuclear cross-sections tend to increase like v^{-1} at low energies, detection becomes more efficient the slower the neutron.

4.3.3 Radiation energy losses

When a charged particle traverses matter it can lose energy by radiative collisions, especially with nuclei. The electric field of a nucleus will accelerate and decelerate the particles as they pass, causing them to radiate photons, and hence lose energy. This process is called *bremsstrahlung* (literally ‘braking radiation’ in German) and is a particularly important contribution to the energy loss for electrons and positrons.

The dominant Feynman diagrams for electron bremsstrahlung in the field of a nucleus, i.e.

$$e^- + (Z, A) \rightarrow e^- + \gamma + (Z, A), \quad (4.16)$$

are shown in Figure 4.11, where it can be seen that they involve three electromagnetic vertices, one with the nuclear charge Ze , and so their contributions are of order $Z^2\alpha^3$. There are also contributions from bremsstrahlung in the fields of the atomic electrons, each of order α^3 . Since there are Z atomic electrons for each nucleus, these give a total contribution of order $Z\alpha^3$, which is small compared to the contribution from the nucleus for all but the lightest elements. A detailed calculation shows that for relativistic electrons with $E \gg mc^2/\alpha Z^{1/3}$, the average rate of energy loss is given by

$$-\frac{dE}{dx} = E/L_R. \quad (4.17)$$

The constant L_R , the *radiation length*, is a function of Z and n_a , the number density of atoms in the medium, and is proportional to m_p^2 for an arbitrary charged particle of mass m_p . Integrating (4.17) gives

$$E = E_0 \exp(-x/L_R), \quad (4.18)$$

where E_0 is the initial energy. It follows that the radiation length is the average thickness of material that reduces the mean energy of an electron or positron by a factor e . For example, the radiation length in lead is 0.56 cm.

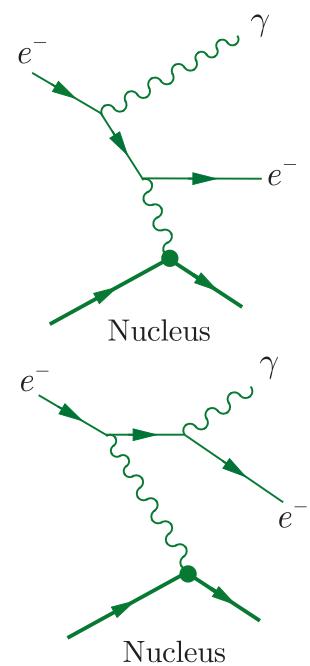


Figure 4.11 Dominant Feynman diagrams for the bremsstrahlung process $e^- + (Z, A) \rightarrow e^- + \gamma + (Z, A)$.

Radiation losses are proportional to E/m_P^2 . On the other hand, ionisation energy losses are only weakly dependent on the projectile mass and energy at very high energies. Consequently, except at relatively low energies, radiation energy losses completely dominate ionisation energy losses for electrons and positrons,¹⁰ but are much smaller than ionisation losses for all particles other than electrons and positrons at all but the highest attainable energies.

Taking into account the above and the results of Section 4.3.2, we see that at low energies particles with the same kinetic energy but different masses can have substantially different ranges. Thus, for example, an electron of 5 MeV has a range that is several hundred times that of an alpha particle of the same kinetic energy.

4.3.4 Interactions of photons in matter

In contrast to heavy charged particles, photons have a high probability of being absorbed or scattered through large angles by the atoms in matter. Consequently, a collimated mono-energetic beam of I photons per second traversing a thickness dx of matter will lose

$$dI = -I \frac{dx}{\lambda} \quad (4.19)$$

photons per second, where

$$\lambda = (n_a \sigma_\gamma)^{-1} \quad (4.20)$$

is the mean free path before absorption or scattering out of the beam and σ_γ is the total photon interaction cross-section with an atom. Integrating (4.19) gives

$$I(x) = I_0 e^{-x/\lambda} \quad (4.21)$$

for the intensity of the beam as a function of distance, where I_0 is the initial intensity.

The main processes contributing to σ_γ are *Rayleigh scattering*, in which the photon scatters coherently from the atom; the *photoelectric effect*, in which the photon is absorbed by the atom as a whole with the emission of an electron; *Compton scattering*,¹¹ where the photon scatters from an atomic electron; and *electron–positron pair production* in the field of a nucleus or of an atomic electron. The corresponding cross-sections on carbon and lead are shown in Figure 4.12, where it can be seen that above a few MeV the cross-section is dominated by pair production from the nucleus. The pair production process is closely related to electron

¹⁰In condensed matter with atomic number Z , for all except the lightest elements, the critical energy E_C at which radiation energy losses begin to exceed ionisation energy losses for electrons and positrons is given approximately by $E_C = 600/Z$ MeV.

¹¹Arthur Compton shared the 1927 Nobel Prize in Physics for the discovery of the increase in wavelength that occurs when photons with energies of around 0.5–3.5 MeV interact with electrons in a material – the original *Compton effect*.

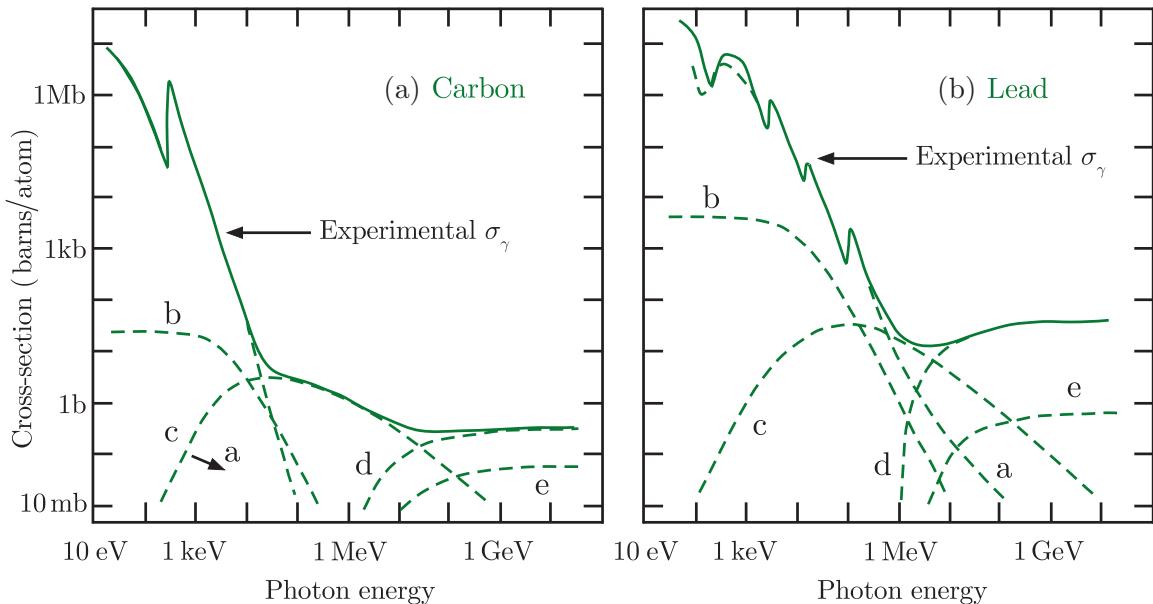


Figure 4.12 Total experimental photon cross-section σ_γ on (a) a carbon atom and (b) a lead atom, together with the contributions from (a) the photoelectric effect, (b) Rayleigh (coherent atomic) scattering, (c) Compton scattering, (d) pair production in the field of the nucleus, and (e) pair production in the field of the atomic electrons. Source: Adapted from Tanabashi et al. (Particle Data Group) (2018).

bremsstrahlung, as can be seen by comparing the Feynman diagrams shown in Figures 4.11 and 4.13. Both are third-order with one vertex involving a nucleus with charge Ze .

The cross-section for pair production rises rapidly from threshold, and is given to a good approximation by

$$\sigma_{\text{pair}} = \frac{7}{9} \frac{1}{n_a L_R}, \quad (4.22)$$

for $E_\gamma \gg mc^2/\alpha Z^{1/3}$. Substituting (4.20) and (4.22) into (4.21), gives

$$I(x) = I_0 \exp(-7x/9L_R), \quad (4.23)$$

so that at high energies, photon absorption, like electron radiation loss, is characterised by the radiation length L_R .

4.3.5 Ranges and interaction lengths

For nonrelativistic charged particles ($\beta \ll 1$), ionisation energy losses are very large and the particles come to rest relatively quickly, as we saw in Section 4.3.2. However, as β becomes larger, the range (4.15) becomes comparable with the collision and absorption lengths (4.8) and (4.9) and the radiation length L_R defined in (4.17). The situation is complicated, but for highly relativistic particles, a fairly simple picture emerges for

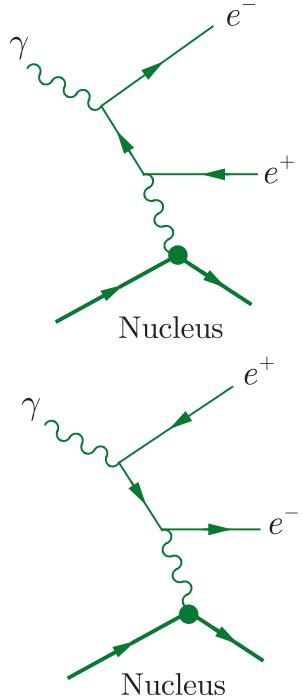


Figure 4.13 The pair production process $\gamma + (Z, A) \rightarrow e^- + e^+ + (Z, A)$.

any given material. We shall briefly summarise the situation here by discussing classes of particles in order of their increasing penetration power in condensed matter, since this is an important factor in the design of experiments at high energies, as we shall see in Section 4.5.

4.3.5(a) *Electrons and photons*

For energies above 1 GeV, these are the least penetrating particles, whose interactions are controlled by their radiation lengths.

4.3.5(b) *Hadrons*

Radiation energy losses are unimportant for charged hadrons, because they are suppressed, relative to those of electrons, by the square of their mass. Their strong interactions with atomic nuclei are controlled by their interaction lengths, which are typically much larger than the radiation lengths for electrons in the same material.

4.3.5(c) *Muons*

Again, radiation energy losses are suppressed by the square of their mass and muons have no strong interactions with nuclei. They are easily the most penetrating of all charged particles.

4.3.5(d) *Neutrinos*

As these only interact by the weak interaction, they can pass through enormous distances of matter without interacting, as discussed earlier.

4.4 Particle detectors

The detection of a particle means more than simply its localisation. To be useful this must be done with a resolution sufficient to enable particles to be separated in both space and time in order to determine which are associated with a particular event. We also need to be able to identify each particle and measure its energy and momentum. No single detector is optimal with respect to all these requirements, although some are multifunctional. For example, calorimeters, primarily used for making energy measurements, can also have very good space and time resolution. Many of the devices discussed below are commonly used in both nuclear and particle physics, but in the former a small number of types of detector is often sufficient, whereas in particle physics, both at fixed-target machines and colliders, modern experiments commonly use very large multicomponent detectors, which integrate many different subdetectors in a single device. In this section we will briefly introduce some of the most important individual detectors currently operational, but detector development is a

rapidly moving area of research and new devices are frequently developed, so the list below is by no means exhaustive.¹²

We start by discussing the large family of *gaseous detectors*, which convert the ionisation produced by the passage of a charged particle through a gas into an electronic signal. Such detectors are primarily used to provide accurate measurements of a particle's position or, by a sequence of such measurements, a record of the particle's trajectory. In this context they are also called *tracking detectors*.

Tracking detectors are very often placed in a magnetic field, in which case they can provide a measurement of the particle's momentum from the resulting curvature of its track. An apparatus that is dedicated to measuring momentum is called a *spectrometer*. It consists of a magnet and a series of detectors to track the passage of the particles. The precise design depends on the nature of the experiment being undertaken. For example, in a fixed-target experiment at high energies, the reaction products are usually concentrated in a narrow cone about the initial beam direction, whereas in colliding beam experiments spectrometers must completely surround the interaction region to obtain full angular coverage. However, in this case the beam will also be deflected, and so at colliders so-called 'compensating magnets' are added to correct for this.

Next we discuss four more types of charged particle detectors: *scintillation counters*, *solid-state detectors*, *Čerenkov counters*, and *transition radiation detectors (TRDs)*. Scintillation counters have excellent time resolution and are sometimes used for 'triggering' other devices in multicomponent detector systems, i.e. to decide whether or not to activate other detectors, or whether to record the information from a particular event. Solid-state detectors exploit the properties of semiconductors. They are in some respects the solid-state analogue of gas detectors and have, to some extent, replaced the latter in current experiments. Čerenkov counters measure the velocity of a charged particle and can be used to distinguish between different particles having a given very high momentum by using their velocities to determine their masses. Transition radiation detectors utilise the fact that electromagnetic radiation is emitted when a charged particle crosses the boundary between two materials, and the probability of emission depends on the relativistic γ -factor. For a given energy, this allows lighter particles to be distinguished from heavier ones of the same energy.

All the above detectors can only detect charged particles and all leave the nature of the particle unchanged as it passes through the detector. In contrast, the final detectors we discuss, *calorimeters*, can detect both neutral and charged particles. They totally absorb the detected particle to yield a measurement of its energy. Calorimeters can also have good

¹²More detailed discussions of particle detectors are given in the references of Footnote 1.

spatial and time resolutions, which is particularly important for neutral particles, which often cannot be detected in any other way.

4.4.1 Gaseous ionisation detectors

Most gaseous detectors detect the ionisation produced by the passage of a charged particle through a gas, typically an inert one such as argon, either by collecting the ionisation products or induced charges on to electrodes or (historically) by making the ionisation track visible in some form. The average energy needed to produce an electron–ion pair is 30 ± 10 eV, with a weak dependence on the gas used and the energy of the incident particle. In practice, the output is a pulse at the anode, which is amplified by electronic means. For a certain range of applied voltages, the so-called ‘proportional region’ (see below), these devices are primarily used to provide accurate measurements of a particle’s position. As position detectors, gaseous detectors largely replaced earlier detectors that used visual techniques, such as cloud chambers, bubble chambers, and stacks of photographic emulsions. Although historically important, none of these visual devices are now in general use, and have been superceded by electronic detectors.¹³

To understand the principles of gaseous detectors we refer to Figure 4.14, which shows the number of ion pairs produced per incident charged particle (the *gas amplification factor*) as a function of the applied voltage V for two cases: a heavily ionising particle (e.g. an alpha particle – upper curve) and a lightly ionising particle (e.g. an electron – lower curve).

4.4.1(a) Ionisation chamber

At low applied voltages, the output signal is very small because electron–ion pairs recombine before reaching the electrodes, but as the voltage increases the number of pairs increases to a saturation level representing complete collection. This is the region of the *ionisation chamber*. The simplest type of chamber is a parallel plate condenser filled with an inert gas and having an electric field $E = V/d$, where d is the distance between the plates. In practice, the gas mixture must contain at least one ‘quenching’ component that absorbs ultraviolet light and stops a plasma forming and spreading throughout the gas.

¹³These early detector techniques produced many notable discoveries and their importance has been recognised by the award of five Nobel Prizes in Physics: a share of the 1927 Prize to Charles Wilson for the invention and use of the cloud chamber; the 1948 Prize to Patrick Blackett for further developments of the cloud chamber and discoveries made with it, the 1950 Prize to Cecil Powell for development of the photographic emulsion technique and its use to discover pions, the 1960 Prize to Donald Glaser for the invention of the bubble chamber, and the 1968 Prize to Luis Alvarez for developing the bubble chamber and associated data analysis techniques, resulting in the discovery of a large number of hadronic resonances.

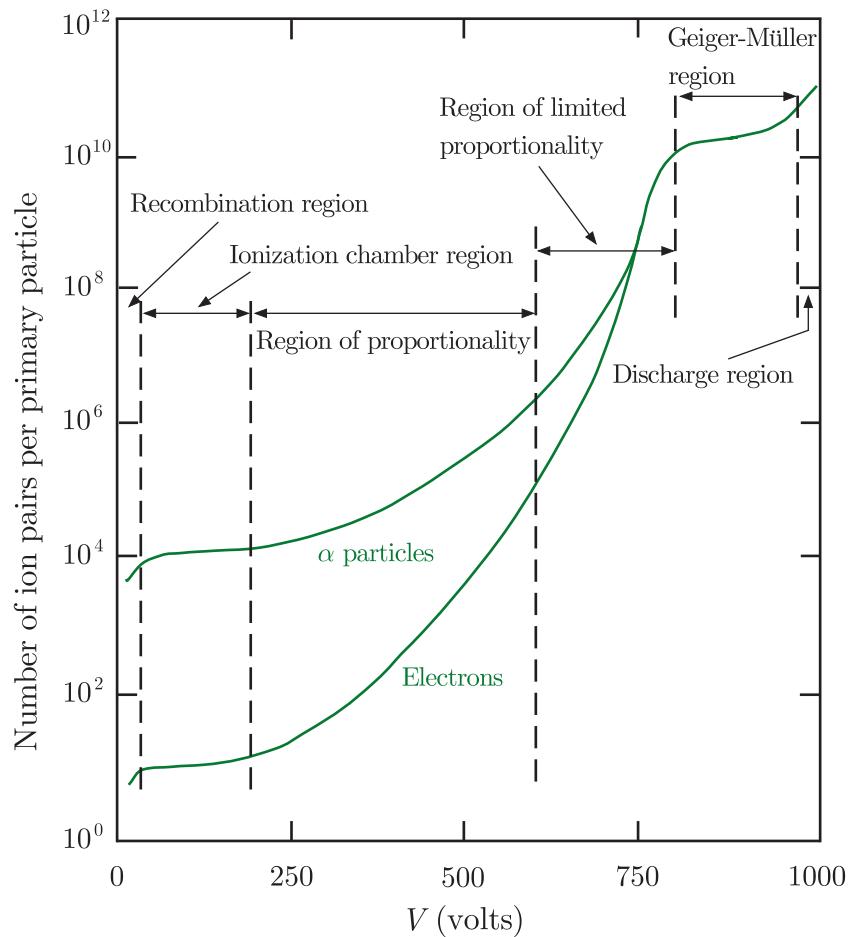


Figure 4.14 Gas amplification factor as a function of voltage V applied in a single-wire gas detector, with a wire radius typically $20\ \mu\text{m}$, for strongly ionising particles (α particles) and weakly ionising particles (electrons).

Another arrangement is cylindrical with an inner anode of radius r_a and an outer cathode of radius r_c , giving an electric field

$$E(r) = \frac{V}{r \ln(r_c/r_a)} \quad (4.24)$$

at a radial distance r from the centre of the anode wire. The output signal is proportional to the number of ions formed and hence the energy deposited by the radiation, but is independent of the applied voltage. However, the signal is very small compared to the noise of all but the slowest electronic circuits and requires considerable amplification to be useful. Overall, the energy resolution and the time resolution of the chamber are relatively poor and ionisation chambers are of very limited use in recording individual pulses. They are used, for example, as beam monitors, where the particle flux is very large, and in medical environments to calibrate radioactive sources.

As mentioned previously, neutrons cannot be directly detected by ionisation methods, but neutron flux measurements can be made with

ionisation chambers (or proportional chambers, see below) filled with BF_3 by utilising the neutron activation reactions mentioned in Section 4.3.1.

4.4.1(b) Wire chambers

If the voltage is increased beyond the region of operation of the ionisation chamber, we move into the *proportional region*. In this region, a cylindrical arrangement as used in the ionisation chamber will produce electric field strengths of order $(10^4 - 10^5)$ V/cm near the wire and this is strong enough for electron–ion pairs released in the primary ionisation to gain sufficient energy to cause secondary ionisation. The rapid increase in amplification due to secondary ionisation is called a *Townsend avalanche*. The output signal at the anode is still proportional to the energy lost by the original particle. There are a number of different types of device working in the proportional region and they are sometimes generically referred to as *track chambers* or simply *wire chambers*.

The earliest detector using this idea was the *proportional counter*, which consists of a cylindrical tube filled with a gas (again a quenching component in the gas is required) and maintained at a negative potential, and a fine central anode wire at a positive potential. Again, neutrons can be detected indirectly by using the direct nuclear reaction ${}^3\text{He}(n, p){}^3\text{H}$ mentioned in Section 4.3.2 in a proportional chamber filled with a mixture of ${}^3\text{He}$ and krypton. Subsequently, the resolution of proportional counters was greatly improved as a result of the discovery that if many anode wires were arranged in a plane between a common pair of cathode plates, each wire acts as an independent detector. This device is called a *multiwire proportional chamber* (MWPC) and was introduced in 1968.¹⁴ An MWPC can achieve spatial resolutions of $200\ \mu\text{m}$ or less, and has a typical time resolution of about 3 ns.

A schematic diagram of an MWPC is shown in Figure 4.15. The planes (a) have anode wires into the page and those in plane (b) are at right angles. The wire spacings are typically 2 mm. The cathodes are the faces of the chambers. A positive voltage applied to the anode wires generates a field as shown in the upper corner. A particle crossing the chamber ionises the gas and the electrons drift along the field lines to the anode wires. In this particular example, there would be signals from one wire in the upper (a) chamber and two in the lower (c) chamber.

Even better spatial resolutions are obtained in a related device called a *drift chamber*, which has now largely replaced the MWPC as a general detector. This uses the fact that the liberated electrons take time to drift from their point of production to the anode. Thus the time delay between the passage of a charged particle through the chamber and the creation of a pulse at the anode is related to the distance between the particle trajectory and the anode wire. In practice, additional wires are incorporated to

¹⁴The MWPC was invented by Georges Charpak and for this and other developments in particle detectors he was awarded the 1992 Nobel Prize in Physics.

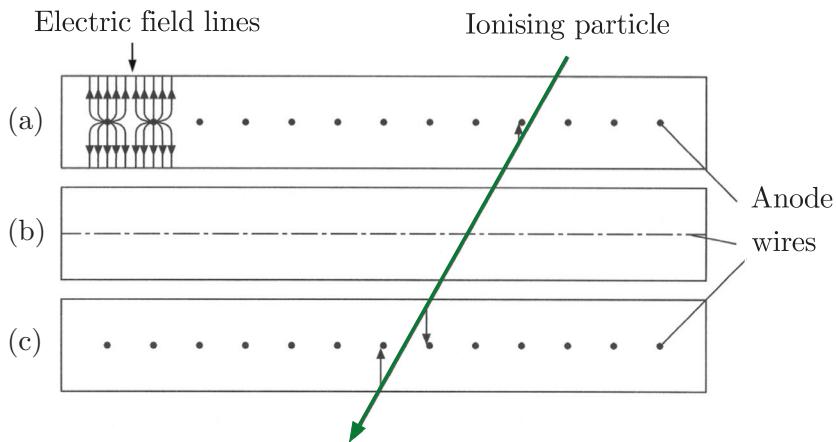


Figure 4.15 A group of three planes of an MWPC (see text for details). Source: From Povh et al. (1999), with kind permission of Springer Science and Business Media.

provide a relatively constant electric field in each cell in a direction transverse to normal incidence. A reference time has to be defined, which, for example, could be done by allowing the particle to pass through a scintillator positioned elsewhere in the experiment. (Scintillation counters are discussed in Section 4.4.2 below.) The electrons drift for a time and are then collected at the anode, thus providing a signal that the particle has passed. If the drift time can be measured accurately (to within a few ns) and if the drift velocity is known, then spatial and temporal resolutions similar to those of the MWPC can easily be achieved, and specialised detectors can reduce this still further.

Drift chambers are constructed in a variety of geometries to suit the nature of the experiment, and arrangements where the wires are in planar, radial, or cylindrical configurations have all been used. The latter type is also called a ‘jet chamber’. Such a chamber was used by the JADE collaboration at an e^+e^- collider at DESY, Hamburg. It was a cylindrical array of drift chambers with the beam direction as the axis and the collision region, at which the e^+e^- interactions occurred, at the centre. The anode wires ran parallel to the axis and the whole detector was divided into 24 segments, with 64 anode wires in each. A ‘two-jet’ event in this jet chamber was shown in Figure 3.9 as evidence for the existence of quarks.

One of the most advanced applications of proportional and drift chamber principles is embodied in the *time projection chamber* (TPC) and an example that illustrates its mode of operation is shown schematically in Figure 4.16. This example consists of a cylindrical barrel, typically 2 m long and 1 m in diameter, surrounding the beam pipe of a collider. At each end of the chamber is a segmented layer of proportional counters. The electric drift field \mathbf{E} , due to a negative high-voltage electrode plane at the centre of the chamber, and a strong magnetic field \mathbf{B} are aligned parallel and antiparallel to each other in the two sections of the chamber with respect to the axis of the cylinder. Electrons formed along the track of an ionising particle emerging from the interaction point at the centre of the barrel drift under the action of the electric field towards one of the

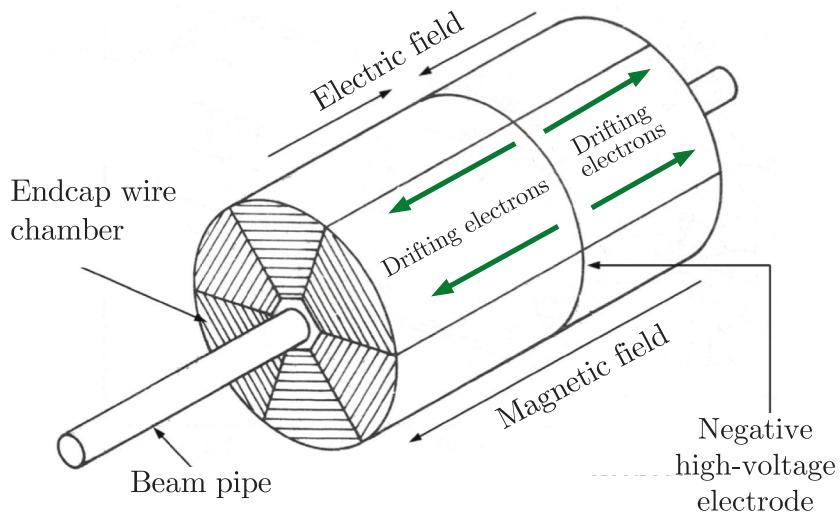


Figure 4.16 Schematic diagram of a time projection chamber. Source: Adapted from Kleinknecht (1986). Copyright Cambridge University Press, reprinted with permission.

endcaps along helical trajectories whose direction is parallel to the axis of the barrel. Their locations are measured by a set of anode wires located between rectangular cathodes in the endcaps. The remaining third coordinate necessary to reconstruct the position of a point on the track is found from the time it takes for the electrons to drift from the point of production to the endcaps where they are detected. The TPC has excellent spatial resolution and has been used in e^+e^- annihilation experiments. A TPC is also at the heart of the STAR detector at the RHIC heavy-ion collider, which we will use as an illustration of multicomponent detector systems in Section 4.5.

Finally, a more robust form of chamber, in which the wires are replaced by conductive metal strips on a printed circuit board, is the *microstrip gaseous chamber* (MSGC). This is being incorporated in experiments running at the new generation of accelerators currently operational or being planned.

4.4.1(c) Resistive plate chamber

In a single-gap resistive plate chamber, gas at atmospheric pressure is contained in the narrow space between two parallel plates of high resistivity, typically glass or bakelite, and a high potential difference is maintained between the two plates. As shown in Figure 4.17, the backs of the two electrodes are coated with a lower resistivity material and beyond these are read-out strips. When a charged particle passes through the detector it produces ionisation, which drifts towards the electrodes, producing an avalanche. The high resistivity of the electrodes ensures that the discharge does not spread out too much and the read-out strips on the two electrodes are orientated at right angles to each other, enabling the particle to be located in the plane of the detector. Resistive plate chambers have good time resolution. With gaps of order 0.2 mm, resolutions of order

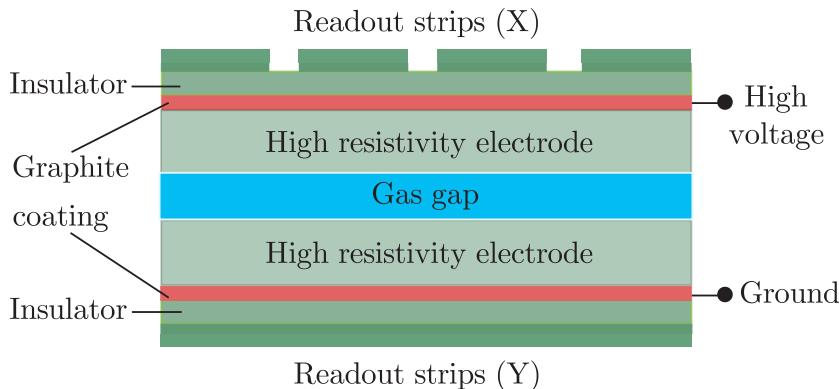


Figure 4.17 Schematic diagram of a resistive-plate chamber.

100 ps can be achieved in multigap detectors, while detectors with larger gaps, of order 2 mm, are used to trigger other detectors, i.e. its signal is used to decide whether or not to record information from them. In this respect it is a cheaper alternative to the scintillation counters discussed below when large areas need to be covered.

4.4.1(d) Beyond the region of proportionality

Referring again to Figure 4.14, by increasing the external voltage still further one moves into a region where the output signal ceases to be proportional to the number of ion pairs produced and hence the incident energy. This is the region of *limited proportionality*. In this region a type of gaseous detector called a *streamer tube* operates. Eventually the process runs out of control and we enter the *Geiger–Müller region* where the output signal is independent of the energy lost by the incident particle. In this region a quenching agent is not used. Detectors working in this region are called Geiger–Müller counters. Physically they are similar to the simple cylindrical proportional counter and are widely used as portable radiation monitors in the context of health and safety regulations.

For completeness, we can mention that if the gas amplification factor is taken beyond the Geiger–Müller region, the avalanche develops moving plasmas or streamers. Recombination of ions then leads to visible light that can be made to generate an electrical output. Eventually complete breakdown occurs and a spark is emitted as the incident particle traverses the gas. Detectors in this region, called *streamer* and *spark chambers* (of parallel plate construction, rather than cylindrical), were widely used in the 1970s and 1980s and played an important role in hadron physics, but are no longer in general use.

4.4.2 Scintillation counters

For charged particles we have seen that energy losses occur due to excitation and ionisation of atomic electrons in the medium of the detector. In

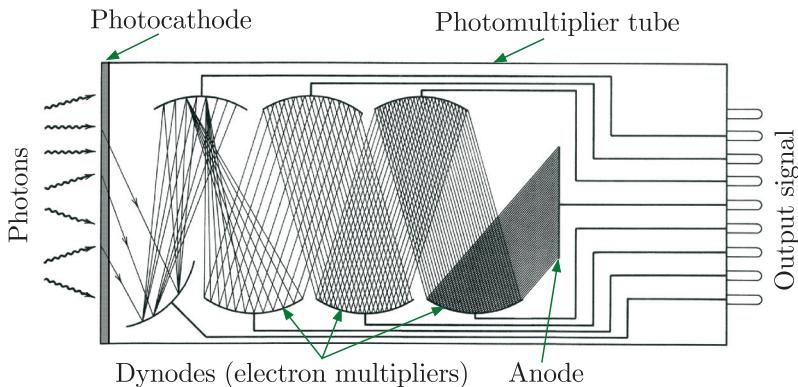


Figure 4.18 Schematic diagram of the main elements of a photomultiplier tube.
Source: Adapted from Krane (1988). Copyright (1988) John Wiley & Sons, reprinted with permission.

suitable materials, called *scintillators*, a small fraction of the excitation energy re-emerges as visible light (or sometimes in the UV region) during de-excitation. In a scintillation counter, this light passes down the scintillator and on to the face of a *photodetector* – a device that converts a weak photon signal to a detectable electric impulse. (Short-wavelength light may initially be collected on to the material of a *wavelength shifter*, which is a device that shifts the wavelength to higher values that are better matched to the frequency sensitivity of the photodetector.)

An important example of a photodetector is the *photomultiplier tube*, a schematic diagram of which is shown in Figure 4.18. Electrons are emitted from the cathode of the photomultiplier by the photoelectric effect and strike a series of focusing dynodes. These amplify the electrons by secondary emission at each dynode (by typically a factor of 5–10) and accelerate the particles to the next stage. The final signal is extracted from the anode at the end of the tube. The electronic pulse can be shorter than 10 ns if the scintillator has a short decay time. The scintillation counter is thus an ideal timing device and it is widely used for ‘triggering’ other detectors, i.e. its signal is used to decide whether or not to record information from the event. Commonly used scintillators are inorganic single crystals (e.g. caesium iodide) or organic liquids and plastics, and some modern complex detectors in particle physics have used several tons of detector in combination with thousands of photomultiplier tubes.¹⁵ The robust and simple nature of the scintillation counter has made it a mainstay of experimental nuclear and particle physics since the earliest days of the subject.

Just as direct detection of neutrons is not possible by ionisation methods, so the same is true using scintillators. However, the α particle and the ${}^3\text{H}$ nucleus from the direct nuclear reaction ${}^6\text{Li}(n, \alpha){}^3\text{H}$, mentioned

¹⁵For example, the SuperKamiokande detector, shown in Figure 3.3, which first detected neutrino oscillations, although not using scintillation counters, has 13 000 photomultiplier tubes.

in Section 4.3.2, can produce light in an LiI crystal scintillator and forms the basis for detecting neutrons with energies up to about 20 MeV.

4.4.3 Semiconductor detectors

Solid-state detectors operate through the promotion of electrons from the valence band of a solid to the conduction band as a result of the entry of the incident particle into the solid. The resulting absence of an electron in the valence band (a ‘hole’) behaves like a positron. Semiconductor detectors are essentially solid-state ionisation chambers with the electron–hole pairs playing the role of electron–ion pairs in gas detectors. In the presence of an electric field, the electrons and holes separate and collect at the electrodes, giving a signal proportional to the energy loss of the incident charged particle. Most semiconductor detectors use the principle of the junction diode. Since the band gap in some solids is as small as 1 eV and the energy loss required to produce a pair is only 3–4 eV on average (cf. the 30 eV required in a gas detector), a very large number of electron–hole pairs with only a small statistical fluctuation will be produced by a low-energy particle. Solid-state detectors are therefore very useful in detecting such particles. Semiconductors (usually silicon or germanium, although there is also some interest in the use of diamond) are used as a compromise between materials that have residual conductivity sufficient to enable conduction pulses due to single particles to be distinguished above background and those in which the charge carriers are not rapidly trapped in impurities in the material.

Such detectors have long been used in nuclear physics, where, for example, their excellent energy resolution and linearity, plus their small size and consequent fast response time, make them ideal detectors in gamma-ray spectroscopy. Only more recently have thin planar detectors become important in particle physics, because of the expense of covering large areas. Nevertheless, several hundred square metres of semiconductor detector are being used in experiments at the LHC.

One example of a solid-state detector is a *silicon microstrip detector*, where narrow strips of active detector are etched on to a thin slice of silicon, with gaps of order 10 μm, to give a tiny analogue of an MWPC. Arrays of such strips can then be used to form detectors with resolutions of order 5 mm. These are often placed close to the interaction vertex in a colliding beam experiment, with a view to studying events involving the decay of very short-lived particles. Another example is the *pixel detector*. A single-plane strip detector only gives position information in one dimension (orthogonal to the strip). A pixel detector improves on this by giving information in two dimensions from a single plane. Solid-state ‘vertex detectors’ have become increasingly important in particle physics and have been incorporated in most of the multicomponent detectors used in experiments at the highest energy colliders. Their main advantage is their superb spatial resolution; a disadvantage is their limited ability to withstand radiation damage, although newer designs are improving this aspect.

4.4.4 Čerenkov counters and transition radiation

When a charged particle traverses a dispersive medium with a velocity greater than the speed of light in the medium, it produces radiation, called Čerenkov radiation, which is somewhat analogous to the shock wave produced by an aircraft travelling faster than the speed of sound in the atmosphere. Also, when a highly relativistic particle crosses from one medium to another with different optical properties, *transition radiation*, usually in the X-ray region, is produced. Although neither of these processes is important in the context of energy loss, both are used in high-energy particle detectors to detect and identify charged particles.

4.4.4(a) Čerenkov counters

Methods of identifying particles are usually based on determining the mass of the particle by a simultaneous measurement of its momentum together with some other quantity. At low values of $\gamma = E/mc^2$, measurements of the rate of energy loss dE/dx can be used, while muons are characterised by their unique penetrating power in matter, as we have seen. Alternatively, the velocity could be measured. The simplest method for low-energy particles is to measure the time-of-flight between, for example, two scintillation counters, but at high energies this method ceases to be practical and an alternative method based on the Čerenkov effect is used.

When a charged particle with velocity v traverses a dispersive medium of refractive index n , excited atoms in the vicinity of the particle become polarised, and if v is greater than the speed of light in the medium c/n , a part of the excitation energy reappears as coherent radiation emitted at a characteristic angle θ to the direction of motion. The necessary condition $v > c/n$ implies $\beta n > 1$ and by considering how the waveform is produced¹⁶ it can be shown that $\cos\theta = 1/\beta n$ for the angle θ , where $\beta = v/c$ as usual. A determination of θ is thus a direct measurement of the velocity.¹⁷

Čerenkov radiation appears as a continuous spectrum and may be collected on to a photosensitive detector. Its main limitation from the point of view of particle detection is that very few photons are produced. The number of photons $N(\lambda) d\lambda$ radiated per unit path length in a wavelength interval $d\lambda$ can be shown to be

$$N(\lambda) d\lambda = 2\pi\alpha \left(1 - \frac{1}{\beta^2 n^2}\right) \frac{d\lambda}{\lambda^2} < 2\pi\alpha \left(1 - \frac{1}{n^2}\right) \frac{d\lambda}{\lambda^2} \quad (4.25)$$

and so vanishes rapidly as the refractive index approaches unity. The maximum value occurs for $\beta = 1$, which, for a particle with unit charge, corresponds to about 200 photons/cm in the visible region in water and

¹⁶This is Huygen's construction in optics (see Problem 4.17).

¹⁷For the discovery and interpretation of this effect, Pavel Čerenkov, Ilya Frank, and Igor Tamm were awarded the 1958 Nobel Prize in Physics.

glass. These numbers should be compared to the 10^4 photons/cm emitted by a typical scintillator. Because the yield is so small, appreciable lengths are needed to give enough photons, and gas Čerenkov counters for fixed-target experiments can be several metres long.

Čerenkov counters are used in two different modes. The first is as a *threshold counter* to detect the presence of particles whose velocities exceed some minimum value. Suppose that two particles with β values β_1 and β_2 at some given momentum p are to be distinguished. If a medium can be found such that $\beta_1 n > 1 \geq \beta_2 n$, then particle 1 will produce Čerenkov radiation but particle 2 will not. Clearly, to distinguish between highly relativistic particles with $\gamma \gg 1$ also requires $n \approx 1$, so that from (4.25) very few photons are produced. Nevertheless, common charged particles can be distinguished in this way up to at least 30 GeV/c.

Another device is the so-called *ring-image Čerenkov detector* and is very important at both fixed-target machines and colliders. If we assume that the particles are not all travelling parallel to a fixed axis, then the radiating medium can be contained within two concentric spherical surfaces of radii R and $2R$ centred on the target or interaction region where the particles are produced, as illustrated in Figure 4.19. The outer surface is lined with a mirror, which focuses the Čerenkov radiation into a ring at the inner detector surface. The radius of this ring depends on the angle θ at which the Čerenkov radiation is emitted, and hence on the particle velocity. It is determined by constructing an image of the ring electronically. This was the technique used in the SuperKamiokande detector, discussed in Chapter 3, to detect relativistic electrons and muons produced by neutrino interactions. In that experiment the radiating medium was very pure water.

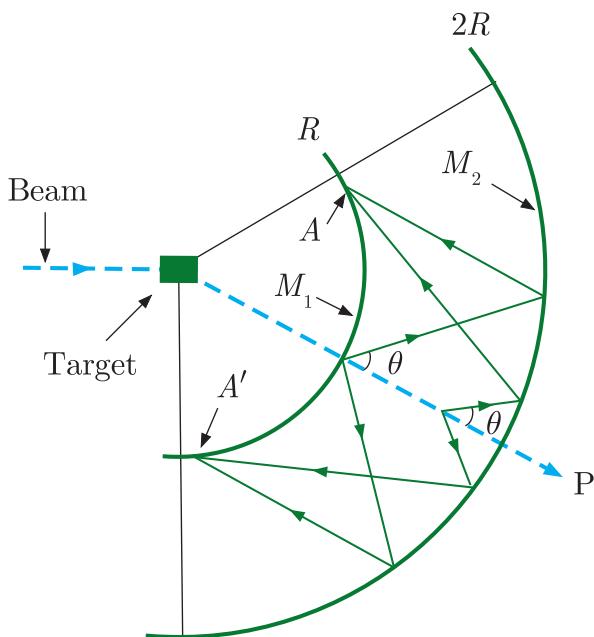


Figure 4.19 A particle P , produced from the target, emits Čerenkov radiation on traversing a medium contained between two spheres of radius R and $2R$. The mirror M_2 on the outer sphere focuses the radiation into a ring image at AA' on the inner detector sphere M_1 . The radii of the ring image depend on the angle of emission of the Čerenkov radiation and hence on the velocities of the particles.

4.4.4(b) Coherent Čerenkov radiation

When two charged particles traverse a dispersive medium, their radiations will normally add incoherently. However, if the distance between them is smaller than the wavelengths emitted, the Čerenkov radiation emitted by the two particles will be coherent, so that interference can occur. Hence, if the electron and positron of a pair produced by a high-energy photon in the field of a nucleus are sufficiently close together, the Čerenkov signal will be suppressed relative to the sum of the two separate signals that would be seen if they were incoherent. However, in some circumstances a shower of electrons and positrons can be produced with a significant net charge. Suppose, for example, that an extremely high-energy photon is incident on a dense material that is transparent to microwaves or radio waves, and produces an initial electron–positron pair. Each of these particles may produce a high energy photon by bremsstrahlung, which can in turn produce another pair, or alternatively convert an atomic electron to a high-energy electron by Compton scattering, creating a charge asymmetry. These processes can be repeated, generating a shower of many charged particles with an excess of electrons typically of order 20%. Of course, as the shower develops it will spread laterally, but if the Čerenkov radiation is observed in the longer wavelength microwave or radio wave region, an enhanced coherent signal will be produced. This effect was first predicted by Askaryan in 1965 and is called the *Askaryan effect*.

The Askaryan effect was first observed in 2000 and has been used, for example, in an experiment called ANITA based in the Antarctic. The experiment, which has been in operation in various forms since 2006, is designed to shed light on the origin of ultrahigh-energy cosmic rays that have been observed with energies up to 10^{20} eV. It does this by using the Askaryan effect to detect the ultrahigh-energy neutrinos, up to energies of 10^{18} eV, which result from collisions of the cosmic rays with photons in the cosmic microwave background radiation that pervades the universe. The resulting neutrinos pass through the Antarctic ice sheet and produce charge particles, leading to a strong electromagnetic pulse that propagates through the ice, because the latter is transparent to radio waves up to a frequency of 1.5 GHz. The ice sheet thus converts neutrino energy to radio waves. The ANITA detector system is mounted on a platform suspended from a helium balloon at a height of about 35 km above the Antarctic ice shelf. It traverses a circumpolar flight path due to the continuous wind circulation around the South Pole and ‘sees’ the ice below out to the horizon at about 700 km, resulting in an effective telescope ‘lens’ with an area of approximately 1.5×10^6 m².

4.4.4(c) Transition radiation

When a highly relativistic charged particle of mass m and energy γmc^2 crosses the boundary between a vacuum and a dispersive medium, electromagnetic radiation called *transition radiation* is emitted at a

characteristic angle of $1/\gamma$ to the particle direction. If the particle has charge ze , the total energy emitted is given by $\alpha z^2 \gamma \hbar \omega_p / 3$, where the plasma frequency¹⁸ ω_p is given by

$$\hbar \omega_p = \frac{1}{\alpha} (4\pi N_e r_e^3)^{1/2} m_e c^2. \quad (4.26a)$$

Here N_e is the electron number density and $r_e = e^2 / (4\pi\epsilon_0 m_e c^2)$ is the classical radius of the electron, so that numerically

$$\hbar \omega_p = [\rho(\text{g/cm}^3) \langle Z/A \rangle]^{1/2} \times 28.81 \text{ eV}, \quad (4.26b)$$

where ρ is the density of the medium having atomic number Z and mass number A .

For typical low- Z materials like styrene, $\hbar \omega_p \approx 20$ eV, and for $\gamma \approx 1000$, the radiation takes the form of soft X-rays with energies between 2 and 40 keV. As can be seen from Figure 4.12, photons of this energy can be detected in high- Z materials by the photoelectric effect. However, the radiation yield from a single interface is small so that simple TRDs typically consist of a series of low- Z foil radiators separated by high- Z detector layers made of proportional counters filled with a gas mixture containing xenon atoms. The radiator materials, for example mylar or polypropylene, are chosen to minimise absorption. In such multiple-layer detectors, interference effects cause the amount of energy emitted, which is proportional to the Lorentz factor γ for a single boundary crossing, to eventually saturate for γ values of order 2000 in a typical detector. Below this limit, the radiation from electrons will be more than two orders of magnitude greater than that from heavier particles like muons and pions, because their much smaller masses lead to much larger values of $\gamma = E/mc^2$. Hence such detectors can be operated in ‘threshold mode’ to distinguish electrons with momenta between about 1 and 150 GeV/c from heavier particles with the same momenta. Such devices have been used in many detectors, such as the ALICE detector at the LHC. Finally, we note that transition radiation can also be used as part of a tracking detector by, for example, distributing a large number of small-diameter ‘straw tube’ detectors through the radiating material, as is done in the ATLAS detector at the LHC.

4.4.5 Calorimeters

Calorimeters are an important class of detector used for measuring the energy and position of a particle by its total absorption and are widely used. They differ from most other detectors in that the nature of the particle is changed by the detector and the fact that they can detect neutral as well as charged particles. A calorimeter may be a homogeneous absorber/detector, such as a CsI scintillator. Alternatively, it can be a

¹⁸This is actually the plasma frequency that would pertain if all the electrons could move freely through the material.

sandwich construction with separate layers of absorber (for example, a metal such as lead) and detector (scintillator, MWPC, etc.). The latter are also known as ‘sampling calorimeters’. During the absorption process, the particle will interact with the material of the absorber, generating secondary particles, which will themselves generate further particles, and so on, so that a cascade or shower develops. For this reason calorimeters are also called ‘shower counters’. The shower is predominantly in the longitudinal direction due to momentum conservation, but will be subject to some transverse spreading due both to multiple Coulomb scattering and the transverse momentum of the produced particles. Eventually all, or almost all, of the primary energy is deposited in the calorimeter and gives a signal in the detector part of the device.

There are several reasons why calorimeters are important, especially at high energies:

- (i) they can detect neutral particles, by detecting the charged secondaries;
- (ii) the absorption process is statistical (and governed by the Poisson distribution), so that the relative precision of energy measurements $\Delta E/E$ varies like $E^{-1/2}$ for large E , which is a great improvement on high-energy spectrometers where $\Delta E/E$ varies like E^2 ;
- (iii) the signal produced can be very fast, of order (10–100) ns, and is ideal for making triggering decisions.

Although it is possible to build calorimeters that preferentially detect just one class of particle (electrons and photons or hadrons), it is also possible to design detectors that serve both purposes. Since the characteristics of electromagnetic and hadronic showers are somewhat different it is convenient to describe each separately. In practice, in particle physics it is common to have both types in one experiment, with the hadron calorimeter stacked behind the electromagnetic one.

4.4.5(a) *Electromagnetic showers*

When a high-energy electron or positron interacts with matter we have seen that the dominant energy loss is due to bremsstrahlung, and for the photons produced the dominant absorption process is pair production. Thus the initial electron will, via these two processes, lead to a cascade of e^\pm pairs and photons, and this will continue until the energies of the secondary electrons fall below the critical energy E_C where ionisation losses equal those from bremsstrahlung. This energy is roughly given by $E_C \approx 600 \text{ MeV}/Z$.

Most of the correct qualitative features of shower development may be obtained from the following very simple model. We assume:

- (i) each electron with $E > E_C$ travels one radiation length and then gives up half of its energy to a bremsstrahlung photon;

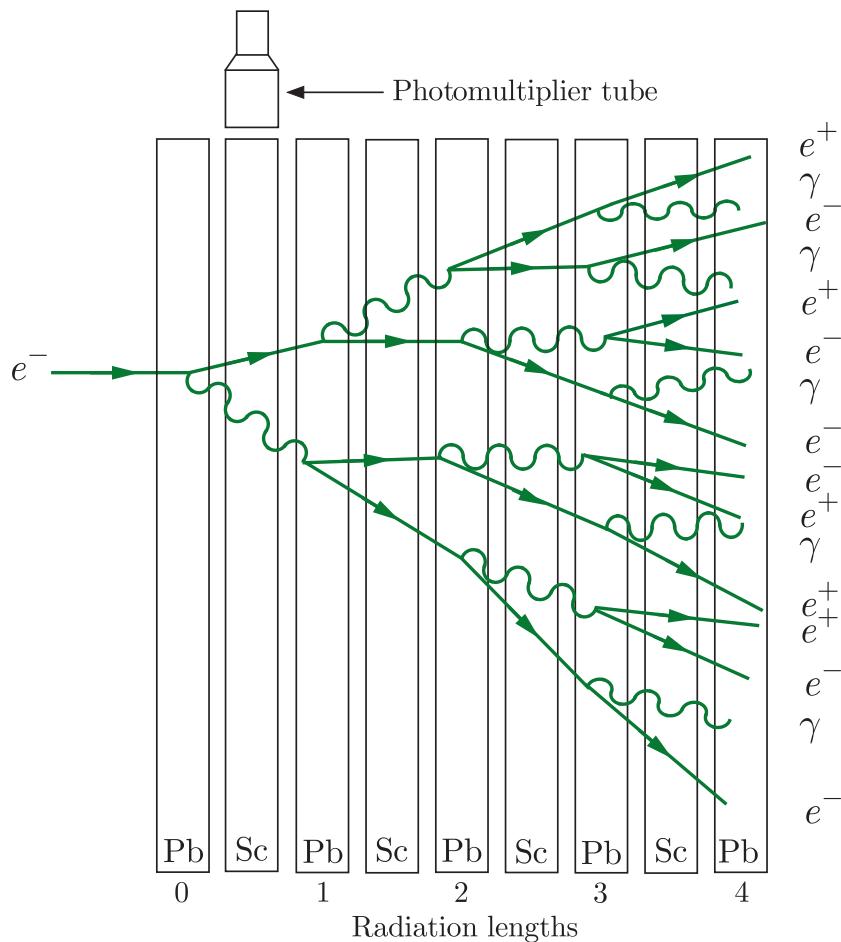


Figure 4.20 Approximate development of an electromagnetic shower in a sampling calorimeter assuming the simple model of the text. The calorimeter consists of alternate layers of lead (Pb) and a scintillator (Sc), the latter attached to photomultipliers (one only shown).

- (ii) each photon with $E > E_C$ travels one radiation length and then creates an electron–positron pair with each particle having half the energy of the photon;
- (iii) electrons with $E < E_C$ cease to radiate and lose the rest of their energy by collisions;
- (iv) ionisation losses are negligible for $E > E_C$.

A schematic diagram of the approximate development of a shower in an electromagnetic calorimeter assuming this simple model is shown in Figure 4.20.

If the initial electron has energy $E_0 \gg E_C$, then after t radiation lengths the shower will contain 2^t particles, which consist of approximately equal numbers of electrons, positrons, and photons, each with an average energy

$$E(t) = E_0/2^t.$$

The multiplication process will cease abruptly when $E(t) = E_C$, i.e. at $t = t_{\max}$ where

$$t_{\max} = t(E_C) \equiv \frac{\ln(E_0/E_C)}{\ln 2} \quad (4.27)$$

and the number of particles at this point will be

$$N_{\max} = \exp(t_{\max} \ln 2) = E_0/E_C. \quad (4.28)$$

The main features of this simple model are observed experimentally, and in particular the maximum shower depth increases only logarithmically with primary energy. Because of this, the physical sizes of calorimeters need to increase only slowly with the maximum energies of the particles to be detected. The energy resolution of a calorimeter, however, depends on statistical fluctuations, which are neglected in this simple model, but for an electromagnetic calorimeter, typically $\Delta E/E \approx 0.05/E^{1/2}$, where E is measured in GeV.

4.4.5(b) Hadronic showers

Although hadronic showers are qualitatively similar to electromagnetic ones, shower development is far more complex because many different processes contribute to the inelastic production of secondary hadrons. The scale of the shower is determined by the nuclear absorption length defined earlier. Since this absorption length is larger than the radiation length that controls the scale of electromagnetic showers, hadron calorimeters are thicker devices than electromagnetic ones. Another difference is that some of the contributions to the total absorption may not give rise to an observable signal in the detector. Examples are nuclear excitation and leakage of secondary muons and neutrinos from the calorimeter. The loss of ‘visible’ or measured energy for hadrons is typically 20–30% greater than for electrons.

The energy resolution of calorimeters is in general much worse for hadrons than for electrons and photons because of the greater fluctuations in the development of the hadron shower. Depending on the proportion of π^0 's produced in the early stages of the cascade, the shower may develop predominantly as an electromagnetic one because of the decay $\pi^0 \rightarrow \gamma\gamma$. These various features lead to an energy resolution typically a factor of 5–10 poorer than in electromagnetic calorimeters.

4.5 Detector Systems

In earlier sections we discussed the physics of individual detectors. However, as we mentioned in the introduction to Section 4.4, modern experiments in practice commonly use very large multicomponent detectors that integrate many different subdetectors in a single device. Such systems rely heavily on fast electronics and computers to monitor and control

the subdetectors and to coordinate, classify, and record the vast amount of information flowing in from different parts of the apparatus. This is particularly true for particle physics experiments. In this final section we will illustrate this by looking briefly at some examples and the physics questions that are addressed.

We start with two examples from particle physics, both located at accelerators. The first is the ATLAS detector, one of two multipurpose detectors located at the LHC at CERN. It has played a pivotal role in detecting the elusive Higgs boson and is now involved, among other things, in experiments to search for ‘superparticles’ predicted by theories that extend the standard model (these will be discussed in Section 10.3.2). The detector measures about 25 m in diameter and 46 m long, with an overall weight of approximately 7000 tonnes. It is shown in Figure 4.21. The upper figure shows the detector under construction and the lower

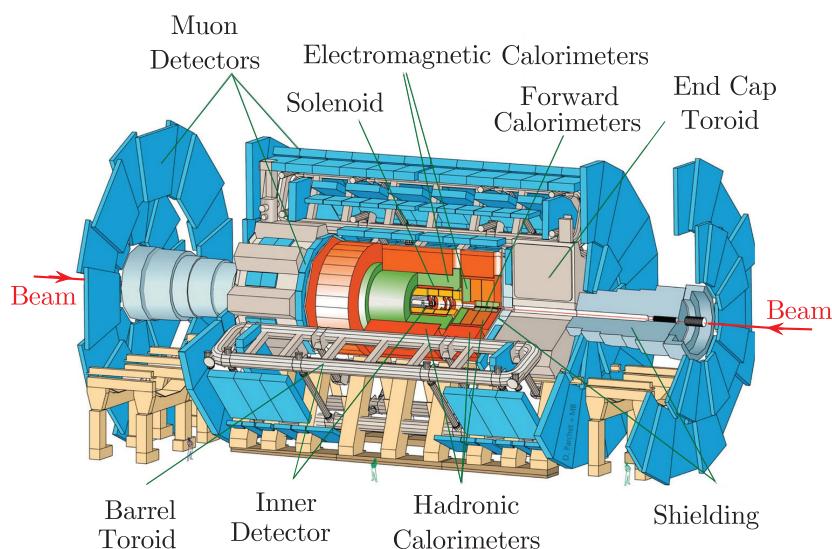
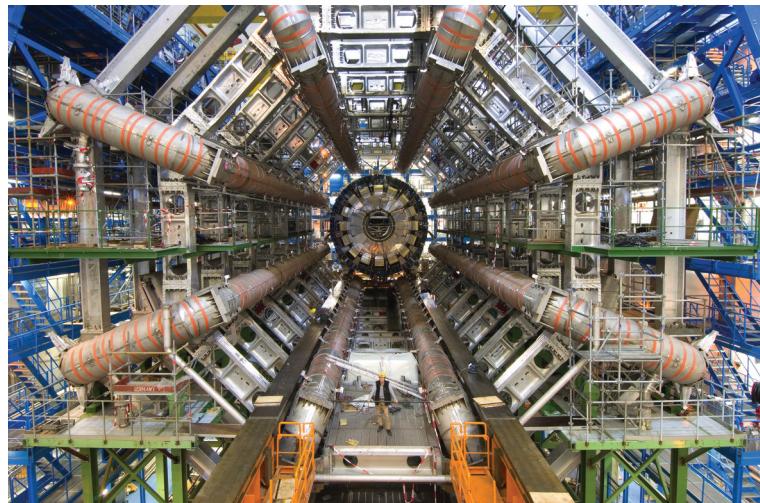


Figure 4.21 The ATLAS detector at the pp collider LHC at CERN, Geneva. The upper picture is a view along the beam direction during the construction phase showing the eight barrel toroids installed, with a calorimeter at the end, before it is moved into the middle of the detector. In the completed detector, the central volume is filled with subdetectors. The lower picture is a schematic diagram of the completed detector. Note the size of the human figures in the figures. Source: CERN, reproduced with permission.

figure is a schematic of the completed detector. Its scale can be gauged by the human figures shown in these figures. As in all detectors at colliders, the subdetectors are arranged in concentric layers surrounding the beam pipe. The inner detector consists of a silicon vertex detector very close to the interaction region and a number of tracking detectors. It is within a magnetic field to measure the momenta of all charged particles. Outside this field are electromagnetic and hadronic calorimeters to measure the energies of particles by absorption. Very large muon detectors are positioned at the outermost parts of the apparatus, because muons are the most penetrating charged particles produced.

The second example is the STAR detector at the RHIC collider at BNL, USA. It is much smaller than ATLAS and one of several detectors at RHIC that detect events resulting from the collisions of heavy ions, typically those of fully stripped gold nuclei, where the final state may contain many thousands of particles. An example of such an event is shown in Figure 5.13. It is designed to study questions at the boundary between nuclear and particle physics, among other things the state of matter known as a ‘quark–gluon plasma’, which is of great interest as it is believed to have existed in the earliest times of the universe. (This is discussed in Section 5.3.2.) The detector is shown in Figure 4.22. The upper figure shows it under construction and the lower figure is a schematic of the complete detector. As for Atlas, the subdetectors are arranged in concentric layers surrounding the beam pipe, each with a specific role in the overall event selection. At the core is a silicon vertex detector to detect very short-lived particles and a large time-projection chamber. At a greater radius are scintillators, providing further event selection by time-of-flight. The whole detector system is enclosed in a magnetic field to give momentum tracking information.

The third example is from nuclear structure physics. There are many questions of interest here relating to the properties of nuclei at the limits of stability on the edge of the ‘valley of stability’ (see Figure 2.13), such as the structure of super-heavy nuclei and why some highly deformed nuclei with high spin are relatively stable. Many of these questions are of great interest to astrophysicists. (We will return to these topics in Section 9.2.3.) To study these, and other, questions requires the study of the excited states of heavy nuclei. These can be produced, for example, by the fusion of two heavy ions. The states produced initially decay very rapidly (10^{-18} s) by ‘boiling off’ a few nucleons and the resulting unstable nuclei then decay via γ emission. Modern γ -ray detectors use high-purity germanium at the temperature of liquid nitrogen ($\sim 77\text{ K}$) connected to electronic systems that provide energy and time signals for the detected γ -rays.

An important example of a modern γ -ray spectrometer is the Gammasphere, which is located at the ATLAS at Argonne National Laboratory. Gammasphere is much smaller than ATLAS (CERN) or STAR (BNL). It consists of a pair of hemispheres, each about 2 m tall overall and weighing about 6 tons. It contains up to 110 Compton-suppressed gamma-ray

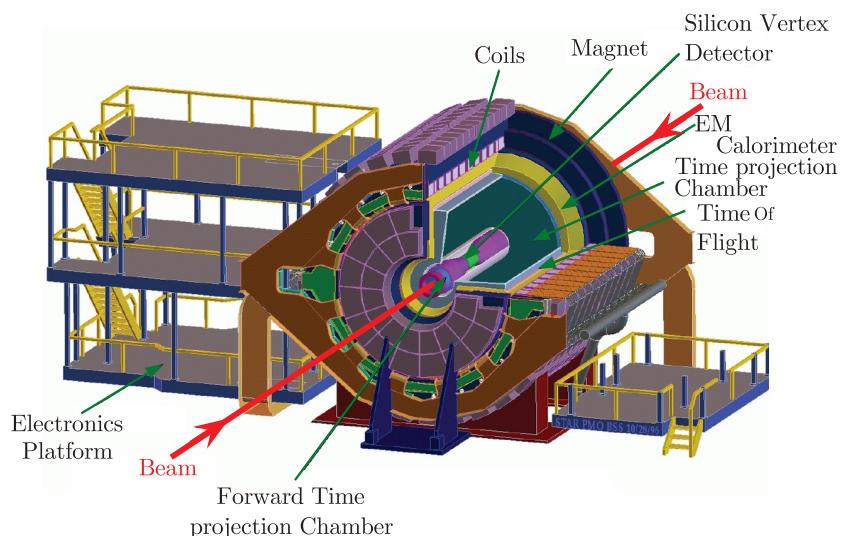
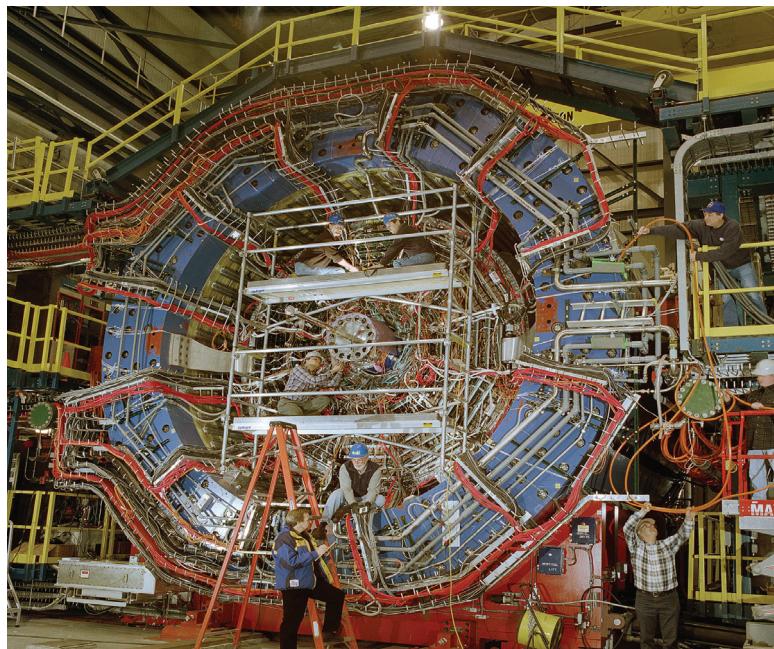


Figure 4.22 The STAR detector at the RHIC accelerator at Brookhaven National Laboratory (BNL), USA. The upper picture was taken during its construction and shows the time projection chamber (blue outer ring) and the installation of the smaller forward TPC at the centre. The lower picture is a schematic diagram of the completed detector. Source: Courtesy of Brookhaven National Laboratory.

detectors,¹⁹ each of which contains a single crystal of high-purity germanium with dimensions of a few centimetres, all pointing to the centre of the device. The spectrometer is shown in the upper picture of Figure 4.23. In the lower picture, the sphere is opened for maintenance and surrounds

¹⁹Photons that interact with the germanium do not always deposit all their energy in the semiconductor. The most common loss of energy is due to Compton scattering. Therefore, the germanium detector is shielded by a scintillator compound that suppresses the scattered γ -rays. Therefore, when a γ -ray is detected simultaneously by the germanium and the scintillator, it is rejected.

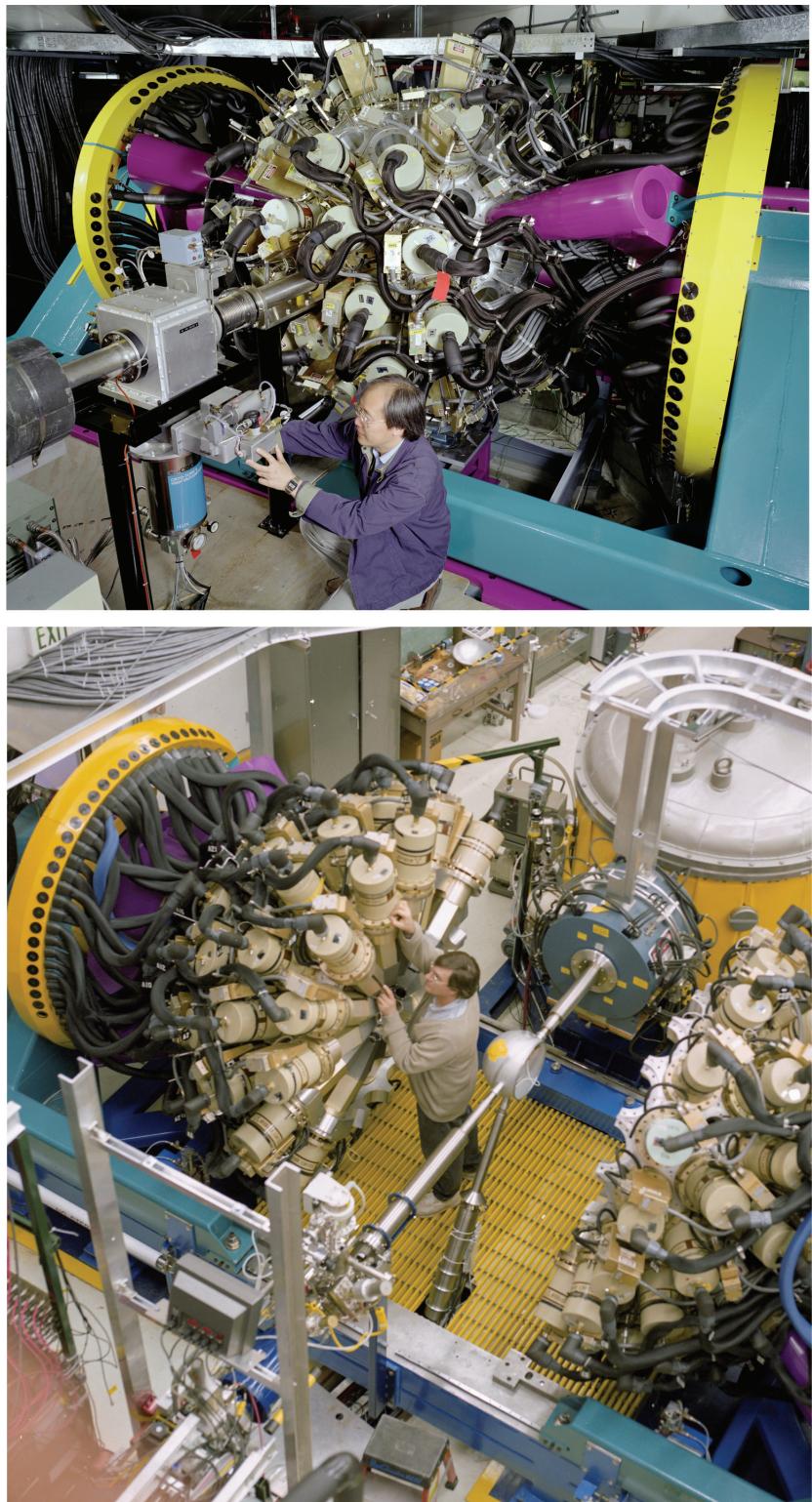


Figure 4.23 The upper figure shows the Gammasphere hemispheres closed; the lower one shows the hemispheres open and the beam line passing through to the Fragment Mass Analyser, the yellow cylinder at the top right of the figure. Source: Courtesy Argonne National Laboratory.

the interaction region of the beam pipe, which is clearly visible. The characteristic pattern of γ -rays emitted by the excited nuclei being studied are accompanied by γ -rays from the decays of numerous other excited nuclei; to distinguish the required events, the nuclei pass into another detector called the Fragment Mass Analyser (FMA), which contains a pixilated silicon detector that measures their masses and records their eventual decays, thus enabling different events to be distinguished. The FMA is shown at the top right of the lower picture at the end of the beam line.

In contrast to collider experiments, which usually involve high event rates with a very large number of particles produced in each event, nonaccelerator experiments in particle physics are frequently designed to detect rare events in which small numbers of particles are produced. This results in marked differences in the nature of the detectors required, which are less elaborate than the multicomponent, multipurpose detectors used at colliders, since relatively few particles are produced. In addition, their design is very dependent on the particular reactions studied. Nonetheless, they often have certain characteristic features. Firstly, in order to increase the event rate, there is a very large ‘active volume’, in which events can occur. Secondly, because event rates are intrinsically small, it is necessary to be especially careful to minimise ‘accidental’ events that might mimic real events. The detectors are therefore often situated deep underground to shield them from cosmic rays. In addition, very pure materials are required in the construction of the detectors themselves, to eliminate traces of radioactive impurities that might give rise to spurious events, and shielding or other methods are required to deal with ambient radioactivity from the surroundings of the detector.

Our final example, the KamLAND detector, is an important example of such a detector. Here the neutrinos originate from a nuclear reactor, whose beta-decay processes are a copious source of low-energy neutrinos. The detector is located deep underground in the Kamioka mine in Japan, and is shown in Figure 4.24. At the centre is a plastic balloon filled with 1000 tons of liquid scintillator. This is surrounded by a layer of buffer oil contained between an acrylic sphere and a stainless steel sphere together with about 2000 photomultipliers, as shown. The buffer oil shields the scintillator from radioactivity emanating from the photomultipliers. Antineutrinos are detected by observing the reaction $\bar{\nu}_e + p \rightarrow n + e^+$ in the scintillator, which acts as both target and detector. When this occurs, the scintillation light generated by the positron passes through the buffer oil and is detected by the photomultipliers. The neutron typically loses energy by multiple collisions until it is essentially at rest. It can then undergo the neutron capture reaction $n + p \rightarrow d + \gamma$ to give rise to a ‘delayed coincidence’ photon of energy 2.2 MeV, which is detected by the photomultipliers about 200 microseconds after the positron is detected. Finally, the steel sphere is itself surrounded by a water Čerenkov counter, which is used to detect any cosmic ray muons that reach the detector, despite its location deep underground.

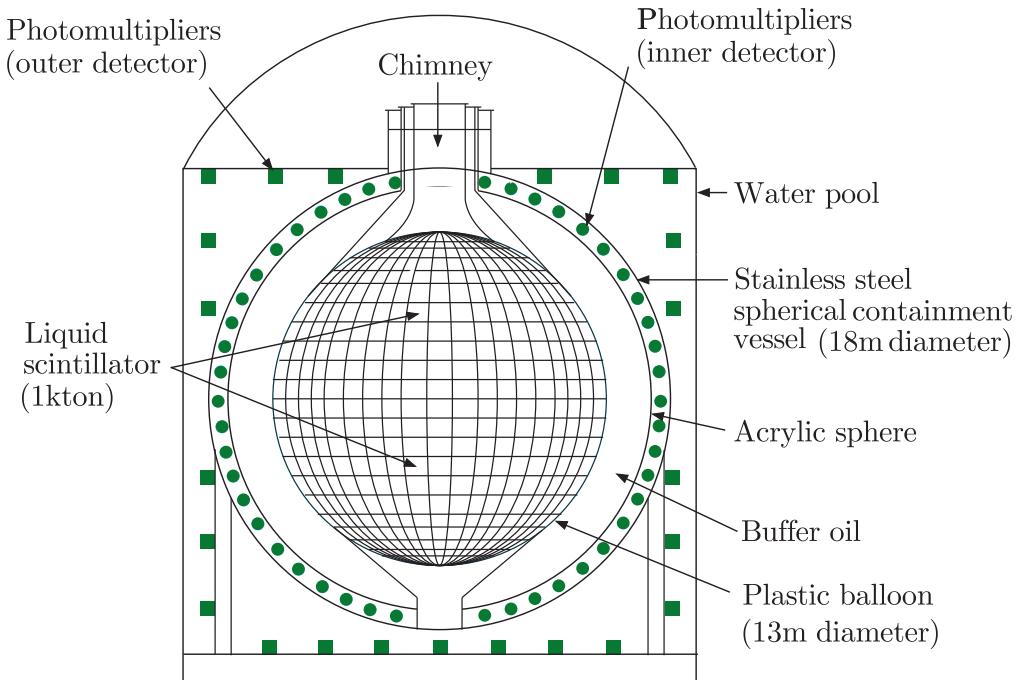


Figure 4.24 Schematic diagram of the KamLAND detector. Source: Adapted from Suzuki (2005), © The Royal Swedish Academy of Sciences. Reproduced by permission of IOP Publishing. All rights reserved.

Problems 4

- 4.1** At the HERA collider (which was operational until 2007) at the DESY Laboratory in Hamburg, a 20 GeV electron beam collided with a 300 GeV proton beam at a crossing angle of 10° . Evaluate the total centre-of-mass energy and calculate what beam energy would be required in an electron accelerator with a fixed-target proton to achieve the same total centre-of-mass energy.
- 4.2** What is the length L of the longest drift tube in a linac, which, operating at a frequency $f = 20$ MHz, is capable of accelerating ^{12}C ions to a maximum energy of $E = 100$ MeV?
- 4.3** Alpha particles are accelerated in a cyclotron operating with a magnetic field of magnitude $B = 0.8$ T. If the extracted beam has an energy of 12 MeV, calculate the extraction radius and the orbital frequency of the beam (the so-called *cyclotron frequency*).
- 4.4** Show that the momentum p in GeV/c for a relativistic particle of charge e moving in a circular orbit of radius ρ metres in a magnetic field of magnitude $B = |\mathbf{B}|$ tesla is given by $p = 0.3B\rho$.
- 4.5** Protons with momentum 50 GeV/c are deflected through a collimator slit 2 mm wide by a bending magnet 1.5 m long that produces a field of 1.2 T. How far from the magnet should the slit be placed so that it accepts particles with momenta in the range 49–51 GeV/c ?

- 4.6** An e^+e^- collider has a diameter of 8 km and produces beams of energy 45 GeV. Each beam consists of 12 bunches each containing 3×10^{11} particles. The bunches have a cross-sectional area of 0.02 mm^2 . What is the luminosity of the machine in units of $\text{cm}^{-2}\text{s}^{-1}$?
- 4.7** Estimate the thickness of iron through which a beam of neutrinos with energy 300 GeV must travel if 1 in 10^9 of them is to interact. Assume that at high energies the neutrino–nucleon total cross-section is given approximately by $\sigma_\nu \approx 10^{-38} E_\nu \text{ cm}^2$, where E_ν is given in GeV. The density of iron is $\rho = 7.9 \text{ g cm}^{-3}$.
- 4.8** An electron with an initial energy of 2 GeV traverses 10 cm of water with a radiation length of 36.1 cm. Calculate its final energy. How would the energy loss change if the particle were a muon rather than an electron?
- 4.9** A beam of neutrons with kinetic energy 0.1 eV and intensity 10^6 s^{-1} is incident normally on a thin foil of $^{235}_{92}\text{U}$ of effective density $10^{-1} \text{ kg m}^{-2}$. The beam can undergo (i) isotropic elastic scattering, with a cross-section $\sigma_{\text{el}} = 3 \times 10^{-2} \text{ b}$, (ii) radiative capture, with a cross-section $\sigma_{\text{cap}} = 10^2 \text{ b}$, or (iii) it can fission a $^{235}_{92}\text{U}$ nucleus, with a cross-section $\sigma_{\text{fission}} = 3 \times 10^2 \text{ b}$. Calculate the attenuation of the beam and the flux of elastically scattered particles 5 m from the foil.
- 4.10** A positron with laboratory energy 50 GeV interacts with the atomic electrons in a lead target to produce $\mu^+\mu^-$ pairs. If the cross-section for this process is given by
- $$\sigma = 4\pi\alpha^2\hbar^2c^2/3E_{\text{CM}}^2,$$
- calculate the positron's effective interaction length. The density of lead is $\rho = 1.14 \times 10^7 \text{ kg m}^{-3}$.
- 4.11** A liquid hydrogen target of volume 125 cm^3 and density $\rho = 0.071 \text{ g cm}^{-3}$ (to two significant figures) is bombarded with a mono-energetic beam of negative pions with a flux $2 \times 10^7 \text{ m}^{-2} \text{ s}^{-1}$ and the reaction $\pi^- + p \rightarrow \pi^0 + n$ observed by detecting the photons from the decay of the π^0 . Calculate the rate of photons emitted from the target per second if the cross-section is 40 mb.
- 4.12** Assuming the Bethe–Bloch formula is valid for low energies, show that the rate of ionisation has a maximum (the Bragg peak) and find the kinetic energy of protons in iron for which this maximum would occur.
- 4.13** Derive the result (4.15) for the range R .
- 4.14** A cylindrical proportional chamber has a central anode wire of radius 0.02 mm and an outer cathode of radius 10 mm with a voltage of 500 V applied between them. What is the electric field at the surface of the anode? If the threshold for ionisation by collision is 750 kV m^{-1} and the mean free path of the particles being detected is $4 \times 10^{-6} \text{ m}$, estimate the number of ion pairs produced per primary particle.
- 4.15** Two particles of masses m_1 and m_2 , and common momentum p , travel between two scintillation counters that are a distance L apart. Show that the difference in their flight times decreases like p^{-2} for large momenta. Calculate the minimum flight path necessary to distinguish pions from kaons if they have momentum 3 GeV/c and the time-of-flight can be measured with an accuracy of 200 ps.
- 4.16** Estimate the minimum length of a gas Čerenkov counter that could be used in threshold mode to distinguish between charged pions and charged kaons

with momentum 20 GeV/c. Assume that a minimum of 200 photons needs to be radiated to ensure a high probability of detection. Assume also that the radiation covers the whole visible spectrum between 400 and 700 nm and neglect the variation with wavelength of the refractive index of the gas.

- 4.17** A charged particle with speed v moves in a medium of refractive index n . By considering the wavefronts emitted at two different times, derive a relation for the angle θ of the emitted Čerenkov radiation relative to the particle's direction in terms of $\beta = v/c$ and n . What is the maximum angle of emission and to what limit does it correspond? If the momentum p of the particle is known, show that the mass squared x of the particle is given by

$$x = (mc^2)^2 = p^2 c^2 (n^2 \cos^2 \theta - 1).$$

If the error on the momentum is negligible, show, by taking derivatives of this expression, that for highly relativistic particles, the standard error σ_x on x is approximately

$$\sigma_x \approx 2p^2 c^2 \sqrt{(n^2 - 1)} \sigma_\theta,$$

where σ_θ is the standard error on θ .

- 4.18** Derive (4.26b) from (4.26a).

- 4.19** The reaction $e^+ e^- \rightarrow \tau^+ \tau^-$ is studied using a collider with equal beam energies of 5 GeV. The differential cross-section is given by

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2 \hbar^2 c^2}{4E_{CM}^2} (1 + \cos^2 \theta),$$

where E_{CM} is the total centre-of-mass energy and θ is the angle between the incoming e^- and the outgoing τ^- . If the detector can only record an event if the $\tau^+ \tau^-$ pair makes an angle of at least 30° relative to the beam line, what fraction of events will be recorded? What is the total cross-section for this reaction in nanobarns? If the reaction is recorded for 10^7 s at a luminosity of $L = 10^{31} \text{ cm}^{-2} \text{s}^{-1}$, how many events are expected?

Suppose the detector is of cylindrical construction and at increasing radii from the beam line there is a drift chamber, an electromagnetic calorimeter, a hadronic calorimeter, and finally muon chambers. If in a particular event the tau decays are

$$\tau^- \rightarrow \mu^- + \bar{\nu}_\mu + \nu_\tau \quad \text{and} \quad \tau^+ \rightarrow e^+ + \bar{\nu}_\tau + \nu_e,$$

what signals would be observed in the various parts of the detector?

5

Quark dynamics: the strong interaction

In Chapter 3 we described the basic properties of quarks and how these are used to construct the simple quark model of hadrons. We now look in more detail at how quarks interact and the role of gluons in the strong interactions. Thus we will be considering dynamical properties and the theoretical framework that describes these interactions.

5.1 Colour

We saw in Chapter 3 that the quark model account of the hadron spectrum is very successful. However, it begs several questions. One is: why are the observed states overwhelmingly of the form qqq , $\bar{q}\bar{q}\bar{q}$, and $q\bar{q}$? Another arises from a particular assumption that was implicitly made in Chapter 3. This is: if two quarks of the same flavour uu , dd , ss , ... are in the same spatial state, they must also be in the same spin state, with their spins parallel. This can be seen very easily by considering the baryon state Ω^- that was shown in Table 3.3 and Figure 3.16.¹ From its decay products, it may be deduced that this state has strangeness $S = -3$ and spin $J = 3/2$ and thus in the quark model the simplest composition is $\Omega^- = sss$, where all three quarks have their spins parallel and there is no orbital angular momentum between them. This means that all three like quarks have the same space and spin states, i.e. the overall wave function must be symmetric, which violates the fundamental requirement of the Pauli principle. The latter states that a system of identical fermions has a wave function that is overall antisymmetric under the interchange of any two of the particles, so that identical fermions cannot simultaneously be

¹The discovery of the Ω^- was a crucial step in gaining acceptance of the quark model of hadron spectroscopy. The experiment is described in Chapter 15 of Trigg (1975).

Table 5.1 Values of the colour isospin I_3^C and the colour hypercharge Y^C for the colour states of quarks and antiquarks

Quarks	I_3^C	Y^C	Antiquarks	I_3^C	Y^C
r	$1/2$	$1/3$	\bar{r}	$-1/2$	$-1/3$
g	$-1/2$	$1/3$	\bar{g}	$1/2$	$-1/3$
b	0	$-2/3$	\bar{b}	0	$2/3$

in the same quantum state. The three s quarks in the Ω^- therefore *cannot* be in the same state. So how do they differ?

The Ω^- is an obvious example of the contradiction, but it turns out that for the predictions of the quark model to agree with the observed spectrum of hadron multiplets, it is necessary to assume that overall baryon wavefunctions are symmetric under the interchange of like quarks.² In order to resolve this contradiction, it is necessary to assume a new degree of freedom exists for quarks, but not leptons, which is somewhat whimsically called *colour*.³

The basic properties of colour are as follows:

- (a) Any quark u, d, s, \dots can exist in three different colour states. We shall see later that there is direct experimental evidence that just three such states exist, which we denote r, g, b for ‘red’, ‘green’, and ‘blue’, respectively.
- (b) Each of these states is characterised by the values of two conserved *colour charges*, called *colour isospin* and *colour hypercharge*, denoted I_3^C and Y^C , which are strong interaction analogues of the electric charge in electromagnetic interactions.⁴ These charges depend only on the colour states r, g, b and *not* on the flavours u, d, s, \dots . The particular values for quarks and antiquarks are given in Table 5.1, and are a consequence of a fundamental symmetry of the strong interaction (called SU(3) colour symmetry), which we will not pursue here. For multiparticle states, the colour charges of the individual states are simply added.
- (c) Only states with zero values for the colour charges are observable as free particles; these are called *colour singlets*. This is the hypothesis of *colour confinement*. It can be derived, at least approximately, from the theory of strong interactions we shall describe.

Returning to the quark model, it can be seen from Table 5.1 that a state qqq can only have both $I_3^C = 0$ and $Y^C = 0$ if it has one quark in each of the three colour states r, g , and b . Hence, in the Ω^- , for example, all three s quarks are necessarily in different colour states, and thus the

²In Problem 3.9 it is shown explicitly that otherwise the predicted hadron spectrum contradicts experiment.

³Needless to say, nothing to do with ‘real’ colour!

⁴There are actually eight colour charges, but we will not need the others in what follows.

Pauli principle can be satisfied. Formally, we are assuming that the total wavefunction Ψ is the product of a spatial part $\psi_{\text{space}}(\mathbf{r})$ and a spin part ψ_{spin} , as usual, but also a colour wavefunction ψ_{colour} , i.e.

$$\Psi = \psi_{\text{space}}(\mathbf{r}) \psi_{\text{spin}} \psi_{\text{colour}}. \quad (5.1)$$

The Pauli principle is now interpreted as applying to the total wavefunction including the colour part ψ_{colour} . The combined space and spin wavefunctions can then be symmetric under the interchange of quarks of the same flavour (to agree with experiment) provided the colour wavefunction is antisymmetric. The structure of ψ_{colour} is therefore

$$\psi_{\text{colour}} = \frac{1}{\sqrt{6}}[r_1 g_2 b_3 + g_1 b_2 r_3 + b_1 r_2 g_3 - r_1 b_2 g_3 - b_1 g_2 r_3 - g_1 r_2 b_3], \quad (5.2)$$

where r , g , and b represent quarks with colour red, green, and blue, respectively.⁵

One can also see from Table 5.1 part of the answer to the first question of this section. Free quarks and fractionally charged combinations like qq and $qq\bar{q}$ are forbidden by colour confinement, in accordance with experimental observation. On the other hand, the combinations $q\bar{q}$ and qqq used in the simple quark model are allowed. More unusual combinations like $qq\bar{q}\bar{q}$ and $qqqq\bar{q}$ are also allowed and will be discussed further in Section 5.3.1 below.

5.2 Quantum chromodynamics (QCD)

The theory that describes strong interactions in the standard model is called *quantum chromodynamics*, or QCD for short (chromos = colour in Greek). Although QCD is not tested to the same extent or precision as quantum electrodynamics (QED), the quantum theory of electromagnetic interactions, it is nevertheless in impressive agreement with a large body of experimental data. QCD is similar to QED in that both describe interactions that are mediated by massless spin-1 bosons; gluons in the former case and photons in the latter. Both theories are of the type called *gauge theories*, which, as mentioned in Chapter 1, refers to a particular symmetry of the theory called gauge invariance.⁶

Gluons, the force carriers of the strong interaction, have zero electric charge, like photons, but unlike photons, which couple to electric charge, gluons couple to *colour* charges. This leads immediately to the flavour independence of strong interactions discussed in Section 3.3.1; that is, the different quark flavours $a = u, d, s, c, b$ and t have identical strong interactions. We now see that this is because they are postulated to exist in the same three colour states r , g , b , with the same possible values

⁵This choice is not only allowed by colour confinement, but is *required* by it. See Problem 5.2.

⁶A brief discussion of gauge invariance and its consequences is given in Appendix D.

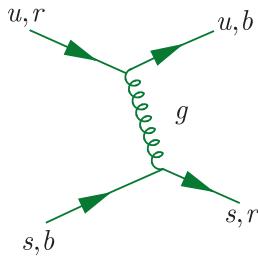


Figure 5.1 Example of quark–quark scattering by gluon exchange. In this diagram, the quark flavours u and s are unchanged, but their colour states can change, as shown.

of the colour charges. This has its most striking consequences for u and d quarks, which have almost equal masses, where it leads to the phenomenon of isospin symmetry. This results, among other things, in the near equality of the masses of the proton and neutron, and of the masses of the different charge states within other multiplets, such as pions and kaons, as we saw in Section 3.3.1. In the same way, because the mass difference between the s quark and the u and d quarks is small on the scale of hadron masses, it leads to the existence of the supermultiplets of particles with similar masses like those shown in Figures 3.15 and 3.16. Flavour independence also implies the equality of the potentials in the charmonium and bottomonium systems, which emerged from experiment, as discussed in Section 3.3.5(c).

Although QED and QCD both describe interactions, albeit of very different strengths, that are mediated by massless spin-1 bosons that couple to conserved charges, there is a crucial difference between them that profoundly affects the characters of the resulting forces. While the photons that couple to the electric charge are themselves electrically neutral, gluons have nonzero values of the colour charges to which they couple. This is illustrated in Figure 5.1, which shows a particular example of a quark–quark interaction mediated by gluon exchange. In this diagram, the colour states of the two quarks are interchanged and the gluon has nonzero values of the colour quantum numbers, whose values follow from colour charge conservation at the vertices, i.e.

$$I_3^C(\text{gluon}) = I_3^C(\text{red quark}) - I_3^C(\text{blue quark}) = 1/2 \quad (5.3)$$

and

$$Y^C(\text{gluon}) = Y^C(\text{red quark}) - Y^C(\text{blue quark}) = 1. \quad (5.4)$$

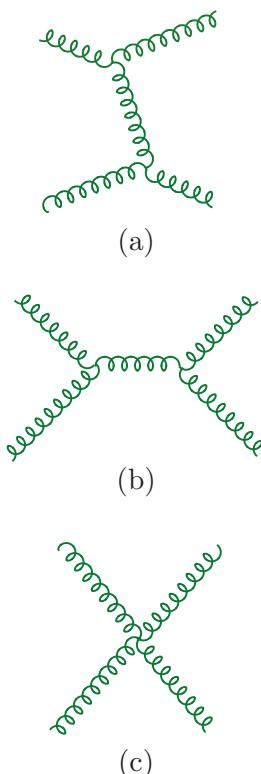


Figure 5.2 The three lowest-order contributions to gluon–gluon scattering in QCD: (a) and (b) one-gluon exchange, (c) contact interaction.

Just as quarks can exist in three colour states, gluons can exist in eight colour states, although we will not need the details of these. The first thing implied by the nonzero values of the colour charges is that gluons, like quarks, are confined and cannot be observed as free particles. The second is that since gluons couple to particles with nonzero colour charges, and since gluons themselves have nonzero colour charges, it follows that gluons couple to other gluons. The two types of gluon self-coupling that occur in QCD are illustrated in Figure 5.2, which shows the three lowest-order contributions to gluon–gluon scattering. The strength of these interactions is characterised by the strong coupling constant α_s , which is analogous to α , the fine structure constant (1.46) in QED. As discussed in Section 1.6.1, each electron–photon vertex in QED corresponds to a basic process with a transition probability of order α . In the same way, each quark–gluon vertex and three-gluon vertex in Figures 5.1 and 5.2a and b, respectively, corresponds to a basic process with a transition probability of order α_s . However, the zero-range ‘contact’ interaction corresponding to the four-gluon vertex in Figure 5.2c is of order α_s^2 , so that all three diagrams in Figure 5.2 are second-order diagrams, even though Figure 5.2c has only one vertex.

The gluon–gluon interactions of Figure 5.2 have no analogue in QED, and it can be shown that they lead to properties of the strong interaction that differ markedly from those of the electromagnetic interaction. These properties are *colour confinement* and *asymptotic freedom*. Colour confinement is the requirement that observed states have zero colour charges. This was discussed in Section 5.1, where, for example, it was shown to imply that all quark bound states (hadrons) must have integral electric charges. It also implies that gluons, like quarks, cannot be observed as isolated free particles, because they have non-zero values of the colour charges. Bound states of two or more gluons with zero colour charges overall can be formed in principle, due to the strong interaction between gluons themselves. Such states are called *glueballs*. Similarly, gluons may bind to, for example, $q\bar{q}$ pairs, to form colourless states, called *hybrids*, composed of both quarks and gluons. Both glueballs and hybrids will be discussed further in Section 5.3.1.

Asymptotic freedom means that the interaction gets weaker at short distances,⁷ and at distances less than about 0.1 fm the lowest-order diagrams dominate. At these distances, quark–quark scattering, for example, is given approximately by one-gluon exchange diagrams like Figure 5.1. However, as the distance between the quarks increases, the interaction gets stronger, and many higher-order diagrams become important. In this strong interaction regime perturbation theory is no longer applicable and it has not yet been possible to evaluate the theory precisely. We therefore have to rely on approximate results obtained by numerical simulations of the theory on very large computers and the demonstration of confinement in QCD rests largely on such simulations. They are done using an approach called *lattice gauge theory* in which space (and sometimes time) is approximated by a finite lattice of discrete points. The exact theory can then in principle be recovered by letting the lattice spacing go to zero, and the number of lattice points become infinite.⁸ In practice, the number of lattice points that can be handled is limited by the computing power available, but nonetheless good results have been obtained for several static properties, for example the masses and decay constants of the lower lying hadrons.⁹

The above features are conveniently illustrated by considering the static potential between a heavy quark and its antiquark in a colour singlet state. This is the appropriate potential for a non-relativistic discussion of charmonium and bottomonium, and was determined empirically for the limited range $0.2 \leq r \leq 0.8$ fm in Section 3.3.5(c). Here we are concerned with the behaviour predicted on the basis of QCD. At short interquark distances $r \leq 0.1$ fm, the interaction is dominated by one-gluon exchange

⁷Asymptotic freedom was derived for QCD in 1973 by David Gross, David Politzer, and Frank Wilczek, who were subsequently awarded the 2004 Nobel Prize in Physics.

⁸A simple account of such calculations is given in Weingarten (1996).

⁹See, for example, Amsler et al. (2017).

and we might expect a Coulomb-like potential analogous to that arising from one-photon exchange in QED, so that the potential is given by

$$V(r) = -\frac{4}{3} \frac{\hbar c \alpha_s}{r} \quad (r \leq 0.1 \text{ fm}), \quad (5.5a)$$

where α_s is the strong coupling constant discussed above.¹⁰ Because of asymptotic freedom, the strength of the interaction, and hence α_s , decreases with decreasing r , but for $r \leq 0.1 \text{ fm}$ this variation is slight and can in many applications be neglected, as we shall see below. At distances beyond 0.1 fm, however, the strength of the interaction increases more rapidly and one-gluon exchange no longer dominates. In this region, we have to rely mainly on lattice gauge theory calculations of limited precision. These are consistent with the empirical potential of Figure 3.24 in the range where this is determined and show that at large distances the potential increases approximately linearly as

$$V(r) \approx \lambda r \quad (r \geq 1 \text{ fm}), \quad (5.5b)$$

where the constant λ is of order 1 GeV fm^{-1} . This is an example of a confining potential in that it does not die away with increasing separation, and the force between the quark and antiquark cannot be neglected, even when they are very far apart.

5.2.1 The strong coupling constant

The strong interaction derives its name from the strong forces acting at distances of order 1 fm that, among other things, bind quarks in hadrons. However, many of the remarkable phenomena discussed in this chapter depend on the fact that the interaction gets weaker at short distances; that is, on asymptotic freedom. Such short-distance interactions are associated with large momentum transfers $|\mathbf{q}|$ between the particles, with

$$|\mathbf{q}| = O(\hbar/r), \quad (5.6)$$

where $r = |\mathbf{r}|$ is the distance at which the interaction occurs. For example, the amplitude (1.47) for scattering from a spherically symmetric potential $V(r)$ becomes

$$\mathcal{M}(q) = 4\pi \int_0^\infty V(r) \left(\frac{\sin(qr)}{qr} \right) r^2 dr \quad (5.7)$$

on integrating over all angular directions. The dominant contributions arise from r values of order q^{-1} as asserted, since for smaller r the integrand is suppressed by the factor r^2 , while for large r it is suppressed by the average over the rapidly oscillating sine factor. Hence in discussing

¹⁰The numerical factor multiplying α_s (i.e. $-4/3$ in this case) depends on the colour state chosen, and we will not discuss it further.

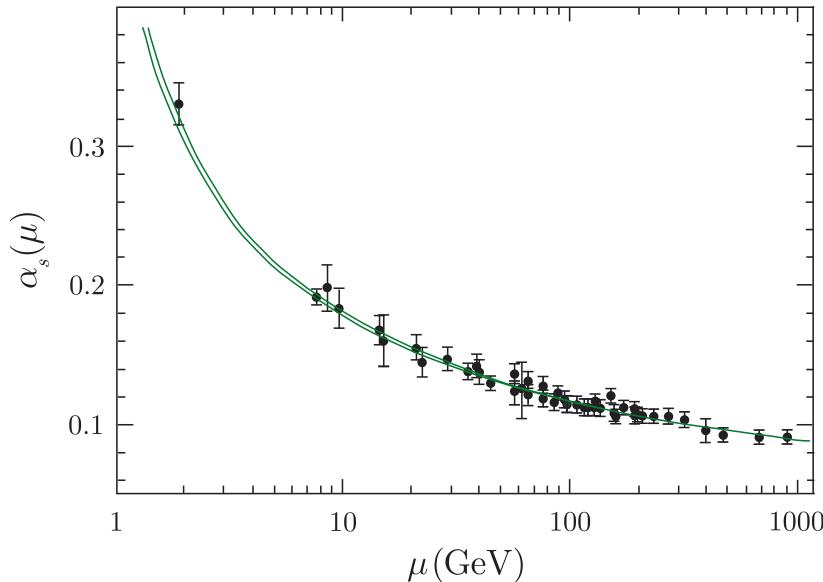


Figure 5.3 Values of the running coupling constant α_s obtained from the following sources: hadronic τ decays; lattice QCD; jets in deep inelastic lepton–nucleon scattering; heavy quarkonia decays; hadronic final states of e^+e^- annihilations; hadron collider jets; and electroweak precision fits to Z^0 decay. The solid curves show the evolution of α_s with μ , as predicted by QCD, assuming the value (5.10) for $\alpha_s(M_Z c^2)$. Source: Adapted from Bethke et al. (2017).

scattering from the static potential (5.5a), in which the strong coupling constant α_s decreases with decreasing r , we can equally well regard it as decreasing with increasing momentum transfer, since the one implies the other by (5.6).

In this discussion, we have considered scattering from a static potential, leaving the energy of the particle unchanged. Particle energies are also unchanged in elastic scattering in the centre-of-mass frame, but in other reference frames, and in other processes, energy as well as momentum can be exchanged between the particles. In such cases, the strength of the interaction can be shown to depend on the Lorentz-invariant quantity

$$\mu^2 \equiv |\mathbf{q}^2 c^2 - E_q^2|, \quad (5.8)$$

which reduces to $\mathbf{q}^2 c^2$ when the energy exchanged E_q is zero. More specifically, it can be shown that for $\mu^2 \gg 1 \text{ GeV}^2$, the QCD coupling constant α_s is given to a good first approximation by

$$\alpha_s(\mu) = \alpha_s(\mu_0) \left[1 + \frac{(33 - 2N_f)}{6\pi} \alpha_s(\mu_0) \ln(\mu/\mu_0) \right]^{-1} \quad (\mu^2 \gg 1 \text{ GeV}^2). \quad (5.9)$$

Here N_f is the number of quark flavours u, d, s, \dots ¹¹ with $4m_q^2 c^4 < \mu^2$, and $\alpha_s(\mu_0)$ is the value of the coupling at a chosen reference value μ_0 , which is usually taken to be $\mu_0 = M_Z c^2$, the rest energy of the Z boson. Measured values of $\alpha_s(\mu)$ obtained from a variety of different processes are shown in Figure 5.3, where the curves show the predicted behaviours

¹¹The change in $\alpha_s(\mu)$ at $\mu^2 = 4m_c^2 c^4, 4m_b^2 c^4$ is, of course, not really discontinuous as implied by the approximation (5.9), but is ‘smoothed out’ over a threshold region.

corresponding to the ‘best-fit’ value

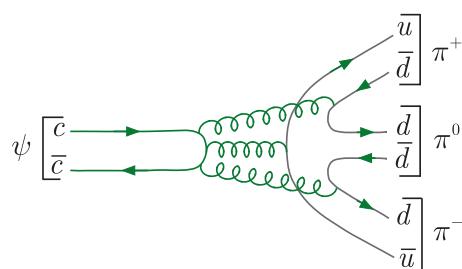
$$\alpha_s(M_Z c^2) = 0.1181 \pm 0.0011 \quad (5.10)$$

at the reference value $\mu_0 = M_Z c^2$. The decrease in $\alpha_s(\mu)$ as μ increases, corresponding to shorter distances, is clearly seen, and because of this variation, $\alpha_s(\mu)$ is often referred to as the *running coupling constant*. Conversely, $\alpha_s(\mu)$ increases more and more rapidly as μ decreases towards 1 GeV, until it is no longer small compared to unity, when the interaction becomes strong and perturbative ideas are no longer applicable.

We are now able to discuss the origin of the OZI rule, which was used in Section 3.3.5(a) to explain the narrowness of the charmonium and bottomonium states below the charm and bottom thresholds, respectively. When a quark–antiquark pair annihilates to gluons, the appropriate mass scale is the invariant mass of the pair. Hence light quark pairs are easily produced in the region where the interaction is strong, and typical hadron decays, like $K^*(890)$ decay (cf. Figure 3.18a), or the decay of charmonium states above the charmed threshold (cf. Figure 3.21a) proceed via the creation of one or more such pairs, with typical widths of order tens of MeV. On the other hand, the annihilation or creation of $c\bar{c}$ or $b\bar{b}$ pairs is relatively suppressed because of the running of the coupling constant, which becomes small enough for perturbative arguments to be used.

Now let us consider the decays of charmonium or bottomonium states below the charm or bottom threshold, shown by the annihilation mechanism illustrated in Figure 3.21b, and let us suppose that the decaying state has charge parity $C = -1$, like the J/ψ or the lightest Υ states. Then, because the initial heavy quark–antiquark pair is colourless, it cannot annihilate to a single gluon without violating colour conservation, and because the initial state has $C = -1$ and n gluons, like n photons, have C -parity $(-1)^n$, it must annihilate to at least three gluons to conserve C -parity. This is illustrated for a $c\bar{c}$ state decaying to a three-pion final state in Figure 5.4, and because of the additional powers of $\alpha_s(\mu)$ required to produce the three gluons, the corresponding decay rates are heavily suppressed. Of course, more quantitative results require more detailed arguments. In the case of the corresponding Υ states, as they are heavier, more accurate results are available because perturbation theory converges more rapidly than for charmonium states. The corresponding value of $\alpha_s(\mu)$ is smaller, and this leads to one of the empirical values of α_s shown in Figure 5.3.

Figure 5.4 OZI-suppressed decay of a $C = -1$ charmonium state below the $D\bar{D}$ threshold. The corresponding quark diagram, in which all virtual gluon lines are by convention suppressed, has already been given in Figure 3.21b.



5.2.2 Screening, antiscreening and asymptotic freedom

Although there are other forces that increase with increasing separation (for example the force between two particles connected by a spring or elastic string), the difference between those and the present case is that in the former cases eventually something happens (for example the string breaks) so that the particles (or the ends of the string) become free. This does not happen with the strong force. Instead, the energy stored in the colour field increases until it becomes sufficiently large enough to create $q\bar{q}$ pairs and eventually combinations of these will appear as physical hadrons. This latter process is called *fragmentation* and is rather poorly understood. The behaviour of the strong interaction as a function of distance (or equivalently momentum transfer) is so unlike the behaviour of other forces we are familiar with (e.g. gravity and electromagnetism) that it is worth looking at why this is the case.

In QED, single electrons are considered to emit and reabsorb photons continually, as shown in Figure 5.5a. Such a process is an example of a so-called *quantum fluctuation*, i.e. one particle converting to two or more particles for a finite time. This is allowed provided the time and the implied violation of energy conservation are compatible with the energy–time uncertainty principle. Of course if another electron is nearby, then it may absorb the photon and we have the usual one-photon exchange scattering process of Figure 5.5b. The emitted photon may itself be subject to quantum fluctuations, leading to more complicated diagrams like those shown in Figure 5.6a. Thus the initial electron emits not only photons but also indirectly electron–positron pairs. These are referred to as a ‘sea’ of virtual electrons and positrons. The equivalent contribution to elastic electron–electron scattering is shown in Figure 5.6b.

These virtual processes are collectively referred to as *vacuum polarisation effects*.¹² The production of virtual e^+e^- pairs produces a shielding effect, so that the charge and the strength of the interaction α , as seen from a distance, will appear altered. Detailed calculations show that if we write the Coulomb potential as

$$\phi_{\text{eff}}(r) = \frac{\alpha_{\text{eff}}(r)\hbar c}{r}, \quad (5.11)$$

then

$$\alpha_{\text{eff}} = \alpha \approx 1/137 \quad (5.12)$$

for

$$r \gg r_C \equiv \hbar/m_e c = 3.9 \times 10^{-13} \text{ m}, \quad (5.13)$$

but for $r \leq r_C$, the value of α is somewhat larger and increases as r becomes smaller. In other words, the strength of the interaction increases at very

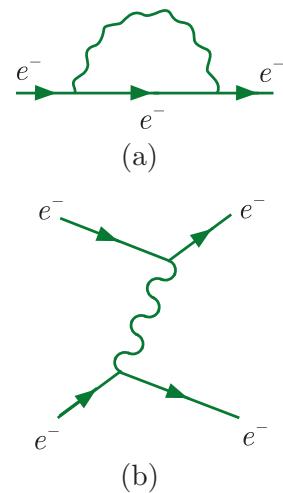


Figure 5.5 (a) The simplest quantum fluctuation of an electron and (b) the associated exchange process.

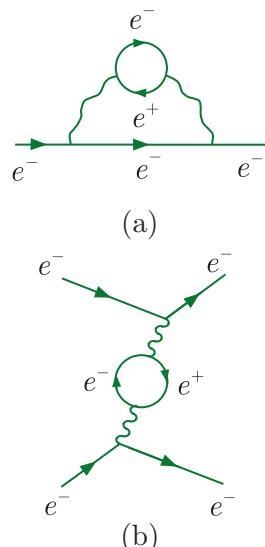


Figure 5.6 (a) A more complicated quantum fluctuation of the electron and (b) the associated exchange process.

¹²The name arises from the analogy of placing a charge in a dielectric medium. This aligns the particles of the medium and produces a net polarisation.

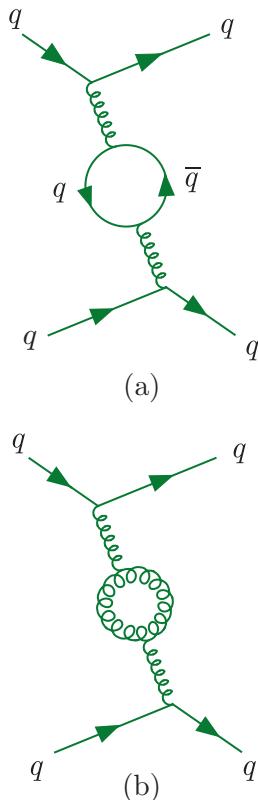


Figure 5.7 The two lowest-order vacuum polarisation corrections to one-gluon exchange in quark–quark scattering.

short distances. Formally, without proof, the QED coupling $\alpha_{\text{em}}(\mu)$ is given to a good approximation by

$$\alpha_{\text{em}}(\mu) = \alpha(\mu_0) \left[1 - \frac{2}{3\pi} \alpha(\mu_0) \ln \left(\frac{\mu}{\mu_0} \right) \right]^{-1}, \quad (5.14)$$

where μ_0 is a value of μ at which the value of α is known. Thus the electromagnetic coupling increases with μ , but only very slowly.

Vacuum polarisation effects have measurable consequences. For example, the 2s state in hydrogen is predicted to be more tightly bound than it would be in a pure Coulomb potential. The increased binding energy is only 1.1×10^{-7} eV but nevertheless is confirmed by extremely accurate measurements on the hydrogen spectrum. There are also very small corrections to the magnetic moment of the electron that have been verified experimentally to extraordinary precision.

Quantum fluctuations also exist in QCD and give rise to a variation of the interaction strength with distance. If, by analogy with QED, we consider quark–quark scattering, then the two lowest-order vacuum polarisation corrections are shown in Figure 5.7. The first of these (Figure 5.7a) is analogous to virtual e^+e^- production in QED and also leads to a screening effect. However, the second diagram (Figure 5.7b) has no counterpart in QED, because there are no direct photon self-couplings. Calculations show that this diagram leads to an *antiscreening* effect that is larger than the screening effect from Figure 5.7a and so the net effect is that the interaction grows *weaker* at short distances, i.e. asymptotic freedom.

5.3 New forms of matter

In the rest of this chapter, we will explore applications of QCD, concentrating mainly on perturbative aspects, which are a consequence of asymptotic freedom. However, in this section we first consider the evidence for two new forms of matter suggested by QCD: exotic hadrons whose existence cannot be accounted for by the simple quark model and the so-called *quark–gluon plasma*, which is expected to form at the very high nuclear energy densities and that can be explored in high-energy collisions of heavy nuclei.

5.3.1 Exotic hadrons

Any hadrons that lie outside the simple quark model (so not having the structure $q\bar{q}$, qqq , or $\bar{q}\bar{q}\bar{q}$) are termed *exotic*, whether or not their quantum numbers are exotic, i.e. whether or not their quantum numbers are forbidden to occur in the simple quark model. As we have seen, such states are in principle allowed in QCD, and lattice gauge theory calculations, for example, suggest that while conventional quark model states dominate at lower masses, exotic states do play a role in hadron spectroscopy.

The evidence for states with exotic internal quantum numbers has already been discussed in Section 3.3.6. There we saw that the overwhelming majority of known hadrons conform to the values required by the simple quark model, and while there is some evidence for one exception, it is not conclusive and not a single example of a hadron with exotic internal quantum numbers can yet be regarded as firmly established. However, many possible exotic states do not have exotic quantum numbers and in this section we discuss the evidence for such states. We will start by considering glueballs and hybrids involving gluons and light quarks u , d , s , and then go on to consider exotic mesons and baryons containing heavy quark–antiquark pairs.¹³

5.3.1 (a) *Glueballs and hybrids*

An interesting point about the gluon–gluon interactions of Figure 5.2 is that if the net forces were attractive and sufficiently strong, they could in principle lead to bound states of two or more gluons. These are the *glueball* states referred to earlier, and since they are composed of gluons alone, they are predicted to be strongly interacting neutral bosons with $I = S = C = \tilde{B} = 0$. Moreover, if they exist, their electromagnetic interactions would be expected to be much weaker than those of ‘ordinary’ mesons composed of charged quarks and antiquarks. Unfortunately, precise theoretical calculations of glueball properties are not yet possible because of our limited understanding of confinement. However, approximate lattice gauge theory calculations that ignore the existence of quarks yield glueball spectra similar to that shown in Figure 5.8, where the lightest glueball is a scalar meson with $J^{PC} = 0^{++}$ and a mass of around $1.5 - 1.7 \text{ GeV}/c^2$, depending on the details of the approximations made. However, when quarks are included in the theory, such states can usually mix with $q\bar{q}$ mesons with the same quantum numbers and similar masses, so that the observed states are unlikely to be pure glueballs, but states with both glueball and $q\bar{q}$ components. These mixed states are difficult to distinguish from ordinary $q\bar{q}$ mesons. There is at present no conclusive evidence for pure glueballs, despite many experimental searches, but there is some evidence for mixed states containing both gluon and $q\bar{q}$ components, as we shall see shortly.

Firstly, however, we note that the problem of glueballs mixing with conventional $q\bar{q}$ mesons would be avoided if the J^{PC} values of the observed state could not occur in the simple quark model. Since the parity of a fermion and its antifermion is opposite, we have $P = (-1)^{L+1}$ for a bound state of a quark with its own antiquark, while by (1.18) its C -parity is given by $C = (-1)^{L+S}$. Hence for $J = 0$ there are just two possibilities: $L = S = 0$ with $J^{PC} = 0^{-+}$ and $L = S = 1$ with $J^{PC} = 0^{++}$. The remaining possibilities, 0^{--} and 0^{+-} , do not occur in the simple quark model.

¹³For a comprehensive review of exotic hadrons see, for example, Olsen (2015).

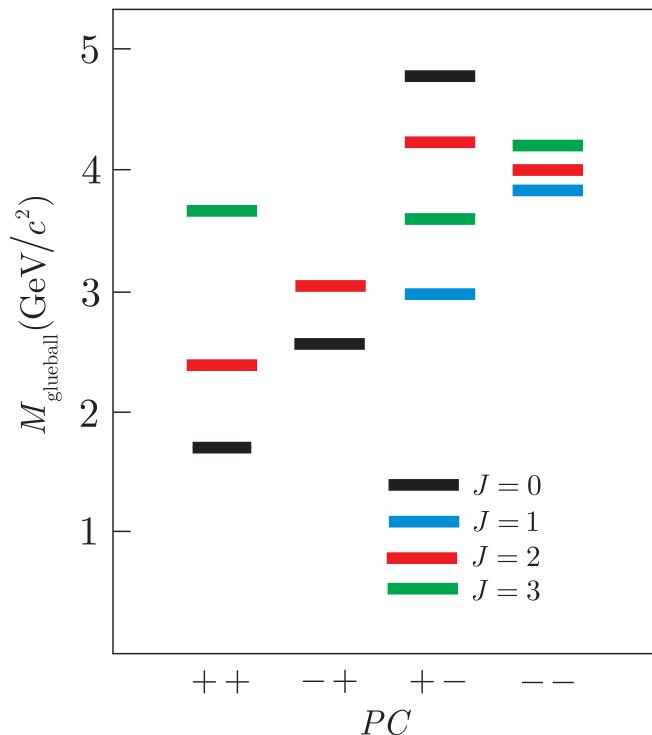


Figure 5.8 Predicted glueball mass spectrum from lattice QCD calculations in the absence of quarks. Source: Adapted from Chen et al. (2006). Copyright, American Physical Society (2006), reproduced by permission.

Similarly, for arbitrary $J > 0$ one finds that the combination $P = (-1)^J$, $C = (-1)^{J+1}$ does not occur (cf. Problem 5.7). Hence the quantum number combinations

$$B = Q = S = C = \tilde{B} = 0 \quad (5.15a)$$

and

$$J^{PC} = 0^{--} \text{ and } J^{PC} = 0^{+-}, 1^{-+}, 2^{+-}, 3^{-+}, \dots \quad (5.15b)$$

are ‘exotic’, i.e. they are forbidden in the simple quark model, and a glueball with these quantum numbers cannot mix with a $q\bar{q}$ state. Unfortunately, as can be seen from Figure 5.8, the lightest glueball with exotic quantum numbers is predicted to be a 2^{+-} state with a mass of about $4 \text{ GeV}/c^2$. Because it is so heavy, such a state is likely to be very unstable and difficult to identify, and has not been observed.

However, while there are no established examples of pure glueballs, as mentioned above, there is evidence for hybrid states, which contain one or more gluons in addition to a $q\bar{q}$ pair, and for mixed states, which are superpositions of $q\bar{q}$ and glueball components. In particular, the 2018 edition of the Particle Data Group compilation lists two isotriplets, the $\pi_1(1400)$ and the $\pi_1(1600)$, as established states, with $S = C = \tilde{B} = 0$ and members with $Q = I_3 = +1, 0, -1$ in both cases. These states cannot be glueballs because they have nonzero isospin and cannot be $q\bar{q}$ states because, in both cases, the neutral member has the exotic quantum numbers $J^{PC} = 1^{-+}$. They could be four-quark states or hybrids. However, four-quark states formed from light quarks are expected to occur in

supermultiplets containing members with exotic internal quantum numbers, and no such states have been identified.¹⁴ Hence the simplest explanation is that they are hybrids. In addition, there is a ‘surplus’ of observed states in the range $(1 - 2) \text{ GeV}/c^2$ with the quantum numbers $I = S = C = \tilde{B} = 0$ and $J^{PC} = 0^{++}$, when compared to the predictions of the simple quark model. The latter predicts just two such states with $L = S = 1$, corresponding to the quark combinations $(u\bar{u} + d\bar{d})/\sqrt{2}$ and $s\bar{s}$, while, as we have seen, the lightest glueball is expected to lie in this region. Experimentally four states are observed, called $f^0(980)$, $f^0(1370)$, $f^0(1500)$, and $f^0(1710)$, which are plausibly interpreted as mixtures of the two predicted quark states, the predicted glueball and a hybrid state. Clearly, states do occur that cannot be interpreted as $q\bar{q}$ pairs, but when they do, unless they have exotic quantum numbers, they are likely to be quantum mechanical superpositions of allowed and exotic states.

5.3.1 (b) Heavy quarkonia

The narrow, low-lying states of charmonium, below the charm threshold, are well described by the simple quark model, as we have seen in Section 3.3.5. However, there are also states above the charm threshold, and while many of these exhibit the behaviour expected in the simple quark model, others do not. In particular, in 2005 the BaBar collaboration at SLAC identified a neutral vector meson, conventionally called $X(4260)$, whose existence was subsequently confirmed by other experiments. It was first observed as a resonance peak in the $(\pi^+\pi^-J/\psi)$ mass spectrum seen in the reaction

$$e^+ + e^- \rightarrow \gamma + \pi^+ + \pi^- + J/\psi \quad (5.16a)$$

and has also been observed in the reaction

$$e^+ + e^- \rightarrow X(4260) \rightarrow \pi^+ + \pi^- + J/\psi, \quad (5.16b)$$

corresponding to the production mechanism of Figure 3.19, characteristic of a vector with $J^{PC} = 1^{--}$. The observed mass and width are $M = (4230 \pm 8) \text{ MeV}/c^2$, $\Gamma = (55 \pm 19) \text{ MeV}$.¹⁵ The $X(4260)$ has been observed to decay to $\pi^0\pi^0J/\psi$ and to K^+K^-J/ψ , as well as to $\pi^+\pi^-J/\psi$, and since the $\pi^0\pi^0J/\psi$ can only have isospin 0 or 2 (cf. Problem 5.8) and K^+K^-J/ψ can only have isospin 0 or 1, the $X(4260)$ can only have isospin zero.

In the simple quark model, the $X(4260)$ would be interpreted as a $c\bar{c}$ state lying above the charm threshold. However, such states are expected to decay predominantly by the mechanism of Figure 3.21a into final states

¹⁴For example, one would expect a tetraquark $u\bar{u}u\bar{u}$ with allowed quantum numbers $Q = 0$, $S = 0$ to belong to a supermultiplet that included, among others, states $u\bar{d}u\bar{d}$ and $u\bar{d}u\bar{s}$ with exotic quantum numbers $Q = 2$, $S = 0$ and $Q = 2$, $S = 1$, respectively.

¹⁵Nonetheless, the now anomalous name $X(4260)$, presumably arising from early less precise mass measurements, has stuck.

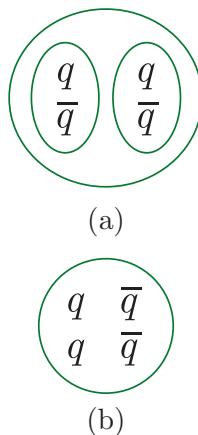


Figure 5.9 Schematic diagram of possible tetraquark structures: (a) a molecular quark state; (b) a more tightly bound four-quark state.

that contain particles with a nonzero charm quantum number, and decay only rarely to give a J/ψ particle in the final state. This is precisely the behaviour observed for the vector charmonium states $\psi(3770)$, $\psi(4040)$, $\psi(4160)$, and $\psi(4415)$, corresponding to the four broad peaks in the total cross-section seen in Figure 3.20. In contrast, the $X(4260)$ couples very weakly to e^+e^- , and open charm decays, far from dominating, have so far not been detected. This behaviour is very difficult to understand if the $X(4260)$ is a $c\bar{c}$ state, but can be understood if it is a hybrid state $c\bar{c}g$ or a *tetraquark-charmonium* state $c\bar{c}q\bar{q}$, where $q = u, d, s$, or a mixture of the three. The first of these possibilities – that it is a hybrid – has the advantage that such states obviously have isospin zero. On the other hand, if it is a four-quark state, one would expect it to belong to a super-multiplet together with so far undetected states with nonzero isospin or strangeness, such as $c\bar{c}u\bar{d}$ or $c\bar{c}s\bar{u}$, although it is possible that this might not be the case if it were a very loosely bound state of a pair of ‘normal’ hadrons that retain, to a significant degree, their individual identities. Such states, whose masses are approximately equal to the sum of the masses of their constituents, are referred to as *molecular states*. They are illustrated schematically, for the case of a di-meson, in Figure 5.9, where they are compared to a more tightly bound four-quark state, which cannot be approximated as a bound state of two mesons.

The ambiguity in the interpretation of the $X(4260)$ only arises because it is electrically neutral; a charged state containing a $c\bar{c}$ pair could not be a simple $c\bar{c}$ or a hybrid, but must necessarily contain extra quarks. That such states do exist was first established in 2013 independently by two experiments that studied the $J/\psi\pi^+$ and $J/\psi\pi^-$ pairs produced in the decay $X(4260)\rightarrow\pi^+\pi^-J/\psi$. The BESIII group, working at the Beijing Electron Positron Collider (BEPCC) studied the reaction $e^+e^-\rightarrow\pi^+\pi^-J/\psi$ at the centre-of-mass energy corresponding to the mass of the $X(4260)$. For each event, the invariant mass of the π^+J/ψ and π^-J/ψ pairs were calculated and the invariant mass distribution of the heaviest pairs found. It is shown in Figure 5.10 and exhibits a clear resonance peak above a background (shown by the dotted line) arising from uncorrelated pairs produced by some other, nonresonant, mechanism. The mass and width of the peak obtained from the fit shown are $M = 3899.0 \pm 3.6 \pm 4.9 \text{ MeV}/c^2$ and $\Gamma = 46 \pm 10 \pm 20 \text{ MeV}$, respectively, where the first errors are statistical and the second are systematic. The Belle collaboration, working at the KEK laboratory near Tokyo, studied the reaction (5.16a) at e^+e^- centre-of-mass energies between 10 and 11 GeV. Events were selected with $\pi^+\pi^-J/\psi$ masses between 4.15 and $4.45 \text{ GeV}/c^2$ to obtain a sample of 689 $X(4260)\rightarrow\pi^+\pi^-J/\psi$ events, with an estimated background of 139 events. The invariant mass distribution of the heavier $\pi^\pm J/\psi$ pair is shown in Figure 5.10 and again there is a clear resonance peak with parameters $M = 3894 \pm 6.6 \pm 4.5 \text{ MeV}/c^2$, $\Gamma = 63 \pm 24 \pm 26 \text{ MeV}$, in agreement with those obtained by the BESIII collaboration. These experiments clearly establish the existence of particles, now called $X^\pm(3900)$, decaying to $J/\psi\pi^\pm$. Subsequently their spin-parity has been determined to be $J^P = 1^+$ and a neutral isospin partner $X^0(3900)$ decaying to $J/\psi\pi^0$.

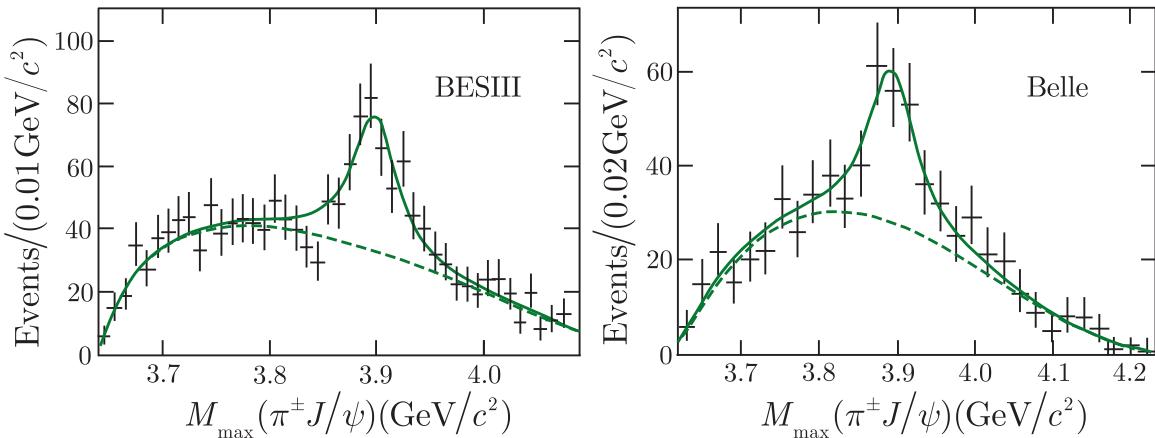


Figure 5.10 Invariant mass distribution of the heaviest pairs $\pi^\pm J/\psi$ observed in the reaction $e^+e^- \rightarrow \pi^+\pi^- J/\psi$ by the BESIII collaboration (Source: Adapted from Ablikim et al. 2013) and in the reaction $e^+e^- \rightarrow \gamma\pi^+\pi^- J/\psi$ by the Belle collaboration (Source: Adapted from Liu et al. 2013). Copyright, American Physical Society, reproduced with permission.

has also been identified.¹⁶ These particles must contain a $c\bar{c}$ pair, since otherwise the observed decays would be heavily suppressed by the OZI rule, and would make a tiny contribution to the decay rate compared to the very many possible OZI allowed final states that do not contain charmed quarks¹⁷; their minimal quark compositions must be $c\bar{c}u\bar{d}$ and $c\bar{c}d\bar{u}$ for the positive and negatively charged states respectively.

In the above, we focussed on the $X(4260)$ because its study led to the discovery of the first four-quark states $X^\pm(3900)$ to be unambiguously identified. However, several other neutral meson states that do not fit into the simple quark model have also been observed. These include the $X(4360)$ and $X(4660)$, both with $J^{PC} = 1^{--}$, somewhat analogous to the $X(4260)$, and the $X(3872)$, which is a 1^{++} state with a decay width measured to be less than 1.2 MeV, even though it lies well above the charm threshold.¹⁸ The $X(3872)$ sits very close to the $D\bar{D}^*$ threshold and is known to decay to $D^0\bar{D}^{*0}$ with a branching ratio of at least 24%. It is therefore a natural candidate for a molecular $D\bar{D}^*$ state, although other explanations are undoubtedly possible at present. In addition, several other anomalous particles have been reported in both the charmonium

¹⁶The notation for these various states is not settled and the letters X , Y , and Z are used in the literature for suspected exotic $c\bar{c}$ states in a way that is not always mutually consistent. Here we refer to all candidate exotic mesons thought to include charmed quarks as X . The Particle Data Group includes them in a list entitled *c̄c mesons (including possible non-q̄q states)* without distinguishing between them, referring to the $X(4260)$ and the $X(3900)$ discussed above as the $\psi(4260)$ and the $Z_c(3900)$.

¹⁷The decay width of a $q\bar{q}$ state of this mass composed of light quarks would be expected to be so large it is very unlikely such a state could ever be detected.

¹⁸Historically, this was the first exotic candidate to be discovered in the charmonium sector, by the BELLE collaboration in 2003. It is referred to by the Particle Data Group as the $\chi_{c1}(3872)$.

and bottomonium spectra, but need further verification before they can be regarded as firmly established states. In short, while some exotic states clearly exist, much more experimental data are needed to clarify their nature and to establish the existence of others. Hopefully, the much higher statistics available at the recently upgraded facilities at KEK(SuperBelle) and BEPC, as well as data from other facilities, including LHCb, will provide these in the next few years.

5.3.1 (c) Exotic baryons

From time to time, experiments have claimed evidence for the existence of baryons with exotic quantum numbers, but as more experiments have been performed and more data accumulated, the evidence has receded. The present consensus is that there is no convincing evidence for such baryons. However, following the discovery of the tetraquark–charmonium states described above, it was natural to search for the analogous *pentaquark–charmonium* states $c\bar{c}qqq$, where $q = u, d, s$, which would not have exotic quantum numbers. The first evidence for such states was published in 2015 by the LHCb collaboration at CERN, who used a detector designed specifically to study the decays of short-lived particles containing heavy quarks, which are produced in large numbers at the LHC.¹⁹ In this case, it was used to measure about 26 000 decays of the form

$$\Lambda_b^0 \rightarrow K^- + p + J/\psi. \quad (5.17a)$$

Here Λ_b^0 denotes $\Lambda_b^0(5624) = ubd$, the lightest baryon with a nonzero bottom quantum number. The main contribution to the decay (5.17a) is expected to arise from the sequence

$$\Lambda_b^0 \rightarrow \Lambda^* + J/\psi; \quad \Lambda^* \rightarrow K^- + p, \quad (5.17b)$$

where Λ^* is any resonant excitation of the $\Lambda(1116) = uds$ state discussed in Section 3.3.3 (cf. (3.72)). This is confirmed by the observed invariant mass distribution of the $K^- p$ pairs shown in Figure 5.11a, which shows a prominent peak corresponding to the $\Lambda(1520) = uds$ resonance and a broader enhancement at higher masses to which several known Λ^* states can contribute. In addition, the $J/\psi p$ mass distribution, shown in Figure 5.11b, exhibits a peak at about $4.5 \text{ GeV}/c^2$, which is naturally interpreted as a $J/\psi p$ resonance P_c^+ occurring in the decay sequence

$$\Lambda_b^0 \rightarrow K^- + P_c^+; \quad P_c^+ \rightarrow J/\psi + p. \quad (5.17c)$$

However, the peak sits upon a large background arising mostly from reaction (5.17b), and in order to clarify the situation, the authors carried out a very detailed analysis including contributions to the reaction (5.17b)

¹⁹The detector used by the LHCb collaboration is described in Section 4.5.2 of Martin and Shaw (2017).

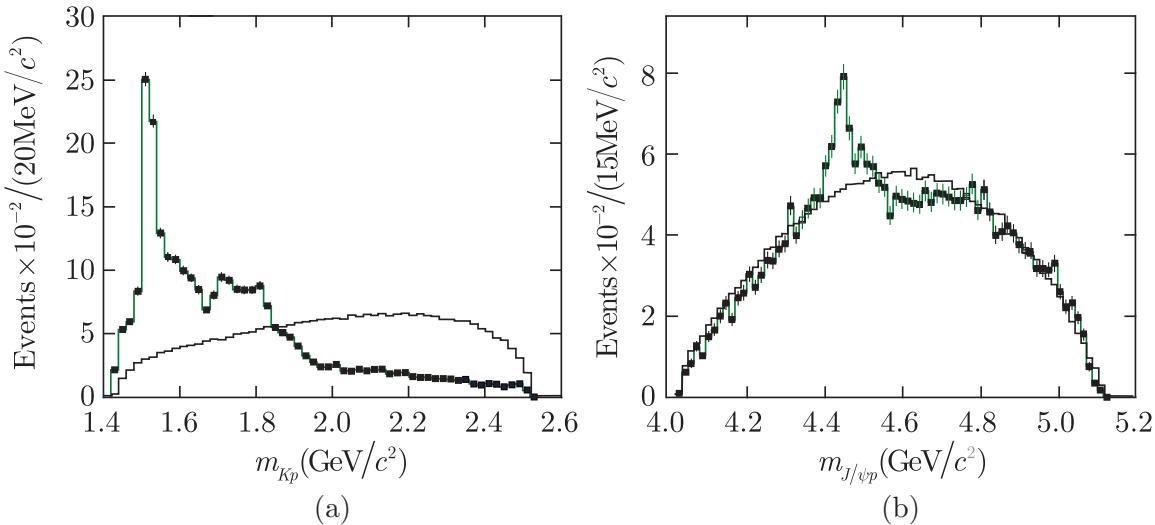


Figure 5.11 Invariant mass of (a) $K^- p$ and (b) $J/\psi p$ combinations from $\Lambda_b^0 \rightarrow J/\psi K^- p$ decays. The black line is the expectation from phase space; the green line is the data. Source: Adapted from Aaij et al. (2015).

from all 14 known Λ^* resonances. This confirmed that the data could not be understood without contributions of the form (5.17c) and suggested that to understand the detailed mass and angular distributions required not one, but two resonances P_c^+ , which were found to have the masses and widths

$$M = (4380 \pm 8 \pm 29) \text{ MeV}/c^2, \quad \Gamma = (205 \pm 18 \pm 86) \text{ MeV}$$

and (5.17d)

$$M = (4449.8 \pm 1.7 \pm 2.5) \text{ MeV}/c^2, \quad \Gamma = (35 \pm 5 \pm 19) \text{ MeV},$$

respectively. They have the same internal quantum numbers as the proton, but, as in the case of the tetraquark–charmonium states discussed above, the presence of the J/ψ in their decay products is a strong indication of a $c\bar{c}$ pair, so the obvious interpretation is that they are pentaquark–charmonium states with the quantum composition $P_c^+ = uudc\bar{c}$, although they could well be molecular states. Finally, while the above evidence strongly suggests contributions of the form (5.17c), further experimental confirmation of their detailed properties (5.17d) is required before they can be regarded as being fully established.²⁰

5.3.2 The quark–gluon plasma

In ordinary matter, quarks are confined within hadrons. However, as the density and/or temperature is increased, a phase transition can in

²⁰For a fuller discussion, see Karliner and Skwarnicki (2016).

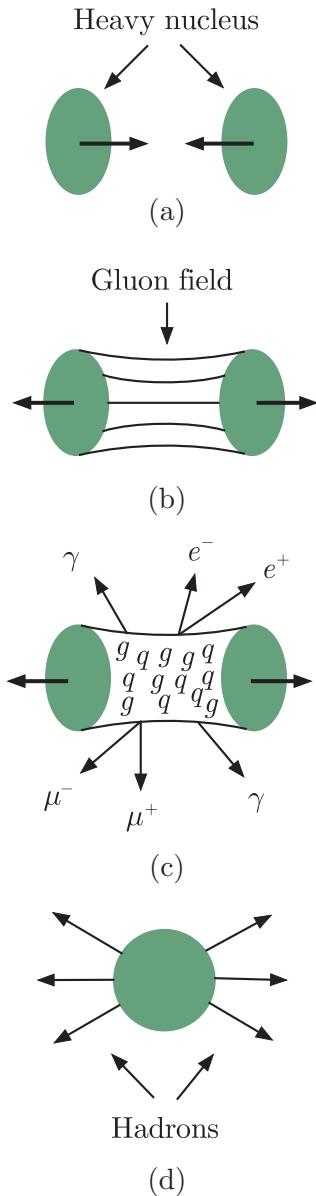


Figure 5.12 Stages in the formation of a quark–gluon plasma and subsequent hadron emission: (a) two heavy nuclei collide at high energies and (b) interact via the colour field; (c) the very high energy-density produced causes the quarks and gluons to deconfine and form a plasma that can radiate photons and lepton pairs; and (d) finally, as the plasma cools, hadrons condense and are emitted.

principle occur to a state in which individual hadrons lose their identities, and quarks and gluons become free to move across a volume that is large compared to a hadron. Lattice gauge theory and other approximate methods suggest that in the low-temperature limit, the transition should occur at densities about five times higher than that at the centre of a heavy nucleus, while in the limit of low nucleon density, it should occur at temperatures of order 150–200 MeV, which translates to a conventional temperature of order 10^{12} K, about five orders of magnitude greater than the temperature at the centre of the Sun. The resulting new state of matter is called a *quark–gluon plasma (QGP)*. A quark–gluon plasma is believed to have existed in the first few microseconds after the Big Bang, and it may exist today at the centre of neutron stars. More prosaically, a quark–gluon plasma may be created briefly in collisions between heavy ions, if the collision energy is large enough. The collisions in the relativistic ion colliders described below are characterised by length and timescales typically of order 10 fm and 10^{-23} s, and if any QGP is created in the collision, the initial ‘fireball’ rapidly cools and expands under its own pressure, first forming a plasma emitting lepton and photon radiation. Then, as it cools further, it will pass through the transition temperature and a gas of hadrons will be formed, resulting in large numbers of hadrons being observed in the final states. This is illustrated schematically in Figure 5.12. The existence of a new state of matter can then be inferred from the properties of the emitted particles, and particularly the differences from what would be expected if they were simply produced by collisions between the individual nucleons contained in the heavy ions.

High-energy collisions between heavy ions have been studied in several experiments, at the Relativistic Heavy Ion Collider (RHIC) at the Brookhaven National Laboratory, starting in 2000, and more recently using ALICE and other detectors at the LHC, starting in 2010. RHIC typically collides two counter-circulating beams of fully stripped gold ions at a maximum energy of 200 GeV per nucleon, i.e. a total collision energy of about 32 TeV. At the LHC, the collisions are between lead ions at a much higher total energy of 574 TeV. If the ions collide centrally (i.e. head-on) several thousand final-state particles are produced. An example of an event seen in the STAR detector (which was shown in Figure 4.22) is shown in Figure 5.13.

The key question is whether the energy-density in these collisions is sufficient to have created a quark–gluon plasma and its subsequent cooling phases. There are many signatures for this, including the relative abundances of different final-state particle types. For example, the large numbers of gluons in the plasma would lead to copious production of $s\bar{s}$ pairs via gluon fusion $gg \rightarrow s\bar{s}$, and hence production of strange particles in excess of that expected if the colliding nuclei behaved simply as a source of nucleon–nucleon collisions. On the other hand, the production of J/ψ and upsilon states would be suppressed because the c and \bar{c} , or b and \bar{b} , quarks produced (also from gluon fusion) would be separated by many

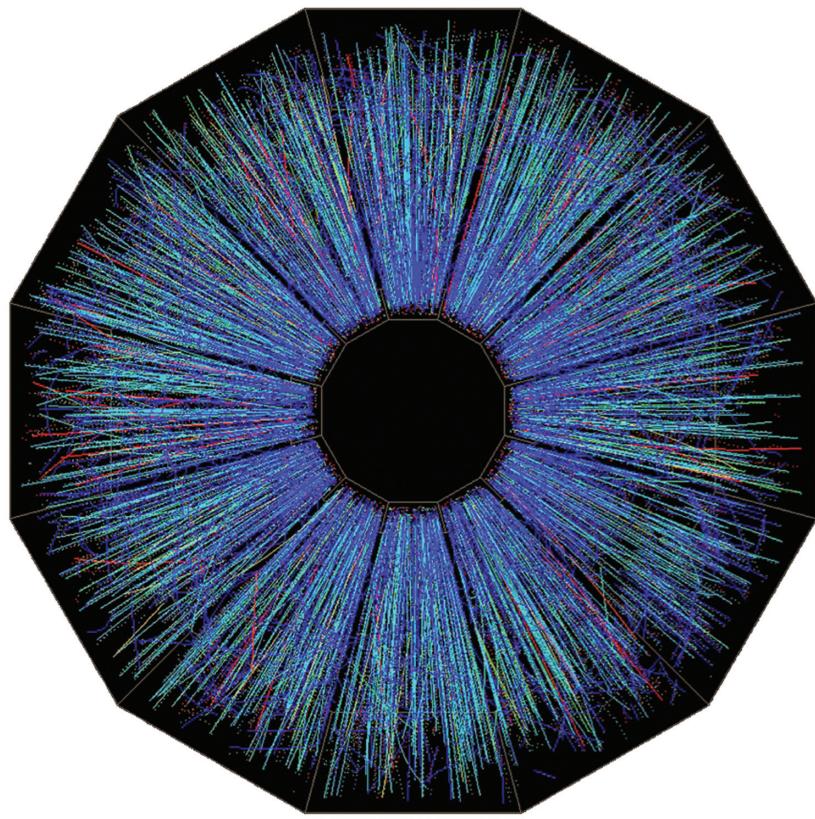


Figure 5.13 View of a 200 GeV gold–gold interaction in the STAR detector at the RHIC accelerator. Source: Courtesy of Brookhaven National Laboratory.

quarks of other flavours, leading instead to the production of mesons with nonzero charm and bottom quantum numbers. Data on the $\mu^+\mu^-$ pairs produced in both heavy ion and proton–proton collisions are shown in Figure 5.14, where the $\mu^+\mu^-$ invariant mass region covers the mass range of the lowest three upsilon states $\Upsilon(9.46)$, $\Upsilon(10.02)$, and $\Upsilon(10.36)$. The three states are significantly suppressed in the Pb–Pb ion collisions, with the $\Upsilon(9.46)$ rate about a factor 2 below what would be expected if the ions behaved as just a collection of nucleons. This is in accord with theoretical expectations, and in general, these and many other measurements support the existence of a new phase of matter and confirm approximate predictions for the transition temperature. However, whether the transition is a true phase transition, or a more gradual transition as the temperature decreases, is not yet clear. The experiments also suggest that, presumably because of the strength of the colour force between quarks, the plasma observed at present temperatures behaves more like a dense liquid than as a gas of quasi-free particles, as in a true plasma formed of electrons and positive ions. Future experiments at RHIC and at the LHC will play a crucial role in resolving these and other detailed questions.

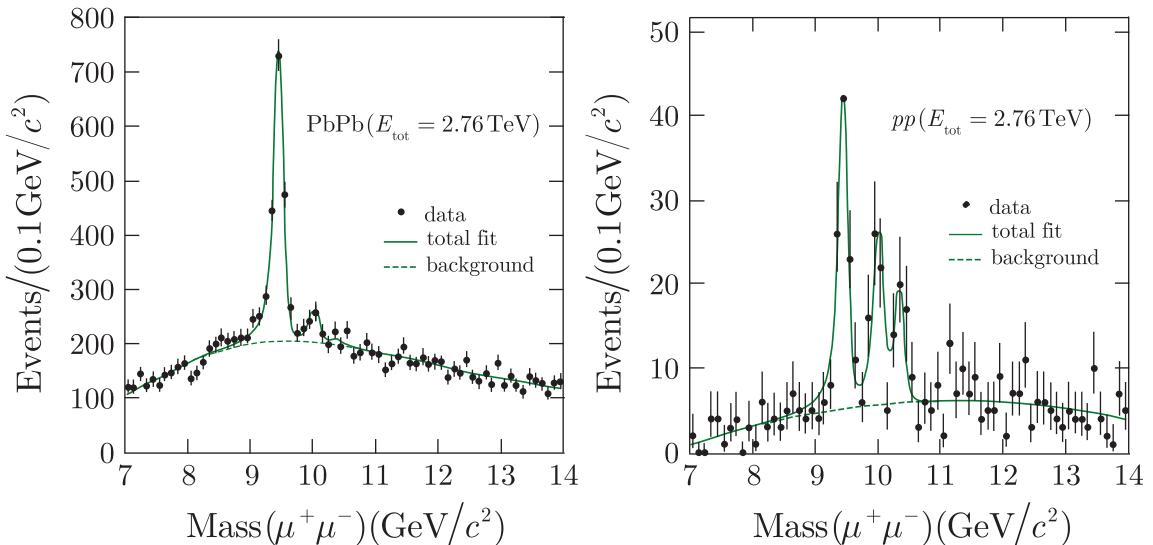


Figure 5.14 Dimuon invariant-mass distributions in PbPb (left) and pp (right) data at a total centre-of-mass energy of 2.76 TeV. Source: Adapted from Chatrchyan et al. (2012). The same reconstruction and analysis selection criteria are applied to both datasets. The solid (signal + background) and dashed (background) curves show the results of the simultaneous fit to the two datasets.

5.4 Jets and gluons

A striking feature of many high-energy particle collisions is the occurrence of jets of hadrons in the final state. We have already mentioned these in Section 3.2.1 when we discussed the experimental evidence for quarks, and they are a feature of the fragmentation process mentioned in Section 5.2.2. Jets have been extensively studied in the reaction



at high energies using colliding beam experiments. High-energy electrons and positrons collide head-on, with equal and opposite momenta, so that the total momentum of the hadrons produced cancels out to zero in order to conserve momentum. This is a particularly ‘clean’ reaction, because the initial particles are elementary, without internal structure. In the centre-of-mass energy range 15–40 GeV, electron–positron annihilation into hadrons is dominated by the production of jets. This can be regarded as occurring in two stages: a primary electromagnetic process $e^+ + e^- \rightarrow q + \bar{q}$ (due to photon exchange), leading to the production of a quark–antiquark pair, followed by fragmentation, which converts the high-energy $q\bar{q}$ pair into two jets of hadrons. These are illustrated in Figure 5.15.

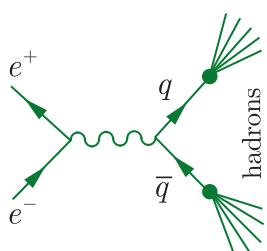


Figure 5.15 Basic mechanism of a two-jet production in electron–positron annihilation.

The fragmentation process that converts the quarks into hadrons is very complicated, and the composition of the jets – i.e. the numbers and types of particles in the jet and their momenta – varies from event to event.

However, the direction of a jet, defined by the total momentum vector

$$\mathbf{P} = \sum_i \mathbf{p}_i, \quad (5.19)$$

where the sum extends over all the particles within the jet, reflects closely the parent quark or antiquark direction. This is because the QCD interaction is relatively weak at very short distances (asymptotic freedom), and the quark and antiquark do not interact strongly until they are separated by a distance r of order 1 fm. At these relatively large distances only comparatively small momenta can be transferred, and hence the jets that subsequently develop point almost exactly in the initial quark and antiquark directions. That is, the jet angular distribution relative to the electron beam direction reflects the angular distributions of the quark and antiquark in the basic reaction $e^+ + e^- \rightarrow q + \bar{q}$. The latter can be easily calculated in QED as it is a purely electromagnetic process, and is in excellent agreement with the observed angular distribution of the jets. This is one of the pieces of evidence for the existence of quarks that was cited in Section 3.2.

Although the dominant process in electron–positron annihilation into hadrons is the formation of two ‘back-to-back’ jets, occasionally we would expect a high momentum gluon to be emitted by the quark or antiquark before fragmentation occurs, in much the same way as a high-energy electron sometimes emits a photon (i.e. bremsstrahlung). The quark, antiquark, and gluon then all fragment into hadrons, leading to a three-jet event. A computer reconstruction of such an event in a jet chamber is shown in Figure 5.16. The solid lines in this figure indicate the reconstructed charged particle trajectories taking into account the known magnetic field, which is also parallel to the beam direction; the dotted lines indicate the reconstructed trajectories of neutral particles, which were detected outside the chamber by other means.

Events like those of Figure 5.16 provided the first unambiguous evidence for gluons, because the angular distributions of the jets was found to be in good agreement with the theoretical expectation for spin-1 gluons, but are inconsistent with what would be expected if, for example, the third jet originated from a particle of spin 0. The ratio of three-jet to two-jet events can also be calculated, assuming that the third jet is a gluon, because the probability that a quark or antiquark will emit a gluon is determined by the strong coupling α_s , in the same way that the probability that an electron or positron will emit a photon is determined by the fine structure constant α . This leads to a value of α_s that is consistent with those found from other determinations and lends further support for the picture of coloured quarks interacting via the exchange of coloured gluons.

5.4.1 Colour counting

What evidence is there that quarks exist in just three colour states? This question can be answered by comparing the total cross-section for electron

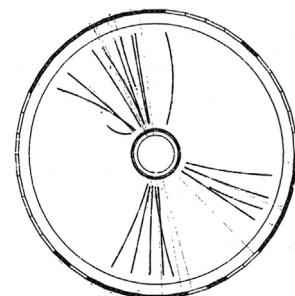


Figure 5.16 Computer reconstruction of a typical ‘three-jet’ event in electron–positron annihilation observed in the JADE jet chamber. Source: After Wu (1984), copyright Elsevier, with permission.

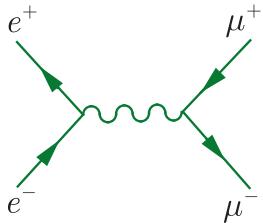


Figure 5.17 Dominant mechanism for electron–positron annihilation to muon pairs.

positron annihilation into hadrons with the total cross-section for annihilation into muon pairs for centre-of-mass energies in the range between 10 and 50 GeV. In this energy region, these cross-sections both decrease rapidly with energy, but their ratio

$$R \equiv \frac{\sigma(e^+e^- \rightarrow \text{hadrons})}{\sigma(e^+e^- \rightarrow \mu^+\mu^-)} \quad (5.20)$$

is almost energy independent. The near constancy of this ratio follows from the dominance of the two-step mechanism of Figure 5.15, with the total annihilation rate being determined by that of the initial reaction $e^+ + e^- \rightarrow q + \bar{q}$, while the total cross-section for muon pair production is dominated in this energy range by the almost identical mechanism²¹ shown in Figure 5.17. The value of the ratio R then directly confirms the existence of three colour states, each with the same electric charge, for each quark flavour.

To understand this, let us suppose that each quark flavour $f = u, d, s, \dots$ exists in N_C colour states, so that $N_C = 3$ according to QCD, while $N_C = 1$ if the colour degree of freedom does not exist. Since the different colour states all have the same electric charge, they will all be produced equally readily by the mechanism of Figure 5.15, and the rate for producing quark pairs of any given flavour $f = u, d, s, \dots$ will be proportional to the number of colours N_C . The cross-section is also proportional to the squared charge of the produced pair (because this is a first-order electromagnetic process), and since muon pairs are produced by an identical mechanism, we obtain

$$\sigma(e^+e^- \rightarrow q_f \bar{q}_f) = N_C e_f^2 \sigma(e^+e^- \rightarrow \mu^+\mu^-), \quad (5.21)$$

where e_f is the electric charge, in units of e , on a quark q_f of flavour f . The cross-section for $e^+ + e^- \rightarrow \text{hadrons}$ will receive an additional contribution of the form (5.21) when the energy passes a threshold for a new quark flavour to be produced. Thus R at low energies will have a series of ‘steps’ corresponding to the production of pairs of new quarks and this is what is observed experimentally. At high energies above the threshold for the production of $b\bar{b}$ pairs, and assuming that hadron production is completely dominated by the two-step process of Figure 5.15, we would have²²

$$R = R_0 \equiv N_C (e_u^2 + e_d^2 + e_s^2 + e_c^2 + e_b^2) = 11N_C/9. \quad (5.22)$$

When the small contribution from the three-jet events and other corrections of order α_s are taken into account, this expression for R is modified to

$$R = R_0(1 + \alpha_s/\pi), \quad (5.23)$$

²¹At higher energies it is also necessary to include the effect of the weak interaction. This will be discussed in Section 6.4.2.

²²There is no contribution from the top quark because it is too heavy to be produced, even at the high energies we are considering.

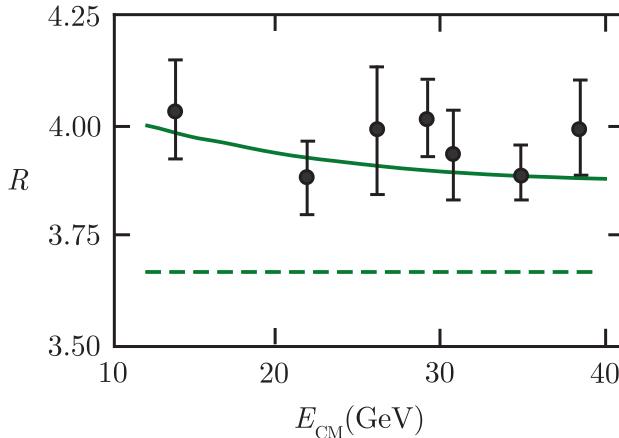


Figure 5.18 Measured values of the cross-section ratio R and the theoretical prediction (5.23) from QCD for $N_C = 3$ colours. The dashed line shows the prediction without QCD corrections.

giving rise to a weak energy dependence of R from the energy dependence of α_s , discussed earlier (cf. (5.9)). Although these corrections of order α_s are small compared to the dominant contribution, they must be included if the experimental data on R are to be accounted for. The data shown in Figure 5.18 are in excellent agreement with the theoretical prediction for the value $N_C = 3$, and hence prove that quarks exist in just three colour states.

5.5 Deep inelastic scattering and nucleon structure

In this section we will consider the dynamical effects of hadron substructure as seen in scattering experiments. Such effects are most directly seen in high-energy inelastic scattering experiments in which beams of leptons, which are themselves structureless, are used to probe the structures of the proton and neutron at high resolution. This leads to a detailed picture of the internal structure of nucleons, which is an essential tool for understanding many other reactions.

5.5.1 Scaling

We begin by studying the inelastic scattering reactions

$$\ell^- + p \rightarrow \ell^- + X \quad (\ell = e, \mu),$$

where X as usual denotes any set of hadrons allowed by the conservation laws. These reactions are dominated by the one-photon exchange mechanism of Figure 5.19, where the energies and momenta of the leptons and proton are also defined. Unlike elastic scattering, where at a given lepton energy E there is only one free variable (e.g. the scattering angle), in inelastic scattering the excitation energy of the nucleon adds a further degree of freedom, so we can define two independent variables. These are usually taken to be ν , defined by

$$2Mc^2\nu \equiv W^2c^4 + Q^2 - M^2c^4 \quad (5.24)$$

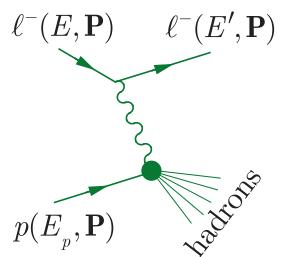


Figure 5.19 Dominant one-photon exchange mechanism for inelastic lepton–proton scattering, where $\ell = e$ or μ .

and a dimensionless quantity (called the *scaling variable*) given by

$$x \equiv Q^2/2Mc^2\nu. \quad (5.25)$$

Here M is the proton mass, W is the invariant mass of the final-state hadrons, and Q^2 is related to the squared 4-momentum transfer q^2 by

$$Q^2 = -c^2 q^2 = -[(E - E')^2 - (\mathbf{p}c - \mathbf{p}'c)^2]. \quad (5.26)$$

In the rest frame of the initial proton, ν reduces to

$$\nu = E - E' \quad (5.27)$$

and so is the Lorentz-invariant generalisation for the energy transferred from the lepton to the proton. For elastic scattering, the final hadron state is just a proton, with the invariant $W = M$, corresponding to the fixed value $x = 1$. Here we are concerned with the so-called *deep inelastic scattering* region in which Q^2 is large and the scaling variable x varies between 0 and 1, depending on the invariant mass of the produced hadron state W .

In Figure 5.19, the interaction of the photon with the lepton is well-understood and can be precisely calculated, but its interaction with the hadrons, denoted by the ‘blob’ at the bottom vertex is not precisely known. However, provided the electron energies are sufficiently large for their masses to be neglected, it can be shown using Lorentz invariance and appropriate conservation laws that the differential cross-section takes the general form

$$\begin{aligned} \frac{d\sigma}{dE'd\Omega} &= \frac{\alpha^2}{4E^2 \sin^4(\theta/2)} \frac{(\hbar c)^2}{\nu} \\ &\times \left[\cos^2(\theta/2) F_2(x, Q^2) + \sin^2(\theta/2) \frac{Q^2}{xM^2c^4} F_1(x, Q^2) \right] \end{aligned} \quad (5.28)$$

where the dimensionless *structure functions* $F_1(x, Q^2)$ and $F_2(x, Q^2)$ parameterise the interaction at the lower vertex in Figure 5.19. For values of $W \leq 2.5 \text{ GeV}/c^2$, the measured cross-sections show considerable structure due to the excitation of nucleon resonances. However, above this mass they are smoothly varying and the measured structure functions become approximately independent of Q^2 ; that is,

$$F_{1,2}(x, Q^2) = F_{1,2}(x) \quad (Q^2 \gg M^2c^4) \quad (5.29)$$

is a good approximation over quite large ranges of Q^2 . The behaviour (5.29) is called *Bjorken scaling* or *scale invariance*,²³ because it implies that the structure functions are left unchanged by a scale transformation;

²³The pioneering work on deep inelastic scattering, leading to the discovery of scaling, was done by groups led by Jerome Friedman, Henry Kendall, and Richard Taylor, and resulted in them receiving the 1990 Nobel Prize in Physics.

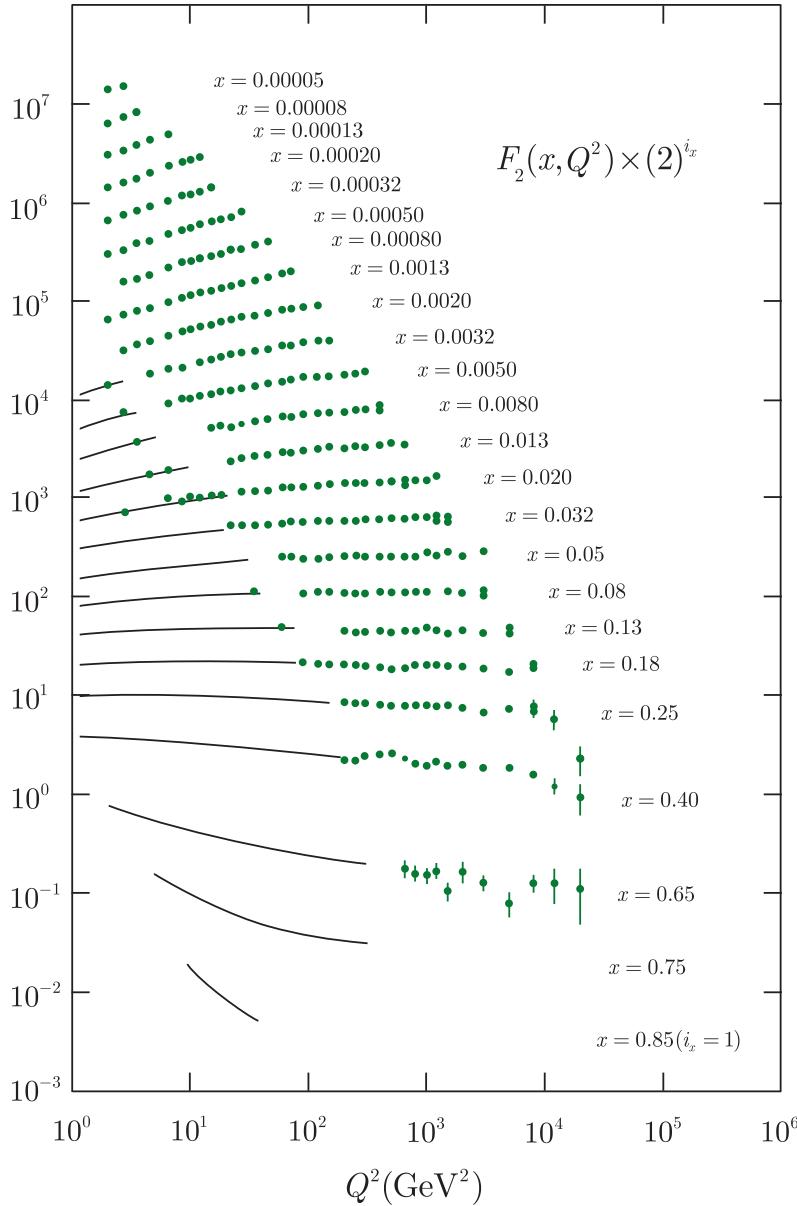


Figure 5.20 Measured values $2^{i_x} \times F_2(x, Q^2)$ for $Q^2 > 1 \text{ GeV}^2$, where $i_x = 1, 2, \dots, 24$ is an integer labelling the various x values in turn starting from $x = 0.85$. The green points are data obtained using the HERA electron–proton collider at DESY, Hamburg, while the curves indicate the behaviour obtained in several earlier experiments using lepton scattering from a fixed proton target. Source: Adapted from the review of Foster et al. (2018).

that is, by a transformation in which all particle masses, energies, and momenta are multiplied by a scale factor κ , so that $Q^2 \rightarrow \kappa^2 Q^2$ but x remains unchanged. Scale invariance is not exact as we have said; deviations from it are called *scaling violations*.

These features can be clearly seen in Figure 5.20, which shows the measured Q^2 dependence of the proton structure function F_2 for $Q^2 > 1 \text{ GeV}^2$ at various fixed values of x .²⁴ In contrast to the corresponding

²⁴This figure should be treated with care, because to fit all the data shown on one graph, log scales have been used and, at each fixed x , the data have been multiplied by a scale factor that varies from 2 at $x = 0.85$ to $2^{24} \approx 1.7 \times 10^7$ at the smallest value of x .

contribution from elastic scattering, which varies by about five orders of magnitude over the much smaller range $1 < Q^2 < 30 \text{ GeV}^2$, the curves in Figure 5.20 are slowly varying, i.e. there is approximate scale invariance. On the other hand, deviations from exact scaling are clearly seen and are largest at very small and very large x . In what follows, we shall initially consider the origin of scaling, deferring discussion of the deviations from it to Section 5.5.3.

5.5.2 The quark–parton model

The interpretation of scaling is simplest in a reference frame where the target nucleon is moving with a very high velocity, so that the transverse momenta and rest masses of its constituents may be neglected. The structure of the nucleon is then given by the longitudinal momentum of its constituents. This approach was first adopted by Feynman and Bjorken, who called the constituents *partons*. (We now identify charged partons with quarks and neutral partons with gluons.) In the *parton model*, deep inelastic scattering is visualised as shown in Figure 5.21. The target nucleon is a stream of partons each with 4-momentum zP , where $P = (p, \mathbf{p})$ is the 4-momentum of the nucleon and \mathbf{p} is very large, so that the nucleon and parton masses may be neglected.

Suppose now that one parton of mass m is scattered elastically by the exchanged photon of 4-momentum q . Then

$$(zP + q)^2 = (z^2 P^2 + 2zP \cdot q + q^2) = m^2 c^2 \approx 0.$$

If $c^2 |z^2 P^2| = z^2 M^2 c^4 \ll Q^2$, then

$$z = -\frac{q^2}{2P \cdot q} = \frac{Q^2}{2Mc^2\nu} \equiv x, \quad (5.30)$$

where the invariant scalar product has been evaluated in the laboratory frame in which the energy transfer is ν and the nucleon is at rest. Thus, we see that the physical interpretation of the scaling variable x is the fractional 3-momentum of the struck parton in the reference frame where the nucleon has a very high velocity. This is a key result in interpreting

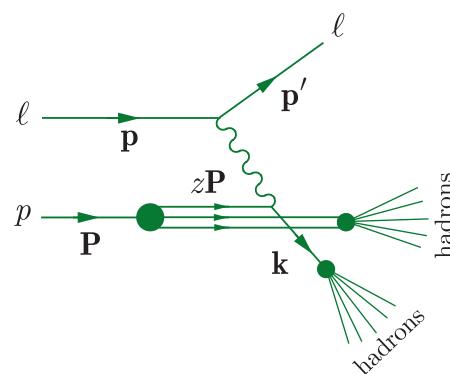


Figure 5.21 Dominant one-photon exchange mechanism for inelastic lepton–proton scattering where $\ell = e$ or μ .

the data. It implies that the measured cross-section at a given value of $x = Q^2/2Mc^2\nu$ is proportional to the probability of finding a parton with a fraction $z = x$ of the proton momentum. The full calculation is more complicated, and we will simply quote below the results obtained for the structure functions in the parton model.²⁵ We will do this for both spin-0 and spin-1/2 partons with arbitrary electric charges in order to see which spin and charge values, if any, are selected by the data.

We start with the structure function F_2 , which is given for both spin-0 and spin-1/2 partons by

$$F_2(x, Q^2) = \sum_a e_a^2 x f_a(x), \quad (5.31)$$

where $f_a(x) dx$ is the probability of finding a parton a of charge $e e_a$ with fractional momentum between x and $x+dx$, and the sum extends over all parton types a . The parton distributions $f_a(x)$ are not determined by the model, so that the x -dependence of the structure functions must be determined by experiment. However, they are predicted to be the same at all Q^2 ; that is, one has exact scaling for point-like partons, irrespective of their spin. In contrast, the predictions obtained for the structure function F_1 do depend on the spin of the partons. They are

$$F_1(x, Q^2) = 0 \quad (\text{spin} - 0) \quad (5.32a)$$

and

$$2xF_1(x, Q^2) = F_2(x, Q^2) \quad (\text{spin} - 1/2), \quad (5.32b)$$

where the latter result is called the *Callan–Gross relation*. The data rule is decisively in favour of (5.32b), as one would expect if the partons are quarks.

Finally, if the partons are quarks, their squared charges e_a^2 must correspond to the squared quark charges $1/9$ and $4/9$ in units of e . Unfortunately, this cannot be confirmed from measurements of the proton structure functions alone, because of the occurrence of the unknown probability distributions $f_a(x)$ in (5.31). However, by combining measurements of the proton structure functions with those for neutron structure functions, and with similar results obtained in neutrino scattering, it is possible to check that the squared parton charges are consistent with those expected for quarks, as we shall illustrate in Section 5.5.4.

In summary: the scaling behaviour of F_2 is evidence for point-like constituents within the proton; the Callan–Gross relation shows that they are spin-1/2 particles; and the checks on their squared charges confirm their identification with quarks beyond all reasonable doubt.

5.5.3 Scaling violations and parton distributions

The parton model described above was developed before the advent of QCD and completely ignores the quark–gluon interaction. This

²⁵The derivation is given, for example, in pp. 191–192 of Halzen and Martin (1984).

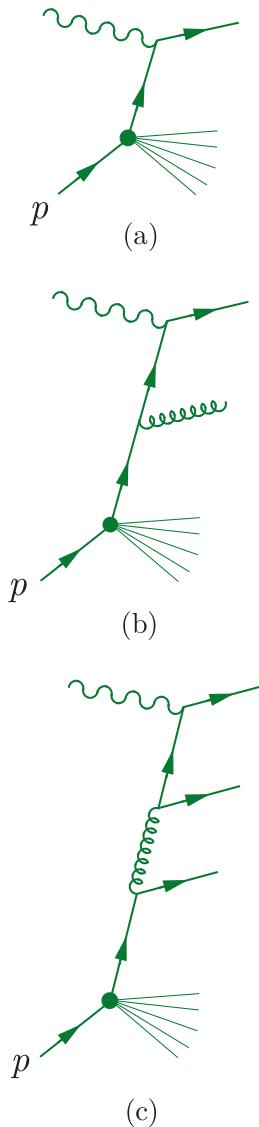


Figure 5.22 (a) The interaction of the exchanged photon with the struck quark in the parton model, together with (b) and (c), two of the additional processes that occur when quark–gluon interactions are taken into account.

interaction is, of course, what holds the proton together, and it manifests itself in deep inelastic scattering in three important ways. The first is that the leptons not only scatter from the three so-called *valence* quarks *uud* that give rise to the internal quantum numbers

$$B = Q = 1, \quad S = C = \bar{B} = T = 0$$

of the proton, but also from a so-called *sea* of quark–antiquark pairs *a* \bar{a} that are created by the interaction.²⁶ The second is that these quarks and antiquarks together carry a little less than one half of the momentum of a high-energy proton, the rest being accounted for by gluons. Finally, the resulting structure functions become weakly dependent on Q^2 , as seen experimentally in Figure 5.20. Here we shall discuss in turn how each of these features arises.

In the absence of quark–gluon interactions, the interaction of the exchanged photon with the struck quark in the nucleon may be depicted by Figure 5.22a. When the quark–gluon interactions are taken into account, other processes besides those of Figure 5.22a can contribute, like those shown in Figures 5.22b,c. In Figure 5.22b the quark emits a gluon before interacting with the exchanged photon. In Figure 5.22c the emitted gluon is converted to a quark–antiquark pair, which forms part of the sea of quark–antiquark pairs mentioned above, and it is one of these particles that absorbs the exchanged photon. Detailed calculations show that the parton model remains a good first approximation when such processes are taken into account, provided the parton distribution functions $f_a(x)$ are extended to include contributions from antiquarks as well as quarks. That is, (5.31) becomes

$$F_2(x, Q^2) = \sum_{a, \bar{a}} [e_a^2 x f_a(x) + e_{\bar{a}}^2 x f_{\bar{a}}(x)], \quad (5.33)$$

where the sum extends over all quark flavours $a = u, d, s, \dots$. The Callan–Gross relation (5.32b) remains unchanged, since both quarks and antiquarks have spin 1/2.

Electron and muon scattering cannot separate the quark and antiquark contributions because the structure functions depend on the squared charges of the constituents, which are the same for quarks and antiquarks. However, neutrino scattering experiments can distinguish between quarks and antiquarks, and are sensitive to the same distribution functions $f_a(x)$ and $f_{\bar{a}}(x)$ as electron and muon scattering.²⁷ The resulting quark and antiquark distributions

$$x q(x) \equiv \sum_a x f_a(x), \quad x \bar{q}(x) \equiv \sum_{\bar{a}} x f_{\bar{a}}(x), \quad (5.34a)$$

²⁶They should be compared to the electron–positron pairs contributing to the vacuum fluctuations of the electron in Figure 5.6a. They also play a crucial role in asymptotic freedom, as discussed in Section 5.2.2, following Figure 5.7.

²⁷Neutrino scattering is discussed briefly in the next section.

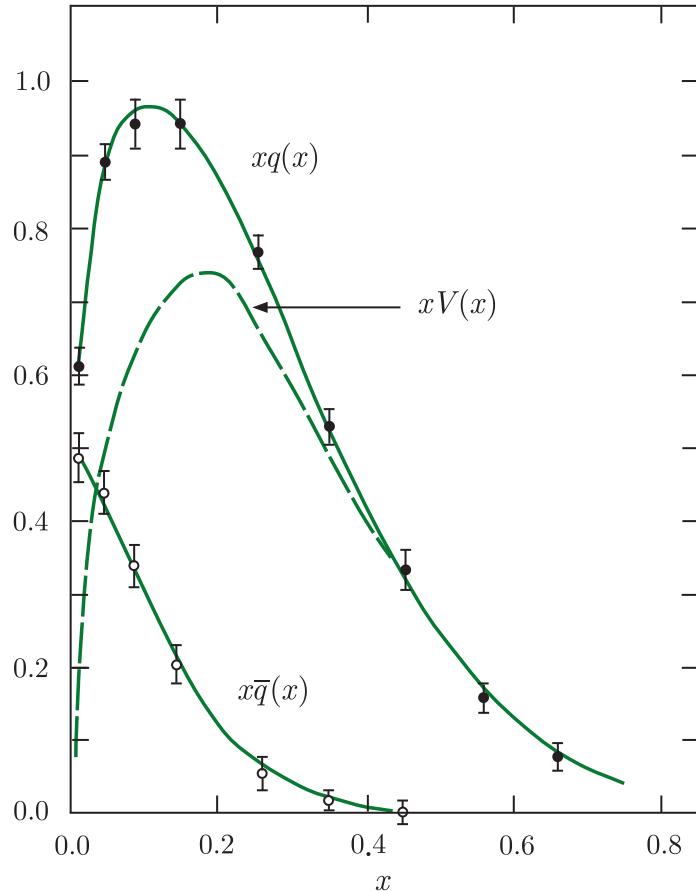


Figure 5.23 Quark and antiquark distributions (5.34a), together with the valence quark distribution (5.34b), measured at a Q^2 value of about 10 GeV^2 , from neutrino experiments at CERN and Fermilab.

where the sum extends over all quark flavours $a = u, d, s, \dots$, are shown in Figure 5.23. Also shown is the difference

$$xV(x) \equiv xq(x) - x\bar{q}(x), \quad (5.34b)$$

which corresponds to the valence quark distribution, if we assume that the quarks belonging to the sea of $q\bar{q}$ pairs produced by the gluons have the same momentum distribution as the antiquarks.²⁸ As can be seen, the valence quarks dominate the distributions for $x \geq 0.3$, while the sea of $q\bar{q}$ pairs contributes mainly at small x values.

A second important result follows directly from the curves in Figure 5.23. In a reference frame in which the proton has a very large momentum, x is the fraction of the proton's momentum carried by the struck parton, as we saw in the previous section. Hence the integral

$$\int_0^1 [xq(x) + x\bar{q}(x)] dx$$

²⁸This follows from C invariance if we assume $q\bar{q}$ pairs are produced in the basic process $g \rightarrow q\bar{q}$.

is the fraction of the proton's momentum carried by the quarks and anti-quarks, and a quick glance at the area under the relevant curves in Figure 5.23 is sufficient to see that this fraction is considerably less than one. The remaining momentum is carried away by gluons, like that emitted in Figure 5.22b and other similar diagrams. These gluons, like the quarks and antiquarks, are also referred to as partons and account for about half the proton's momentum. Since they only have strong interactions, their x distribution cannot be measured directly in lepton scattering experiments, but can be determined by other means, as we shall see in Section 5.6.

The third important result of the quark–gluon interaction is that the effective parton distributions, which are shown in Figure 5.23 for Q^2 values of about 10 GeV^2 , depend on Q^2 , albeit weakly. At larger x values, where the valence quarks dominate, the most important correction to the parton model diagram in Figure 5.22a is the process shown in Figure 5.22b in which the struck quark emits a gluon before interacting with the exchanged photon. Detailed calculations show that as Q^2 increases, this process increases in importance,²⁹ and that the emitted gluon carries off some of the initial quark momentum. Thus the effective quark distribution shrinks towards smaller values of the fractional momentum x as Q^2 increases. This is the reason for the decrease in the structure functions with increasing Q^2 seen for fixed $x > 0.2$ in Figure 5.20. In addition, because gluons are emitted more readily as Q^2 increases, the sea distribution of quark–antiquark pairs is also found to increase with Q^2 , leading to a rise in the structure functions with Q^2 at small x values, where the sea is important. From Figure 5.22b and c it is clear that the magnitude of these deviations from exact scaling depends on the magnitude of the strong coupling constant $\alpha_s(Q^2)$, and their measurement is one of the classic means of determining its value.

Finally, it is worth noting that the neutron structure functions, and hence the quark densities, are found from lepton scattering experiments using a range of different nuclear targets. We have seen in Chapter 2 that the average binding energy of nucleons in heavy nuclei is of order 7–8 MeV per nucleon. As this energy is much smaller than those used in deep inelastic scattering experiments, it might be thought safe to ignore nuclear effects (except those due to the internal motion of the nucleons – the Fermi momentum – that are typically about 200 MeV/c). However, experiments have shown that the structure functions do in fact depend slightly on the nuclear medium. Although the effects are very small and not enough to alter the conclusions of this chapter, it is a reminder that there are still things to be learnt about the role of nuclear matter and that this may hold information on the nuclear force in terms of the fundamental quark–gluon interaction. We will return briefly to this in Section 10.2.1.

²⁹This is analogous to the situation in electromagnetism. In QED one finds that the more an electron is accelerated, the more likely it is that the process will be accompanied by photon emission. Similarly in QCD, the more a quark is accelerated, the more likely it is that the process will be accompanied by gluon emission.

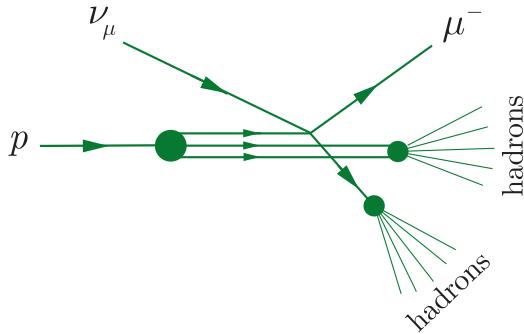


Figure 5.24 Dominant contribution to deep inelastic neutrino scattering in the parton model.

5.5.4 Inelastic neutrino scattering

We now turn to inelastic neutrino reactions such as

$$\nu_\mu + p \rightarrow \mu^- + X^{++} \quad (5.35a)$$

and

$$\nu_\mu + n \rightarrow \mu^- + X^+, \quad (5.35b)$$

where X^+ and X^{++} are as usual any hadronic states allowed by the conservation laws. At high energies these reactions are well described by the parton model, and they enable the contributions of different types of partons to be separated, as we noted in the previous section.

In the parton model, the reactions (5.35a) and (5.35b) are assumed to proceed via two-step processes like that shown in Figure 5.24. In the first step, the neutrino scatters from one of the quark constituents in a weak interaction like

$$\nu_\mu + d \rightarrow \mu^- + u, \quad (5.36)$$

while the second step is a fragmentation process that converts the recoil quark and the remaining constituents of the nucleons into hadrons. This mechanism is similar to that shown in Figure 5.21 for inelastic electron and muon scattering. As in that case, we shall only discuss experiments in which the hadrons are not identified, and we will not need to consider the fragmentation process further.

To extract the different quark and antiquark distributions from neutrino data, it is necessary to measure differential cross-sections analogous to those given in (5.28) for charged lepton scattering. The corresponding formulas for neutrino and antineutrino scattering are somewhat similar to (5.28), but in addition to structure functions $F_1^{\nu N}$, $F_2^{\nu N}$ analogous to F_1 and F_2 in (5.28), there is a third structure function $F_3^{\nu N}$ that occurs because of the existence of parity violation in weak interactions. We will not discuss this here, but simply cite without proof the results obtained for the neutrino and antineutrino structure functions analogous to F_2 . In addition, we shall confine ourselves for simplicity to the region $x \geq 0.3$, where the sea of quark–antiquark pairs can to a good approximation be neglected. In this region, the only nonvanishing quark distributions are

$$u(x) \equiv f_u(x) \quad \text{and} \quad d(x) \equiv f_d(x) \quad (5.37)$$

arising from the valence quarks uud within the proton, and the proton structure function (5.33) measured in electron or muon scattering reduces to

$$F_2(x) = e_u^2 x u(x) + e_d^2 x d(x) \quad (x \geq 0.3). \quad (5.38)$$

The corresponding results for neutrino and antineutrino scattering are

$$F_2^{\nu p}(x) = 2x d(x) \quad \text{and} \quad F_2^{\bar{\nu} p}(x) = 2x u(x) \quad (x \geq 0.3) \quad (5.39)$$

and can be understood, apart from their normalisation, from the conservation of electric charge, baryon number, and muon number. The reaction (5.36) and the corresponding reaction

$$\bar{\nu}_\mu + u \rightarrow \mu^+ + d \quad (5.40)$$

satisfy all these conservation laws. However, other hypothetical reactions like

$$\nu_\mu + u \rightarrow \mu^+ + d \quad \text{and} \quad \bar{\nu}_\mu + d \rightarrow \mu^- + u,$$

which satisfy charge and baryon number conservation, violate the conservation of muon number defined in Section 3.1.1, and there are no $\nu_\mu u$ or $\bar{\nu}_\mu d$ scattering reactions that satisfy all three conservation laws. Hence neutrinos can scatter from d quarks, as implied by (5.36), but not from u quarks, while antineutrinos can scatter from u quarks, as implied by (5.40), but not from d quarks. It is for this reason that neutrino and/or antineutrino beams are ideal tools for separating u and d quark contributions, among others.

The above discussion can be extended to small x by including the sea contributions, but we shall not pursue it further.³⁰ Instead, we shall continue to work at $x > 0.3$, and show how the comparison of neutrino with electron or muon scattering on nuclear targets leads to a determination of the sum of quark charges ($e_u^2 + e_d^2$). To do this, we first note that, according to isospin symmetry, a neutron $n(940) = ddu$ differs from a proton $p(938) = uud$ in that u and d quarks are interchanged. Hence the quark distributions $u^n(x)$ and $d^n(x)$ of the neutron are related to those of the proton by

$$u^n(x) = d^p(x) \equiv d(x) \quad \text{and} \quad d^n(x) = u^p(x) \equiv u(x)$$

to a very good approximation. The neutron structure function F_n^2 measured in electron or muon scattering is then given by

$$F_2^n(x) = e_u^2 x u^n(x) + e_d^2 x d^n(x) = e_u^2 x d(x) + e_d^2 x u(x) \quad (x > 0.3), \quad (5.41a)$$

³⁰See, for example, Section 9.3 of Halzen and Martin (1984).

in analogy to (5.38), and the corresponding neutrino structure functions are given by

$$F_2^{\nu n}(x) = 2xd^n(x) = 2xu(x) \quad \text{and} \quad F_2^{\bar{\nu} n}(x) = 2xu^n(x) = 2xd(x) \quad (5.41\text{b})$$

in analogy to (5.39). From these results, one can easily show that

$$\frac{F_2^p(x) + F_2^n(x)}{F_2^{\nu p}(x) + F_2^{\bar{\nu} n}(x)} = \frac{e_u^2 + e_d^2}{2} = \frac{5}{18}, \quad (5.42)$$

where we have substituted the values of the quark charges. This prediction can be tested by taking data on nuclei that contain equal numbers of protons and neutrons. It is verified within experimental errors of order 10%, and was historically important in confirming that the charges of the ‘charged partons’ found in deep inelastic scattering were indeed the same as the charges of the quarks with which they are now identified.

Finally, we note that the mechanism of Figure 5.24 leads directly to the rapid increase of the neutrino total cross-section at high energy noted at the end of Section 3.1.3. Figure 5.24 is a two-step process in which the first step, which determines the reaction rate, is a simple neutrino–quark scattering process like (5.36). The energy dependence of the total cross-section at high energy then follows from a simple dimensional argument that rests on three points. The first is that the cross-sections are proportional to G_F^2 , since the amplitudes for weak interactions like (5.36) are proportional to G_F . The second is that in the parton model these basic scattering processes only involve point-like particles (i.e. leptons and quarks) with no length scales associated with them. The third point is that, if we work in the centre-of-mass frame, it is reasonable to neglect the proton mass as well as the lepton masses, since the proton will also have a high energy in this frame. Hence, in this approximation, apart from \hbar and c , the only relevant physical constants are G_F and the total centre-of-mass energy E_{CM} , and a simple dimensional argument gives

$$\sigma_\nu = KG_F^2 E_{\text{CM}}^2 / (\hbar c)^4 \quad (5.43\text{a})$$

for the total cross-section, where K is a dimensionless constant and we have used (1.55) to determine the dimensions of G_F . The dependence on the neutrino energy in the laboratory frame E_L then follows directly from (B.19a) and (B.19b), finally giving

$$\sigma \propto E_L \quad (5.43\text{b})$$

at high energies $E_L \gg Mc^2$.

5.6 Other processes

In many experiments on high-energy colliders, in which proton and/or antiproton beams are accelerated, the main interest lies in the interactions

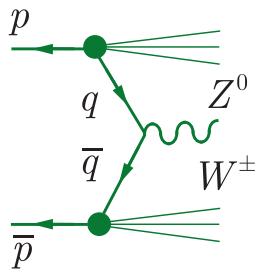


Figure 5.25 Mechanism for producing W^\pm and Z^0 bosons in proton–antiproton annihilations.

between the fundamental constituents within the protons and/or antiprotons, and in the particles or jets these interactions produce. For example, W^\pm and Z^0 bosons were discovered in proton–antiproton collisions via the mechanism of Figure 5.25, in which a quark from the proton annihilates an antiquark from the antiproton to produce the bosons, which were then identified from their decay products. In order to understand the rate and other characteristics of this and many other reactions, in addition to calculating the fundamental interactions between the partons, it is necessary to know the momentum distributions of the partons within the initial hadrons. Conversely, the data from such reactions can be used to extend our knowledge of the parton distributions, and our most precise knowledge of these is based on the simultaneous analysis of many reactions,³¹ including the deep inelastic lepton scattering processes already discussed. Recent results from one such analysis are shown in Figure 5.26 at two different mass scales $\mu^2 = Q^2$, using a log scale for x in order to be able to show the detailed information at very small x . Several points should be noted:

- (i) The valence distributions, which dominate at large x , vary relatively slowly as μ^2 increases.

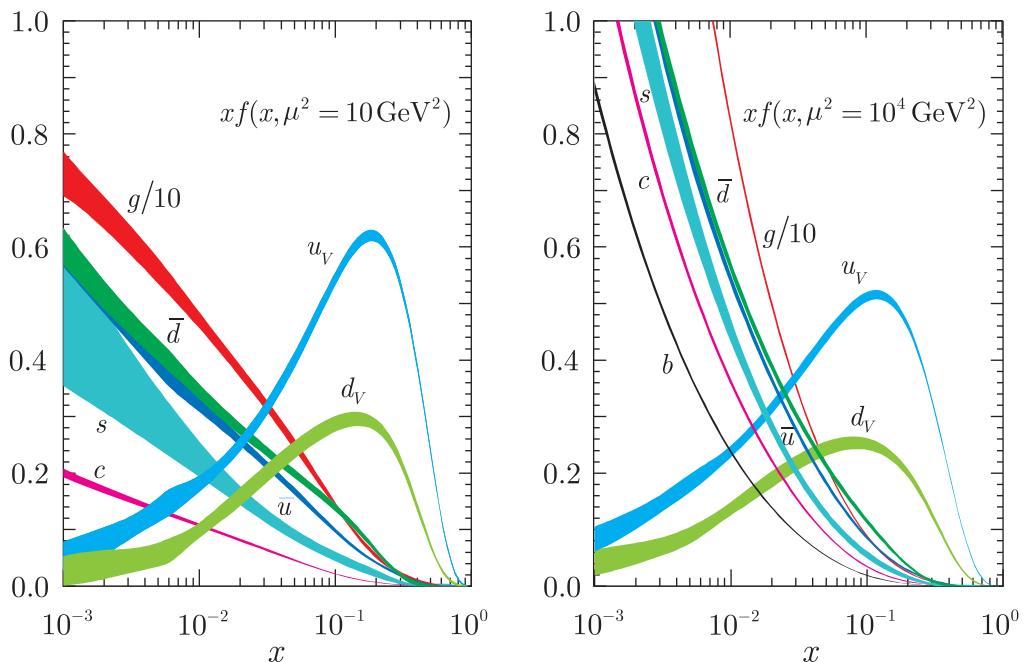


Figure 5.26 Parton distribution functions from a global analysis of data at two different energy scales $\mu^2 = Q^2$. Source: Adapted from Foster et al. (2018).

³¹The Particle Data Group lists about 25 reactions as important sources of information on parton distributions. See Table 19.1 in Foster et al. (2015).

- (ii) At small x , the antiquark distributions arising from the sea depend on the quark flavour and increase rapidly as μ^2 increases.
- (iii) The gluon distribution, which also increases rapidly as μ^2 increases, is very large at small x . Because of this, gluons play a very important, and often dominant, role in many reactions at the LHC where, for example, a parton with x as small as 10^{-2} has an energy of 80 GeV.

5.6.1 Jets

In this and the following section we shall consider two important topics in proton–proton scattering. However, before doing so it will be useful to comment briefly on the jet structure in deep inelastic electron and muon scattering. As we have seen, the dominant term is the parton model diagram in Figure 5.21, in which the final state comprises two parts. When a quark is knocked out of the proton, the remaining partons continue on their way and are converted by a fragmentation process into hadrons with relatively small momenta transverse to the beam direction. This is referred to as *beam fragmentation* or sometimes *target fragmentation*. On the other hand, the quark recoils against the lepton and fragments into a jet of hadrons in the same way as the quarks and gluons produced in electron–positron annihilation, described in Section 5.4. The global properties of this or any other jet can then be summarised by the total energy E and momentum \mathbf{p} of its constituents. In addition, it will be useful to set up Cartesian axes with associated unit vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$ with the z axis in the beam direction, and to separate the momentum into the *longitudinal momentum* $p_z \mathbf{k}$ and the transverse momentum $\mathbf{p}_T = p_x \mathbf{i} + p_y \mathbf{j}$. For example, in deep inelastic electron and muon scattering, when an event is viewed along the beam line, so that one sees the projection of the event in the xy plane, the charged lepton and the quark jet appear ‘back to back’, as illustrated in Figure 5.27. This is because the internal transverse momenta of the quarks within the proton can be neglected,³² so that by momentum conservation, the lepton and the struck quark in Figure 5.21 recoil against each other with equal and opposite transverse momenta. They do not in general emerge back-to-back in three dimensions, because they do not have equal and opposite longitudinal momenta p_z , and as we will see, it is often convenient to replace the longitudinal momentum by the *rapidity*, defined by

$$y \equiv \frac{1}{2} \ln \left(\frac{E + p_z c}{E - p_z c} \right). \quad (5.44)$$

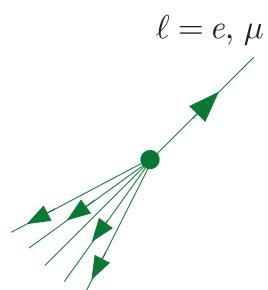
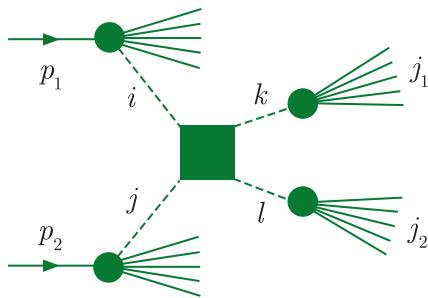


Figure 5.27 Sketch of a final-state charged lepton and the associated jet produced by the mechanism of Figure 5.21, viewed along the initial lepton beam direction. The green circle indicates the collision point.

³²The internal momenta of the quarks within a proton at rest are of order 300 MeV (because their de Broglie wavelengths are of the order of the proton diameter), which is very small compared to the typical values of the transverse momenta of the recoil electron in deep inelastic scattering. The same result applies in the electron–proton centre-of-mass, or any frame related by a Lorentz transformation along the z axis, since such a transformation leaves transverse momenta invariant.

Figure 5.28 The dominant mechanism for di-jet production in proton–proton collisions. The dashed lines represent partons of type i , j , k , and l , and the square box represents a short-range interaction between them.



If the jet has zero invariant mass $M_j c^4 \equiv E^2 - p^2 c^2 = 0$, then $p_z c = E \cos \theta$, where θ is the polar angle of the jet momentum and the rapidity reduces to the *pseudorapidity*

$$\eta = -\ln(\tan \theta/2). \quad (5.45)$$

For high-energy jets, it is usually a good approximation to neglect the invariant mass of the jet compared to its energy, so that the pseudorapidity is a good approximation to the rapidity itself. It is often used as it depends only on one angle, and is therefore relatively easy to measure experimentally.

In proton–proton scattering, di-jet production arises from a large number of analogous mechanisms, summarised generically in Figure 5.28. In this figure, a parton i from a proton interacts with a parton j from another proton to produce two partons labelled by k and l , which are observed as jets in the final state. These jets are usually required to have large transverse momentum, which serves two purposes: it enables them to be easily distinguished from the beam fragmentation, which produces hadrons with relatively low transverse momenta, and it ensures that the parton scattering process $i + j \rightarrow k + l$ is dominated by short-range interactions and can be calculated using perturbative QCD. Since the initial partons i and j contained in the protons have negligible transverse momenta, the two jets must have equal and opposite transverse momenta. Hence the jets appear ‘back-to-back’ when observed along the beam axis, as illustrated in Figure 5.29. This is confirmed by experimental observations. However, this is not in general true when the event is viewed in other projections, since their longitudinal momenta need not be equal and opposite.

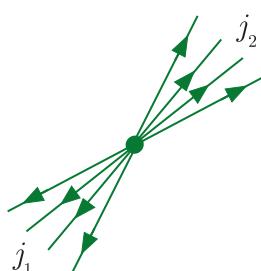


Figure 5.29 Sketch of the two high- p_T jets j_1 and j_2 produced by the mechanism of Figure 5.28, viewed along the beam direction. The green circle indicates the collision point.

To progress further, it is necessary to evaluate the various possibilities in Figure 5.28. However, while yielding results in excellent agreement with experiment, such calculations are very complicated, not least because the initial partons i and j can each be either gluons, or any of the five quarks u , d , s , c , b , or their antiquarks, giving 121 such terms to be taken into account. In practice, this is done using very large computer programs, and to illustrate the principles underlying such calculations, we turn to a much simpler process, in which the final parton jets are replaced by charged leptons.

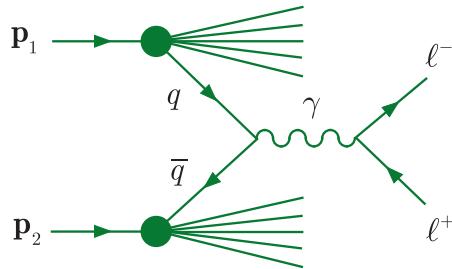


Figure 5.30 The Drell–Yan mechanism for charged lepton-pair production in proton–proton collisions.

5.6.2 Lepton pair production

In high-energy proton–proton collisions, charged lepton pairs with large invariant mass are produced in the so-called *Drell–Yan* reaction

$$p + p \rightarrow \ell^+ + \ell^- + X, \quad (5.46)$$

where $\ell = e, \mu, \tau$ and X represents unobserved hadrons, as usual. For lepton-pair invariant masses with M less than about 50 GeV, the dominant mechanism for this reaction is shown in Figure 5.30, where we assume for the moment that the annihilating quarks are a $u\bar{u}$ pair. The probabilities of finding a u -quark in proton 1 with fractional momentum x_1 to $x_1 + dx_1$ and a \bar{u} -antiquark in proton 2 with fractional momentum x_2 to $x_2 + dx_2$ are $u(x_1) dx_1$ and $\bar{u}(x_2) dx_2$, respectively. Hence the contribution to the cross-section for the reaction (5.46) from Figure 5.30 is

$$d^2\sigma = u(x_1) dx_1 \bar{u}(x_2) dx_2 \sigma(u\bar{u} \rightarrow \ell^+ \ell^-), \quad (5.47)$$

where the momenta of the u and \bar{u} quarks are $x_1 \mathbf{p}_1$ and $x_2 \mathbf{p}_2$, respectively. At high energies, when lepton and quark masses can be neglected, the cross-section for $u\bar{u} \rightarrow \ell^+ \ell^-$ can be deduced from the cross-section

$$\sigma(e^+e^- \rightarrow \mu^+\mu^-) = \frac{4\pi\alpha^2}{3} \left(\frac{\hbar}{Mc} \right)^2, \quad (5.48)$$

which arises from the essentially identical mechanism of Figure 5.17. For $u\bar{u} \rightarrow \ell^+ \ell^-$, we must replace e^2 by $e_u^2 e^2$ and divide by a factor three, because the quark and antiquark must be in the same colour state to annihilate.³³ Hence (5.47) becomes

$$d\sigma^2 = \frac{4\pi\alpha^2}{9} \left(\frac{\hbar}{Mc} \right)^2 e_u^2 u(x_1) \bar{u}(x_2) dx_1 dx_2.$$

³³This should be contrasted to the discussion in Section 5.4.1, where in obtaining the cross-section for $e^+e^- \rightarrow q\bar{q}$, we had to multiply by three, because the quark pairs could be produced in all three colour states.

Finally, summing over all quark flavours $a = u, d, s, c, b$ in protons and taking account of the fact that the quark in Figure 5.30 could originate equally well in the second proton, we obtain

$$d^2\sigma = \frac{4\pi\alpha^2}{9} \left(\frac{\hbar}{Mc} \right)^2 \sum_a e_a^2 [f_a(x_1)f_{\bar{a}}(x_2) + f_{\bar{a}}(x_1)f_a(x_2)] dx_1 dx_2. \quad (5.49)$$

It remains to relate the variables x_1 and x_2 to variables that are directly measured in experiments. To do this, we shall assume that the proton and quark energies are large enough for their masses to be neglected, and that we have colliding beams such that the energy-momenta $cP = (E, \mathbf{p}c)$ of protons 1 and 2 are

$$(E, 0, 0, E) \quad \text{and} \quad (E, 0, 0, -E), \quad (5.50a)$$

respectively. The quark and antiquark energy-momenta are then

$$E(x_1, 0, 0, x_1) \quad \text{and} \quad E(x_2, 0, 0, -x_2), \quad (5.50b)$$

respectively, and by energy-momentum conservation, the energy-momentum of the lepton pair is

$$(\tilde{E}, \tilde{\mathbf{p}}c) = E(x_1 + x_2, 0, 0, x_1 - x_2). \quad (5.50c)$$

For a fixed proton energy E , the squared centre-of-mass energy of the proton-proton system is

$$s = 4E^2, \quad (5.51)$$

while the squared invariant mass of the lepton pair $M_{\ell\ell}$ is given by

$$M_{\ell\ell}^2 c^4 = \tilde{E}^2 - \tilde{\mathbf{p}}^2 c^2 = 4x_1 x_2 E^2 = x_1 x_2 s. \quad (5.52)$$

The rapidity (5.44) of the lepton pair follows from (5.50c) and is

$$y = \frac{1}{2} \ln \left(\frac{\tilde{E} + \tilde{p}_z c}{\tilde{E} - \tilde{p}_z c} \right) = \frac{1}{2} \ln \left(\frac{x_1}{x_2} \right), \quad (5.53)$$

while the inverse relations

$$x_1 = \frac{M_{\ell\ell} c^2}{\sqrt{s}} e^y, \quad x_2 = \frac{M_{\ell\ell} c^2}{\sqrt{s}} e^{-y} \quad (5.54)$$

follow from (5.52) and (5.53), leading to the Jacobian

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial M_{\ell\ell}} & \frac{\partial x_1}{\partial y} \\ \frac{\partial x_2}{\partial M_{\ell\ell}} & \frac{\partial x_2}{\partial y} \end{vmatrix} = -\frac{2M_{\ell\ell} c^4}{s}.$$

Hence,

$$dx_1 dx_2 = |J| dM_{\ell\ell} dy = \frac{2M_{\ell\ell} c^4}{s} dM_{\ell\ell} dy \quad (5.55)$$

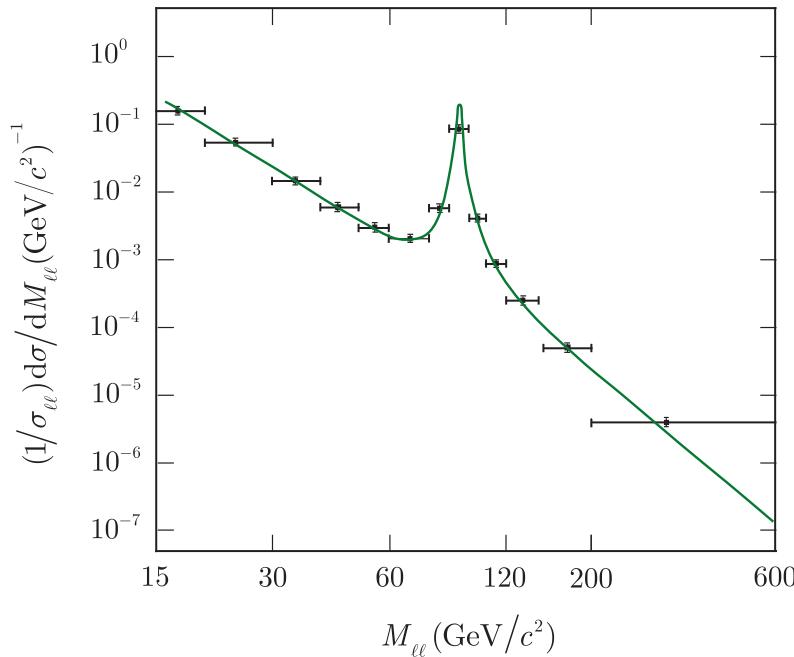


Figure 5.31 Dilepton invariant mass spectrum, in the mass region $15 < M < 600 \text{ GeV}/c^2$. The vertical bars indicate the small experimental uncertainties (statistical and systematic); the horizontal bars indicate bin sizes. The green line is the prediction of the Drell–Yan mechanism, normalised to the Z resonance region (Chatrchyan et al. 2011).

and (5.49) becomes

$$\frac{d^2\sigma}{dM_{\ell\ell}dy} = \frac{8\pi\alpha(\hbar c)^2}{9M_{\ell\ell}s} F(x_1, x_2), \quad (5.56a)$$

where

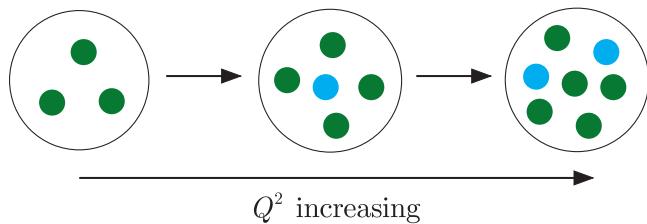
$$F(x_1, x_2) = e_a^2 \sum_a [f_a(x_1)f_{\bar{a}}(x_2) + f_{\bar{a}}(x_1)f_a(x_2)] \quad (5.56b)$$

and x_1, x_2 are given by (5.54).

The Drell–Yan mechanism shown in Figure 5.30, with the corresponding cross-section formula (5.56), is the dominant contribution to the production of lepton pairs with masses between about 15 and 50 GeV³⁴. Above the latter value, one must also take into account a second Drell–Yan production mechanism, in which the intermediate photon in Figure 5.30 is replaced by an intermediate Z^0 boson. For $M_{\ell\ell} = M_Z$, this corresponds to the production and decay of a physical Z^0 , leading to a corresponding sharp peak in the $\ell^+\ell^-$ mass spectrum. In addition, while these are the dominant mechanisms, if very precise comparisons with data are to be made it is necessary to include corrections from more complicated higher-order diagrams. The resulting predictions are compared with data from the CMS experiment at the LHC for the dilepton mass range $15 < M_\ell < 600 \text{ GeV}/c^2$ in Figure 5.31. As can be seen, there is excellent agreement between theory and experiment, including a sharp peak at the Z^0 mass, as expected.

³⁴Below this range, there are peaks corresponding to the production and decay of other particles, such as Υ and J/ψ mesons.

Figure 5.32 Schematic representation of the evolution of the quark content of the proton with increasing resolution. Quarks and antiquarks are denoted by green and blue circles, respectively; gluons are not shown.



5.7 Current and constituent quarks

In previous sections, we have established that when the proton is probed with high spatial resolution, corresponding to large Q^2 , it is found to be composed of three valence quarks uud , plus gluons and a sea of quark–antiquark pairs. In this section we will briefly digress to consider the relation between this description and the simple quark model discussed earlier.

From Figure 5.23 we saw that, at $Q^2 = 10 \text{ GeV}^2$, gluons account for about half the proton's momentum and most of the rest is carried by the valence quarks. The contribution of the sea of quark–antiquark pairs is far smaller and increases/decreases in importance as Q^2 gets larger/smaller. This trend is indicated schematically in Figure 5.32, where we have assumed that, for very low Q^2 , corresponding to low resolution, the sea contribution can be neglected and the proton can be approximated by the three valence quarks moving in a gluon field. The success of the simple quark model in this regime implies that this can in turn be approximated by three constituent quarks moving in a simple potential. This is a very useful approximation, but not exact, as shown by the fact that the masses of the constituent quarks depend on their environment, as seen in Section 3.3.4, where we found that they had heavier masses in baryons than in mesons.

The quarks and antiquarks of the QCD-improved parton model are, together with gluons, the fundamental constituents of hadrons and are often called *current quarks* to distinguish them from the constituent quarks of the simple quark model. Just like constituent quarks, current quarks are confined within hadrons, so their masses cannot be defined as a measure of the inertia of a free quark. However, unlike constituent quarks masses, the masses of current quarks appear as parameters in the fundamental Hamiltonian that defines QCD. Their values can then be estimated by relating them to observables, like hadron masses, using various approximations, including lattice gauge theory. Two complicating features must be noted before doing this. Firstly, because quarks are confined in QCD, their masses cannot be directly measured. Rather, they are parameters that occur in the QCD Hamiltonian. This is in contrast to QED, where we usually define the electron mass as the inertial mass of a free electron, which can be directly measured. Hence its value is independent of any theoretical considerations. Secondly, QCD can be formulated in various different ways, using slightly different definitions of the quark masses.

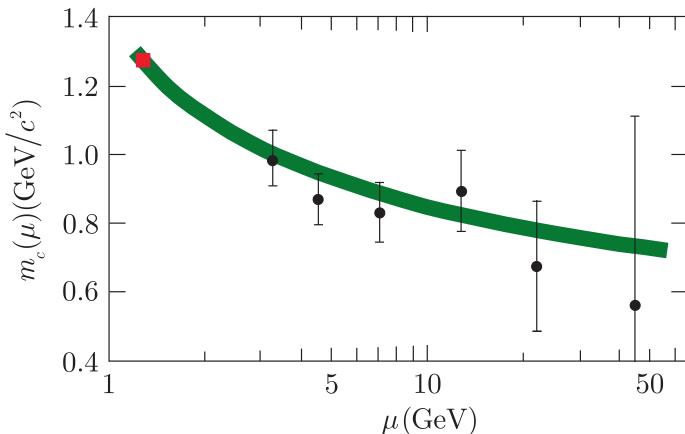


Figure 5.33 Measured charm mass $m_c(\mu)$ in the $\overline{\text{MS}}$ scheme. Source: Behnke et al. (2015), reproduced with permission from Elsevier Ltd. The red square at scale m_c is the PDG world average and the green band is its expected running.

These are just different accounts of the same theory and the different mass definitions are related to one another, but a choice has to be made when quoting mass values. Here we will use the most popular scheme, called the $\overline{\text{MS}}$ scheme.³⁵ In addition, like the running coupling constant $\alpha_s(\mu)$, the quark masses depend on the momentum scale μ (cf. (5.8) and (5.9)) at which they are measured. These *running masses*³⁶ $m_a(\mu)$ ($a = u, d, s, \dots$) decrease with increasing μ and their values are quoted at a chosen reference value. For the u, d, s quarks this is taken to be $\mu = 2 \text{ GeV}$ and for the c and b quarks it is chosen to be m_c and m_b , respectively, i.e. the value for which $m_c(\mu) = \mu$, for example. The resulting values are

$$m_u = 2.2^{+0.6}_{-0.4}, \quad m_d = 4.7^{+0.5}_{-0.4}, \quad m_s = 96^{+8}_{-4}, \quad (5.57\text{a})$$

in units of MeV/c^2 , and

$$m_c = 1.28 \pm 0.03, \quad m_b = 4.18^{+0.04}_{-0.03}, \quad (5.57\text{b})$$

in units of GeV/c^2 , while the predicted behaviour of $m_c(\mu)$ as a function of μ is shown in Figure 5.33.

The most striking feature of these results is that the mass of the light quarks, u, d, s , are small compared to the typical hadron mass scale of $1 \text{ GeV}/c^2$. In the limit that their masses vanish, it can be shown that QCD possesses a new symmetry, called *chiral symmetry*, and the pions and kaons would also be massless. The symmetry is broken by the non-zero quark masses in a well-defined way, yielding relations between the pion, kaon, and quark masses. These relations, together with the complementary results from lattice gauge theories and the measured meson masses, lead directly to the quark masses (5.57a).

³⁵The $\overline{\text{MS}}$ scheme and an alternative are defined in Section 15.3.1 of Mandl and Shaw (2010).

³⁶The running mass is defined formally in Section 15.3.2 of Mandl and Shaw (2010).

Finally, we note that (5.57a) implies that

$$2m_u + m_d \simeq 9 \text{ MeV}/c^2 \ll m_{\text{proton}}.$$

Even when allowing for the increase in the quark masses with decreasing μ , one can conclude that the three valence quarks in the proton make only a small contribution to the mass of the proton. The bulk of the proton mass arises from the interaction energy of the quarks and the gluon field that holds the proton together. This is in marked contrast to more familiar examples of bound states, like atoms and nuclei, whose masses are to a good approximation equal to the sum of the masses of their constituent particles.

Problems 5

- 5.1** The general combination of m quarks and n antiquarks $q^m \bar{q}^n$, with baryon number $B > 0$, has a colour wavefunction that may be written $r^\alpha g^\beta b^\gamma \bar{r}^{\bar{\alpha}} \bar{g}^{\bar{\beta}} \bar{b}^{\bar{\gamma}}$, where r^α means that there are α quarks in the r colour state, etc. By imposing the condition of colour confinement, show that $m - n = 3p$, where p is a nonnegative integer, and hence show that states with the structure qq are not allowed.
- 5.2** The colour quark states are eigenstates of only two of the eight colour charges \hat{F}_i ($i = 1, 2, \dots, 8$). These are $\hat{F}_3 \equiv \hat{I}_3^C$ and $\hat{F}_8 \equiv Y^C$. The other six operators mix the states. For example,

$$\hat{F}_1 r = \frac{1}{2}g, \quad \hat{F}_1 g = \frac{1}{2}r, \quad \hat{F}_1 b = 0.$$

Use these relations in the general baryon colour wavefunction

$$\chi_B^C = \alpha_1 r_1 g_2 b_3 + \alpha_2 g_1 r_2 b_3 + \alpha_3 b_1 r_2 g_3 + \alpha_4 b_1 g_2 r_3 + \alpha_5 g_1 b_2 r_3 + \alpha_6 r_1 b_2 g_3,$$

where α_i are constants, to show that colour confinement implies

$$\alpha_1 = -\alpha_2, \quad \alpha_3 = -\alpha_4, \quad \alpha_5 = -\alpha_6.$$

(The complete set of such relations for all the colour charges leads to the antisymmetric form (5.2).)

- 5.3** Draw the lowest-order Feynman diagrams that contribute to the processes:
 (a) $g + q \rightarrow g + q$, (b) $q + \bar{q} \rightarrow g + g$, (c) $q + \bar{q} \rightarrow g + \gamma$.
- 5.4** Make a rough estimate of the relative magnitude of the static potential between two quarks separated by distances of 0.1 and 0.001 fm.
- 5.5** The 3γ decay of positronium (the bound state of $e^+ e^-$) has a width that in QED is predicted to be

$$\Gamma(3\gamma) = 2(\pi^2 - 9)\alpha^6 m_e c^2 / 9\pi,$$

where α is the fine structure constant. If the hadronic decay of the $c\bar{c}$ bound state $J/\Psi(3100)$ proceeds via an analogous mechanism, but involving three gluons, use the experimental hadronic width $\Gamma(3g) = 80 \text{ keV}$ to estimate the strong interaction coupling constant α_s . Use an analogous assumption to estimate α_s from the radiative width $\Gamma(gg\gamma) = 0.16 \text{ keV}$ of the $b\bar{b}$ bound state $\Upsilon(9460)$.

- 5.6** Would you expect the widths of the three χ_c states listed in Table 3.8 to be larger, smaller, or about the same as the widths of the $J/\psi(3097)$ and $\psi(3686)$? Check your answer with the experimental widths given in the PDG tables.
- 5.7** Show that mesons with the quantum numbers (5.15) are forbidden in the simple quark model.
- 5.8** Use the methods of Appendix A.4 to show that a two-pion state with $I = 1$, $I_3 = 0$ has no $\pi^0\pi^0$ component, so that a $\pi^0\pi^0$ and hence the state $\pi^0\pi^0 J/\psi$ can only have isospin 0 or 2, as stated in the text.
- 5.9** Draw two quark diagrams corresponding to the decay sequence (5.17b), given that the b quark couples to lighter quarks almost exclusively via the vertex bcW analogous to the udW vertex of Figure 3.11.
- 5.10** Draw quark diagrams for the decay sequence (5.17c), assuming P_c^+ is the charmonium–pentaquark state $P_c^+ = uudcc\bar{c}$.
- 5.11** Estimate the cross-section ratio R defined in (5.20) at centre-of-mass energies $E_{CM} = 2.8 \text{ GeV}$ and 15 GeV .
- 5.12** The lowest Feynman diagram for inelastic electron–proton scattering at high energies

$$e^-(E, \mathbf{pc}) + p(E_P, \mathbf{P}_P c) \rightarrow e^-(E', \mathbf{p}' c) + X(\text{hadrons})$$

is shown in (5.19). Use energy–momentum conservation to show that the variable ν defined in (5.24) becomes $\nu = E - E'$ in the rest frame of the proton. Hence show that the variable x defined in (5.25) lies in the range $0 \leq x \leq 1$ if the mass of the electron is neglected.

- 5.13** What is the value of the scaling variable x for the special case of elastic scattering? Hence show that in the rest frame of the proton, the initial (E) and final (E') electron energies are related by

$$Mc^2(E - E') = EE'(1 - \cos \theta),$$

where M is the proton mass, θ is the lepton scattering angle and lepton masses are assumed to be negligible compared to their energies.

- 5.14** Derive the Gottfried sum rule,

$$\int_0^1 [F_2^p(x) - F_2^n(x)] \frac{dx}{x} = \frac{1}{3} + \frac{2}{3} \int_0^1 [\bar{u}(x) - \bar{d}(x)] dx,$$

where the quark distributions refer to the proton.

- 5.15** Rough but simple first approximations to the momentum distributions of valence quarks in the proton are:

$$xu(x) = ax^{1/2}(1-x)^3; \quad xd(x) = bx^{1/2}(1-x)^3.$$

Find the values of a and b and use them to estimate: (a) the position and size of the maximum value of the valence quark contribution to the proton structure function $xV(x)$ and (b) the fractional contribution of the valence quarks to the proton's momentum. Compare your results with the measured values for $xV(x)$ at $Q^2 = 10 \text{ GeV}^2$ shown in Figure 5.23.

- 5.16** Show that: (a) the difference between two rapidities y_1 and y_2 is invariant under a Lorentz transformation along the beam direction and (b) the

rapidity reduces to the pseudorapidity if the mass of the jet or particle referred to can be neglected compared to its energy.

- 5.17** A $p\bar{p}$ collider with equal beam energies is used to produce a pair of top quarks. Draw a Feynman diagram for this process that involves a single gluon. Make a rough estimate of the minimum beam energy required to produce significant numbers of $t\bar{t}$ pairs.

- 5.18** (a) The cross-section $\sigma(u\bar{d} \rightarrow W^+)$ near the mass of the W^+ is given by a Breit–Wigner formula (1.84a). For $\sigma(u\bar{d} \rightarrow W^+)$ this may be written as

$$\sigma_{u\bar{d}} = \frac{12\pi(\hbar c)^2 \Gamma \Gamma_{u\bar{d}}}{M^2 c^4 [4(E - Mc^2)^2 + \Gamma^2]},$$

where (M, Γ) are the mass and total width of the W^+ , $\Gamma_{u\bar{d}}$ is the partial width for $W^+ \rightarrow u\bar{d}$, and E is the total centre-of-mass energy of the $u\bar{d}$ pair. Find σ_{\max} , the maximum value of σ , in practical units, given that the branching ratio for $W^+ \rightarrow u\bar{d}$ is 1/3.

- (b) Find an expression for the cross-section $\sigma(p\bar{p} \rightarrow W^+ + \dots)$ via the mechanism shown in Figure 5.25 in terms of the quark distributions in the proton, assuming that the $p\bar{p}$ total centre-of-mass energy \sqrt{s} is sufficiently small for the contributions of sea quarks to be neglected. Use the narrow width approximation

$$\sigma_{u\bar{d}}(E) = \frac{\pi \Gamma}{Mc^2} \sigma_{\max} \delta \left(1 - \frac{E^2}{M^2 c^4} \right)$$

for the cross-section (see Section 1.6.4), where δ is the Dirac delta function, to find an expression for the cross-section $\sigma(p\bar{p} \rightarrow W^+ + \dots)$ in terms of σ_{\max} , the $p\bar{p}$ total centre-of-mass energy squared s , and a single integral over the quark distribution functions of the proton.

- 5.19** Generalise the Drell–Yan production formula (5.56) to apply to the reaction $\pi^- p \rightarrow \mu^+ \mu^- X$. Then show that under reasonable assumptions, the ratio

$$R = \frac{d^2\sigma(\pi^- + C \rightarrow \mu^+ + \mu^- + X)}{d^2\sigma(\pi^+ + C \rightarrow \mu^+ + \mu^- + X)} \rightarrow 4$$

as $M^2 c^4 \rightarrow s$, where C is a carbon-12 nucleus and M is the invariant mass of the lepton pair.

6

Weak interactions and electroweak unification

We have discussed various aspects of weak interactions in previous chapters, for example when we reviewed nuclear instability in Chapter 2 and again when we introduced the basic properties of leptons in Chapter 3 and discussed neutrino scattering in Chapter 5. In this chapter we will consider wider aspects of the weak interaction and also its unification with electromagnetism to produce the spectacularly successful *electroweak theory*.

6.1 Charged and neutral currents

Like the strong and electromagnetic interactions, the weak interaction is also associated with elementary spin–1 bosons, which act as ‘force carriers’ between quarks and/or leptons. Until 1973 all observed weak interactions were consistent with the hypothesis that they were mediated by the exchange of charged bosons W^\pm only.¹ However, in the 1960s, a theory was developed that unified electromagnetic and weak interactions in a way that is often compared to the unification of electric and magnetic interactions by Maxwell a century earlier. This new theory made several remarkable predictions, including the existence of the charm quark, prior to its discovery in 1974, and of a heavy neutral vector boson Z^0 and the weak reactions arising from its exchange. The latter processes are called *neutral current* reactions (the word ‘neutral’ referring to the charge of the exchanged particle) to distinguish them from the so-called *charged*

¹The idea that weak interactions are due to the exchange of massive charged bosons has a long history dating back to the work of Oskar Klein: Klein (1938) and Klein (1948).

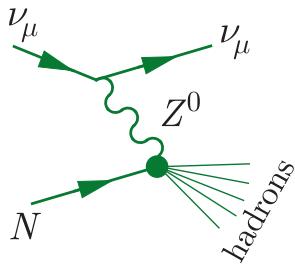


Figure 6.1 Feynman diagram for the weak neutral current reaction $\nu_\mu + N \rightarrow \nu_\mu + X$.

current reactions arising from W^\pm boson exchange. In particular, neutral current reactions of the type

$$\nu_\mu + N \rightarrow \nu_\mu + X \quad (6.1)$$

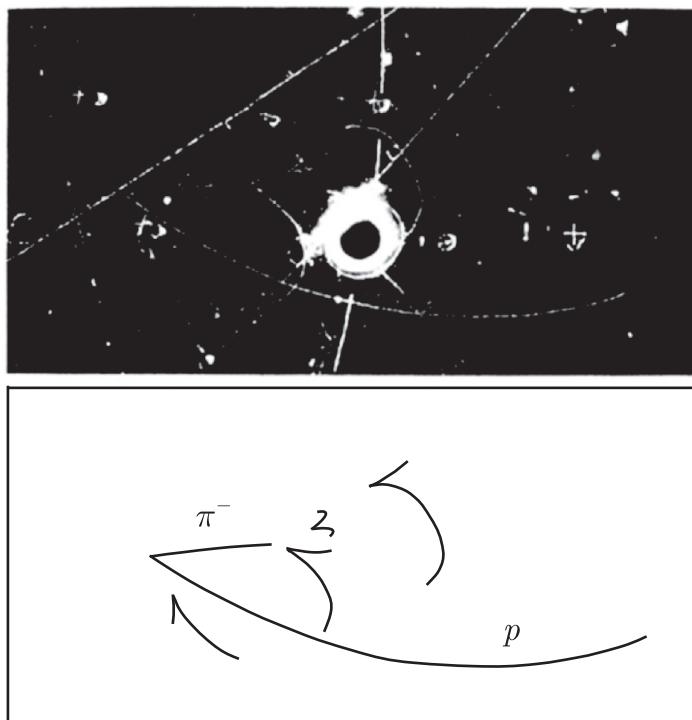
were predicted to occur via the mechanism of Figure 6.1, where N is a nucleon and X is any set of hadrons allowed by the conservation laws. Such reactions were first observed in a bubble chamber experiment in 1973 and an example of such an event is shown in Figure 6.2. In this event, an incoming muon neutrino interacts with a nucleon inside the chamber, producing several hadrons, but no muon. The presence of a neutrino in the final state is inferred from the law of conservation of muon number discussed in Chapter 3.

The unified theory predicted the masses of the W^\pm and Z^0 bosons prior to the long-awaited detection of these particles. The latter was done in 1993 by two experiments at CERN, called UA1 and UA2.² They studied the reactions

$$\bar{p} + p \rightarrow W^+ + X^-, \quad \bar{p} + p \rightarrow W^- + X^+, \quad \text{and} \quad \bar{p} + p \rightarrow Z^0 + X^0, \quad (6.2)$$

where X^\pm and X^0 are arbitrary hadronic states allowed by the conservation laws. The beams of protons and antiprotons were supplied by a

Figure 6.2 Neutral current reactions of the type (6.1) observed in a heavy-liquid bubble chamber at CERN in 1973. The neutrino is incident from the left and interacts with an atomic nucleus inside the chamber to produce a negative pion and a proton. Three photons are also detected via their conversion to e^+e^- pairs. They almost certainly arise from the decays of neutral pions produced in the initial neutrino interaction and in the subsequent interaction of the negative pion with another atomic nucleus. Source: CERN copyright, with permission.



²One of the experiments (UA1) is described in detail in Section 4.5.1 of Martin and Shaw (2017).

proton–antiproton collider that was specifically built for this purpose.³ The best current values for their masses are

$$M_W = 80.379 \pm 0.012 \text{ GeV}/c^2, \quad M_Z = 91.1876 \pm 0.0021 \text{ GeV}/c^2, \quad (6.3)$$

which from (1.40) gives $R_W \approx R_Z \approx 2 \times 10^{-3}$ fm for the ranges of the weak interaction. These are very small distances, even when compared with the size of the nucleon, and at low energies the weak interaction can be treated as a zero-range interaction, as discussed in Section 1.6.1. In this chapter we shall extend the boson-exchange theory to include quarks and also consider high-energy weak interactions, where the zero-range approximation is no longer appropriate.

In general, the unified theory is in agreement with all data on both weak and electromagnetic interactions, which are now referred to collectively as the *electroweak interaction*, in the same way that electric and magnetic interactions are referred to collectively as electromagnetic interactions. Furthermore, the theory predicted the existence of a new spin-0 boson, the so-called *Higgs boson*. The Higgs boson plays a central role in understanding the origin of particle masses in the theory and was finally discovered, more than forty years later, in 2012. (These experiments will be described in Section 6.5.4.)

The unification of weak and electromagnetic interactions only becomes manifest at high energies, and at low energies weak and electromagnetic interactions can still be clearly separated. This follows from the general form of the amplitude (1.51):

$$\mathcal{M}(q^2) = \frac{g^2 \hbar^2}{q^2 - M_X^2 c^2}, \quad (6.4)$$

where M_X^2 is the mass of the exchanged particle and g is the appropriate coupling. For weak interactions $M_X = M_{W,Z} \approx 80 \text{ GeV}/c^2$ and for the electromagnetic interaction $M_X = M_\gamma = 0$. Thus, even with $g_{\text{weak}} \sim g_{\text{em}}$, the amplitudes for the two interactions will only become of comparable size at very high values of q^2 . We therefore start by discussing the weak interaction at low energies, initially for charged currents, and deduce some of its general properties that are valid at all energies. Later we will consider neutral currents and how unification arises, together with some of its consequences, including the existence and properties of the Higgs boson.

6.2 Charged current reactions

The simplest charged current reactions are purely leptonic processes like muon decay

$$\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu \quad (6.5)$$

³Simon van der Meer, who led the team that built the accelerator, and Carlo Rubbia, the leader of the UA1 experiment, shared the 1984 Nobel Prize in Physics.

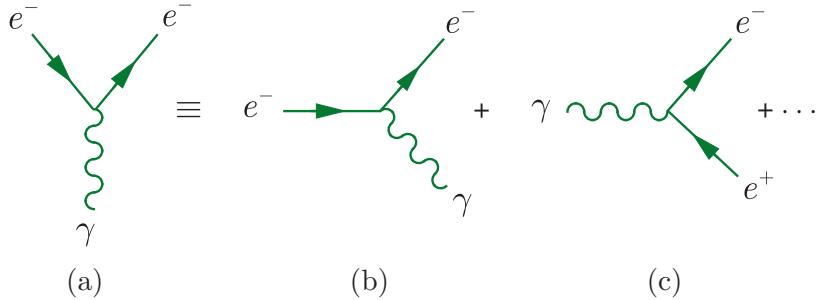


Figure 6.3 The basic vertex for electron–photon interactions, together with two of the basic processes derived from it.

and we discuss these first. We then discuss purely hadronic processes, such as lambda decay

$$\Lambda \rightarrow p + \pi^- \quad (6.6)$$

and *semileptonic* reactions, such as neutron decay

$$n \rightarrow p + e^- + \bar{\nu}_e, \quad (6.7)$$

which involve both hadrons and leptons. As usual, the interactions of the hadrons are interpreted in terms of their quark constituents, and we will see that there is a remarkable symmetry between the weak interactions of leptons and quarks.

6.2.1 W^\pm –lepton interactions

In Chapter 1, we saw that the Feynman diagrams for electromagnetic processes involving electrons and processes could be constructed from a single three-point vertex. This vertex is shown in Figure 6.3, together with two of the virtual processes to which it gives rise. In each case, there is one line pointing towards the vertex and one away to guarantee charge and lepton number conservation at the vertex.

Leptonic weak interaction processes like muon decay (6.5) and inverse muon decay

$$\nu_\mu + e^- \rightarrow \mu^- + \nu_e \quad (6.8)$$

can be built from a limited number of basic vertices in the same way. In the case of the charged current reactions, for each lepton type $\ell = e, \mu, \tau$ there are two such vertices. The first is shown in Figure 6.4, together with two of the basic processes to which it can give rise. The second differs

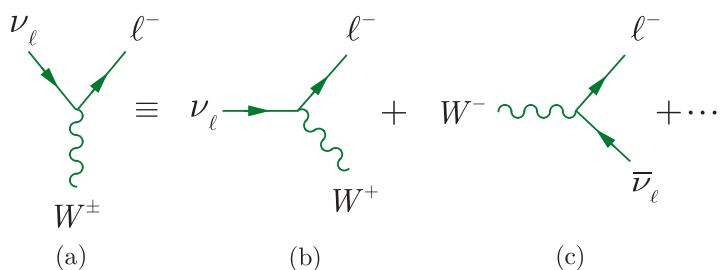


Figure 6.4 One of the two basic vertices for W^\pm –lepton interactions, together with two of the basic processes derived from it.

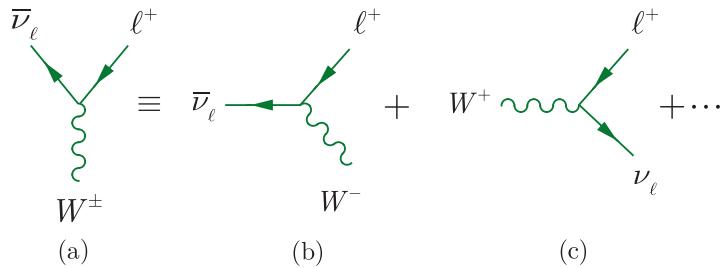


Figure 6.5 The second basic vertex for W^\pm -lepton interactions, together with two of the basic processes derived from it.

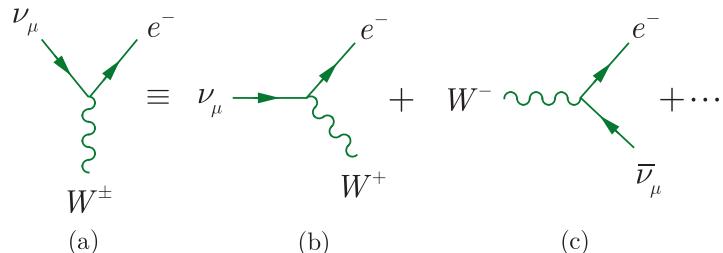


Figure 6.6 Example of a vertex that violates lepton number conservation, together with two of the forbidden processes to which it would give rise.

from it in that lepton lines are replaced by antileptons and is shown in Figure 6.5. In both cases, a W boson can be emitted or absorbed with a charge determined by electric charge conservation, at the vertex, as illustrated in the diagrams. Conservation of the lepton numbers L_e , L_μ , and L_τ , defined in Section 3.1, is guaranteed by the fact that in each of the vertices in Figures 6.4a and 6.5a there is one arrow pointing into the vertex and one pointing out of it, and the lepton label ℓ is the same on both lines. In contrast, vertices like those in Figure 6.6a, which gives rise to processes that violate lepton number, like those shown in Figure 6.6b,c, are excluded from the scheme.⁴

Although the processes of Figures 6.4 and 6.5 conserve lepton numbers, this does not mean that they can occur as isolated reactions in free space. As we saw in Chapter 1, the basic processes of Figure 6.3 violate energy conservation if momentum conservation is assumed, and they can only contribute to physical processes if two or more are combined in such a way that energy is conserved overall. The same applies to most of the processes derived from the two fundamental W^\pm -lepton vertices, but in this case there are some important exceptions. In particular, by working in the rest frame of the W boson one can easily show that Figures 6.4c and 6.5c are compatible with both energy and momentum conservation if

$$M_W > M_\ell + M_{\nu_\ell} \quad (\ell = e, \mu, \tau),$$

which is obviously the case. Figure 6.4c and its ‘antiparticle’ reaction Figure 6.5c are the dominant mechanisms for the leptonic decays

$$W^- \rightarrow \ell^- + \bar{\nu}_\ell \quad \text{and} \quad W^+ \rightarrow \ell^+ + \nu_\ell,$$

⁴In fact the diagrams of Figure 6.4 are the only possible W^\pm -lepton vertices that satisfy both charge and lepton number conservation (see Problem 6.3).

which were used to detect the W^\pm bosons when they were discovered at CERN in 1993.

We turn now to the strength of the reactions. In Section 3.1.2 we showed that the weak and electromagnetic interactions of electrons, muons, and tauons are identical, provided their mass differences are taken into account. Correspondingly, the vertices of Figures 6.4a and 6.5a are all characterised by the same dimensionless strength parameter α_w , analogous to the fine structure constant α in QED, independent of which lepton type is involved. We can obtain an order-of-magnitude estimate for α_w by applying the method of dimensions to the partial width for the leptonic decays shown in Figures 6.5a and b for $\ell = e$, which is measured to be

$$\Gamma(W \rightarrow e\nu) \simeq 0.223 \pm 0.007 \text{ GeV}. \quad (6.9)$$

Since it involves only a single basic process, Figures 6.4c or 6.5c, this decay width is proportional to α_w , and if we neglect lepton masses, the only other parameter is the mass of the W boson. Hence a simple dimensional estimate gives

$$\Gamma(W \rightarrow e\nu) \simeq \alpha_w M_w c^2 \simeq 80 \alpha_w \text{ GeV},$$

which combined with (6.9) gives $\alpha_w \simeq 1/350$. Alternatively, a more detailed calculation⁵ gives

$$\Gamma(W \rightarrow e\nu) = 2\alpha_w M_w c^2 / 3,$$

resulting in the value

$$\alpha_w = g_w^2 / 4\pi\hbar c = 0.0042 \pm 0.0002, \quad (6.10)$$

which is not very different from the value $\alpha = 0.0073$ for the fine structure constant. We thus see that the weak and electromagnetic interactions are actually of comparable intrinsic strengths. However, at low energies weak interactions are well-described by the zero-range approximation described in Section 1.6.1, and reaction rates are small, giving the weak interaction its name, because of the large mass of the W boson, which enters the low-energy effective coupling constant (1.54) as the inverse square. At high energy and momentum transfers, when the zero-range approximation breaks down, weak and electromagnetic rates become comparable.

6.2.2 Lepton–quark symmetry and mixing

The weak interactions of hadrons are understood in terms of basic processes in which W^\pm bosons are emitted or absorbed by their constituent quarks. These can give rise to semileptonic processes like neutron decay (6.7), for which the dominant diagram is shown in Figure 6.7 and purely hadronic decays such as Λ decay (6.6), for which the dominant diagrams are shown in Figure 6.8. Purely hadronic decays are not as well understood

⁵See, for example, Section 16.6.3 of Mandl and Shaw (2010).

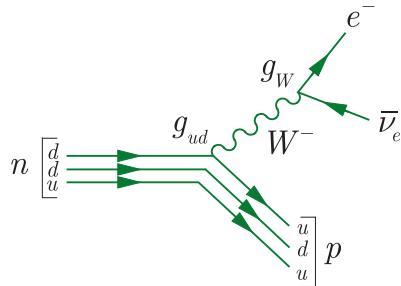


Figure 6.7 Dominant diagram for neutron decay (6.7).

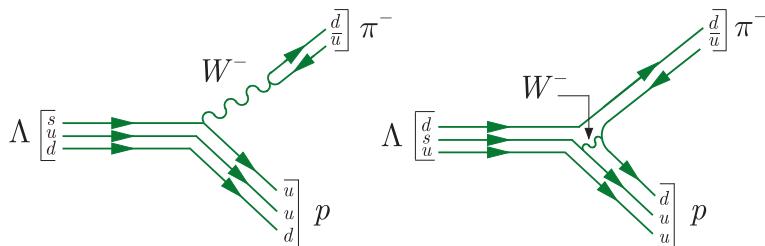


Figure 6.8 Dominant diagrams for Λ decay (6.6).

as the semileptonic ones, because the final-state particles interact strongly with each other, leading to effects that are difficult to calculate. We shall therefore concentrate mainly on the semileptonic interactions in what follows, and we shall initially restrict ourselves to the first two generations of quarks

$$\begin{pmatrix} u \\ d \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} c \\ s \end{pmatrix}, \quad (6.11)$$

of which the majority of known hadrons are composed.

The weak interactions of quarks are best understood in terms of two ideas: *lepton–quark symmetry* and *quark mixing*. Here we shall introduce these ideas and use them to deduce the form of the W^\pm -boson–quark interactions. They may also be used to deduce the form of the Z^0 -boson–quark interaction, as we shall see in Section 6.4, and their justification lies in the fact that the interactions so deduced successfully explain the observed weak interactions of hadrons in both cases.

We start with lepton–quark symmetry. In its simplest form, this asserts that the two generations of quarks (6.11) and the two generations of leptons

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix} \quad (6.12)$$

have identical weak interactions. That is, one obtains the basic W^\pm –quark vertices by making the replacements $\nu_e \rightarrow u$, $e^- \rightarrow d$, $\nu_\mu \rightarrow c$, $\mu^- \rightarrow s$ in the basic W^\pm –lepton vertices of Figures 6.4a and 6.5a, leaving the coupling constant g_W unchanged. In this way one obtains the vertices of Figure 6.9, where the coupling constants

$$g_{ud} = g_{cs} = g_W. \quad (6.13)$$

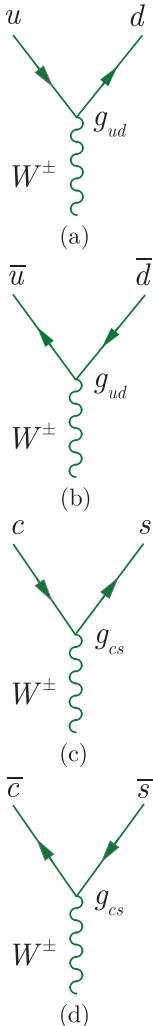


Figure 6.9 The W^\pm quark vertices obtained from lepton–quark symmetry when quark mixing is ignored.

This works quite well for many reactions, like pion decay

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu, \quad (6.14a)$$

which corresponds to

$$d\bar{u} \rightarrow \mu^- + \bar{\nu}_\mu \quad (6.14b)$$

at the quark level, and is interpreted by the Feynman diagram of Figure 6.10a. However, many other decays that are experimentally observed are forbidden in this simple scheme. An example is the kaon decay

$$K^- \rightarrow \mu^- + \bar{\nu}_\mu, \quad (6.15a)$$

which corresponds to

$$s\bar{u} \rightarrow \mu^- + \bar{\nu}_\mu \quad (6.15b)$$

at the quark level. This decay would be naturally explained by the Feynman diagram of Figure 6.10b, which includes a usW vertex, but the latter is not included in the vertices of Figure 6.9. It can be incorporated, however, by introducing an hypothesis due to Cabibbo called ‘quark mixing’, mentioned above. According to this idea, the d and s quarks participate in the weak interactions via the linear combinations

$$d' = d \cos \theta_C + s \sin \theta_C \quad (6.16a)$$

and

$$s' = -d \sin \theta_C + s \cos \theta_C, \quad (6.16b)$$

where the parameter θ_C is called the *Cabibbo angle*. That is, lepton–quark symmetry is assumed to apply to the doublets

$$\begin{pmatrix} u \\ d' \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} c \\ s' \end{pmatrix}. \quad (6.17)$$

This is illustrated, for the first of these, in Figure 6.11. As can be seen, the coupling at the previously allowed udW vertex is suppressed by a factor $\cos \theta_C$ relative to (6.13), while the previously forbidden usW vertex is now allowed with a coupling $g_W \sin \theta_C$. The same argument applies to the other three vertices in Figure 6.9, so that in addition to the four vertices of Figure 6.9 with the couplings

$$g_{ud} = g_{cs} = g_W \cos \theta_C, \quad (6.18a)$$

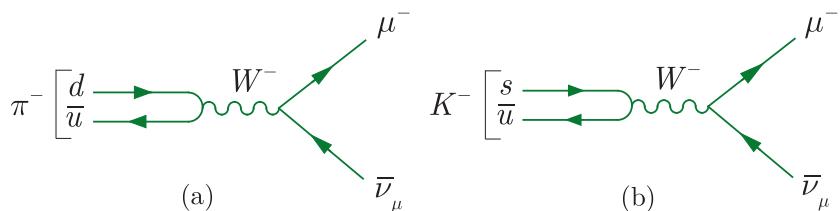


Figure 6.10 Feynman diagrams for the semileptonic decays: (a) $\pi^- \rightarrow \mu^- + \bar{\nu}_\mu$ and (b) $K^- \rightarrow \mu^- + \bar{\nu}_\mu$.

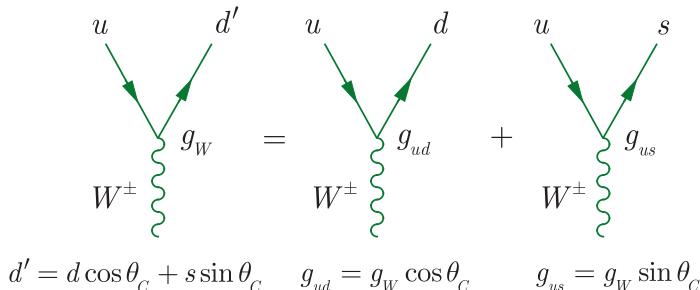


Figure 6.11 The $ud'W$ vertex and its interpretation in terms of udW and usW vertices.

one has the vertices of Figure 6.12 with the couplings

$$g_{us} = -g_{cd} = g_W \sin \theta_C. \quad (6.18b)$$

The set of W^\pm -quark couplings corresponding to (6.18a) and (6.18b) successfully accounts for the charged current interactions of hadrons. Since the lepton coupling g_W is already fixed, these couplings are specified in terms of a single free parameter, the Cabibbo angle θ_C . This is determined by deducing the values of the couplings g_{ud} and g_{us} from the measured rates of various hadron decays. One way of doing this is to compare the rates of decays like those shown in Figure 6.10. As can be seen, they only differ in that a d quark has been replaced by an s quark in the initial state and a coupling g_{ud} has been replaced by a coupling g_{us} , so that the ratio of their partial widths is

$$\frac{\Gamma(K^- \rightarrow \mu^- \bar{\nu}_\mu)}{\Gamma(\pi^- \rightarrow \mu^- \bar{\nu}_\mu)} \propto \frac{g_{us}^2}{g_{ud}^2} = \tan^2 \theta_C,$$

by (6.18a) and (6.18b). Of course the difference in the d and s quark masses will also have an effect on the widths and this must be taken into account. We omit the details and merely quote the mean value obtained from this and other similar determinations, which is

$$g_{us}/g_{ud} = \tan \theta_C = 0.2313 \pm 0.0007, \quad (6.19)$$

corresponding to a Cabibbo angle of

$$\theta_C = 13.02 \pm 0.04 \text{ degrees}. \quad (6.20)$$

A similar value is obtained by comparing the rates for neutron and muon decay, which depends on the ratio

$$(g_{ud}/g_W)^2 = \cos^2 \theta_C.$$

It remains to consider the charmed quark couplings g_{cd} and g_{cs} . These are measured most accurately in neutrino scattering experiments and yield a value consistent with (6.20), but with a larger error. However, the most striking result is for charmed particle decays, which almost always yield a strange particle in the final state. In order to understand this we first note that decays that involve the couplings (6.18b) are called

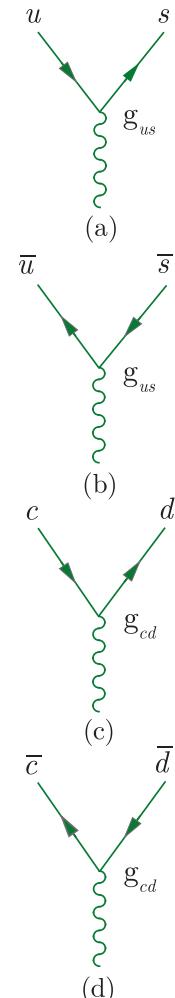


Figure 6.12 The additional vertices arising from lepton–quark symmetry when quark mixing is taken into account.

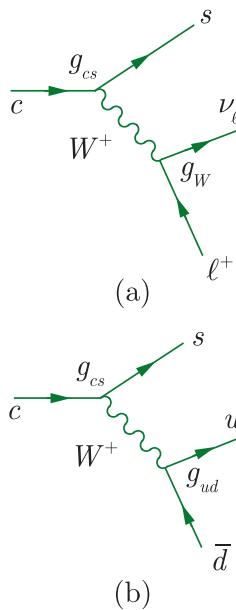


Figure 6.13 Cabibbo-allowed decays from (6.22a) and (6.22b) of a charmed quark.

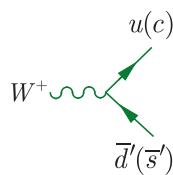


Figure 6.14 Feynman diagram for the decays $W^+ \rightarrow u\bar{d}', c\bar{s}'$.

Cabibbo-suppressed because their rates are typically reduced by a factor of order

$$g_{us}^2/g_{ud}^2 = g_{cd}^2/g_{cs}^2 = \tan^2 \theta_C \approx 1/20 \quad (6.21)$$

compared with similar *Cabibbo-allowed* decays, which involve the couplings (6.18a). The Cabibbo-allowed decays

$$c \rightarrow s + \ell^+ + \nu_\ell \quad (\ell = e, \mu) \quad (6.22a)$$

and

$$c \rightarrow s + u + \bar{d} \quad (6.22b)$$

of a charmed quark to lighter quarks and leptons are shown in Figure 6.13 and they necessarily produce an *s* quark in the final state, as can be seen. Of course, both these quarks must be bound into hadrons, and the implication is therefore that the decay of a charmed hadron will almost always lead to a strange hadron in the final state. This is verified experimentally.

6.2.3 W-boson decays

In Section 6.2.1 we saw that the W^+ boson can decay to lepton pairs $\ell^+\nu_\ell$ ($\ell = e, \mu, \tau$), by the mechanism of Figure 6.5c. Lepton–quark symmetry implies that it can also decay to quark–antiquark pairs

$$W^+ \rightarrow u\bar{d}', c\bar{s}', \quad (6.23)$$

by the corresponding mechanism as Figure 6.14 with the same coupling strength α_W . Here d' and s' are given by (6.16a) and (6.16b). Thus, for example, using

$$u\bar{d}' = u\bar{d} \cos \theta_C + u\bar{s} \sin \theta_C,$$

(6.23) implies that $\bar{u}d$ and $\bar{u}s$ final states will occur with relative probabilities $\cos^2 \theta_C$ and $\sin^2 \theta_C$, respectively. In the rest frame of the decaying W^+ , the quark and antiquark are emitted back-to-back with high energies $E \approx M_W c^2/2 \approx 40$ GeV. They fragment into hadron jets, like those shown in Figure 5.15, so that the observed decay reaction

$$W^+ \rightarrow \text{hadrons} \quad (6.24)$$

is dominated by the two-step mechanism of Figure 6.15. However, the rate for hadron decays is determined by the initial decays (6.23), in the

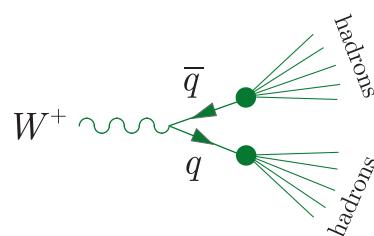


Figure 6.15 The dominant mechanism for the decay of W^\pm bosons into hadrons.

same way that the rate for electron–positron annihilation into hadrons is determined by the rate for the initial quark reaction $e^+e^- \rightarrow q\bar{q}$ (cf. Section 5.4.1).

The relative decay widths for the decays to quarks (6.23) and the leptonic decays

$$W^+ \rightarrow \ell^+ + \nu_\ell \quad (\ell = e, \mu, \tau) \quad (6.25)$$

are easily estimated in the good approximation that the final-state lepton and quark masses are neglected. We then have

$$\Gamma(W^+ \rightarrow u\bar{d}') = \Gamma(W^+ \rightarrow c\bar{s}') = 3\Gamma(W^+ \rightarrow e^+\nu_e), \quad (6.26a)$$

since the mechanisms of these reactions are identical, but the $q\bar{q}$ pairs can be produced in three colour states with equal probabilities, while universality gives

$$\Gamma(W^+ \rightarrow e^+\nu_e) = \Gamma(W^+ \rightarrow \mu^+\nu_\mu) = \Gamma(W^+ \rightarrow \tau^+\nu_\tau). \quad (6.26b)$$

Because these are the only first-order weak decays possible and there are two quark combinations (6.23) contributing to the hadron decays (6.24), we immediately arrive at the predicted branching ratios

$$B(W^+ \rightarrow \text{hadrons}) \approx 2/3, \quad (6.27a)$$

together with

$$B(W^+ \rightarrow \ell^+\nu_\ell) \approx 1/9 \quad (\ell = e, \mu, \tau), \quad (6.27b)$$

for each of the three leptonic decay modes. This simple approximation is in good agreement with experiment.

6.2.4 Charged current selection rules

Many observations about the weak decays of hadrons are explained by W^\pm exchange without the need for detailed calculation. For example, the decays

$$\Sigma^- \rightarrow n + e^- + \bar{\nu}_e \quad (6.28)$$

and

$$\Sigma^+ \rightarrow n + e^+ + \nu_e \quad (6.29)$$

seem very similar, where $\Sigma^+(1189) = uus$ and $\Sigma^-(1197) = dds$ are long-lived hadrons belonging to the $J^P = 1/2^+$ baryon octet of Figure 3.16. However, while reaction (6.28) is observed, reaction (6.29) is not, and the experimental upper limit on its width relative to that of the observed decay is

$$\frac{\Gamma(\Sigma^+ \rightarrow n + e^+ + \nu_e)}{\Gamma(\Sigma^- \rightarrow n + e^- + \bar{\nu}_e)} < 0.004.$$

The reason for this is that reaction (6.28) is allowed via the mechanism of Figure 6.16, whereas no diagram with a single W -exchange can be drawn for reaction (6.29), which at the quark level is



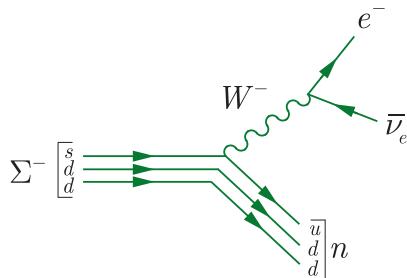


Figure 6.16 Mechanism for the decay $\Sigma^- \rightarrow n + e^- + \bar{\nu}_e$.

and would require two separate quark transitions. It would be allowed via a mechanism involving the emission and absorption of two W bosons, but this contribution is of a higher order in the weak interaction, and is negligibly small.

Reaction (6.29) is just one of many that cannot proceed via single W^\pm exchange and are therefore not observed, despite the fact that they satisfy all the appropriate conservation laws for weak interactions. These ‘forbidden’ reactions can be identified systematically using a number of *selection rules* for single W^\pm exchange processes, which can be deduced from the vertices of Figures 6.9 and 6.12. We shall illustrate this by considering the allowed changes of strangeness.

We consider firstly semileptonic decays, like those of Figures 6.10 and 6.16. Since these involve a single W^\pm -quark vertex, the changes in the strangeness and electric charge of the hadrons are given by the possible changes in S and Q at this vertex. There are just two possibilities. If no strange quarks are involved at the vertex as in Figures 6.9a and b and 6.12c and d, there is obviously no change in strangeness, while the quark charge changes by ± 1 depending on the charge of the W boson. Hence the changes ΔS and ΔQ in the strangeness and the electric charge of the hadrons satisfy

$$\Delta S = 0, \quad \Delta Q = \pm 1. \quad (6.30)$$

On the other hand, those vertices like Figures 9.9c and d and 6.12a and b, which do involve a strange quark, give rise to processes like

$$u \rightarrow s + W^+ \quad \text{or} \quad W^- \rightarrow s + \bar{c},$$

in which the total quark charge and strangeness both decrease, giving $\Delta S = \Delta Q = -1$, or processes like

$$s \rightarrow u + W^- \quad \text{or} \quad W^+ \rightarrow \bar{s} + c,$$

in which the total quark charge and strangeness both increase, giving $\Delta S = \Delta Q$. Thus the allowed semileptonic decays are characterised by the selection rules (6.30) and

$$\Delta S = \Delta Q = \pm 1, \quad (6.31)$$

where ΔQ is the change in the charge of the hadrons only. This is called the $\Delta S = \Delta Q$ rule for strangeness-changing decays, and decays with

$$\Delta S = -\Delta Q = \pm 1 \quad (6.32)$$

are forbidden. Reaction (6.29) is a typical example of a forbidden $\Delta S = -\Delta Q$ reaction requiring changes (6.31) since the Σ^+ has strangeness $S = -1$ and $Q = +1$ while the neutron has $S = 0$ and $Q = 0$. Other examples are discussed in Problem 6.9.

Interesting results are also obtained for purely hadronic decays. In such decays, the exchanged W boson must be both emitted and absorbed at W^\pm -quark vertices, as illustrated in Figure 6.9. Hence the selection rules for strangeness can be inferred by applying the selection rules (6.30) and (6.31) to each individual vertex, subject to the constraint that the change in the hadron charge must now be $\Delta Q = 0$ overall since no leptons are involved and the total charge must of course be conserved. If two vertices satisfying (6.30) are involved, strangeness is conserved and $\Delta S = 0$, while if one satisfies (6.30) and the other (6.31), we obviously have $\Delta S = \pm 1$. Finally, if two vertices satisfying (6.31) are involved, we again obtain $\Delta S = 0$ overall because of the charge conservation condition $\Delta Q = 0$. We thus obtain the selection rule

$$\Delta S = 0, \quad \pm 1 \quad (6.33)$$

for hadronic weak decays, and the same selection rule also holds for semileptonic decays since it is guaranteed by the semileptonic selection rules (6.30) and (6.31).

The selection rule (6.33), which holds for all weak decays, has its most important implications for the decays of the so-called *cascade* particles

$$\Xi^0(1532) = ssu, \quad \Xi^-(1535) = ssd \quad (S = -2), \quad (6.34)$$

and the omega-minus baryon

$$\Omega^-(1672) = sss \quad (S = -3), \quad (6.35)$$

which we met in Sections 3.3.2 and 3.3.4. (See, for example, Figure 3.16 and Tables 3.5 and 3.7). Because the baryon number is conserved, these must ultimately decay to give a proton in the final state, since this is the only stable baryon. However, this cannot occur directly because of the selection rule (6.33), but must proceed via a series of successive decays, in which strangeness changes by at most one unit in each step. Thus the Ξ^- particle frequently decays by the sequence

$$\Xi^- \rightarrow \Lambda^0 + \pi^- \downarrow p + \pi^-, \quad (6.36)$$

giving rise to a characteristic pattern of charged particle tracks. In contrast, decay processes like $\Xi^- \rightarrow n + \pi^-$ that satisfy all known conservation

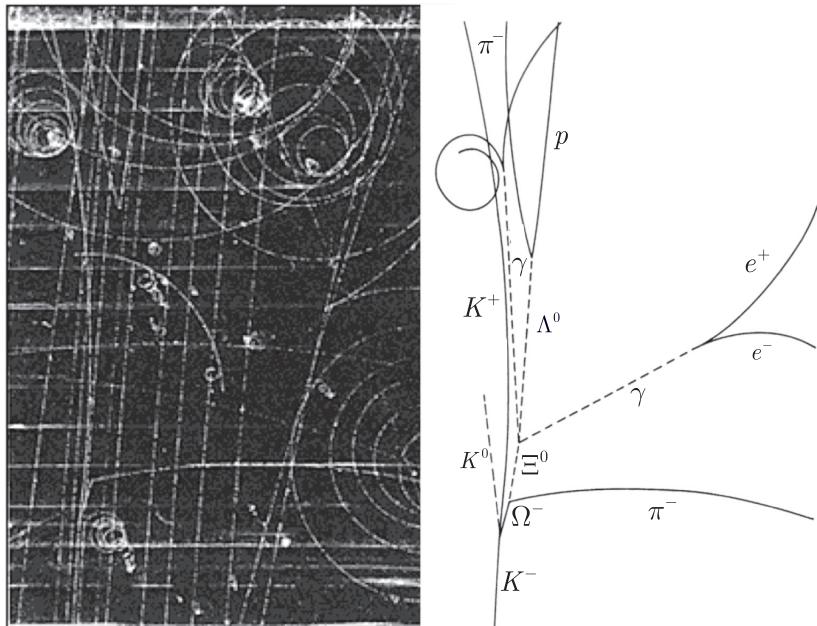


Figure 6.17 Characteristic pattern of tracks produced by the production and decay of the Ω^- . Source: From Barnes et al. 1964, photo courtesy of Brookhaven National Laboratory.

laws, but violate the selection rule (6.33), are never observed. Similar phenomena occur in Ω^- decays, and it is a remarkable fact that when the Ω^- particle was first observed, both its production and the whole decay sequence

$$\begin{aligned}
 K^- + p &\rightarrow \Omega^- + K^+ + K^0 \\
 &\quad \downarrow \Xi^0 + \pi^- (\Delta S = 1 \text{ weak decay}) \\
 &\quad \downarrow \pi^0 + \Lambda (\Delta S = 1 \text{ weak decay}) \\
 &\quad \downarrow \pi^- + p (\Delta S = 1 \text{ weak decay}) \\
 &\quad \downarrow \gamma + \gamma \text{ (electromagnetic decay)} \\
 &\quad \downarrow \quad \downarrow \\
 &\quad e^+ e^- e^+ e^-
 \end{aligned}$$

was captured in a single bubble chamber picture, as shown in Figure 6.17.

6.3 The third generation

In Section 6.2.2 we deduced the weak couplings of the first two generations of quarks (u, d) and (c, s) using the ideas of lepton–quark symmetry and quark mixing. Historically, the most remarkable thing about these ideas is that they were formulated before the discovery of the charmed quark. In 1971 only seven fundamental fermions were known: the four leptons $\nu_e, e^-, \nu_\mu, \mu^-$ and the three quarks u, d, s . This led Glashow, Iliopoulos, and Maiani to propose the existence of a fourth quark c to complete the

lepton–quark symmetry and to solve problems associated with neutral currents.⁶ The charmed quark was subsequently discovered in 1974, firstly indirectly from the charmonium states discussed in Section 3.3.5 and later directly by the production of states with nonzero charm quantum number. Its measured weak couplings are consistent with the predictions of lepton–quark symmetry, as we have seen.

Since 1974, events have moved on. By 1975 there were six known leptons

$$\begin{pmatrix} e^- \\ \nu_e \end{pmatrix}, \quad \begin{pmatrix} \mu^- \\ \nu_\mu \end{pmatrix}, \quad \begin{pmatrix} \tau^- \\ \nu_\tau \end{pmatrix}, \quad (6.37)$$

then in 1977 a fifth quark, the bottom quark b , was discovered, again initially indirectly from bottomonium states and later from states with a nonzero beauty quantum number. The result was five known quarks

$$\begin{pmatrix} u \\ d \end{pmatrix}, \quad \begin{pmatrix} c \\ s \end{pmatrix}, \quad \begin{pmatrix} b \end{pmatrix},$$

so that once again an extra quark – the top quark – was needed to restore lepton–quark symmetry. By 1994, the mass of this quark had been predicted to be

$$m_t = 170 \pm 30 \text{ GeV}/c^2, \quad (6.38)$$

by arguments based on small effects in the unified theory of electroweak interactions (see Section 6.4). The top quark was finally detected in 1995 by the CDF collaboration using the $p\bar{p}$ collider at Fermilab. Top quarks were first produced in pairs in the reaction $p + \bar{p} \rightarrow t + \bar{t} + X^0$, where X^0 is an arbitrary hadronic state allowed by the conservation laws, and the $t\bar{t}$ pairs identified by their subsequent decay products. The top quark mass was found to be $m_t \simeq 180 \text{ GeV}/c^2$, compatible with the prediction of (6.38). In this section we discuss the third generation of quarks b and t , and their consequences for quark mixing.

6.3.1 More quark mixing

Lepton–quark symmetry was applied to the first two generations (6.11) and (6.12) in Section 6.2.2. In doing so, we allowed for mixing between the d and s quarks according to (6.16a) and (6.16b), which is conveniently rewritten in the matrix form⁷

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \begin{pmatrix} \cos \theta_C & \sin \theta_C \\ -\sin \theta_C & \cos \theta_C \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix} \quad (6.39)$$

⁶Neutral currents are discussed in Section 6.4, and these problems in particular in Section 6.4.1.

⁷This is the most general form of mixing possible between two generations (see Problem 6.11).

When the third generation of quarks (t , b) is included, the mixing scheme becomes more complicated, as we must allow for the possibility of mixing between all three ‘lower’ quarks d , s , and b instead of just the first two, and more parameters are needed. This is done by generalising (6.39) to

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \quad (6.40)$$

where the so-called *CKM matrix*⁸ $V_{\alpha\beta}$ ($\alpha = u, c, t$; $\beta = d, s, b$) must be unitary to ensure that d' , s' , and b' are orthonormal single-quark states, like d , s , and b . The matrix elements V_{ij} are all obtainable from various decay processes and values exist for them, although the smaller off-diagonal terms are not all very well measured.⁹ Lepton–quark symmetry is then applied to the doublets

$$\begin{pmatrix} u \\ d' \end{pmatrix}, \quad \begin{pmatrix} c \\ s' \end{pmatrix}, \quad \begin{pmatrix} t \\ b' \end{pmatrix} \quad (6.41)$$

in order to obtain values for the $\alpha\beta W$ couplings $g_{\alpha\beta}$, where again $\alpha = u, c, t$ and $\beta = d, s, b$. The argument is analogous to that used earlier following (6.16a) and (6.16b) and gives

$$g_{\alpha\beta} = g_W V_{\alpha\beta} \quad (\alpha = u, c, t; \beta = d, s, b). \quad (6.42)$$

To progress further, we must consider the values of the CKM matrix elements $V_{\alpha\beta}$. However, in the limit that the mixing between the b quark and the (d , s) quarks can be neglected, it reduces to

$$\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \approx \begin{pmatrix} \cos \theta_C & \sin \theta_C & 0 \\ -\sin \theta_C & \cos \theta_C & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (6.43)$$

which is equivalent to our previous result (6.39) together with

$$b' = b. \quad (6.44)$$

This must be at least a good first approximation, since it reproduces the predictions (6.18) for the couplings of the u , d , c , and s quarks, which agree with the experimental data within errors,¹⁰ as we saw in Section 6.2.2.

⁸The initials stand for Cabibbo, Kobayashi, and Maskawa. The last two authors extended the original Cabibbo scheme to include the third generation of quarks and were awarded a half share of the 2008 Nobel Prize in Physics.

⁹A review is given in Ceccucci et al. (2018).

¹⁰If the predictions (6.18a) and (6.18b) were exact, then (6.43) would follow exactly since the CKM matrix $V_{\alpha\beta}$ is a unitary matrix.

The accuracy of the approximation (6.43) can be very sensitively tested by considering decays of the b quark and of hadrons containing it. The b quark has a mass of approximately $4.5 \text{ GeV}/c^2$. It can decay to lighter quarks by the mechanisms of Figure 6.18. As can be seen, all of these decay modes have rates that are proportional to one of the squared couplings

$$|g_{ub}|^2 = |V_{ub}|^2 g_W^2 \quad \text{or} \quad |g_{cb}|^2 = |V_{cb}|^2 g_W^2. \quad (6.45)$$

Hence the b quark is stable in the approximation (6.43), since both these couplings vanish.

Experimentally, b decays occur at rates that can be inferred from the decays of hadrons containing b quarks, and imply

$$\tau_b \approx 10^{-12} \text{ s} \quad (6.46)$$

for the b quark lifetime. This is very long compared with what one would expect if the couplings (6.45) were not suppressed; i.e. if

$$g_{ub} = g_{cb} = g_W. \quad (6.47)$$

To see this, we compare the decays of the b quark shown in Figure 6.18 with the decays of the lighter τ lepton shown in Figure 6.19. As can be seen, all the mechanisms are identical in the approximation (6.47) and the difference in the total decay rates of the τ and b can only arise from the different masses of the particles involved. This contributes in two ways. Firstly, by counting the numbers of different final states in Figures 6.18 and 6.19, it can be seen that for each τ lepton decay there are N decays of the b quark possible, where $N = 3$ and 4 for diagrams (a) and (b), respectively. Secondly, and more importantly, weak decay rates increase rapidly with the energy released in the decay. If the masses of the final-state particles can be neglected, the decay rates increase as the fifth power of the mass of the decaying particle, as shown for leptonic τ decay in Section 3.1.2. In this simple approximation¹¹ we obtain

$$\tau_b \approx \frac{1}{N} \left(\frac{m_\tau}{m_b} \right)^5 \tau_\tau \approx 10^{-15} \text{ s}$$

on substituting the experimental τ lifetime $\tau_\tau \approx 3 \times 10^{-13} \text{ s}$. This is much shorter than the observed lifetime (6.46), implying that the couplings (6.45) are heavily suppressed relative to (6.47), as stated. Furthermore, this qualitative conclusion is confirmed by a more thorough analysis of the various decay modes, which yields

$$|V_{ub}|^2 = (1.55 \pm 0.28) \times 10^{-5} \quad \text{and} \quad |V_{cb}|^2 = (1.78 \pm 0.07) \times 10^{-3}. \quad (6.48)$$

In other words, while the neglect of mixing between the b quark and the (d, s) quarks is not exact, it is a very good first approximation.

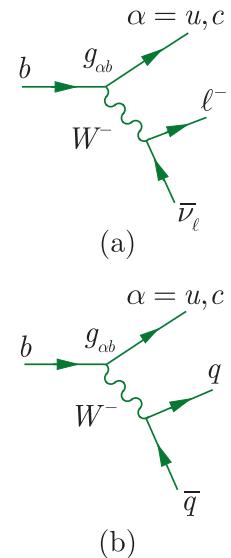


Figure 6.18 The dominant decays of the b quark to lighter quarks and leptons. Here $\ell = e, \mu$ or τ and $q\bar{q} = d\bar{u}, d\bar{c}, s\bar{u}$ or $s\bar{c}$.

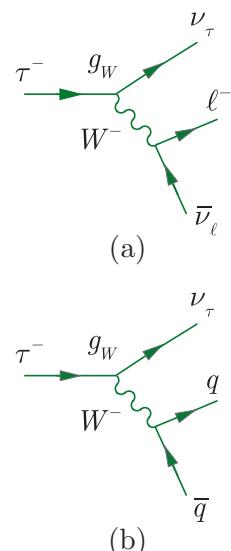


Figure 6.19 The dominant decays of the τ lepton to quarks and leptons. Here $\ell = e$ or μ and $q\bar{q} = d\bar{u}$ or $s\bar{u}$.

¹¹A more accurate treatment of the leptonic decay modes is given in Problem 6.10.

At this point, we note that while the elements of the two-component mixing matrix (6.39) can be chosen to be real, this is not the case for that of the three-component mixing matrix (6.40), whose elements are in general complex. This has important consequences for the behaviour of the weak interactions under the space–time symmetries P , C , and T , discussed in Section 1.3, and when we discuss this topic in Chapter 7, we will need to consider the general form of the mixing matrix (6.40) in more detail, taking the phases of its elements $V_{\alpha\beta}$ into account. However, for the moment we need only consider their magnitudes, and it is sufficient to know that the elements that are neglected in the approximation (6.43), like those in (6.48), are very small.

6.3.2 Properties of the top quark

The properties of the top quark differ markedly from those of the other quarks because it is so much heavier. In particular, it is much heavier than the W^\pm bosons. Hence, it can decay by the first-order weak interaction

$$t \rightarrow q + W^+ \quad (q = d, s, b),$$

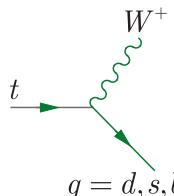


Figure 6.20 The mechanism for t quark decays. The decays that lead to b quarks are overwhelmingly the most important.

as shown in Figure 6.20, with rates proportional to the squared couplings $|g_{tq}|^2$. In the approximation (6.43), these couplings are predicted by (6.47) to be

$$g_{td} = 0, \quad g_{ts} = 0, \quad g_{tb} = g_W.$$

Hence the only significant decay mode is¹²

$$t \rightarrow b + W^+, \quad (6.49)$$

with a rate proportional to α_W , given by (6.10). A crude dimensional estimate of the decay width $\Gamma \sim \alpha_W m_t c^2 \sim 1$ GeV is enough to establish that the top quark is highly unstable. A full calculation of Figure 6.20 for $m_t = 180$ GeV/c² leads to the prediction $\Gamma \approx 1.7$ GeV, with a corresponding lifetime of

$$\tau = \hbar/\Gamma \approx 4 \times 10^{-25} \text{ s}. \quad (6.50)$$

The predicted lifetime (6.50) is at the crux of top quark physics. By relativity, a hadron state of diameter $d \approx 1$ fm cannot form in a time less than $t \approx d/c = O(10^{-23})$ s. The other five quarks u , d , s , c , and b have lifetimes of order 10^{-12} s or more, and there is plenty of time for them to form hadrons that can be observed in the laboratory. In contrast, when top quarks are created they decay too rapidly to form observable hadrons. Instead they decay by (6.49) to give a b quark and a W boson, which in turn decays predominantly to either leptons

$$W^+ \rightarrow \ell^+ + \nu_\ell \quad (\ell = e, \mu, \tau). \quad (6.51a)$$

¹²When deviations from the approximation (6.43) are taken into account, about 2 decays in 1000 lead to a d or s quark instead.

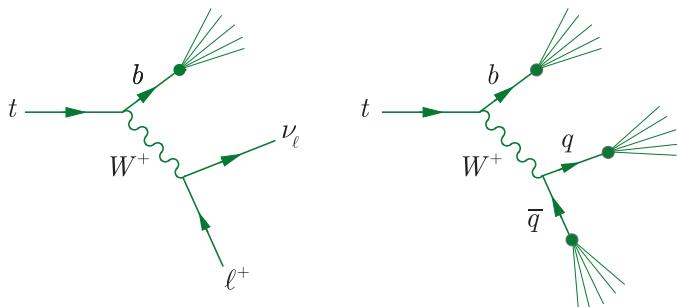


Figure 6.21 Production of hadron jets from the decay $t \rightarrow b + W^+$ where the W boson decays to give either leptons or hadrons.

or light quarks

$$W^+ \rightarrow q_1 + \bar{q}_2 \quad (q_1 \bar{q}_2 = u\bar{d}, u\bar{s}, c\bar{d} \text{ or } c\bar{s}). \quad (6.51b)$$

In the latter case, decays into quark pairs $u\bar{b}$ and $c\bar{b}$ are also allowed by energy conservation, but can only occur through mixing between the b and (d, s) quarks, and so only occur very rarely. Furthermore, the quarks released in these decays are not seen directly, but ‘fragment’ into jets of hadrons. This is illustrated in Figure 6.21, which shows the observable final states resulting from top quark decay. These characteristically contain a single b quark jet, which is very useful in separating top quark events from background events containing hadron jets produced in other ways, and it was these final states that were studied by the CDF collaboration and led to the discovery of the t quark.¹³

6.4 Neutral currents and the unified theory

We now turn from charged current reactions involving the emission, absorption, or exchange of W bosons to neutral currents, which involve the emission, absorption, or exchange of Z^0 bosons. As mentioned earlier, the unified electroweak theory predicted the existence of such reactions before their discovery in 1973 and they play a central role in the theory.¹⁴ In this section we start by looking at the origins of electroweak unification and its prediction of the W^\pm and Z^0 masses. Then we construct the basic Z^0 vertices by analogy with the construction of the W^\pm vertices given earlier and examine the experimental consequences of unification.

6.4.1 Electroweak unification

The electroweak theory was originally proposed mainly to solve problems associated with Feynman diagrams in which more than one W

¹³The CDF experiment and the analysis of the data are described in detail in Section 9.2.3 of Martin and Shaw (2017).

¹⁴The originators of the electroweak theory, Sheldon Glashow, Abdus Salam, and Steven Weinberg, shared the 1979 Nobel Prize in Physics for their contributions to its formulation and the prediction of weak neutral currents.

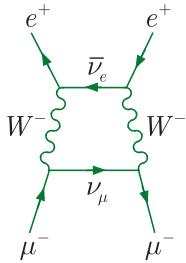


Figure 6.22 Higher-order contribution to the reaction $e^+\mu^- \rightarrow e^+\mu^-$ from the exchange of two W bosons.

boson was exchanged, such as that shown in Figure 6.22 for the reaction $e^+\mu^- \rightarrow e^+\mu^-$. Such contributions are expected to be small because they are of a higher order in the weak interaction and this appears to be confirmed by experimental data, which are in good agreement with theoretical predictions that neglect them entirely. However, when these higher-order contributions are explicitly calculated, they are found to be proportional to divergent integrals, i.e. they are infinite. In the unified theory, this problem is solved when diagrams involving the exchange of Z^0 bosons and photons are taken into account. Separately, these also give infinite contributions, but when all the diagrams of a given order are added together the divergences cancel, giving a well-defined and finite contribution overall.¹⁵

This cancellation is not accidental, but follows if two fundamental relations, called the *unification condition* and the *anomaly condition*, hold. The proof of this result is formidable and relies on a fundamental symmetry of the theory called *gauge invariance*.¹⁶ However, the form of the equations is simple. The unification condition is

$$\frac{e}{2(2\varepsilon_0)^{1/2}} = g_W \sin \theta_W = g_Z \cos \theta_W, \quad (6.52)$$

where the *weak mixing angle* θ_W is given by

$$\cos \theta_W = M_W / M_Z \quad (0 < \theta_W < \pi/2) \quad (6.53)$$

and g_Z is a coupling constant that characterises the strength of the neutral current vertices.¹⁷ This condition explicitly relates the weak and electromagnetic coupling constants. In contrast, the anomaly condition relates the electric charges Q_ℓ and Q_a of the leptons ℓ and quarks a , and is

$$\sum_{\ell} Q_{\ell} + 3 \sum_a Q_a = 0. \quad (6.54)$$

The sums extend over all leptons ℓ and all quark flavours $a = u, d, s, \dots$, and the factor 3 arises because there is also a sum over the three quark colour states, which has been explicitly carried out. On substituting the quark and lepton charges, one easily finds that the anomaly condition is satisfied by the six known leptons and the six known quarks.

The unification condition (6.52) relates the strengths of the various interactions to the W and Z^0 masses, and historically was used to predict the latter from the former before the W^\pm and Z^0 bosons were discovered. In the low-energy limit, the charged current reactions are characterised

¹⁵The first person to demonstrate this was Gerardus 'tHooft. He and Martinus Veltman shared the 1999 Nobel Prize in Physics for their fundamental work on the electroweak theory.

¹⁶Gauge invariance is discussed briefly in Section 6.5 and in more detail in Appendix D.

¹⁷The strengths are not all equal, but are given by g_Z multiplied by known constants that depend on θ_W and the specific vertex.

by the Fermi constant (1.54), and on substituting for g_W from (6.52) one obtains

$$M_W^2 c^4 = \frac{\sqrt{2}(\hbar c)^2 g_W^2}{G_F} = \frac{\pi \alpha (\hbar c)^3}{\sqrt{2} G_F \sin^2 \theta_W}, \quad (6.55a)$$

which together with (6.53) implies

$$M_Z^2 c^4 = \frac{\pi \alpha (\hbar c)^3}{\sqrt{2} G_F \sin^2 \theta_W \cos^2 \theta_W} \quad (6.55b)$$

for the Z^0 mass. The weak mixing angle itself can be determined by comparing neutral and charged current processes at low energies $E \ll M_W c^2$, $M_Z c^2$. In this regime, neutral current processes are characterised by an effective zero-range coupling constant G_Z in exactly the same way that charged current reactions are characterised by the Fermi coupling constant G_F . The neutral current coupling is given by

$$\frac{G_Z}{(\hbar c)^3} = \frac{\sqrt{2} g_Z^2}{\hbar c (M_Z c^2)^2}, \quad (6.56)$$

by analogy with the Fermi constant G_F , and the ratio of these low-energy couplings can be expressed in the form

$$\frac{G_Z}{G_F} = \frac{g_Z^2}{g_W^2} \cdot \frac{M_W^2}{M_Z^2} = \sin^2 \theta_W, \quad (6.57)$$

using (6.52) and (6.53). Hence the weak mixing angle θ_W may be found by comparing the measured rates of charged and neutral current reactions at low energies, and by 1981 its value was determined in this way to be

$$\sin^2 \theta_W = 0.227 \pm 0.014.$$

This value was used to predict the W^\pm and Z^0 masses by substitution into (6.55a) and (6.55b) and using the values of α and G_F given in (1.46) and (1.55) respectively. The resulting values were

$$M_W = 78.3 \pm 2.4 \text{ GeV}/c^2 \text{ and } M_Z = 89.0 \pm 3.1 \text{ GeV}/c^2$$

and the subsequent discovery of the W^\pm and Z^0 bosons with masses compatible with these predictions (as mentioned in Section 6.1) is perhaps the greatest triumph of the unified theory.

Nowadays, the best value of the weak mixing angle,

$$\sin^2 \theta_W = 0.23155 \pm 0.00004, \quad (6.58)$$

is obtained by comparing the predictions of the unified theory with a wide range of measurements on different neutral current reactions. However, on substituting into (6.55a) and (6.55b), this gives

$$M_W = 77.504 \pm 0.017 \text{ GeV}/c^2 \text{ and } M_Z = 88.413 \pm 0.039 \text{ GeV}/c^2, \quad (6.59)$$

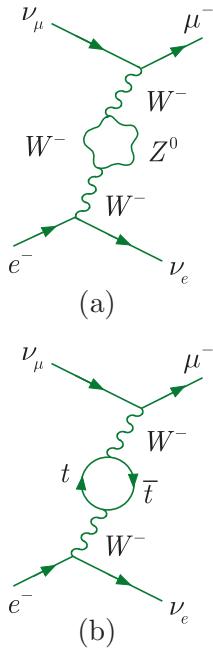


Figure 6.23 Two of the higher-order contributions to inverse muon decay that were neglected in obtaining the Fermi coupling constant G_F as a low-energy limit.

which are not in very good agreement with the current best experimental values (6.3). The reason for this is well understood. It arises because in deriving (6.56) we used the relation (1.54) for the Fermi constant G_F . This latter relation was obtained by taking the low-energy limit of single W^\pm exchange only, whereas strictly we should have also included the small contributions arising from higher-order diagrams. Two of the most important of these are shown in Figure 6.23; one involves the interaction between the W^\pm and Z^0 bosons that is predicted by the unified theory, whereas the other involves the t quark. Hence the magnitude of the higher-order corrections to (6.56), and also to other predictions obtained using just lowest-order diagrams involving single W^\pm and Z^0 exchange, depends on the mass of the t quark. We shall not discuss this in detail, but merely state the important result that when higher-order corrections are taken into account, the predictions of the theory agree with experiment in all cases.

6.4.2 The Z^0 vertices and electroweak reactions

Just as all the charged current interactions of leptons can be understood in terms of the basic W^\pm -lepton vertices, in the same way all known neutral current interactions can be accounted for in terms of basic Z^0 -lepton vertices shown in Figure 6.24a and b. The corresponding quark vertices can be obtained from the lepton vertices by using lepton–quark symmetry and quark mixing, in the same way that W^\pm -quark vertices are obtained from the W^\pm -lepton vertices. Thus, making the replacements

$$\nu_e \rightarrow u, \quad \nu_\mu \rightarrow c, \quad e^- \rightarrow d', \quad \mu^- \rightarrow s' \quad (6.60)$$

in the lepton vertices

$$\nu_e \nu_e Z^0, \quad \nu_\mu \nu_\mu Z^0, \quad e^- e^- Z^0, \quad \mu^- \mu^- Z^0 \quad (6.61)$$

leads to the quark vertices

$$uuZ^0, \quad ccZ^0, \quad d'd'Z^0, \quad s's'Z^0. \quad (6.62)$$

Finally, we interpret the latter two of these using (6.16a) and (6.16b). Thus, for example,

$$\begin{aligned} d'd'Z^0 &= (d \cos \theta_C + s \sin \theta_C)(d \cos \theta_C + s \sin \theta_C)Z^0 \\ &= ddZ^0 \cos^2 \theta_C + ssZ^0 \sin^2 \theta_C + (dsZ^0 + sdZ^0) \sin \theta_C \cos \theta_C. \end{aligned} \quad (6.63)$$

When all the terms in (6.62) are evaluated, one obtains a set of vertices equivalent to

$$uuZ^0, \quad ccZ^0, \quad ddZ^0, \quad ssZ^0, \quad (6.64)$$

which are shown in Figure 6.24c.

One important difference from charged current reactions that follows from Figure 6.24 is that neutral current interactions conserve individual quark numbers. Thus, for example, strangeness-changing weak neutral current reactions are forbidden. An example is the decay $K^0 \rightarrow \mu^+ \mu^-$. This is not seen experimentally, although nothing else forbids it.

In the unified theory, in any process in which a photon is exchanged, a Z^0 boson can also be exchanged. At energies that are small compared to the Z^0 mass, the Z^0 -exchange contributions can be neglected compared to the corresponding photon-exchange contributions, and these reactions can be regarded as purely electromagnetic to a high degree of accuracy. However, at very high energy and momentum transfers, Z^0 -exchange contributions become comparable with those of photon exchange and we are therefore dealing with genuinely electroweak processes that involve both weak and electromagnetic interactions to a comparable degree.

These points are clearly illustrated by the cross-section for the muon pair production reaction $e^+ + e^- \rightarrow \mu^+ + \mu^-$ arising from the mechanisms of Figure 6.25. If we assume that the energy is large enough for the lepton masses to be neglected, then the centre-of-mass energy E is the only quantity in the system that has dimensions. Because a cross-section has dimensions of an area, on dimensional grounds the electromagnetic total cross-section for one-photon exchange is of the form

$$\sigma_\gamma \approx \alpha^2(\hbar c)^2/E^2, \quad (6.65a)$$

where the factor of α^2 follows because the relevant diagram in Figure 6.25a is second order. For Z^0 exchange with $E \ll M_Z c^2$, the zero-range approximation may be used, so that the total cross-section σ_Z arising from Figure 6.25b alone is proportional to the square of the effective low-energy coupling (6.25). Since G_Z has the same dimensions as the Fermi coupling constant (cf. (1.55)), a similar dimensional argument gives

$$\sigma_Z \approx G_Z^2 E^2 / (\hbar c)^4 \quad (6.65b)$$

and hence

$$\sigma_Z/\sigma_\gamma \approx 1/\cos^4 \theta_W,$$

where we have used (6.52) and (6.56), neglecting a numerical factor of order unity. Thus the one-photon exchange diagram dominates at low energies and the cross-section falls as E^{-2} . This is in agreement with the observed behaviour shown in Figure 6.26 and justifies our neglect of the Z^0 -exchange contribution at low energies. However, the relative importance of this contribution increases rapidly with energy and at beam energies of about 50 GeV it begins to make a significant contribution to the total cross-section.¹⁸ At still higher energies, the cross-section is

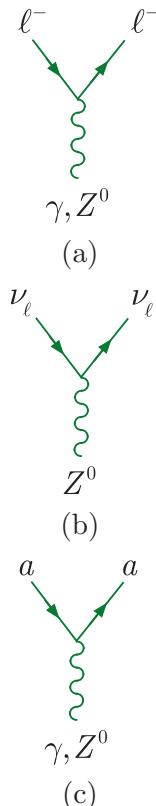


Figure 6.24 Z^0 and γ couplings to leptons and quarks in the unified electroweak theory, where $\ell = e, \mu$ and $a = u, d, s, \dots$ denotes a quark.

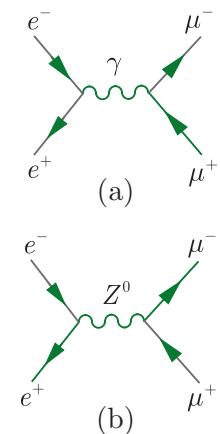


Figure 6.25 The two dominant contributions to the reaction $e^+ + e^- \rightarrow \mu^+ + \mu^-$ in the unified theory.

¹⁸Significant corrections to the differential cross-section, arising from the interference term between Figures 6.25a and b, begin to appear at around 25 GeV, but make a much smaller correction to the total cross-section. This is discussed in detail in, for example, Mandl and Shaw (2010), Section 19.4.

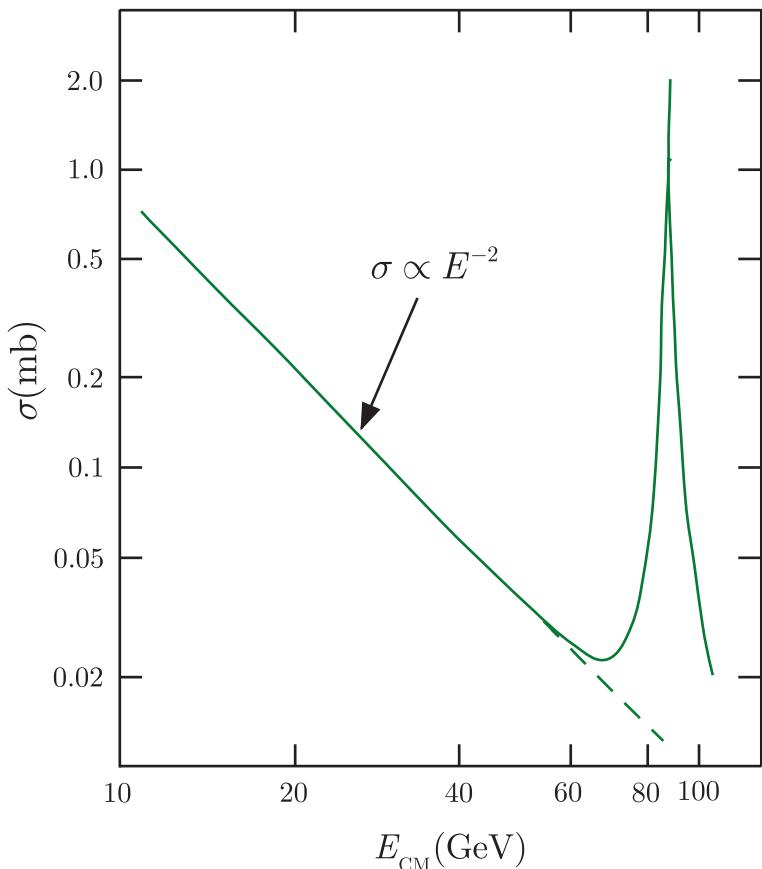


Figure 6.26 Qualitative sketch of the observed total cross-section for the reaction $e^+e^- \rightarrow \mu^+\mu^-$.

dominated by a very large peak at an energy corresponding to the Z^0 mass, as illustrated in Figure 6.26. At this energy the low-energy approximation is irrelevant and Figure 6.26 corresponds to the formation of physical Z^0 bosons in the process $e^+ + e^- \rightarrow Z^0$, followed by the subsequent decay $Z^0 \rightarrow \mu^+ + \mu^-$ to give the final-state muons. Finally, beyond the peak we once again regain the electroweak regime where both contributions are comparable.

6.5 Gauge invariance and the Higgs boson

Gauge invariance is a fundamental symmetry associated with theories in which the force carriers are spin-1 bosons. It plays an important role in the unified electroweak theory, where it is needed to ensure the cancellation of the divergences that occur in individual Feynman diagrams, as mentioned in Section 6.4.1. In addition, because the W^\pm and Z^0 bosons have nonzero masses, it leads to the prediction of a new spin-0 boson – the *Higgs boson* – which, after a huge experimental programme, was finally detected in 2012. This section is devoted to a brief introduction to gauge invariance and its

consequences for weak interactions and the unified theory. A fuller account is given in Appendix D.

There are different forms of gauge invariance, corresponding to the different interactions of particle physics. Their common feature is that the parameters of the corresponding symmetry transformations are allowed to vary with position (\mathbf{r}, t) in space and time.¹⁹ For example, in QED one takes it as a fundamental assumption that the theory must be invariant under gauge transformations of the form

$$\psi(\mathbf{r}, t) \rightarrow \psi'(\mathbf{r}, t) = \exp[-iqf(\mathbf{r}, t)]\psi(\mathbf{r}, t), \quad (6.66)$$

where $\psi(\mathbf{r}, t)$ is the wavefunction of a particle of charge²⁰ $\sqrt{\epsilon_0}q$ and $f(\mathbf{r}, t)$ is an arbitrary continuous function. In other words, one requires that if $\psi(\mathbf{r}, t)$ is a solution of the equation of motion, $\psi'(\mathbf{r}, t)$ must also be a solution. This condition is not satisfied by a free or non-interacting particle. In this case, the equation of motion is the free particle Schrödinger equation

$$i\hbar \frac{\partial \psi(\mathbf{r}, t)}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi(\mathbf{r}, t) \quad (6.67)$$

in the nonrelativistic limit, and it is easy to see that if $\psi(\mathbf{r}, t)$ is a solution, $\psi'(\mathbf{r}, t)$ is not a solution for arbitrary $f(\mathbf{r}, t)$. Hence gauge invariance requires the existence of an interaction. Furthermore, if one adds the minimal interaction required to make (6.67) invariant under (6.66), it can be shown that one arrives at the correct equation of motion for a charge particle interacting with an electromagnetic field. This approach, in which the form of the interaction is inferred by adding the minimal interaction terms needed to make the equation of motion gauge invariant, is called the *principle of minimal gauge invariance*, or the *gauge principle* for short. Here we shall confine ourselves to a qualitative account of its application to weak interactions, omitting all technical details.²¹ In doing so, we shall restrict ourselves for simplicity to the first generation of leptons (e^- , ν_e), since the extension to other leptons, and to quarks using lepton–quark symmetry, is straightforward and involves no new issues of principle.

6.5.1 Unification and the gauge principle

The gauge principle described above begs the question: what are the appropriate forms of the gauge transformations for the weak interactions?

¹⁹This is in contrast, for example, to the cases of translational and rotational invariance, where the parameters that characterise the transformations are assumed to be independent of position and time (see, for example, Sections 5.1 and 5.2 of Martin and Shaw (2017)). Translational and rotational invariance are therefore not examples of gauge invariance.

²⁰We remind the reader that we are using standard SI units. It is more usual to discuss this topic in Heavyside–Lorentz, or rationalised Gaussian units, in which case $\epsilon_0 = 1$ and q is just the electric charge.

²¹A fuller account of the gauge principle is given in Appendix D.1.

In QED, one considers phase transformations (6.66) that change an electron state to a new electron state, symbolically $e^- \rightarrow e^-$. The gauge principle then leads to interactions $e^- \rightarrow e^- \gamma$ in which a gauge boson (the photon) is emitted or absorbed. More generally, one can define gauge transformations in which not only the phase of the wavefunction but also the nature of the particle changes. Specifically, one can define a set of gauge transformations in which electrons and electron neutrinos can transform into themselves, or each other. That is, they incorporate the transformations

$$e^- \rightarrow \nu_e, \quad \nu_e \rightarrow e^-, \quad e^- \rightarrow e^-, \quad \nu_e \rightarrow \nu_e$$

and lead, via the gauge principle, to interactions

$$e^- \rightarrow \nu_e W^-, \quad \nu_e \rightarrow e^- W^+, \quad e^- \rightarrow e^- W^0, \quad \nu_e \rightarrow \nu_e W^0,$$

in which gauge bosons of the appropriate charge are emitted or absorbed. As implied by our notation, the charged bosons can be identified with the observed W^\pm bosons, leading to charged current processes of the type observed. However, a W^0 boson would lead to neutral current processes of the same form and strength as the charge current interactions, which is not what is observed experimentally.

The resolution of this problem lies in unification of the weak interactions with electromagnetism. If electromagnetism were introduced directly, as described above following (6.66), we would simply add the photon to the list of gauge bosons W^+ , W^- , W^0 and the problem of the W^0 would remain unchanged. Instead we regard both the photon γ and the observed neutral boson Z^0 as mixtures of the W^0 with another neutral boson denoted B^0 . That is, we write

$$\gamma = B^0 \cos \theta_W + W^0 \sin \theta_W \quad (6.68a)$$

and

$$Z^0 = -B^0 \sin \theta_W + W^0 \cos \theta_W, \quad (6.68b)$$

where θ_W is the weak mixing angle and the particle B^0 is such that the combination (6.68a) has all the properties of the observed photon. This can be achieved by requiring gauge invariance under the transformations

$$\psi_\ell(\mathbf{r}, t) \rightarrow \psi'_\ell(\mathbf{r}, t) = \exp[-ig_Z y_\ell f(\mathbf{r}, t)] \psi_\ell(\mathbf{r}, t) \quad (\ell = e, \nu_e), \quad (6.69)$$

where g_Z is the weak coupling constant introduced in (6.52) and $y_\ell \equiv y(e^-)$, $y(\nu_e)$ are constants to be determined. The transformations (6.69) are similar to (6.66) so the gauge principle leads to vertices

$$e^- \rightarrow e^- B^0, \quad \nu_e \rightarrow \nu_e B^0,$$

analogous to $e^- \rightarrow e^- \gamma$ but with the charge replaced by the couplings $g_Z y(e^-)$, $g_Z y(\nu_e)$. An explicit calculation²² then shows that these can be

²²This calculation is done in Martin and Shaw (2017), Appendix D, Section D.7.3.

chosen so that the combination (6.68a) has exactly the couplings of a photon, provided the unification condition (6.52)

$$\frac{e}{2(2\varepsilon_0)^{1/2}} = g_W \sin \theta_W = g_Z \cos \theta_W$$

is satisfied.

In this way, one arrives at a unified theory of electroweak interactions that is gauge invariant, as required. The assumed form of gauge invariance is complicated and is chosen to ensure that the resulting electromagnetic and charged current interactions have the form dictated by experiment. Once that has been done, the properties of the weak neutral current interactions arising from the exchange of the Z^0 boson (6.68b) are predicted in detail. The precise agreement between these predictions and experiment is one of the greatest triumphs of the unified theory, as we have seen. However, there remains one problem, to which we now turn.

6.5.2 Particle masses and the Higgs field

The *Higgs boson* is a neutral spin-0 particle whose existence is predicted by the standard model. It is required because gauge invariance implies that the spin-1 gauge bosons have zero masses, as shown explicitly for the electromagnetic case in Appendix D, Section D.1.1. This is acceptable for QED and QCD, where the gauge bosons are the photons and the gluons, which do indeed have zero masses. However, the W^\pm and Z^0 bosons are very heavy and not massless, as they would be if gauge invariance were exact. This problem is overcome by assuming that the various particles interact with a new type of scalar field that differs from other fields in its behaviour in the so-called *vacuum state*, which contains no particles of any kind. Other fields, such as the electromagnetic field, are assumed to be zero in the vacuum state, as one would naively expect. However, this new field has a non-zero value in the vacuum, and this value is not invariant under a gauge transformation. Because of this, the theory is no longer gauge invariant, and the gauge bosons are no longer required to have zero mass. This form of symmetry breaking, in which the gauge invariance of the interaction (as opposed to the gauge invariance of the vacuum) remains exact, is called *spontaneous symmetry breaking*.²³

Spontaneous symmetry breaking of this type is known in other branches of physics. It occurs whenever the vacuum state, defined as the state of lowest energy, is not unique. A familiar example is ferromagnetism. In a ferromagnetic material, the forces that couple the electron spins are rotationally invariant and at high temperatures the net magnetisation \mathbf{M} is zero. Below the Curie temperature the spins in the ground state are aligned in some definite direction, resulting in a magnetisation

²³The demonstration of spontaneous symmetry breaking in a field theoretic formalism was first made by Yoichiro Nambu, who subsequently shared the 2008 Nobel Prize in Physics.

that breaks the rotational invariance. However, \mathbf{M} could equally well point in any other direction and the rotational invariance of the interaction manifests itself in that all the properties, other than the direction of \mathbf{M} , are independent of the actual direction.

In particle physics we are dealing with gauge theories, and the gauge particles acquire mass as a result of their interactions with a nonvanishing vacuum field, called the *Higgs field*. The name is because this mechanism was first proposed in 1964 by Higgs, and independently by Englert and Brout.²⁴ It is now called the *Higgs mechanism*. It is in some ways analogous to the way a conduction electron in a semiconductor acquires an effective mass different from its mass in vacuo as a result of its interactions with the crystal lattice. We omit all details here²⁵ and simply note that applying the Higgs mechanism to the unified theory leads to three main consequences:

- (i) The W^\pm and Z^0 bosons acquire the nonzero masses discussed in Section 6.4.1. These masses arise from the interactions of the gauge fields with the nonzero vacuum expectation value of the Higgs field.
- (ii) There are electrically neutral quanta H^0 associated with the Higgs field, called *Higgs bosons*, in the same way that there are quanta, i.e. photons, associated with the electromagnetic field, except that the Higgs bosons have nonzero mass and have spin-0 rather than spin-1.
- (iii) The Higgs field throws light on the origin of quark and lepton masses.

The explanation of (iii) is that in the absence of a Higgs field, gauge invariance requires that the masses of spin-1/2 fermions with the parity-violating interactions discussed in the next chapter are zero.²⁶ Parity is conserved in strong and electromagnetic interactions, but violated in weak interactions, so that quarks and leptons would be massless in this case. However, interactions with the Higgs field can generate fermion masses arising from the nonzero vacuum expectation value of the Higgs field, as well as interactions with the Higgs bosons. These interactions are of a form shown in Figure 6.27, with a dimensionless coupling constant g_{Hff} related to the fermion mass m_f by

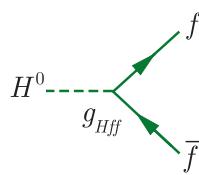


Figure 6.27 The basic vertices for Higgs boson–fermion interactions. The fermion f can be any quark, charged lepton, or neutrino.

$$g_{Hff} = \sqrt{2}g_W(m_f/M_W). \quad (6.70)$$

This theory of fermion masses – that they are generated by interactions with the Higgs field – does not make any predictions for their values m_f . However, it can be tested by measuring the Higgs boson couplings g_{Hff} , and verifies the predictions (6.70), as we shall see.

²⁴Peter Higgs and François Englert (Robert Brout had died two years earlier) shared the 2013 Nobel Prize in Physics for their prediction.

²⁵The Higgs mechanism is discussed in Section D.2 of Appendix D.

²⁶The reason for this is discussed in Martin and Shaw (2017), Appendix D, Section D.7.4.

6.5.3 Properties of the Higgs boson

The basic vertices for the interaction of Higgs bosons with fermions are shown in Figure 6.27. In addition, the unified electroweak theory predicts the existence of the first and second-order boson vertices shown in Figures 6.28 and 6.29, respectively. The strength of the various Higgs boson interactions are completely specified for a given value of the Higgs mass, and, for leptons, quarks, and weak bosons, they increase with the mass of the particle to which it couples. The Higgs boson therefore couples very weakly to light particles like neutrinos, electrons, muons, and u , d , and s quarks, and much more strongly to heavy particles like W^\pm and Z^0 bosons and t quarks. This is reflected in both the decay and production properties of the Higgs boson, as we shall now discuss.

Although the Higgs mass is not predicted in the unified theory, before it was first detected in 2012 its mass was estimated by calculating the contribution to any given electroweak process from higher-order diagrams in which Higgs bosons are emitted and absorbed. These diagrams, together with other higher-order diagrams, give very small corrections to the leading order processes that we have considered throughout this chapter. Nonetheless, the corrections can be measured, and the requirement that the measured and calculated corrections agree led to the restricted range

$$M_H = 94^{+29}_{-24} \text{ GeV}/c^2, \quad (6.71)$$

corresponding to the range $46 < M_H < 152$ (GeV/c^2) at a 95% confidence limit. The observed mass

$$M_H = 125.18 \pm 0.16 \text{ (GeV}/c^2) \quad (6.72a)$$

is compatible with this. In contrast, the decay width of the Higgs boson has not been measured directly, but is predicted to be

$$\Gamma_H = 4.1 \pm 0.2 \text{ MeV}, \quad (6.72b)$$

corresponding to a lifetime of about 1.6×10^{-22} s. This value is obtained by calculating and adding together the rates of all the principal decay modes, which are given in Table 6.1.

Table 6.1 Predicted decay widths Γ and branching ratios B for the main decay modes of the Higgs boson, and for the rare but important decay $H^0 \rightarrow \gamma\gamma$

Decay mode	$\Gamma(\text{MeV})$	B
$b\bar{b}$	2.35 ± 0.11	$(57.5 \pm 1.8) \times 10^{-2}$
$W^\pm f \bar{f}'$	0.88 ± 0.01	$(21.6 \pm 0.9) \times 10^{-2}$
gg	0.35 ± 0.02	$(8.6 \pm 0.8) \times 10^{-2}$
$\tau^+\tau^-$	0.26 ± 0.01	$(6.3 \pm 0.3) \times 10^{-2}$
$c\bar{c}$	0.12 ± 0.01	$(2.9 \pm 0.3) \times 10^{-2}$
$Z^0 f \bar{f}$	$(11.0 \pm 0.1) \times 10^{-3}$	$(2.7 \pm 0.1) \times 10^{-2}$
$\gamma\gamma$	$(9.5 \pm 0.1) \times 10^{-3}$	$(2.3 \pm 0.1) \times 10^{-3}$

The errors arise from uncertainties in the input parameters.

Source: Data from Heinemeyer et al. (2013).

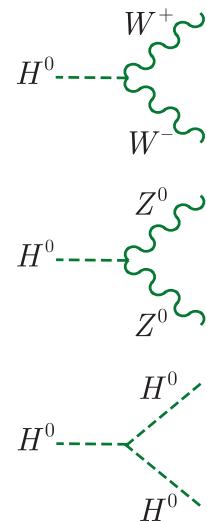


Figure 6.28 The basic first-order vertices for Higgs boson–boson interactions.

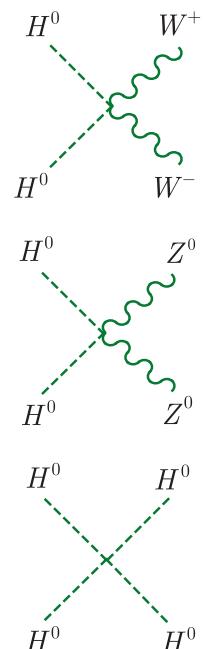


Figure 6.29 The basic second-order vertices for Higgs boson–boson interactions.

The decays listed in Table 6.1 arise as follows. The Higgs boson is much more stable than the W^\pm and Z^0 bosons because its couplings to the relatively light particles to which it can decay are heavily suppressed. In lowest order, the decays allowed by energy conservation are

$$H \rightarrow f + \bar{f}, \quad (6.73)$$

where f can be any lepton or quark except the t quark. These decays occur via the vertices shown in Figure 6.27, and if the fermion masses are neglected relative to the Higgs mass, the corresponding decay widths are given by²⁷

$$\Gamma(H \rightarrow f \bar{f}) = N \alpha_W \left(\frac{m_f^2}{M_W^2} \right) M_H c^2, \quad (6.74)$$

where $\alpha_W^2 = g_W^2/4\pi = 0.0042$ is the characteristic strength of the weak interaction. The suppression factor (m_f^2/M_W^2) comes from the mass factor in the coupling (6.74) and is much less than unity for all f . The factor N in (6.74) is 1 for leptons but 3 for quarks, to allow for the three possible colour states in the latter case, so that

$$\Gamma(H \rightarrow b\bar{b}) : \Gamma(H \rightarrow c\bar{c}) : \Gamma(H \rightarrow \tau\bar{\tau}) = 3m_b^2 : 3m_c^2 : m_\tau^2. \quad (6.75)$$

In the case of the quarks, the masses are those of the current quarks discussed in Section 5.7 evaluated at $\mu = M_H$, and are approximately $0.6 \text{ GeV}/c^2$ and $3 \text{ GeV}/c^2$ for the c and b quarks, respectively. The resulting partial widths and branching ratios for these modes are given in Table 6.1. The decay widths for the other lighter quarks and leptons are very much smaller.

The first-order decays (6.73) are so heavily suppressed by the mass factors in (6.74) that higher-order decays can lead to comparable decay rates. The three higher-order decays that contribute significantly are

$$H^0 \rightarrow W^\pm + f + \bar{f}', \quad (6.76)$$

$$H^0 \rightarrow Z^0 + f + \bar{f}, \quad (6.77)$$

and

$$H^0 \rightarrow g + g, \quad (6.78)$$

where $f\bar{f}$ and $f\bar{f}'$ are any lepton–antilepton or quark–antiquark pair allowed by the conservation laws, and where the gluons and quarks are observed as jets. Reactions (6.76) and (6.77) are second-order electroweak processes. This is illustrated for (6.76) in Figure 6.30, while (6.78) proceeds via an intermediate $t\bar{t}$ pair as shown in Figure 6.31. Their rates are given in Table 6.1, together with that for the rare decay

$$H^0 \rightarrow \gamma + \gamma, \quad (6.79)$$

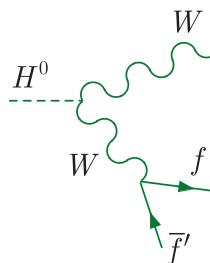


Figure 6.30 The dominant mechanism for the decay $H^0 \rightarrow W^\pm + f + \bar{f}'$.

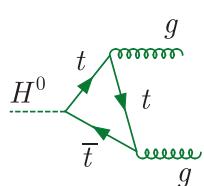


Figure 6.31 The dominant mechanism for the decay $H^0 \rightarrow g + g$.

²⁷The exact first-order result, including fermion masses, is given in Section 19.5.1 of Mandl and Shaw (2010). Note that (6.74) is often expressed in terms of the Fermi coupling, which is related to α_W by (1.54).

the dominant mechanisms that are shown in Figure 6.32. It is included because of its important role in the experimental discovery of the Higgs boson, as we shall see later.

Before the above decays can be studied experimentally, it is necessary to produce Higgs bosons at a measurable rate, and since they couple extremely weakly to the leptons and quarks of which stable beams and targets are composed, this can only occur by processes in which heavy particles, namely W^\pm , Z^0 bosons and top quarks, act as intermediate states. For example, in electron–positron colliding beams, the following reactions are used:

$$e^+ + e^- \rightarrow H^0 + Z^0, \quad (6.80a)$$

$$e^+ + e^- \rightarrow H^0 + \nu_e + \bar{\nu}_e, \quad (6.80b)$$

and

$$e^+ + e^- \rightarrow H^0 + e^+ + e^-. \quad (6.80c)$$

In these cases, the dominant mechanisms involve intermediate vector bosons, as shown in Figure 6.33, and lead to the predicted cross-sections shown in Figure 6.34 for centre-of-mass energies up to 3 TeV. As can be seen, the first of these reactions, often called *Higgsstrahlung*, dominates in the region close to threshold, but its cross-section decreases as the square of the centre-of-mass energy W at high energies, while that for (6.80b), which increases with W , dominates at higher energies. Reaction (6.80c) is similar to (6.80b) but is much smaller.

The e^+e^- reactions of (6.80a), (6.80b), (6.80c) are in principle ideal for searching for the Higgs boson and studying its properties. The bosons are produced in final states containing very few other particles, so that the signal is easy to identify, and background reactions are not too large since they are second-order electroweak processes, like the production processes themselves. Unfortunately, the maximum energy of the largest e^+e^- collider ever built (the Large Electron-Positron (LEPII) collider at CERN) was too low for a Higgs boson to be detected, but experiments were able to set a lower limit of $114 \text{ GeV}/c^2$ on the then-unknown Higgs mass. Things would greatly improve if the proposed International Linear Collider (ILC) or the Circular Electron Positron Collider (CEPC), mentioned in Section 4.2.2(d), is constructed. Detecting Higgs bosons in proton–proton collisions is less straightforward, but in this case sufficiently high energies are already available, as we shall see below.

6.5.4 Discovery of the Higgs boson

Results that first established the existence of the Higgs boson were reported in 2012 from two experiments, ATLAS and CMS, using the LHC at CERN. The ATLAS detector has been described in Section 4.5; CMS is a broadly similar general-purpose detector, although there are significant differences in their details.

At the LHC the reaction of interest is

$$p + p \rightarrow H^0 + X, \quad (6.81)$$

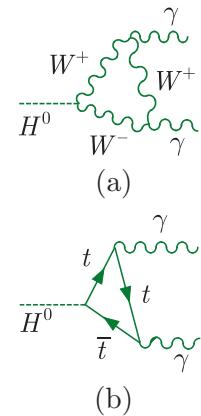


Figure 6.32 The dominant mechanisms for the rare decay $H^0 \rightarrow \gamma + \gamma$.

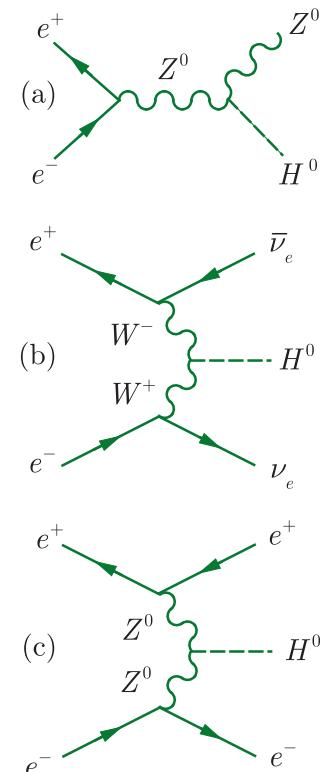


Figure 6.33 The dominant mechanisms for the reactions (6.80a), (6.80b), and (6.80c), respectively.

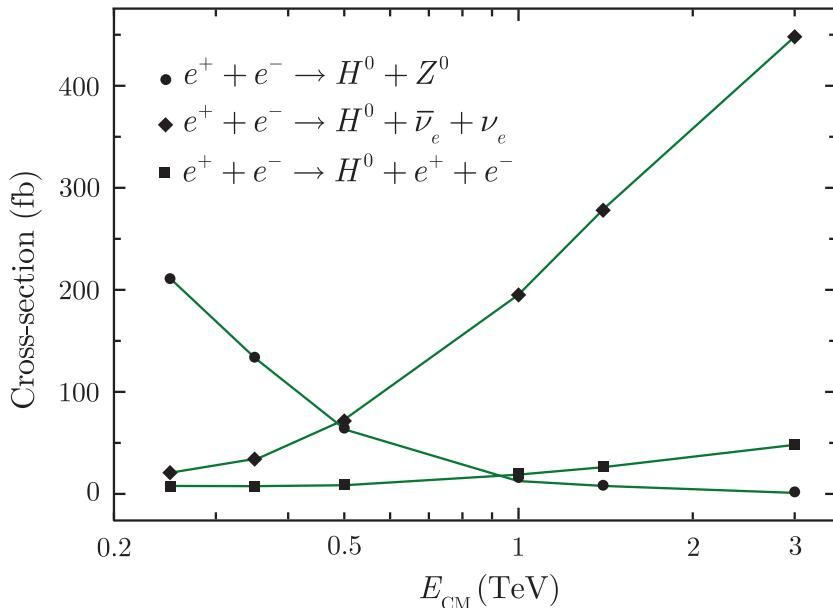


Figure 6.34 Predicted cross-sections for reactions (6.80a), (6.80b), and (6.80c). Source: Based on Table 1.15 of Ajaib et al. (2014).

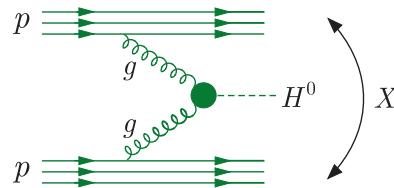


Figure 6.35 Gluon fusion production mechanism for the reaction $p + p \rightarrow H^0 + X$ at the LHC.

where X is any state consistent with conservation of the appropriate quantum numbers. The dominant Higgs production mechanism for this reaction at the LHC is ‘gluon fusion’, as shown in Figure 6.35, and the contribution of this process to the production cross-sections (6.81) is given by

$$\sigma(pp \rightarrow HX) = \int_0^1 \int_0^1 g(x_1)g(x_2)\sigma(gg \rightarrow H)dx_1 dx_2,$$

where $g(x)$ is the gluon distribution function of the proton²⁸ and the cross-section for $\sigma(gg \rightarrow H)$ is dominated by a Feynman diagram with a loop involving top quarks, shown in Figure 6.36a, because of the strong coupling of H^0 to the very heavy top quark. Other production processes are shown in Figure 6.36b to d, where for simplicity the spectator particles are not shown, and the corresponding production cross-sections for all four in the LHC energy range are shown in Figure 6.37.

In all these processes, the Higgs boson is produced together with large numbers of particles resulting from beam fragmentation. These particles

²⁸The use of quark and gluon distribution functions to make analogous calculations was discussed in Section 5.6.2.

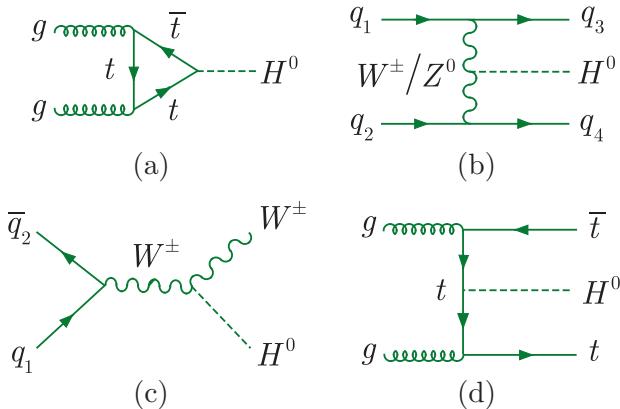


Figure 6.36 Production mechanisms for the Higgs boson at the LHC.

have low transverse momenta, as discussed in Section 5.6.1. In contrast, since the Higgs boson is heavy, its decay products have appreciable transverse momenta, except for the few cases when they are almost collinear with the beam. Hence in selecting Higgs boson candidates, the background from particles produced in beam fragmentation can in principle be greatly reduced by accepting only events in which the decay products have significant transverse momenta. The main problem arises because, on the one hand, most decays of the Higgs boson result in hadron jets either directly from the quarks or gluons produced in the decay or indirectly because most of the weak bosons emitted themselves decay to jets in the final state. On the other hand, at the same collision energies the proton–proton total cross-sections is of order 100 mb, and it is very difficult to distinguish jets arising from Higgs decays from the huge number of such

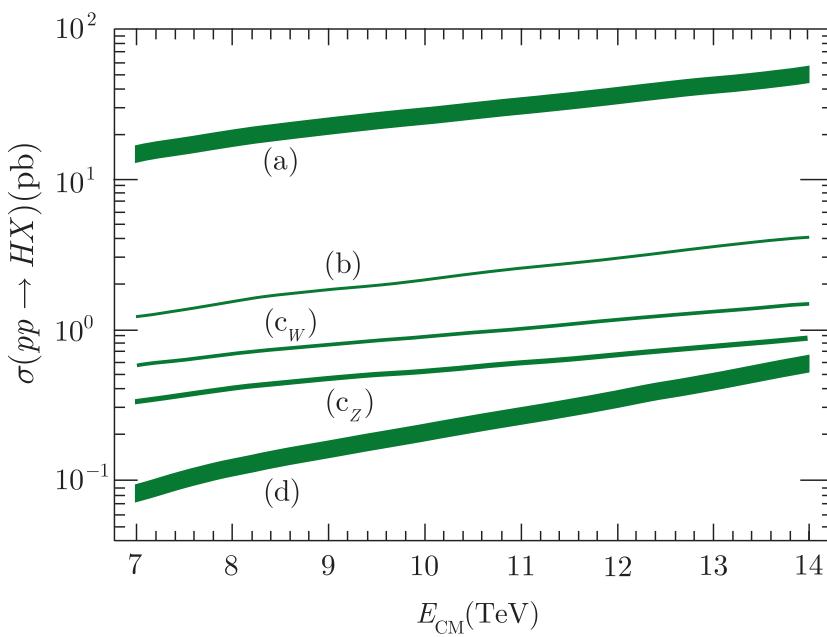


Figure 6.37 Predicted values for Higgs boson production cross-sections in pp collisions as a function of the centre-of-mass energy. The labelling of the curves corresponds to the diagrams in Figure 6.36, $c_{W,Z}$ referring to the production of W and Z bosons with the Higgs, respectively. The bands represent the uncertainties arising from input parameters and theoretical uncertainties. Source: Carena et al. (2013).

jets produced in proton–proton collisions by QCD interactions, like those briefly described in Section 5.6.1. In particular, from Table 6.1 we see that the dominant decay mode is $H^0 \rightarrow b\bar{b}$, with the quarks fragmenting to jets, but even if these are identified as b jets,²⁹ it is very difficult to distinguish those arising from Higgs decays from the much more numerous ones from other sources. This can be partly overcome by requiring that a W or Z boson be produced together with the Higgs boson in the final state,

$$p + p \rightarrow H + (W, Z) + X, \quad (6.82)$$

corresponding to the production mechanism of Figure 6.36c. While this reduces the Higgs production rate, it reduces the jet background much more, since the jets produced in proton–proton collisions are rarely accompanied by W or Z bosons in the final state. However, the jet background still remains very large compared to the Higgs production rate. Because of this, and similar problems with the other prominent decay modes, the best results were obtained by choosing rare decay modes containing only photons and leptons in the final states, so that the problem of background jets is eliminated, albeit at the price of a lower counting rate.

In any experiment, the number N of Higgs bosons produced and decaying to a particular final state F is given by

$$N = L_i \sigma(pp \rightarrow H + \dots) B(H \rightarrow F), \quad (6.83)$$

where L_i is the integrated luminosity given in (4.7b), σ is the production cross-section, and $B(H \rightarrow F)$ is the branching ratio for the final state observed. The experiments that established the existence of the Higgs boson collected data in 2011 and 2012 at proton–proton centre-of-mass energies of 7 and 8 GeV, where the cross-section for Higgs production is roughly 20 pb (1 pb = 10^{-12} b). In addition, because rare decay modes are chosen to reduce background problems, the branching ratio $B(H \rightarrow F)$ is also small, and high luminosity is required to produce a useful counting rate. Hence the LHC was designed as a pp rather than a $p\bar{p}$ collider, since proton beams can be produced with much higher intensity than antiproton beams. In particular, the ATLAS and CMS experiments each had integrated luminosities of about 5 pb⁻¹ at a centre-of-mass energy of 7 TeV and about 20 pb⁻¹ at 8 TeV, giving approximately 500 predicted Higgs boson decays leading to a final state F with a branching ratio of 10^{-3} .³⁰ In practice, two rare decays are of particular importance, in that they both avoid final-state jets and lead to high mass resolution. We shall consider each in turn.

²⁹This can be done, albeit with limited efficiency, by exploiting the fact that b jets almost always contain a fast-moving hadron with nonzero bottom quantum number $B = 0$. Other jets do not usually contain such particles. This is called ‘ b -jet tagging’.

³⁰The discovery of the Higgs boson was initially reported on the basis of integrated luminosities of 5 pb⁻¹ at 7 TeV and of 5 pb⁻¹ at 8 TeV. However, the results discussed here correspond to the full luminosity stated above.

6.5.4(a) $H^0 \rightarrow \gamma\gamma$

This decay has a predicted branching ratio of about 2×10^{-3} and the Higgs boson should appear as a narrow peak in the 2γ mass distribution, sitting on a large background from QCD production of two photons, the principal diagrams for which are shown in Figure 6.38. In the ATLAS experiment the energies and directions of photons are detected in the electromagnetic calorimeters and events were selected in which the transverse energy of the higher energy (lower energy) photon in the pair was required to be above 40 GeV (30 GeV). The invariant mass spectrum of the selected events exhibited a clear peak, as shown in Figure 6.39 and a detailed analysis, taking the backgrounds carefully into account, led to a Higgs mass of $M_H = 125.98 \pm 0.42 \pm 0.28 \text{ GeV}/c^2$, where the first error is statistical and the second systematic. Similar results were also obtained by CMS, corresponding to a mass value of $124.70 \pm 0.31 \pm 0.15 \text{ GeV}/c^2$.

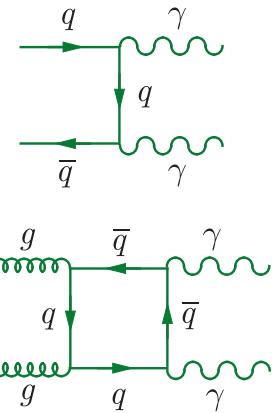


Figure 6.38 Dominant background processes producing 2γ final states.

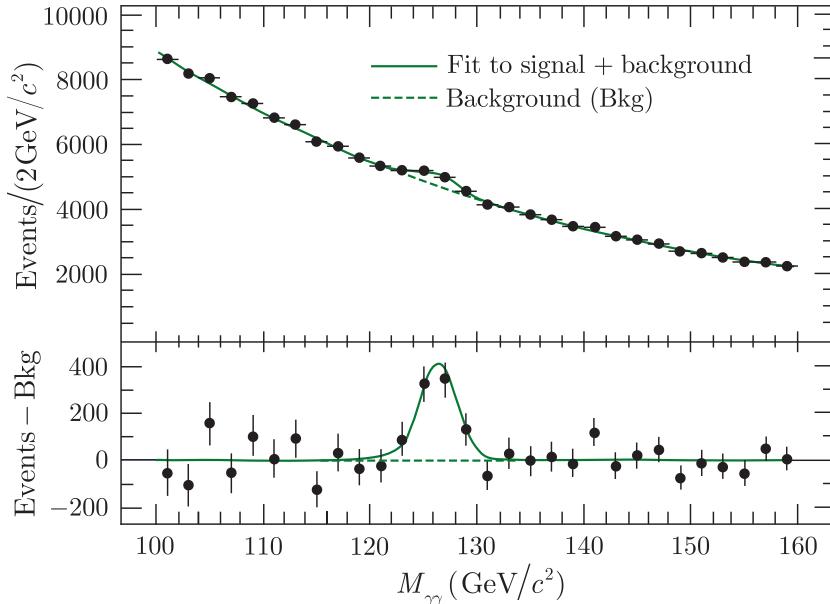


Figure 6.39 Distribution of the invariant mass of 2γ candidates obtained in the ATLAS experiment from runs at $E_{\text{CM}} = 7 \text{ TeV}$ with an integrated luminosity of 4.8 fb^{-1} , and 8 TeV with an integrated luminosity of 20.7 fb^{-1} , taken during 2011 and 2012. Source: Adapted from Aad et al. (2013).

6.5.4(b) $H^0 \rightarrow 4 \text{ charged leptons}$

The principal mechanism for the decays

$$H^0 \rightarrow \ell^+ + \ell^- + \ell'^+ + \ell'^- \quad (\ell = e, \mu; \ell' = e, \mu) \quad (6.84)$$

is shown in Figure 6.40. The main contribution to these decays arises when one of the intermediate Z^0 bosons is a real particle. The resulting branching ratio is predicted (cf. Problem 6.18) to be much smaller than the branching ratio for the 2γ decay, and hence there are far fewer events produced in any given experiment. However, to compensate for this, the signal-to-background ratio is expected to be much better and the energy

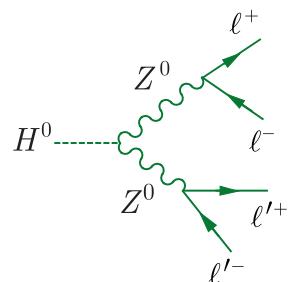


Figure 6.40 Dominant mechanism for the decay $H^0 \rightarrow \ell^+ + \ell^- + \ell'^+ + \ell'^-$ ($\ell = e, \mu; \ell' = e, \mu$).

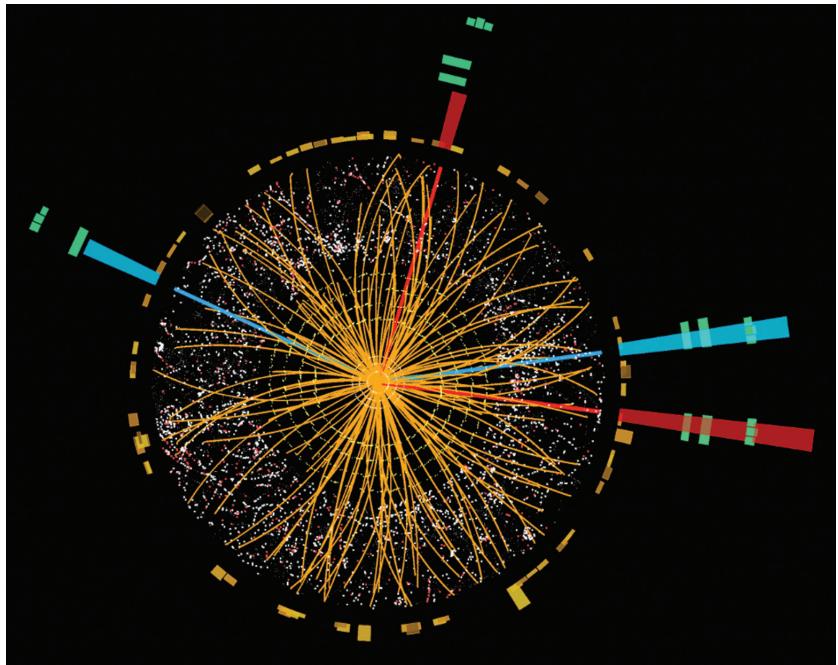


Figure 6.41 ATLAS event display of an $H^0 \rightarrow$ four-electron candidate event with $m_{4\ell} = 124.5 \text{ GeV}$. The masses of the lepton pairs are 70.6 and 44.7 GeV. The tracks and clusters of the two electron pairs are coloured red and blue, respectively.
Source: Reproduced by permission of CERN.

resolution is almost as good as in the 2γ mode. The charged leptons were observed in tracking detectors and in electromagnetic calorimeters. In the CMS experiment, the electrons and muons were required to have a minimum transverse momenta of $7 \text{ GeV}/c$ and $5 \text{ GeV}/c$, respectively; the invariant mass of one of the heaviest of the particle–antiparticle pairs was required to be in the range $40\text{--}120 \text{ GeV}/c^2$ and the other in the range $12\text{--}120 \text{ GeV}/c^2$; and the measured tracks of the pairs originating from Z decay were checked to ensure they were consistent with them originating from the same primary vertex. A typical event is shown in Figure 6.41 and the invariant mass spectrum of the selected events is shown in Figure 6.42. The peak at around $90 \text{ GeV}/c^2$ arises from Z decays to charged lepton pairs in which one of the leptons radiates a photon that converts into a second lepton pair. More significantly, there is a peak, sitting on a relatively small background, corresponding to a mass of $125.6 \pm 0.4 \pm 0.2 \text{ GeV}/c^2$, where again the first error is statistical and the second systematic. Atlas obtained similar results, with a mass of $124.51 \pm 0.52 \pm 0.06 \text{ GeV}/c^2$.

The above results clearly indicate the existence of a new state with a mass of about $125 \text{ GeV}/c^2$, and combining the four mass determinations cited above yields the value given in (6.72a). The widths of the peaks in the invariant masses are compatible with the experimental resolution, which is too large for the relatively small predicted width (6.72b) to be verified. However, the fact that the new particle is observed to decay to two photons implies that its C parity is +1 and that it has integer spin, while a subsequent study of the angular distributions in the four-charged-lepton mode shows that it has spin zero, as required for the Higgs boson.

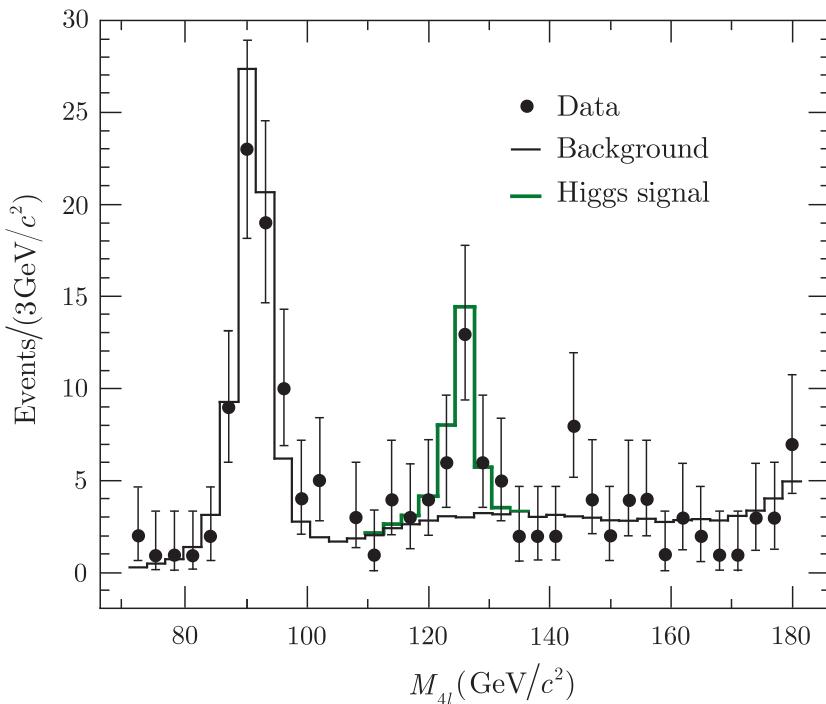


Figure 6.42 Distribution of the invariant mass of the four charged lepton candidates obtained in the CMS experiment from experimental runs at $E_{\text{CM}} = 7\text{ eV}$ with an integrated luminosity of 5.1 fb^{-1} and 8 TeV with an integrated luminosity of 19.6 fb^{-1} . Source: Adapted from Chatrchyan et al. (2014).

6.5.4(c) Further evidence

Further evidence is obtained by the comparison of the number of observed events with the number predicted by the standard model, as noted briefly above. Since the luminosity is known, one sees from (6.83) that the observed rate determines the product of the production cross-section and the branching ratio. In practice, it is usual to give results in terms of the ratio μ of the measured value to that predicted by the standard model, i.e.

$$\mu(F) = \frac{[\sigma(pp \rightarrow H + \dots)B(H \rightarrow F)]_{\text{exp}}}{[\sigma(pp \rightarrow H + \dots)B(H \rightarrow F)]_{\text{SM}}}. \quad (6.85)$$

As well as the two-photon and four-charged-lepton channels discussed above, results of comparable statistical significance for $\mu(F)$ have been obtained for the decay channel

$$H \rightarrow W + W^* \rightarrow \ell + \nu_\ell + \ell' + \nu_{\ell'}, \quad (6.86)$$

where ℓ and ℓ' can be electrons or muons and the neutrinos are such as to conserve lepton numbers in the reaction. This has a much higher branching ratio than the mode with four charged leptons, but a poorer mass resolution because the neutrinos cannot be detected. Values of μ with less precision are also obtained for the decay $H \rightarrow b\bar{b}$, which has the background problems discussed above, and the rare decay $H \rightarrow \tau\bar{\tau}$, in which observed final states again contain neutrinos resulting from the tauon decays. The measured values for these various decays, obtained by

Table 6.2 Measured values of the ratio (6.85) for the observed to predicted signal in Higgs production at the LHC, for individual final states F and for all final states combined

Final state	μ	$+\delta\mu$	$-\delta\mu$
$\ell \ell' \nu \nu$	1.08	0.18	0.16
$\gamma \gamma$	1.16	0.18	0.18
$\ell^+ \ell^- \ell'^+ \ell'^-$	1.14	0.15	0.13
$b\bar{b}$	0.95	0.22	0.22
$\tau^+ \tau^-$	1.12	0.23	0.23
Combined	1.10	0.11	0.11

combining results from both ATLAS and CMS, are shown in Table 6.2. As can be seen, the values for μ from these very varied final states, and the combined value from all the channels, are all statistically consistent with the predictions of the standard model, providing further impressive evidence that the observed state is indeed the long-sought-for Higgs boson.

Finally, in the above discussion we have considered only the simplest case of a single, electrically neutral, Higgs boson, as assumed in the standard model. However, some interesting extensions of the standard model, which are discussed in Chapter 10, require more than one Higgs particle, including electrically charged varieties, as well as other new states. The LHC was designed to search for states up to masses of a few TeV, but none have been observed so far. Whatever the truth, it is clear that experimental investigation of the Higgs sector will play a central role in developments in particle physics for many years to come.

Problems 6

- 6.1** Define charged and neutral current reactions in weak interactions and give an example of each in symbol form. How do they differ in respect of conservation of the strangeness quantum number? Why does observation of the process $\nu_\mu + e^- \rightarrow \nu_\mu + e^-$ constitute unambiguous evidence for weak neutral currents, whereas the observation of $\nu_e + e^- \rightarrow \nu_e + e^-$ does not?
- 6.2** Two identical particles can scatter elastically by both single photon and single Z^0 exchange. Calculate the typical squared energy-momentum transfer q^2 between them for a distance of closest approach of (a) 1 fm and (b) 10^{-3} fm in their centre-of-mass frame. Assuming that the intrinsic strengths of the fundamental weak and electromagnetic interactions are approximately equal, compare the relative sizes of the invariant (scattering) amplitudes for Z^0 and photon exchange processes at these two energy-momentum transfers.
- 6.3** Show that the vertices of Figures 6.4 and 6.5 are the only possible abW vertices allowed by charge and lepton number conservation, where a and b can be any lepton or antilepton.
- 6.4** Draw a Feynman diagram at the quark level for the decay $\Lambda \rightarrow p + \pi^-$. If nature were to double the weak coupling constant and decrease the mass

of the W boson by a factor of four, what would be the effect on the decay rate $\Gamma(\Lambda \rightarrow p + \pi^-)$?

- 6.5** Explain, with the aid of Feynman diagrams, why the decay $D^0 \rightarrow K^- + \pi^+$ can occur as a charged-current weak interaction at lowest order, but the decay $D^+ \rightarrow K^0 + \pi^+$ cannot.
- 6.6** Why is the decay rate of the charged pion much smaller than that of the neutral pion? Draw Feynman diagrams to illustrate your answer.
- 6.7** Draw the lowest-order Feynman diagrams for the decays $\pi^- \rightarrow \mu^- + \bar{\nu}_\mu$ and $K^- \rightarrow \mu^- + \bar{\nu}_\mu$. Use lepton–quark symmetry and the Cabibbo hypothesis with the Cabibbo angle $\theta_C = 13^\circ$ to estimate the ratio

$$R \equiv \frac{\text{Rate}(K^- \rightarrow \mu^- + \bar{\nu}_\mu)}{\text{Rate}(\pi^- \rightarrow \mu^- + \bar{\nu}_\mu)},$$

ignoring all kinematic and spin effects. Comment on your result with respect to the observed ratio.

- 6.8** Classify the following semileptonic decays of the $D^+(1869) = c\bar{d}$ meson as Cabibbo-allowed, Cabibbo-suppressed, or forbidden in lowest-order weak interactions, by finding selection rules for the changes in strangeness, charm, and electric charge in such decays:

- (a) $D^+ \rightarrow K^- + \pi^+ + e^+ + \nu_e$, (b) $D^+ \rightarrow K^+ + \pi^- + e^+ + \nu_e$,
 (c) $D^+ \rightarrow \pi^+ + \pi^+ + e^- + \bar{\nu}_e$, (d) $D^+ \rightarrow \pi^+ + \pi^- + e^+ + \nu_e$.

- 6.9** Which of the following six decays are allowed in lowest-order weak interactions?

- (a) $K^+ \rightarrow \pi^+ + \pi^+ + e^- + \bar{\nu}_e$, (b) $K^- \rightarrow \pi^+ + \pi^- + e^- + \bar{\nu}_e$,
 (c) $\Xi^0 \rightarrow \Sigma^- + e^+ + \nu_e$, (d) $\Omega^- \rightarrow \Xi^0 + e^- + \bar{\nu}_e$,
 (e) $\Xi^0 \rightarrow p + \pi^- + \pi^0$, (f) $\Omega^- \rightarrow \Xi^- + \pi^+ + \pi^-$.

- 6.10** Use lepton universality and lepton–quark symmetry to estimate the branching ratios for the decays $b \rightarrow c + e^- + \bar{\nu}_e$ (where the b and c quarks are bound in hadrons) and $\tau^- \rightarrow e^- + \bar{\nu}_e + \nu_\tau$. Ignore final states that are Cabibbo-suppressed relative to the lepton modes.

- 6.11** Show that an arbitrary $n \times n$ unitary matrix has n^2 real parameters, and hence that

$$\mathbf{U} = e^{-i\alpha} \begin{pmatrix} \cos \theta_C e^{i\beta} & \sin \theta_C e^{i\gamma} \\ -\sin \theta_C e^{-i\gamma} & \cos \theta_C e^{-i\beta} \end{pmatrix} \quad (\text{A})$$

is the most general form of a 2×2 unitary matrix. The most general form of (d, s) mixing is

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \mathbf{U} \begin{pmatrix} d \\ s \end{pmatrix}, \quad (\text{B})$$

where \mathbf{U} is an arbitrary 2×2 unitary matrix, $\mathbf{U}^\dagger \mathbf{U} = 1$. Show that this can be reduced to the form

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \begin{pmatrix} \cos \theta_C & \sin \theta_C \\ -\sin \theta_C & \cos \theta_C \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix}$$

by adjusting the arbitrary phases of the quark states s , s' , and d .

- 6.12** Hadronic strangeness-changing weak decays approximately obey the $\Delta I = 1/2$ rule, i.e. the total isospin changes by $1/2$ in the decay. By assuming a fictitious strangeness zero $I = 1/2$ particle S in the initial state, find the prediction of this rule for the ratio

$$R \equiv \frac{\Gamma(\Xi^- \rightarrow \Lambda + \pi^-)}{\Gamma(\Xi^0 \rightarrow \Lambda + \pi^0)}.$$

Assume that the state $|\Xi^0, S^0\rangle$ is an equal mixture of states with $I = 0$ and $I = 1$.

- 6.13** If the top quark were stable, the low-lying states of toponium $t\bar{t}$ could be approximated by nonrelativistic motion in a Coulomb potential (cf. (5.5a))

$$V(r) = -\frac{4}{3} \frac{\hbar c \alpha_s}{r},$$

with $\alpha_s \approx 0.1$. Use the simple Bohr model to calculate the radius of the ground state and the time taken to complete a single Bohr orbit in the ground state. Compare this with the expected lifetime of the top quark.

- 6.14** Which of the following processes are allowed in electromagnetic interactions and which are allowed in weak interactions via the exchange of a single W^\pm or Z^0 ?

- (a) $K^+ \rightarrow \pi^0 + e^+ + \nu_e$, (b) $K^+ \rightarrow \pi^+ + e^+ + e^-$,
 (c) $\Sigma^0 \rightarrow \Lambda^0 + e^+ + e^-$, (d) $\Sigma^0 \rightarrow n + \nu_e + \bar{\nu}_e$.

- 6.15** The charged-current differential cross-sections for ν and $\bar{\nu}$ scattering from a nucleon are

$$\frac{d\sigma^{CC}(\nu)}{dy} = \frac{1}{\pi} \frac{G^2 H s}{(\hbar c)^4}, \quad \frac{d\sigma^{CC}(\bar{\nu})}{dy} = \frac{d\sigma^{CC}(\nu)}{dy} (1-y)^2,$$

where $s = E_{CM}^2$, $y = \frac{1}{2}(1 - \cos \theta)$, and

$$H \equiv \int_0^1 x[u(x) + d(x)]dx$$

is the integral of the quark density for the target and the contribution from antiquarks is neglected. The corresponding cross-sections for neutral current scattering are

$$\begin{aligned} \frac{d\sigma^{NC}(\nu)}{dy} &= \frac{d\sigma^{CC}(\nu)}{dy} [g_L^2 + g_R^2(1-y)^2], \\ \frac{d\sigma^{NC}(\bar{\nu})}{dy} &= \frac{d\sigma^{CC}(\nu)}{dy} [g_L^2(1-y)^2 + g_R^2], \end{aligned}$$

where the neutral current couplings to u and d quarks are given by

$$g_L(u) = \frac{1}{2} - \frac{2}{3} \sin^2 \theta_W, \quad g_R(u) = -\frac{2}{3} \sin^2 \theta_W,$$

and

$$g_L(d) = -\frac{1}{2} + \frac{1}{3} \sin^2 \theta_W, \quad g_R(d) = \frac{1}{3} \sin^2 \theta_W.$$

Derive expressions for the ratios $\sigma^{NC}(\nu)/\sigma^{CC}(\nu)$ and $\sigma^{NC}(\bar{\nu})/\sigma^{CC}(\bar{\nu})$ in the case of an isoscalar nuclear target assuming that the cross-sections are just the sums of the individual nucleon cross-sections.

- 6.16** One way of looking for the Higgs boson H is in the reaction $e^+ e^- \rightarrow Z^0 H$. If this reaction is studied at a centre-of-mass energy of 500 GeV in a collider operating for 10^7 s per year and the cross-section at this energy is 60 fb, what instantaneous luminosity (in units of $\text{cm}^{-2}\text{s}^{-1}$) would be needed to collect 2000 events in one year if the detection efficiency is 10%? For a Higgs boson with mass $M_H < 120$ GeV, the branching ratio for $H \rightarrow b\bar{b}$ is predicted to be 85%. Why will looking for b quarks help distinguish $e^+ e^- \rightarrow Z^0 Z^0$ from the background reaction $e^+ e^- \rightarrow Z^0 Z^0$?
- 6.17** In our discussion of the reaction $e^+ + e^- \rightarrow \mu^+ + \mu^-$, we completely neglected the Higgs exchange diagram in Figure 6.43, compared with the dominant diagrams in Figures 6.25a and b, where the exchanged particle is a photon or a Z^0 boson. Since the Higgs contribution is obviously largest at $E_{CM} = M_H c^2$, where the contributions of the photon and Z^0 exchanges are expected to be comparable, justify this approximation by using the appropriate H^0 branching ratios to estimate the total cross-section at $E_{CM} = M_H c^2$ that would arise from Figure 6.43 alone and compare with estimates of the cross-sections that would arise from Figure 6.25a alone. Make the same comparisons of the contributions from Figure 6.25a alone and Figure 6.25b alone at $E_{CM} = M_Z c^2$.
- 6.18** Estimate the branching ratios for the four-lepton decays (6.84) and (6.86) of the Higgs boson by relating them to the branching ratios for other reactions listed in Table 6.1. Use your answers to estimate the average total number of Higgs decays (6.84) for an LHC energy of 7–8 TeV, where the Higgs production cross-section is about 20 pb with an error of approximately 15%, and an integrated luminosity of 25 fb^{-1} . Compare your results with the number actually observed as shown in Figure 6.42.
- 6.19** Draw the dominant lowest-order diagrams for the double Higgs boson production reaction $e^+ e^- \rightarrow H^0 H^0 Z^0$ at a total centre-of-mass energy of 500 GeV.

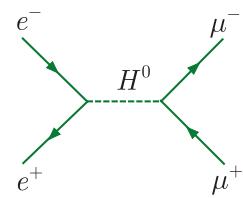


Figure 6.43 Feynman diagram for the reaction $e^+ e^- \rightarrow \mu^+ \mu^-$ via the exchange of a single Higgs boson.

7

Symmetry breaking in the weak interaction

In this chapter we will discuss parity (P) and charge conjugation (C), which are introduced in Sections 1.3.1 and 1.3.2 in the context of the strong and electromagnetic interactions. Here they will be discussed for the case of the weak interactions, where they are not conserved. In the course of this discussion, two interconnected themes will emerge. The first is that the violations have their origin in the spin dependence of weak interactions, which is quite remarkable, as we shall see. The second is that while parity and C parity violations are large effects, there is a weaker combined symmetry, called CP invariance, which is consistent with all current data on the weak interaction of leptons. However, while extending this symmetry to the weak interactions of hadrons initially leads to some striking experimental successes, it also leads to experimental evidence for deviations from CP invariance. This topic is discussed in Sections 7.3 and 7.4, where we also briefly discuss the more fundamental symmetry of CPT invariance, which is believed to be respected in all interactions. Finally in Section 7.5, we discuss the important question of whether CP violation can be fully accounted for within the standard model, or whether some new interaction is required.

7.1 P violation, C violation, and CP conservation

Parity and charge conjugation quantum numbers are conserved in the strong and electromagnetic interactions. This is verified by, for example, the absence of decays of nuclei that would violate this rule. The first indication that parity might be violated in weak interactions came from

observations on the pionic decays of K mesons, $K \rightarrow \pi\pi$ and $\pi\pi\pi$,¹ which led Lee and Yang in 1956 to make a thorough study of previous experiments in which parity conservation had been assumed, or apparently proved. They came to the startling conclusion that, contrary to general belief, there was in fact no firm evidence for parity conservation in weak interactions and they suggested experiments where the assumption could be tested.² This led directly to the classic demonstration of parity violation from a study of the β decay of polarised ${}^{60}\text{Co}$ nuclei. We shall describe just the principles of this experiment.³

The experiment was done in 1957 by Wu and co-workers, who placed a sample of ${}^{60}\text{Co}$ inside a magnetic solenoid and cooled it to a temperature of 0.01 K. At such temperatures, the interaction of the magnetic moments of the nuclei with the magnetic field overcomes the tendency to thermal disorder, and the nuclear spins tend to align parallel to the field direction. The polarised ${}^{60}\text{Co}$ nuclei produced in this way decay to an excited state of ${}^{60}\text{Ni}$ by the β decay



Parity violation was established by the observation of a ‘forward–backward decay asymmetry’, the fact that fewer electrons were emitted in the forward hemisphere than in the backward hemisphere with respect to the spin direction of the decaying nuclei.

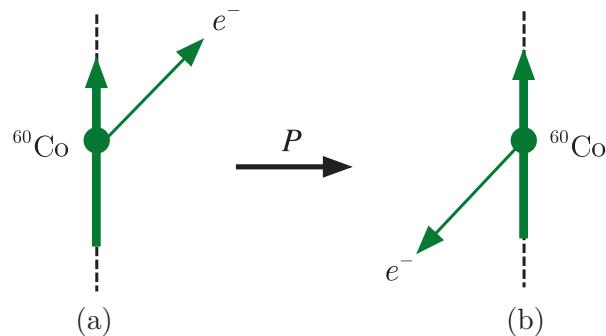


Figure 7.1 Effect of a parity transformation on ${}^{60}\text{Co}$ decay. The thick arrows indicate the direction of the spin of the ${}^{60}\text{Co}$ nucleus, while the thin arrows show the direction of the electron's momentum.

We can show that this implies parity violation as follows. The parity transformation reverses all particle momenta \mathbf{p} while leaving their orbital angular momenta $\mathbf{r} \times \mathbf{p}$, and by analogy their spin angular momenta, unchanged. Hence, in the rest frame of the decaying nuclei its effect is to reverse the electron velocity while leaving the nuclear spins unchanged, as shown in Figure 7.1. Parity invariance would then require that the

¹Two particles, called at that time τ and θ , were observed to decay via the weak interaction to $\pi\pi$ and $\pi\pi\pi$ final states, respectively, which necessarily had different final-state parities. However, the τ and θ had properties, including the near equality of their masses, that strongly suggested that they were in fact the same particle. Analysis of the ‘ $\tau - \theta$ puzzle’ suggested that parity was not conserved in the decays.

²For their work on parity nonconservation, Tsung-Dao Lee and Chen Yang were awarded the 1957 Nobel Prize in Physics.

³This classic experiment is described in readable detail in Chapter 10 of Trigg (1975).

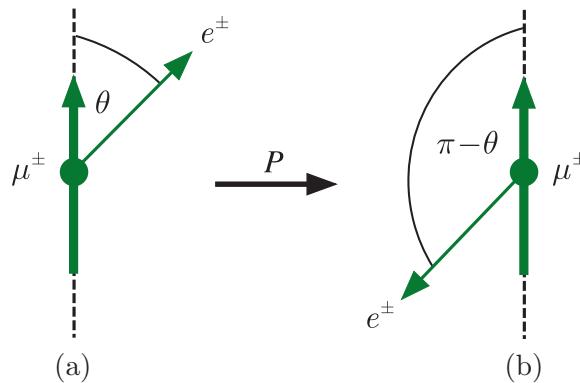


Figure 7.2 Effect of a parity transformation on muon decays. The thick arrows indicate the direction of the muon spin, while the thin arrows indicate the direction of the electron's momentum.

rates for the two processes shown in Figure 7.1a and b were equal, so that equal numbers of electrons would be emitted in the forward and backward hemispheres with respect to the nuclear spins, in contradiction to what was observed. The discovery of parity violation was a watershed in the history of weak interactions because the effect is large, and an understanding of weak interactions is impossible if it is neglected. We will see presently that the charge conjugation operator \hat{C} , that changes all particles to antiparticles, is also not conserved in weak interactions.

7.1.1 Muon decay symmetries

C violation and its relation to P violation are both conveniently illustrated by considering the angular distributions of the electrons and positrons emitted in the decays

$$\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu \quad (7.2a)$$

and

$$\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu \quad (7.2b)$$

of polarised muons.⁴ In the rest frame of the decaying particle these are found to be of the form

$$\Gamma_{\mu^\pm}(\cos \theta) = \frac{1}{2} \Gamma_\pm \left(1 - \frac{\xi_\pm}{3} \cos \theta \right), \quad (7.3)$$

where θ is the angle between the muon spin direction and the direction of the outgoing electron or positron, as shown in Figure 7.2a. The quantities ξ_\pm are called the *asymmetry parameters* and Γ_\pm are the total decay widths, or equivalently the inverse lifetimes, i.e.

$$(\hbar/\tau_\pm) \equiv \int_{-1}^{+1} d\cos \theta \Gamma_{\mu^\pm}(\cos \theta) = \Gamma_\pm, \quad (7.4)$$

as may easily be checked by direct substitution.

⁴Muons produced in the decays of charged pions are naturally polarised, as we shall see in Section 7.2.2.

We consider the consequences of assuming parity and charge conjugation for these decays, starting with the latter as it is the simpler. Under charge conjugation, μ^- decay converts to μ^+ decay. C invariance then implies that the rates, or equivalently the decay widths, and angular distributions for these decays should be the same, i.e.

$$\Gamma_+ = \Gamma_- \quad (C \text{ invariance}) \quad (7.5)$$

and

$$\xi_+ = \xi_- \quad (C \text{ invariance}). \quad (7.6)$$

The parity transformation preserves the identity of the particles, but reverses their momenta while leaving their spins unchanged. Its effect on muon decay is shown in Figure 7.2, where we see that it changes the angle θ to $\pi - \theta$, so that $\cos \theta$ changes sign. Hence P invariance implies

$$\Gamma_{\mu^\pm} (\cos \theta) = \Gamma_{\mu^\pm} (-\cos \theta) \quad (P \text{ invariance}). \quad (7.7)$$

Substituting (7.3) leads to the prediction that the asymmetry parameters vanish,

$$\xi_\pm = 0 \quad (P \text{ invariance}). \quad (7.8)$$

Experimentally, the μ^\pm lifetimes are equal to very high precision and the prediction for the decay widths (7.5) is satisfied, but the measured values of the asymmetry parameters are

$$\xi_- = -\xi_+ = 1.00 \pm 0.04, \quad (7.9)$$

which shows that both C invariance and P invariance are violated.

In view of these results, a question that arises is: why do the μ^+ and μ^- have the same lifetime if C invariance is violated? The answer lies in the principle of CP conservation, which states that the weak interaction is invariant under the combined operation CP , even though both C and P are separately violated. The CP operator transforms particles at rest to their corresponding antiparticles at rest, and CP invariance requires that these states should have identical properties. Thus, in particular, the masses of particles and antiparticles are predicted to be the same. Specifically, if we apply the CP operator to muon decays, the parity operator changes θ to $\pi - \theta$ as before, while the C operator changes particles to antiparticles. Hence CP invariance alone implies that the condition obtained from P invariance is replaced by the weaker condition

$$\Gamma_{\mu^+} (\cos \theta) = \Gamma_{\mu^-} (-\cos \theta). \quad (7.10)$$

Again, substituting (7.3) into this equation gives

$$\Gamma_+ = \Gamma_- \quad (CP \text{ invariance}), \quad (7.11)$$

implying equal lifetimes and also

$$\xi_+ = -\xi_- \quad (CP \text{ invariance}), \quad (7.12)$$

in agreement with the experimental results. Thus CP invariance retains the symmetry between particles and antiparticles as observed by experiment, at least for μ decays. In fact CP invariance has been verified in a wide variety of experiments involving weak interactions and present data are compatible with exact CP invariance for purely leptonic processes (i.e. ones involving only leptons). However, CP invariance is not exact for weak interactions involving hadrons, although it is usually a good approximation and deviations from it have so far only been observed in the decays of neutral kaons and charged and neutral B mesons. These will be discussed in Sections 7.3 and 7.4, where we will see that particles and antiparticles have the same masses and lifetimes even if CP is violated. For the moment, we will ignore CP violation and continue to discuss the implications of CP conservation for several processes, for which it is at least a very good approximation.

7.1.2 Parity violation in electroweak processes

The parity violating reactions described above are all charged current reactions, mediated by the exchange of W^\pm bosons, but parity violation also occurs in weak neutral current reactions. In particular, whenever a process is mediated by photon exchange, a Z^0 can also be exchanged, so that one has an electroweak reaction and parity violation is expected to arise from the neutral current contribution. However, as we saw in Section 6.4.2, in such reactions photon exchange dominates up to quite large energies and the reaction can be treated to a good approximation as an electromagnetic process, which conserves parity. Nonetheless, even in the region where the electromagnetic interaction dominates, tiny parity violation effects can occur arising from interference between the electromagnetic and the neutral current contributions.

The simplest reaction to consider is Möller scattering

$$e^- + e^- \rightarrow e^- + e^-,$$

for which the dominant mechanisms are given in Figure 7.3. This can be investigated in an experiment where a polarised electron beam is incident on an unpolarised electron target and the polarisation states of the final electrons are not detected. Thus the possible polarisation states of the target and final-state electrons are averaged and summed over respectively, and need not be considered further. The initial state is then summarised by Figure 7.4a, which transforms under parity P into Figure 7.4b, in which the momenta are reversed, while the spin remains unchanged. A rotation R_π through 180° about an axis lying in the reaction plane, passing through the interaction point at right angles to the beam direction, then leads to Figure 7.4c, which is identical to Figure 7.4a, except that the helicity of the incoming electron is reversed. Hence, if parity is conserved, the cross-sections $\sigma_R(\sigma_L)$ for incident electrons with their spins aligned opposite

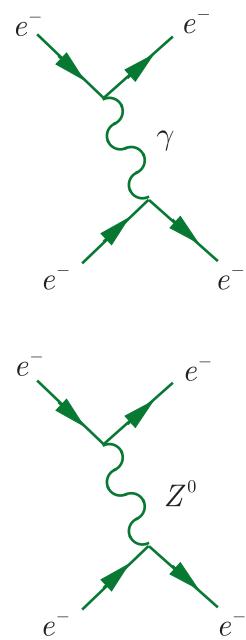
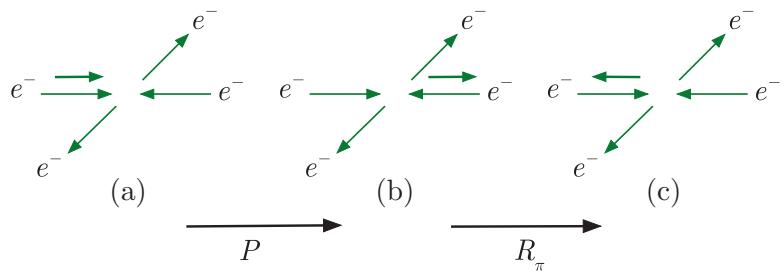


Figure 7.3 Photon and Z^0 exchange in Möller scattering.

Figure 7.4 (a) Möller scattering; (b) the effect of a parity transformation on the configuration (a); and (c) the effect of a rotation of 180° on the configuration (b). The thin lines represent a particle's direction and the short thick lines the spin orientation of the polarised electron.



(parallel) to their direction of motion should be equal, and the detection of a non-zero value of the asymmetry

$$A_{PV} \equiv \frac{\sigma_R - \sigma_L}{\sigma_R + \sigma_L} \quad (7.13)$$

is clear evidence for parity violation.

This phenomenon has been confirmed using polarised electron beams of about 50 GeV primary energy from the SLAC accelerator and scattering them from a liquid hydrogen target. The experiment was able to distinguish final-state electrons scattered from atomic electrons from those scattered from protons, effectively measuring A_{PV} for e^-e^- scattering in the rest frame of the atomic electrons. Then, if E_L is the beam energy in this frame, θ_{CM} is the electron scattering angle in the centre-of-mass frame, and m is the electron mass, the magnitude of the squared four-momentum transfer is given by

$$q^2 = (2m^2c^2 - 2mE_L) \sin^2(\theta_{CM}/2). \quad (7.14)$$

At $E_L = 50$ GeV, $|q^2|$ is very small compared to $M_Z^2c^2$, so that the contribution of the Z^0 is accurately given by the zero range approximation and both it and the predicted parity-violating asymmetry are expected to be extremely small. On the other hand, in the standard model A_{PV} is predicted to be proportional to $(1/4 - \sin^2\theta_W)$, and because $\sin^2\theta_W$ is not very different from $1/4$, even a relatively imprecise determination of A_{PV} leads to a precise determination of the weak mixing angle. The measured value obtained (by Anthony et al. (2005)) was $A_{PV} = (-131 \pm 17) \times 10^{-9}$, leading to a value $\sin^2\theta_W = 0.2330 \pm 0.0015$, consistent with the value (6.58) obtained by averaging over many other experiments.⁵

Before parity violation had been detected in polarised Möller scattering, a similar effect had been observed in deep inelastic scattering from nucleons. As discussed in Section 5.5, and illustrated in Figure 5.21, deep inelastic scattering is essentially elastic scattering

$$e^- + q \rightarrow e^- + q$$

⁵In quoting these results we have for simplicity combined the statistical and systematic errors in quadrature. A future experiment, planned at the Jefferson Laboratory using an 11 GeV polarised electron beam, is expected to yield a determination of $\sin^2\theta_W$ with an error of order 0.0002, comparable with the best determination by any other method.

from quarks within the nucleon, and if the incoming electrons are polarised, a non-zero value of the asymmetry (7.13) is again an unambiguous signal for parity violation. However, in contrast to Möller scattering, where the magnitude of A_{PV} depends only on $\sin^2\theta_w$, in this case it also depends on the quark distributions, with their attendant uncertainties. Parity violation in deep inelastic scattering was first detected at SLAC in 1978 by scattering polarised electrons with energies of about 20 GeV from a deuterium target and measuring the parity-violating asymmetry A_{PV} at a $|q^2|$ of approximately $1.6(\text{GeV}/c)^2$, where it is predicted to be of order 10^{-4} . Subsequently A_{PV} was also measured as a function of the fractional energy loss of the initial electron. This is a function of the weak mixing angle and a value of 0.224 ± 0.020 was found, in agreement with other determinations.⁶ Later experiments have confirmed the result in polarised electron–proton scattering.

These remarkable experiments provide unambiguous evidence for parity violation in ‘electromagnetic’ reactions at the level predicted by theory and hence for the electroweak unification as specified in the standard model. It should also in principle be possible to detect parity-violating effects in atomic physics, where a parity-violating interaction between the electrons and nucleus in an atom could allow atomic transitions that otherwise are forbidden. However, the interpretation of the results requires a detailed knowledge of the atomic theory of the atom and the uncertainties are such that atomic physics experiments do not currently compete with those in particle physics in detecting parity-violating effects and measuring $\sin^2\theta_w$, although this may change in the future.

7.2 Spin structure of the weak interactions

We turn now to the spin structure of the weak interactions, which is closely related to the symmetry properties discussed above. As this spin structure takes its simplest form for zero-mass particles, we will discuss the case of neutrinos and antineutrinos first, assuming that they have zero mass for the purpose of this discussion.

7.2.1 Left-handed neutrinos and right-handed antineutrinos

In discussing neutrinos, it is convenient to use the so-called *helicity states*, in which spin is quantised along the direction of motion of the particle, rather than along some arbitrarily chosen ‘z direction’. For a spin- $1/2$ particle, the spin component along the direction of its motion can be either $+1/2$ or $-1/2$ (in units of \hbar), as illustrated in Figure 7.5, corresponding to positive or negative helicity, respectively.⁷ These states are called

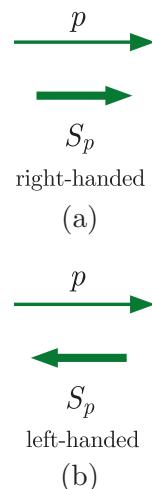


Figure 7.5 Helicity states of a spin- $1/2$ particle. The long thin arrows represent the momenta of the particles and the shorter thick arrows represent their spins.

⁶Prescott et al. (1978, 1979).

⁷More precisely, the helicity of a particle is defined as the spin projection along its direction of motion, divided by its maximum possible value.

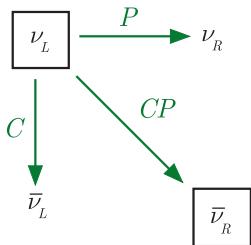
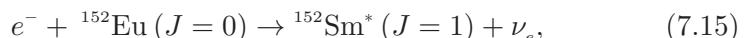


Figure 7.6 Effect of C , P , and CP transformations. Only the states shown in boxes are observed in nature.

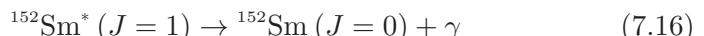
right-handed or *left-handed*, respectively, since the spin direction corresponds to rotational motion in a right-handed or left-handed sense when viewed along the momentum direction.

We will denote these states by a subscript R or L , so that, for example, ν_L means a left-handed neutrino. The remarkable fact about neutrinos and antineutrinos, which only interact via the weak interaction, is that *only left-handed neutrinos ν_L and right-handed antineutrinos $\bar{\nu}_R$ are observed in nature*. This obviously violates C invariance, which requires neutrinos and antineutrinos to have identical weak interactions. It also violates P invariance, which requires the states ν_L and ν_R to also have identical weak interactions, since the parity operator reverses the momentum while leaving the spin unchanged and so converts a left-handed neutrino into a right-handed neutrino. It is, however, compatible with CP invariance, since the CP operator converts a left-handed neutrino to a right-handed antineutrino, as illustrated in Figure 7.6.

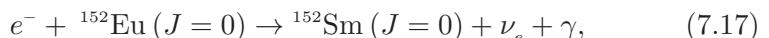
The helicity of the neutrino was first measured in an ingenious experiment by Goldhaber and co-workers in 1958. Again, we will discuss only the principles of the experiment.⁸ They studied electron capture in ^{152}Eu , i.e.



where the spins of the nuclei are shown in parentheses. The excited state of samarium that is formed decays to the ground state by γ emission



and it is these γ -rays that were detected in the experiment. In the first reaction (7.15), the electrons are captured from the K shell and the initial state has zero momentum, so that the neutrino and the ${}^{152}\text{Sm}^*$ nucleus recoil in opposite directions. The experiment selected events in which the photon was emitted in the direction of motion of the decaying ${}^{152}\text{Sm}^*$ nucleus,⁹ so that overall the observed reaction was



where the three final-state particles were co-linear, and the neutrino and photon emerged in opposite directions, as shown in Figure 7.7. The helicity of the neutrino can then be deduced from the measured helicity of the photon by applying angular momentum conservation about the event axis to the overall reaction. In doing this, no orbital angular momentum is involved, because the initial electron is captured from the atomic K shell and the final-state particles all move along the event axis. Hence the spin components of the neutrino and photon, which can be $\pm 1/2$ and ± 1

⁸A description of the experimental setup is given in Section 7.5 of Perkins (1987).

⁹This was done by resonant scattering from a second samarium target. It relies on the fact that those γ -rays travelling in the opposite direction to the neutrino have slightly more energy than those in other directions, and only the former have enough energy to excite the resonance level.

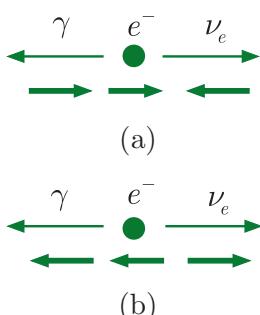


Figure 7.7 Possible helicities of the photon and neutrinos emitted in the reaction $e^- + {}^{152}\text{Eu} (J = 0) \rightarrow {}^{152}\text{Sm} (J = 0) + \nu_e + \gamma$ for those events in which they are emitted in opposite directions. Experiment selects configuration (a).

respectively,¹⁰ must add to give the spin component of the initial electron, which can be $\pm 1/2$. This gives two possible spin configurations, as shown in Figure 7.7a and b. In each case the photon and neutrino have the same helicities. In the actual experiment, the polarisation of the photons was determined by studying their absorption in magnetised iron (which depends on the polarisation of the photon) and the results obtained were consistent with the occurrence of left-handed neutrinos only, corresponding to Figure 7.7a.

A similar experiment for antineutrinos has been carried out involving the emission of a γ -ray following the positron decay of the ^{203}Hg nucleus. The polarisation is consistent with the earlier statement that only right-handed antineutrinos take part in weak interactions.

7.2.2 Particles with mass: chirality

To see the effect of the spin dependence in weak interactions involving particles with mass, we will look at the decays of the pion and muon, which are examples of charged current reactions. The spin dependence is of a special form, called a *V-A interaction*. This name is derived from the behaviour under a parity transformation of the weak interaction analogue of the electromagnetic current. The letter *V* denotes a *proper vector*, which is one whose direction is reversed by a parity transformation. (An example is momentum \mathbf{p} .) The familiar electric current, to which photons couple, transforms as a proper vector under parity. Because parity is not conserved in weak interactions, the corresponding weak current, to which W^\pm bosons couple, has in addition to a vector (*V*) component another component whose direction is unchanged by a parity transformation. Such a quantity is called an *axial-vector* (*A*). (An example of an axial-vector is orbital angular momentum $\mathbf{L} = \mathbf{r} \times \mathbf{p}$.) Since observables are related to the modulus squared of amplitudes, either term would lead by itself to parity conservation. Parity nonconservation is an interference effect between the two components.¹¹

Here we shall consider only the most important characteristic of this spin dependence, which is that the results discussed above for neutrinos hold for all fermions in the ultrarelativistic limit. That is, in the limit that their velocities approach that of light, only left-handed fermions ν_L, e_L^- , etc. and right-handed antifermions $\bar{\nu}_R, e_R^+$, etc. are emitted in charged current interactions. More generally, the eigenstates that take part in the weak interactions are linear combinations of helicity states, called *chiral* states, in which the contributions of the ‘forbidden’ helicity states e_R^-, e_L^+ , etc. are suppressed by factors that are typically of order $(mc^2/E)^2$, where m is the appropriate fermion mass and E its energy. For

¹⁰There are only two possible spin projections for the photons, which correspond to the two possible circular polarisation states for electromagnetic waves.

¹¹We will meet these ideas again in Section 8.7.1 when we discuss the *V-A* theory of beta decay.

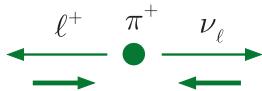


Figure 7.8 Helicities of the charged leptons in pion decays. The short arrows denote spin vectors and the longer arrows denote momentum vectors.

neutrinos, it is always a very good approximation to neglect such terms and for zero masses, chiral states and helicity states are identical. However, for particles with mass, it is only a good approximation for large energies E . These spin properties can be verified most easily for the electrons and muons emitted in weak decays, by directly measuring their spins. Here we shall assume them to hold and use them to understand some interesting features of pion and muon decays.

We start by considering the pion decay mode

$$\pi^+ \rightarrow \ell^+ + \nu_\ell \quad (\ell = e, \mu). \quad (7.18)$$

In the rest frame of the decaying pion, the charged lepton and the neutrino recoil in opposite directions, and because the pion has zero spin, their spins must be opposed to satisfy angular momentum conservation about the decay axis. Since the neutrino (assumed here to be zero mass) is left-handed, it follows that the charged lepton must also be left-handed, as shown in Figure 7.8, in contradiction to the expectations for a relativistic antilepton.

For the case of a positive muon this is unimportant, since it is easy to check that it recoils nonrelativistically and so both chirality states are allowed. However, if a positron is emitted it recoils *relativistically*, implying that this mode is suppressed by a factor that we can estimate from the above to be of order $(m_e/m_\pi)^2 \approx 10^{-5}$. Thus the positron decay mode is predicted to be much rarer than the muonic mode. This is indeed the case, and the measured ratio

$$\frac{\Gamma(\pi^+ \rightarrow e^+ + \nu_e)}{\Gamma(\pi^+ \rightarrow \mu^+ + \nu_\mu)} = (1.230 \pm 0.004) \times 10^{-4} \quad (7.19)$$

is in excellent agreement with a full calculation that takes into account both the above suppression and the difference in the density-of-final states (i.e. the difference in the Q values) for the two reactions.

A second consequence of the chirality argument is that the muons emitted in pion decays are polarised (see Figure 7.8).¹² We have mentioned this earlier in connection with measuring the muon decay asymmetries. These have their origins in the spin structure of the interaction, as we shall illustrate for the highest energy electrons emitted in the decay of the muon. These have energy

$$E = \frac{m_\mu c^2}{2} \left(1 + \frac{m_e^2}{m_\mu^2} \right) \gg m_e c^2 \quad (7.20)$$

and correspond to decays in which the neutrino and antineutrino are both emitted in the direction opposite to the electron. This is illustrated in Figure 7.9 for the two simplest cases in which the electron is emitted

¹²This is in the rest frame of the decaying pion and assumes that the neutrino has zero mass. The degree of polarisation in the laboratory frame is a function of the muon momentum.

in the muon spin direction (Figure 7.9a) and opposite to it (Figure 7.9b). Since the neutrino and antineutrino have opposite helicities, the muon and electron must have the same spin component along the event axis in order to conserve angular momentum, implying the electron helicities shown in Figure 7.9. When combined with the fact that the relativistic electrons emitted must be left-handed, this implies that electrons cannot be emitted in the muon spin direction. We thus see that the spin structure of the interaction automatically gives rise to a forward–backward asymmetry in polarised muon decays. Of course, not all the electrons have the maximum energy and the actual asymmetry, averaged over all electron energies, can only be calculated by using the full form of the $V-A$ interaction.¹³ The resulting prediction is in excellent agreement with the measured values.

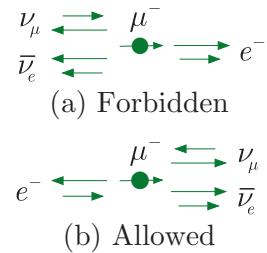


Figure 7.9 Muon decays in which electrons of the highest possible energy are emitted (a) in the muon spin direction and (b) opposite to the muon spin direction.

7.3 Neutral kaons: particle–antiparticle mixing and CP violation

The weak decays of neutral mesons are of particular interest, not only because they enable very sensitive tests of CP conservation to be made, but also because the application of basic quantum mechanics leads to surprising effects that, for example, allow the symmetry between particles and antiparticles to be tested with extraordinary precision. In both cases the crucial ingredient is the phenomenon of *particle–antiparticle mixing*. In this section, we will discuss these phenomena in the context of neutral kaon decays, where they were first discovered. Similar phenomena occur in neutral B and D decays, and the former will be discussed in detail in Section 7.4.

7.3.1 CP invariance and neutral kaons

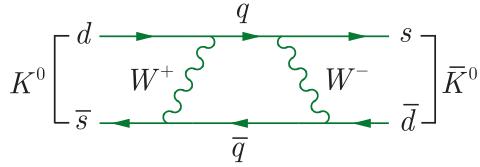
We have seen that there are two neutral kaon states,

$$K^0(498) = d\bar{s} \quad \text{and} \quad \bar{K}^0(498) = s\bar{d}, \quad (7.21)$$

which have strangeness $S = 1$ and $S = -1$, respectively. However, because strangeness is not conserved in weak interactions, these states can be converted into each other by higher-order weak processes like that shown in Figure 7.10. This is in marked contrast to most other particle–antiparticle systems, for which such transitions are forbidden, because the particle and its antiparticle differ by quantum numbers that are conserved in all known interactions. For example, the π^+ and π^- have opposite electric charges and the neutron and antineutron have opposite baryon numbers. For neutral kaons, however, there is no conserved quantum number to distinguish the K^0 and \bar{K}^0 states when weak interactions are taken into account and the observed physical particles with definite masses correspond

¹³See, for example, Chapter 12 of Halzen and Martin (1984).

Figure 7.10 Example of a process that can convert a K^0 state to a \bar{K}^0 state. The intermediate quark states are $q = u, c$, or t and $\bar{q} = \bar{u}, \bar{c}$, or \bar{t} .



not to the K^0 and \bar{K}^0 states themselves but to linear combinations of them. As we will see, this not only leads to very precise tests of CP conservation but also to so-called *flavour oscillations*. The latter are in some ways analogous to neutrino oscillations, which occur because the states ν_e , ν_μ , and ν_τ are linear combinations of the states ν_1 , ν_2 , and ν_3 with definite masses, as discussed in Chapter 3.

We start by assuming that CP invariance is exact and that neutral kaons are eigenstates of the combined CP operator. In this case, using the standard phase convention we can define

$$\hat{C}|K^0, \mathbf{p}\rangle = -|\bar{K}^0, \mathbf{p}\rangle, \quad \hat{C}|\bar{K}^0, \mathbf{p}\rangle = -|K^0, \mathbf{p}\rangle, \quad (7.22)$$

where $|K^0, \mathbf{p}\rangle$ denotes a K^0 state with momentum \mathbf{p} , etc. Since kaons have negative intrinsic parity, we also have, for $\mathbf{p} = \mathbf{0}$,

$$\hat{P}|K^0, \mathbf{0}\rangle = -|K^0, \mathbf{0}\rangle, \quad \hat{P}|\bar{K}^0, \mathbf{0}\rangle = -|\bar{K}^0, \mathbf{0}\rangle, \quad (7.23)$$

so that

$$\hat{C}\hat{P}|K^0, \mathbf{0}\rangle = |\bar{K}^0, \mathbf{0}\rangle, \quad \hat{C}\hat{P}|\bar{K}^0, \mathbf{0}\rangle = |K^0, \mathbf{0}\rangle. \quad (7.24)$$

Thus CP eigenstates $K_{1,2}^0$ are

$$|K_{1,2}^0, \mathbf{0}\rangle = \frac{1}{\sqrt{2}}[|K^0, \mathbf{0}\rangle \pm |\bar{K}^0, \mathbf{0}\rangle] \quad (CP = \pm 1). \quad (7.25)$$

If CP were conserved, then K_1^0 would decay entirely to states with $CP = 1$ and K_2^0 would decay entirely into states with $CP = -1$. We examine the consequences of this for decays leading to pions in the final state.

Consider the state $\pi^0\pi^0$. Since the kaon has spin zero, by angular momentum conservation the pion pair must have zero orbital angular momentum in the rest frame of the decaying particle. Its parity is therefore given by

$$P = P_\pi^2(-1)^L = 1, \quad (7.26)$$

where $P_\pi = -1$ is the intrinsic parity of the pion. The C -parity is given by

$$C = (C_{\pi^0})^2 = 1, \quad (7.27)$$

where $C_{\pi^0} = 1$ is the C -parity of the neutral pion. Combining these results gives $CP = 1$. The same result holds for the $\pi^+\pi^-$ final state.

The argument for three-pion final states $\pi^+\pi^-\pi^0$ and $\pi^0\pi^0\pi^0$ is more complicated, because there are two orbital angular momenta to consider.

If we denote by \mathbf{L}_{12} the orbital angular momentum of one pair (either $\pi^+\pi^-$ or $\pi^0\pi^0$) in their mutual centre-of-mass frame, and \mathbf{L}_3 is the orbital angular momentum of the third pion about the centre-of-mass of the pair in the overall centre-of-mass frame, then the total orbital angular momentum $\mathbf{L} \equiv \mathbf{L}_{12} + \mathbf{L}_3 = \mathbf{0}$, since the decaying particle has spin-0. This can only be satisfied if $L_{12} = L_3$, which implies that the parity of the final state is

$$P = P_\pi^3(-1)^{L_{12}}(-1)^{L_3} = -1. \quad (7.28)$$

For the $\pi^0\pi^0\pi^0$ final state, the C -parity is

$$C = (C_{\pi^0})^3 = 1 \quad (7.29)$$

and combining these results gives $CP = -1$ overall. The same result can be shown to hold for the $\pi^+\pi^-\pi^0$ final state.

7.3.2 CP violation in K_L^0 decay

The experimental position is that two neutral kaons are observed, called K^0 -*short* and K^0 -*long*, denoted K_S^0 and K_L^0 , respectively. They have almost equal masses of about $499 \text{ MeV}/c^2$, but very different lifetimes and decay modes. The K_S^0 has a lifetime of $\tau_S \approx 9 \times 10^{-11} \text{ s}$ and decays overwhelmingly to two pions; the longer-lived K_L^0 has a lifetime of $\tau_L \approx 5 \times 10^{-8} \text{ s}$, with a significant branching ratio to three pions, but not to two. In view of the CP analysis above, this immediately suggests the identification

$$K_S^0 = K_1^0 \quad \text{and} \quad K_L^0 = K_2^0. \quad (7.30)$$

However, in 1964 it was discovered that the K_L^0 also decayed to two pions¹⁴

$$K_L^0 \rightarrow \pi^+ + \pi^-, \quad (7.31)$$

but with a very small branching ratio of order 10^{-3} . This result is clear evidence of CP violation. It was confirmed in later experiments on the decay $K^0 \rightarrow \pi^0\pi^0$. Subsequently, CP violation was observed in the semileptonic decays

$$K^0 \rightarrow \pi^- + e^+ + \nu_e \quad \text{and} \quad \bar{K}^0 \rightarrow \pi^+ + e^- + \bar{\nu}_e. \quad (7.32)$$

For example, suppose we start with a beam of K^0 particles. If we assume CP invariance, then (7.30) holds and the beam will initially have equal amounts of K_S^0 and K_L^0 . Then after a time that is large compared to the K_S^0 lifetime, the K_S^0 component will have decayed leaving just the K_L^0 component, which itself will be an equal admixture of K^0 and \bar{K}^0 components. We would therefore expect to observe identical numbers of

¹⁴The experiment is described in Section 11.2.2 of Martin and Shaw (2017). It was led by James Cronin and Val Fitch who received the 1980 Nobel Prize in Physics for their discovery.

electrons (N^-) and positrons (N^+) from the two decays (7.32). However, if K_L^0 is not an eigenstate of CP , then there will be an asymmetry in these numbers, which will depend on the relative strengths of the K^0 and \bar{K}^0 components in K_L^0 . This is what was observed.

Because CP is not conserved, the physical states K_S^0 and K_L^0 need not correspond to the CP eigenstates K_1^0 and K_2^0 , but can contain small components of states with the opposite CP , i.e. we may write

$$|K_S^0, \mathbf{0}\rangle = \frac{1}{(1 + |\varepsilon|^2)^{1/2}} [|K_1^0, \mathbf{0}\rangle + \varepsilon |K_2^0, \mathbf{0}\rangle] \quad (7.33)$$

and

$$|K_L^0, \mathbf{0}\rangle = \frac{1}{(1 + |\varepsilon|^2)^{1/2}} [\varepsilon |K_1^0, \mathbf{0}\rangle + |K_2^0, \mathbf{0}\rangle], \quad (7.34)$$

where ε is a small complex parameter. (The factor in front of the brackets is to normalise the states.) It is straightforward to show that the asymmetry observed in the semileptonic decays is given by

$$A \equiv (N^+ - N^-)/(N^+ + N^-) = 2\text{Re}\varepsilon \quad (7.35)$$

for a pure K_L^0 beam, if we neglect terms of order $|\varepsilon|^2$. Figure 7.11 shows data on A as a function of proper time (i.e. measured in the rest frame of the decaying particle). After the initial oscillations, there is seen to be an asymmetry, and from (7.35) this gives $2\text{Re}\varepsilon \approx 2.3 \times 10^{-3}$.

The magnitude of ε can be deduced from the measured rates of the two CP -violating decays



and

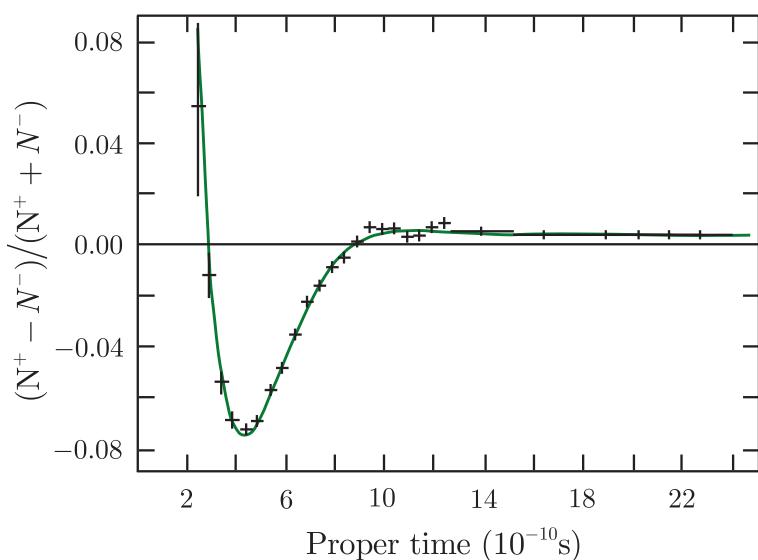


Figure 7.11 The charge asymmetry observed for $K^0 \rightarrow \pi^- e^+ \nu_e$ and $\bar{K}^0 \rightarrow \pi^+ e^- \bar{\nu}_e$ as a function of proper time, for a beam that is initially predominantly K^0 . Source: Adapted from Gjesdal et al. (1974). Copyright (1974) Elsevier, reprinted with permission.

which have both been observed with the branching ratios shown. These can occur in two different ways: either (i) *CP violation by mixing*, in which the *CP*-forbidden K_1^0 component in the K_L^0 decays via the *CP*-allowed processes

$$K_1^0 \rightarrow \pi^+ \pi^- \quad \text{and} \quad K_1^0 \rightarrow \pi^0 \pi^0, \quad (7.37)$$

giving in each case a contribution proportional to the probability

$$|\varepsilon|^2 / (1 + |\varepsilon|^2) \approx |\varepsilon|^2$$

of finding a K_1^0 component in the K_L^0 ; or (ii) *direct CP violation*, in which the K_2^0 component in the K_L^0 decays directly to pion pairs via the *CP*-violating reactions

$$K_2^0 \rightarrow \pi^0 \pi^0 \quad \text{and} \quad K_2^0 \rightarrow \pi^+ \pi^-. \quad (7.38)$$

In practice, *CP* violation by mixing dominates, as can be seen by comparing the ratios of probability amplitudes \mathcal{M} for K_L^0 and K_S^0 to $\pi^0 \pi^0$ and $\pi^+ \pi^-$. Defining

$$\eta_{00} \equiv \frac{\mathcal{M}(K_L^0 \rightarrow \pi^0 \pi^0)}{\mathcal{M}(K_S^0 \rightarrow \pi^0 \pi^0)}, \quad \eta_{+-} \equiv \frac{\mathcal{M}(K_L^0 \rightarrow \pi^+ \pi^-)}{\mathcal{M}(K_S^0 \rightarrow \pi^+ \pi^-)}, \quad (7.39)$$

it is easily seen from (7.33) and (7.34) that if direct *CP* violation in decay is completely neglected, these ratios are given by

$$\eta_{+-} = \eta_{00} = \varepsilon. \quad (7.40)$$

The measured values of the ratios of decay rates yields

$$|\eta_{00}| = (2.220 \pm 0.011) \times 10^{-3}, \quad |\eta_{+-}| = (2.232 \pm 0.011) \times 10^{-3}, \quad (7.41)$$

which are consistent with the prediction of *CP* violation by mixing alone. However, complete agreement with all the data on K_L^0 decays also requires a very small, but nonzero, additional contribution from the direct *CP*-violating term. When this is included (7.39) becomes

$$\eta_{+-} = \varepsilon + \varepsilon', \quad \eta_{00} = \varepsilon - 2\varepsilon', \quad (7.42)$$

where ε' is the additional contribution from direct *CP* violation. A detailed analysis of all the data with this term included yields

$$|\varepsilon| = (2.228 \pm 0.011) \times 10^{-3}, \quad |\varepsilon'| = (3.69 \pm 0.50) \times 10^{-6}, \quad (7.43)$$

confirming that *CP* violation by mixing is the dominant mechanism for *CP* violation in K_L^0 decay.

7.3.3 Flavour oscillations and *CPT* invariance

Although *CP* is not conserved, there is good reason to believe that all interactions are invariant under the combined operation of charge conjugation *C*, parity *P*, and time reversal *T*, taken in any order. This result,

which was mentioned briefly at the end of Section 1.3.3 is called the *CPT theorem* and can be shown to hold in any relativistic quantum theory in which signals cannot propagate faster than the speed of light. Like *CP*, the combined operation of *CPT* converts particles at rest to antiparticles at rest, and invariance under this operation requires particles and antiparticles to have the same masses and lifetimes and equal but opposite charges and magnetic moments. This is in accord with experiment, and is tested by observations on so-called *flavour oscillations*, which have been observed in several neutral meson systems. The most precise results are for the $K^0\bar{K}^0$ system, where we are dealing with *strangeness oscillations*. For example, the neutral kaon produced in the strong interaction

$$\begin{array}{c} \pi^- + p \rightarrow K^0 + \Lambda^0 \\ S = \quad 0 \quad 0 \quad 1 \quad -1 \end{array} \quad (7.44)$$

must necessarily be a K^0 state with $S=1$, in order to conserve strangeness. However, if the produced particle is allowed to travel through free space and its strangeness is measured, one finds that it no longer has a definite strangeness $S=1$, but has components with both $S=1$ and $S=-1$ whose intensities oscillate with time. The phenomenon is very similar mathematically to that describing the flavour oscillations of neutrinos we met in Chapter 3 and enables the mass difference between K_S^0 and K_L^0 particles to be measured with extraordinary precision, as we will now show.

In what follows, we shall measure time in the rest frame of the produced particle, and define $t=0$ as the moment when it is produced. If we ignore the very small *CP* violations, the initial state produced in the $\pi^- p$ reaction above is

$$|K^0, \mathbf{0}\rangle = \frac{1}{\sqrt{2}} [|K_S^0, \mathbf{0}\rangle + |K_L^0, \mathbf{0}\rangle]. \quad (7.45)$$

At later times, however, this will become

$$\frac{1}{\sqrt{2}} [a_S(t) |K_S^0, \mathbf{0}\rangle + a_L(t) |K_L^0, \mathbf{0}\rangle], \quad (7.46)$$

where

$$a_\alpha(t) = e^{-im_\alpha c^2 t/\hbar} e^{-\Gamma_\alpha t/2\hbar} \quad (\alpha = S, L) \quad (7.47)$$

and m_α and Γ_α are the mass and decay rate of the particle concerned. Here the first exponential factor is the usual oscillating time factor $e^{-iEt/\hbar}$ associated with any quantum mechanical stationary state, evaluated in the rest frame of the particle. The second exponential factor reflects the fact that the particles decay and ensures that the probability

$$\left| \frac{1}{\sqrt{2}} a_\alpha(t) \right|^2 = \frac{1}{2} e^{-\Gamma_\alpha t/\hbar} \quad (\alpha = S, L) \quad (7.48)$$

of finding a K_S^0 or K_L^0 decreases exponentially with a mean lifetime $\tau_\alpha = \hbar\Gamma_\alpha^{-1}$ ($\alpha = S, L$). Because $\tau_S \ll \tau_L$, for times t such that $\tau_S \ll t \lesssim \tau_L$ only the K_L^0 component survives, implying equal intensities for the K^0 and

\bar{K}^0 components. Here we are interested in the intensities of the K^0 and \bar{K}^0 components at shorter times, and to deduce these we rewrite the expression (7.46) in the form

$$A_0(t) |K^0, \mathbf{0}\rangle + \bar{A}_0(t) |\bar{K}^0, \mathbf{0}\rangle, \quad (7.49)$$

where

$$A_0(t) = \frac{1}{2}[a_S(t) + a_L(t)] \quad \text{and} \quad \bar{A}_0(t) = \frac{1}{2}[a_S(t) - a_L(t)]. \quad (7.50)$$

The intensities of the two components are then given by

$$\begin{aligned} I(K^0 \rightarrow K^0) &\equiv |A_0(t)|^2 \\ &= \frac{1}{4}[e^{-\Gamma_S t/\hbar} + e^{-\Gamma_L t/\hbar} + 2e^{-(\Gamma_S + \Gamma_L)t/2\hbar} \cos(\Delta m c^2 t/\hbar)] \end{aligned} \quad (7.51a)$$

and

$$\begin{aligned} I(K^0 \rightarrow \bar{K}^0) &\equiv |\bar{A}_0(t)|^2 \\ &= \frac{1}{4}[e^{-\Gamma_S t/\hbar} + e^{-\Gamma_L t/\hbar} - 2e^{-(\Gamma_S + \Gamma_L)t/2\hbar} \cos(\Delta m c^2 t/\hbar)]. \end{aligned} \quad (7.51b)$$

where

$$\Delta m = |m_S - m_L|, \quad (7.52)$$

and we have used (7.25) to explicitly evaluate the amplitudes (7.49). The resulting predictions for $\Delta m c^2 \tau_S / \hbar = 0.5$, where τ_S is the lifetime of the K_S^0 , are shown in Figure 7.12 and the *strangeness oscillations* associated with the cosine terms in (7.51a) and (7.51b) are clearly seen. The corresponding argument starting from \bar{K}^0 leads to the results

$$I(\bar{K}^0 \rightarrow \bar{K}^0) = I(K^0 \rightarrow K^0), \quad I(\bar{K}^0 \rightarrow K^0) = I(K^0 \rightarrow \bar{K}^0), \quad (7.51c)$$

which will be required below.

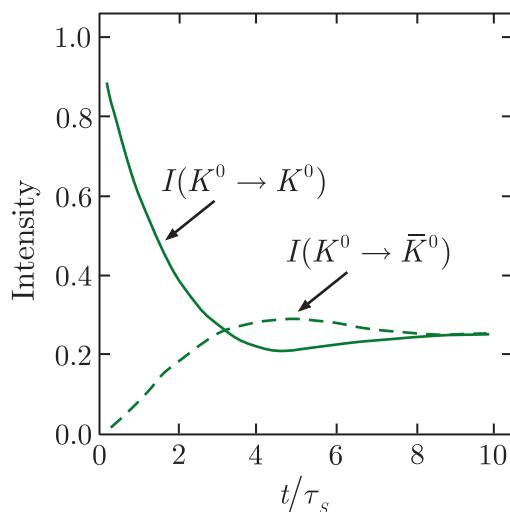
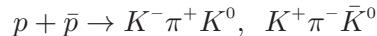


Figure 7.12 Predicted variation with time of the intensities $I(K^0 \rightarrow K^0)$ (solid line) and $I(K^0 \rightarrow \bar{K}^0)$ (dashed line) for an initial K^0 beam. The curves are calculated using (7.51) for $\Delta m c^2 \tau_S / \hbar = 0.5$, where Δm is the mass difference (7.52) and τ_S is the K_S^0 lifetime.

The strangeness oscillations (7.51) have been observed and the mass difference (7.52) measured, in several experiments. At the CPLEAR experiment at CERN, protons and antiprotons were observed to annihilate at rest at the centre of a large 4π detector and events corresponding to the reactions



were selected, where the identity of the neutral kaon produced is deduced using strangeness conservation. In addition, the particles are produced with sufficiently low momenta that the charged particles can be observed in the same detector and the time between production and decay determined.

Because of oscillations, the identity of the decaying particle as either a K^0 or a \bar{K}^0 will not necessarily be the same as that of the kaon produced initially, but can be identified by detecting it via the allowed leptonic decays (7.33). In this way, the number of events in which the nature of the neutral kaon has changed [$N_+(t)$], or not changed [$N_-(t)$], can be measured as a function of the time t between formation and decay. The resulting values for the ratio

$$\alpha_K(t) \equiv \frac{N_+(t) - N_-(t)}{N_+(t) + N_-(t)} \quad (7.53)$$

are plotted in Figure 7.13. The corresponding theoretical predictions are obtained by noting that K^0 and \bar{K}^0 mesons are produced in equal numbers at $t = 0$, so that

$$\alpha_K(t) = \frac{I(K^0 \rightarrow K^0) + I(\bar{K}^0 \rightarrow \bar{K}^0) - I(K^0 \rightarrow \bar{K}^0) - I(\bar{K}^0 \rightarrow K^0)}{I(K^0 \rightarrow K^0) + I(\bar{K}^0 \rightarrow \bar{K}^0) + I(K^0 \rightarrow \bar{K}^0) + I(\bar{K}^0 \rightarrow K^0)},$$

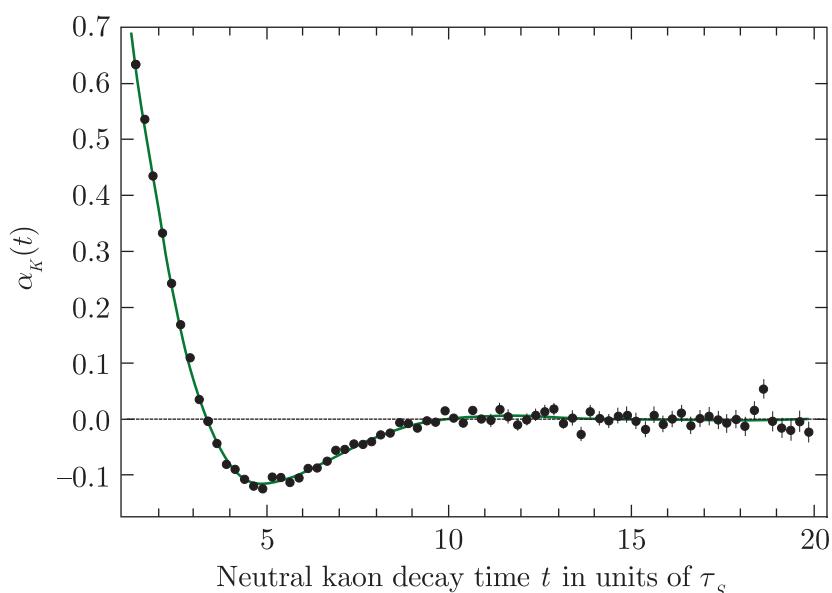


Figure 7.13 Plot of the ratio $\alpha_K(t)$ as a function the neutral kaon decay time in units of τ_S . The theoretical curve for $\alpha_K(t)$ is obtained from (7.54) using a mass difference $\Delta m = (3.483 \pm 0.006)10^{-12} \text{ MeV}/c^2$. Source: Adapted from Angelopoulos et al. (2001).

and substituting from (7.51) gives

$$\alpha_K(t) = \frac{2 \exp[-(\Gamma_S + \Gamma_L)t/2\hbar] \cos(\Delta mc^2 t/\hbar)}{\exp(-\Gamma_S t/\hbar) + \exp(-\Gamma_L t/\hbar)}. \quad (7.54)$$

This is in good agreement with the measurements, as can be seen in Figure 7.13 and, together with other experiments, yields a value

$$\Delta m = (3.483 \pm 0.006) 10^{-12} \text{ MeV}/c^2, \quad (7.55)$$

which is close to the value assumed, for illustrative purposes, in Figure 7.12.

Flavour oscillations have also been observed in the $D^0 - \bar{D}^0$, $B^0 - \bar{B}^0$, and $B_s^0 - \bar{B}_s^0$ meson systems, and will be discussed in the following section for the $B^0 - \bar{B}^0$ case. However, before turning to this system, we will briefly discuss the implications of the above results for the fundamental *CPT* theorem.

The *CPT* theorem relates the properties of particles and antiparticles at rest and the result (7.55) is important because it leads to a very precise confirmation that the masses of particles and antiparticles are equal. The K_S^0 and K_L^0 are not antiparticles, but the K^0 and \bar{K}^0 are, as can be seen from (7.22). The *CPT* theorem requires

$$m_{K^0} \equiv m_{\bar{K}^0}, \quad (7.56)$$

where the masses are defined as the expectation values of the energy for states of zero momentum. If this is so, the mass difference (7.55) can be shown to arise solely from the possibility of transitions $K^0 \leftrightarrow \bar{K}^0$, whose magnitude can be calculated from diagrams like that shown in Figure 7.10. We shall not discuss this further, but merely note that the resulting agreement between the predicted and measured values confirms the identity (7.56) to better than one part in 10^{18} . In contrast, the particle–antiparticle mass relations that have been most precisely tested by direct measurement are $m_{e^+} = m_{e^-}$ and $m_p = m_{\bar{p}}$, which are only verified to within experimental errors of order of one part in 10^8 and one part in 10^9 respectively. Many other tests of *CPT* invariance have been and continue to be performed and so far are all consistent with *CPT*-invariance. For example, the equality of the lifetimes of the μ^+ and μ^- has been confirmed to better than one part in 10^4 , while the equality of the magnitude of the magnetic moments of the electron and positron has been established to one part in 10^{11} and the equality of the magnitudes of the charge-to-mass ratio has been confirmed for protons and antiprotons to one part in 10^{10} .

7.4 CP violation and flavour oscillations in B decays

Until 2001, the evidence for *CP* violation was confined solely to the $K^0 - \bar{K}^0$ system. However, since then many other examples of *CP* violation

have been observed in the decay modes of the lightest meson states with nonzero bottom quantum numbers:

$$B^+(5279) = u\bar{b}, \quad B^0(5279) = d\bar{b}, \quad (\tilde{B} = +1) \quad (7.57a)$$

$$B^-(5279) = b\bar{u}, \quad \bar{B}^0(5279) = b\bar{d}, \quad (\tilde{B} = -1) \quad (7.57b)$$

and the lightest bottom-strange mesons

$$B_s^0(5367) = s\bar{b}, \quad \bar{B}_s^0(5367) = b\bar{s}, \quad (7.57c)$$

with quantum numbers $\tilde{B} = +1$, $S = -1$ and $\tilde{B} = -1$, $S = +1$, respectively. These include the $B^0 - \bar{B}^0$ and $B_s^0 - \bar{B}_s^0$ systems, which are analogous in many ways to the $K^0 - \bar{K}^0$ system, as we shall see in Section 7.4.2 below. However, there are important differences. In particular, direct CP violation is very much larger for mesons containing b quarks, enabling it to be observed much more directly. As we shall see immediately, this applies not only to neutral mesons but also to charged mesons, where there can be no particle–antiparticle mixing without violating charge conservation and direct CP violation is the only form that can occur.

7.4.1 Direct CP violation in decay rates

One method of testing for direct CP violation, for any hadron, is to compare the rates of CP -conjugate reactions. In particular, consider the decay $A \rightarrow f$ of a particle A at rest to a final state f . Then under a CP transformation this will become the antiparticle decay $\bar{A} \rightarrow \bar{f}$, with the momenta of the final state particles reversed. Summing over all possible final state momenta then gives

$$\Gamma(A \rightarrow f) = \Gamma(\bar{A} \rightarrow \bar{f}) \quad (7.58)$$

for the total decay rates if CP is conserved, and a nonzero value of the asymmetry parameter

$$A_{CP}(A \rightarrow f) \equiv \frac{\Gamma(\bar{A} \rightarrow \bar{f}) - \Gamma(A \rightarrow f)}{\Gamma(\bar{A} \rightarrow \bar{f}) + \Gamma(A \rightarrow f)} \quad (7.59)$$

is unambiguous evidence for direct CP violation.

At present, nonzero values of the asymmetry parameter (7.59) have been observed in several decay modes of the B mesons (7.57a) to (7.57c). To illustrate this, consider the decay modes

$$B^0 \rightarrow K^+ + \pi^- \quad \text{and} \quad \bar{B}^0 \rightarrow K^- + \pi^+ \quad (\text{allowed}), \quad (7.60a)$$

which proceed in lowest order by the mechanisms shown in Figure 7.14. In contrast, there are no possible diagrams leading to the decays

$$\bar{B}^0 \rightarrow K^+ + \pi^- \quad \text{and} \quad B^0 \rightarrow K^- + \pi^+ \quad (\text{forbidden}) \quad (7.60b)$$

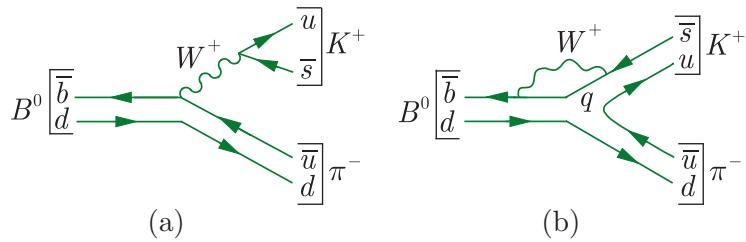


Figure 7.14 Quark diagrams for the decay $B^0 \rightarrow K^+ + \pi^-$ in lowest order weak interactions. The quark $q = \bar{u}, \bar{c}$ or \bar{t} . The corresponding mechanisms for $\bar{B}^0 \rightarrow K^- + \pi^+$ are obtained by replacing all particles by their antiparticles.

in lowest-order weak interactions. Hence neutral B mesons decaying into $K^+\pi^-$ and $K^-\pi^+$ automatically identify themselves as B^0 and \bar{B}^0 mesons, respectively. Comparison of their measured decay rates leads to the asymmetry parameter

$$A_{CP}(B^0 \rightarrow K^+ + \pi^-) = -0.082 \pm 0.006, \quad (7.61a)$$

while other examples of nonzero measured asymmetries include

$$\begin{aligned} A_{CP}(B_s^0 \rightarrow K^+ + \pi^-) &= 0.26 \pm 0.04 \quad \text{and} \\ A_{CP}(B^+ \rightarrow K^+ + \eta) &= -0.37 \pm 0.08, \end{aligned} \quad (7.61b)$$

where $\eta(549)$ is the neutral member of the $J^P = 0^-$ meson nonet shown in Figure 3.15a. These values are very large compared to the contribution of direct CP violation in K_L^0 decays, but were expected on the basis of the standard model. This is discussed in Section 7.5, where we will also see that CP violation in charmed meson decays is predicted to be much smaller than for B -mesons, and has not yet been observed at the time of writing.

7.4.2 $B^0 - \bar{B}^0$ mixing

We now turn to the $B^0 - \bar{B}^0$ system, where B^0 is the meson $B^0(5279) = d\bar{b}$. The bottom quantum number is not conserved in weak interactions and $B^0 - \bar{B}^0$ mixing can occur by diagrams like Figure 7.15, in analogy to $K^0 - \bar{K}^0$ mixing (Figure 7.10), and can be described by the same formalism. In general, for any system $M^0 - \bar{M}^0$ for which mixing of this type can occur, the physical states analogous to the K_S^0 and K_L^0 mesons, which we denote M_a and M_b , can be expanded in the form

$$|M_a\rangle = p|M^0\rangle + q|\bar{M}^0\rangle \quad (7.62a)$$

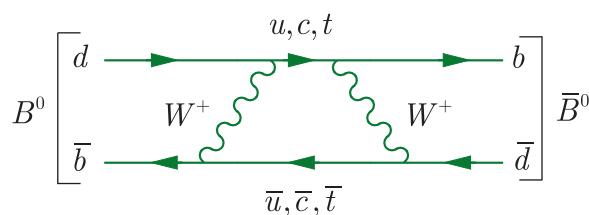


Figure 7.15 Example of a process that can convert a $B^0 = d\bar{b}$ state into a $\bar{B}^0 = \bar{d}\bar{b}$ state.

and

$$|M_b\rangle = p|M^0\rangle - q|\bar{M}^0\rangle, \quad (7.62b)$$

where p and q are complex parameters satisfying

$$|p|^2 + |q|^2 = 1. \quad (7.62c)$$

CP conservation in mixing is defined by

$$|p| = |q|, \quad (7.63)$$

implying equal probabilities for the particle and antiparticle components in the physical states M_a , M_b . On substituting (7.25) into (7.33) and (7.34), one sees that K_S^0 and K_L^0 can be expanded in the form (7.62) with $|p| \neq |q|$, so that *CP violation in mixing* occurs, and in fact dominates over direct *CP violation*, as we have seen.

The $B^0 - \bar{B}^0$ system has much in common with the $K^0 - \bar{K}^0$ system, but there are also important differences. There are a large number of decay modes possible with large Q values for B^0 and \bar{B}^0 mesons, so that the lifetimes are much shorter than for kaons. Furthermore, the total decay rates for B^0 mesons are dominated by reactions $B^0 \rightarrow f$ in which the final state f is such that the corresponding decay $\bar{B}^0 \rightarrow f$ is forbidden. Decays to final states f for which $B^0 \rightarrow f$ and $\bar{B}^0 \rightarrow f$ are both possible are relatively rare and in discussions of mixing they can, to a good approximation, be neglected. If this is done, it can be shown that mixing takes the form¹⁵

$$|B_L^0\rangle = \frac{1}{\sqrt{2}}[|B^0\rangle + \xi|\bar{B}^0\rangle] \quad (7.64a)$$

and

$$|B_H^0\rangle = \frac{1}{\sqrt{2}}[|B^0\rangle - \xi|\bar{B}^0\rangle], \quad (7.64b)$$

where

$$\xi = q/p = e^{-2i\beta}, \quad (7.64c)$$

and β is a real phase. In (7.64a) and (7.64b), B_L^0 and B_H^0 are the physical particles analogous to the K_S^0 and K_L^0 mesons. The notations are L for ‘light’ and H for ‘heavy’, because B_L^0 and B_H^0 have almost identical lifetimes of about 1.5×10^{-12} s. On comparing (7.64) with (7.62) and (7.63), we see that in this approximation, *CP violation by mixing* is completely absent. This is not an exact result, but it shows that, in contrast to the $K^0 - \bar{K}^0$ system, *CP violation in mixing* is a small effect compared to direct *CP violation*, which is a large effect for B mesons, as shown in Section 7.4.1. This does not mean that mixing is unimportant. In those rare decays in which B^0 and \bar{B}^0 decay to the same final state, it leads to a third form of *CP violation* – called *CP violation in interference* – which will be discussed in the next section. As for neutral kaons, it leads to flavour

¹⁵For a derivation of this result, see Section 11.2.8 of Martin and Shaw (2017).

oscillations that can be used to determine the magnitude of the tiny mass difference between the B_L^0 and B_H^0 mesons, as we will immediately see.

To obtain the form of flavour oscillations in the $B^0 - \bar{B}^0$ system, we start from the initial states

$$|B^0, t = 0\rangle = \frac{1}{\sqrt{2}} [|B_L^0\rangle + |B_H^0\rangle]$$

and

$$|\bar{B}^0, t = 0\rangle = \frac{1}{\xi} \frac{1}{\sqrt{2}} [|B_L^0\rangle - |B_H^0\rangle]$$

obtained by inverting (7.64). At later times, these will evolve into states

$$|B^0, t\rangle = \frac{1}{\sqrt{2}} [a_L(t) |B_L^0\rangle + a_H(t) |B_H^0\rangle] \quad (7.65a)$$

and

$$|\bar{B}^0, t\rangle = \frac{1}{\sqrt{2}} \frac{1}{\xi} [a_L(t) |B_L^0\rangle - a_H(t) |B_H^0\rangle], \quad (7.65b)$$

where

$$\begin{aligned} a_L(t) &= \exp(-im_L c^2 t/\hbar) \exp(-\Gamma t/\hbar), \\ a_H(t) &= \exp(-im_H c^2 t/\hbar) \exp(-\Gamma t/\hbar). \end{aligned} \quad (7.66)$$

Here m_L and m_H are the masses of B_L^0 and B_H^0 , $\Gamma = \hbar/\tau = (4.3 \times 10^{-12})$ MeV is their common decay width, and t , as usual, is the proper time. Substituting (7.64a) and (7.64b) into (7.65a) and (7.65b) gives

$$|B^0, t\rangle = A(t) |B^0\rangle + \xi \bar{A}(t) |\bar{B}^0\rangle \quad (7.67a)$$

and

$$|\bar{B}^0, t\rangle = A(t) |\bar{B}^0\rangle + (1/\xi) \bar{A}(t) |B^0\rangle, \quad (7.67b)$$

where

$$A(t) = \frac{1}{2} [a_L(t) + a_H(t)] = e^{-\Gamma t/2\hbar} e^{-iMc^2 t/\hbar} \cos(\Delta m c^2 t / 2\hbar) \quad (7.68a)$$

and

$$\bar{A}(t) = \frac{1}{2} [a_L(t) - a_H(t)] = i e^{-\Gamma t/2\hbar} e^{-iMc^2 t/\hbar} \sin(\Delta m c^2 t / 2\hbar). \quad (7.68b)$$

Here, $M = (m_L + m_H)/2$ and $\Delta m = m_H - m_L$. Since $|\xi| = 1$, these relations imply

$$I(B^0 \rightarrow B^0) = I(\bar{B}^0 \rightarrow \bar{B}^0) = e^{-\Gamma t/\hbar} \cos^2(\Delta m c^2 t / 2\hbar) \quad (7.69a)$$

and

$$I(B^0 \rightarrow \bar{B}^0) = I(\bar{B}^0 \rightarrow B^0) = e^{-\Gamma t/\hbar} \sin^2(\Delta m c^2 t / 2\hbar), \quad (7.69b)$$

where $I(B^0 \rightarrow B^0)$ is the probability of finding a B^0 at a time t in a state evolving from a B^0 at $t = 0$, and so on. The striking thing about (7.69) is that they are independent of the phase β in (7.64c) and would be unchanged if $\beta = 0$, when B_H^0 and B_L^0 would be CP eigenstates. However, in order to test these predictions, we must first confront an experimental problem.

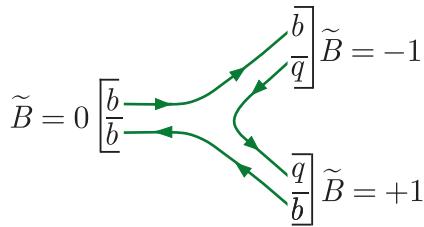


Figure 7.16 Mechanism for the decay of the $4S \tilde{B} = 0$ state $\Upsilon(10.58)$; $q = u$ or d .

Because their lifetimes are so short compared to those of the kaons, it is not possible to form well-defined beams of neutral B mesons, so some other means of studying their decay modes must be found. One way of doing this is by the construction of so-called B -factories. These facilities exploit the properties of the $\Upsilon(4S) = b\bar{b}$ resonance, which has a mass of $10.58 \text{ GeV}/c^2$ and a width of only 20 MeV . This is a state of bottomonium that is just heavy enough to decay to the lightest meson states with nonzero bottom quantum number listed in (7.57) by the mechanism of Figure 7.16, but not heavy enough to decay to any other final states with nonzero bottom by the same mechanism.¹⁶ It therefore decays almost entirely to B^+B^- and $B^0\bar{B}^0$ pairs, in approximately equal numbers. In addition, it has the same quantum numbers $J^{PC} = 1^{--}$ as the photon and so can be produced in e^+e^- collision annihilation by the mechanism of Figure 3.19. There is correspondingly a peak in the e^+e^- annihilation cross-section, and tuning the beam energies to coincide with this peak results in a copious source of B^\pm , B^0 , and \bar{B}^0 mesons.

Two such B factories were constructed to study CP violation in B decays: the PEP-II facility at SLAC, California, which operated from 1999 to 2008, and the KEK-B facility in Japan, which operated from 1999 to 2010. At PEP-II, a 3.1 GeV positron beam was collided head-on with a 9.0 GeV electron beam, whereas at the KEK-B facility a 3.5 GeV positron beam was collided with an 8 GeV electron beam at a small crossing angle of 83 mrad . In each case, the centre-of-mass energy corresponded to the $\Upsilon(4S)$ mass, while the asymmetric beam energies ensured that B mesons were produced with enough momentum to travel a measurable distance before decaying. This is important, because studies of CP violation often require the measurement of the time between the production and decay of the B mesons. This was achieved using dedicated multicomponent detectors, called BaBar at PEP-II and Belle at KEKB, respectively, that surrounded the vertices. Subsequently, KEK-B has been upgraded to SuperKEK-B, with beams of energies of 4.0 and 7.0 GeV , and a projected luminosity about 40 higher than KEK-B. First collisions are expected in 2018 and should lead in the near future to considerably more precise data than are currently available.

Because B_L^0 and B_H^0 have very similar lifetimes, it is not possible to observe a pure sample of one species analogous to the K_L^0 in neutral

¹⁶Compare the discussion of charmonium and bottomonium in Sections 3.3.5(a) and 3.3.5(b).

kaon decays, so that one is always dealing with particle mixtures. This problem was dealt with by using B -meson ‘tagging’. In B factories, as we have seen, pairs of B mesons result from $\Upsilon(4S)$ decays, and recoil against each other in their mutual centre-of-mass frame. Like kaons, the states produced will evolve in time, but which is which is unclear. Let us suppose, however, that one of the mesons is observed to decay by a \bar{B}^0 mode, then the B meson that recoils against it must at that instant be thrown into a B^0 state, to conserve the bottom quantum number. (This is an example of the quantum mechanical phenomenon known as ‘collapse of the wavefunction’.) Away from this instant, the B^0 state will evolve into an oscillating superposition of B^0 and \bar{B}^0 states, as specified by (7.69). A similar result applies if the identified meson decays by a B^0 mode rather than a \bar{B}^0 mode. This technique is called tagging and in both cases one can classify the observed pairs of decays into *mixed events*, in which the pairs both decay by B^0 (or \bar{B}^0) mode, implying mixing has occurred in the proper time t elapsed between the two decays, and *unmixed events*, in which a B^0 decay is associated with a \bar{B}^0 decay, which would always be the case if there were no mixing. If the numbers of each event are measured as a function of time, then the asymmetry

$$\alpha_B(t) = \frac{N(\text{mixed}) - N(\text{unmixed})}{N(\text{mixed}) + N(\text{unmixed})} \quad (7.70\text{a})$$

is given by

$$\alpha_B(t) = \frac{I(B^0 \rightarrow B^0) + I(\bar{B}^0 \rightarrow \bar{B}^0) - I(B^0 \rightarrow \bar{B}^0) - I(\bar{B}^0 \rightarrow B^0)}{I(B^0 \rightarrow B^0) + I(\bar{B}^0 \rightarrow \bar{B}^0) + I(B^0 \rightarrow \bar{B}^0) + I(\bar{B}^0 \rightarrow B^0)} \quad (7.70\text{b})$$

and substituting from (7.69a) and (7.69b) yields the predicted behaviour

$$\alpha_B(t) = \cos(\Delta mc^2 t / \hbar). \quad (7.71)$$

Figure 7.17 shows the asymmetry obtained in the BaBar experiment at PEP-II. The observed oscillations are fully compatible with (7.71) when a slowly varying background is taken into account, and when combined with data from other experiments, yield a best value of

$$m_H - m_L = (3.337 \pm 0.033) \times 10^{-10} \text{ MeV/c}^2.$$

7.4.3 CP violation in interference

We now turn to rare decays of the type $B^0 \rightarrow f$ and $\bar{B}^0 \rightarrow f$, where f is an eigenstate of CP , i.e.

$$\hat{C}\hat{P}|f\rangle = \eta_f|f\rangle, \quad \eta_f = \pm 1, \quad (7.72)$$

and calculate the probability that a particle tagged as, say, a B^0 at $t = 0$ decays to a final state f at time t . Because of oscillations, this can occur

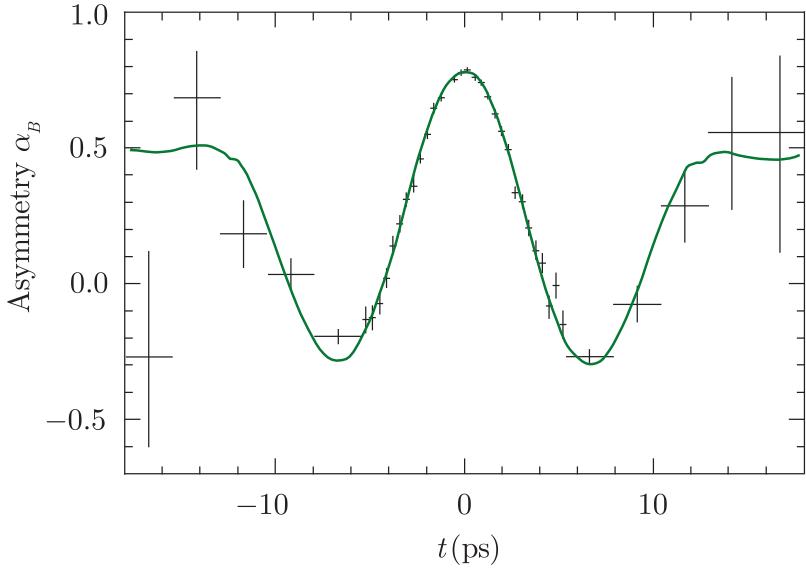


Figure 7.17 The measured asymmetry α_B as a function of the observed time difference t between the two tagged neutral B decays as measured in the BaBar experiment at PEP-II. The error increases at large $|t|$ as the mesons decay, and fewer events are observed. Source: Adapted from Aubert et al. (2006), with permission. Copyright the American Physical Society.

in two ways: $B^0 \rightarrow B^0 \rightarrow f$ or $B^0 \rightarrow \bar{B}^0 \rightarrow f$, where the first step in each case denotes the evolution from $t = 0$ to time t and the second step is the actual decay. It is the interference between these two terms that gives rise to *CP violation in interference*, as we shall see.

We will denote the decay amplitudes for $B^0 \rightarrow f$ and $\bar{B}^0 \rightarrow f$ by

$$\mathcal{M}_f = \langle f | H | B^0 \rangle \quad \text{and} \quad \bar{\mathcal{M}}_f = \langle f | H | \bar{B}^0 \rangle, \quad (7.73)$$

respectively. Then using (7.67), we see that the amplitude for an initial B^0 state to decay to a final state f at time t is

$$\mathcal{M}[B^0(t) \rightarrow f] = \mathcal{M}_f A(t) + \xi \bar{\mathcal{M}}_f \bar{A}(t), \quad (7.74a)$$

while the corresponding result for an initial \bar{B}^0 state is

$$\mathcal{M}[\bar{B}^0(t) \rightarrow f] = \bar{\mathcal{M}}_f A(t) + (1/\xi) \mathcal{M}_f \bar{A}(t). \quad (7.74b)$$

The corresponding decay probabilities are then given by

$$I[B^0(t) \rightarrow f] = |\mathcal{M}[B^0(t) \rightarrow f]|^2 \quad \text{and} \quad I[\bar{B}^0(t) \rightarrow f] = |\mathcal{M}[\bar{B}^0(t) \rightarrow f]|^2. \quad (7.75)$$

However, before evaluating these expressions, it is useful to rewrite (7.74a) and (7.74b) in the forms

$$\mathcal{M}[B^0(t) \rightarrow f] = \mathcal{M}_f [A(t) + \lambda \bar{A}(t)] \quad (7.76a)$$

and

$$\mathcal{M}[\bar{B}^0(t) \rightarrow f] = (\mathcal{M}_f / \xi) [\bar{A}(t) + \lambda A(t)], \quad (7.76b)$$

where the complex parameter

$$\lambda \equiv \xi \bar{\mathcal{M}}_f / \mathcal{M}_f. \quad (7.77)$$

Substituting (7.76) into (7.75) and using (7.68) to evaluate the resulting contributions gives, after some algebra (see Problem 7.11),

$$\begin{aligned} I[B^0(t) \rightarrow f] &= |\mathcal{M}_f|^2 e^{-\Gamma t/\hbar} \\ &\times \left[\frac{1}{2}(1 + |\lambda|^2) + \frac{1}{2}(1 - |\lambda|^2) \cos(\Delta mc^2 t/\hbar) - \text{Im}\lambda \sin(\Delta mc^2 t/\hbar) \right] \end{aligned} \quad (7.78a)$$

and

$$\begin{aligned} I[\bar{B}^0(t) \rightarrow f] &= |\mathcal{M}_f|^2 e^{-\Gamma t/\hbar} \\ &\times \left[\frac{1}{2}(1 + |\lambda|^2) - \frac{1}{2}(1 - |\lambda|^2) \cos(\Delta mc^2 t/\hbar) + \text{Im}\lambda \sin(\Delta mc^2 t/\hbar) \right] \end{aligned} \quad (7.78b)$$

where the terms proportional to $\text{Im}\lambda$ arise from the interference between the first and second terms in (7.74a) and (7.74b), respectively. Finally, when comparing to experiment it is usual to define the asymmetry as

$$\alpha_{fCP} \equiv \frac{I[\bar{B}^0(t) \rightarrow f] - I[B^0(t) \rightarrow f]}{I[\bar{B}^0(t) \rightarrow f] + I[B^0(t) \rightarrow f]} \quad (7.79)$$

and substituting (7.78a) and (7.78b) into this expression gives

$$\alpha_{fCP} = S \sin(\Delta mc^2 t/\hbar) - C \cos(\Delta mc^2 t/\hbar) \quad (7.80a)$$

where

$$S = \frac{2\text{Im}\lambda}{1 + |\lambda|^2} \quad \text{and} \quad C = \frac{1 - |\lambda|^2}{1 + |\lambda|^2}. \quad (7.80b)$$

At this point, we note that we have made no assumption about the amplitudes (7.73) for direct B^0 and \bar{B}^0 decays. If we now assume there is no direct CP violation in these decays, so that $\hat{X}H\hat{X}^{-1} = H$, where $\hat{X} = \hat{C}\hat{P}$ and H is that part of the Hamiltonian that contributes to these decays, then

$$\mathcal{M}_f = \langle f | H | B^0 \rangle = \langle f | \hat{X}^{-1} \hat{X} H \hat{X}^{-1} \hat{X} | B^0 \rangle = \eta_f \langle f | H | \bar{B}^0 \rangle = \eta_f \bar{\mathcal{M}}_f,$$

so, recalling that $\eta_f^2 = 1$,

$$\lambda = \eta_f \xi = \eta_f e^{-2i\beta}. \quad (7.81)$$

Hence the coefficient $C = 0$ and (7.80a) reduces to

$$\alpha_{fCP}(t) = -\eta_f \sin(2\beta) \sin(\Delta mc^2 t/\hbar). \quad (7.82)$$

Finally, if CP were exactly conserved, then the physical states (7.64) would be eigenstates of CP , requiring $\xi = \pm 1$ and $\alpha_{fCP} = 0$.

Such oscillating asymmetries have been observed in several decay modes, including $f = J/\psi K_S^0$ and $J/\psi K_L^0$, for which the standard model predicts that (7.81) and hence (7.82) hold to a very good approximation, as we shall see in Section 7.5. These modes have a combined branching ratio of about 10^{-3} and the final states are readily identified if the kaons are detected by the 2π and 3π decay modes, respectively. Since CP violation is very small for neutral kaons, K_S^0 and K_L^0 are to a good approximation identical to the K_1^0 and K_2^0 states with $CP = +1, -1$, respectively. The J/ψ state has $C = P = -1$, so that $CP(J/\psi) = +1$, and it has spin-1. Then the final state must have orbital angular momentum $L = 1$, leading to a factor $(-1)^L = -1$ under CP from the parity transformation. Hence

$$\eta_f (J/\psi K_S^0) = CP(K_S^0) \times CP(J/\psi) \times (-1)^L = -1$$

and similarly

$$\eta_f (J/\psi K_L^0) = CP(K_L^0) \times CP(J/\psi) \times (-1)^L = +1.$$

The asymmetries measured in the Belle experiment are shown in Figure 7.18. As can be seen, they are compatible with (7.82) and, together with other experiments, give

$$\sin 2\beta = 0.682 \pm 0.019.$$

This does not, of course, mean that there is no contribution from direct CP violation for all final states f , and, for example, for the very rare decay mode $f = \pi^+ \pi^-$, the more general result (7.80a) is required, with

$$C = -0.31 \pm 0.05 \quad \text{and} \quad S = -0.67 \pm 0.06.$$

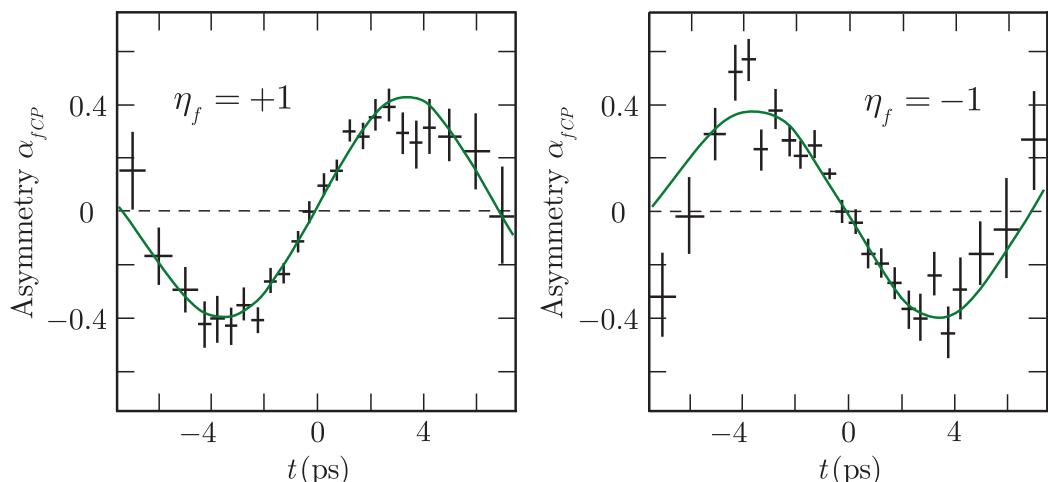


Figure 7.18 The asymmetry parameter α_{fCP} as measured by the Belle collaboration at KEK-B. The left panel is for the channel $f = J/\psi K_L^0$ (CP eigenstate $\eta_f = +1$) and the right panel is for $f = J/\psi K_S^0$ (CP eigenstate $\eta_f = -1$). Source: Adapted from Adachi et al. (2012). Copyright American Physical Society, reprinted with permission.

7.5 CP violation in the standard model

We conclude this chapter by reviewing whether the experimental observations on CP violation can be accommodated in the standard model, or whether there is some other mechanism necessary. In the standard model, CP violation has its origins in the quark mixing described by (6.40), where the elements of the CKM matrix

$$V \equiv \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}$$

determine the relative strengths of the various $\alpha\beta W$ couplings, where $\alpha = u, c, t$ and $\beta = d, s, b$. The CKM matrix in general contains nine complex elements. However, the unitary nature of the matrix implies that there are nine relations between the elements, such as

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0.$$

Using these, and exploiting the freedom to define the phases of the basic quark states, the CKM matrix may be parameterised by just four quantities, three mixing angles and one phase angle, and is conveniently written in the standard form:

$$V = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{13}s_{23}e^{i\delta} & c_{12}c_{23} - s_{12}s_{13}s_{23}e^{i\delta} & c_{13}s_{23} \\ s_{12}s_{23} - c_{12}s_{13}c_{23}e^{i\delta} & -c_{12}s_{23} - s_{12}s_{13}c_{23}e^{i\delta} & c_{13}c_{23} \end{pmatrix}. \quad (7.83)$$

Here $c_{ij} \equiv \cos \theta_{ij}$, $s_{ij} \equiv \sin \theta_{ij}$ and the angles $\theta_{ij} = \theta_{12}, \theta_{13}$, and θ_{23} can be chosen to lie in the first quadrant, so that $s_{ij}, c_{ij} \geq 0$. Since the time-reversal operator acting on a state involves complex conjugation (cf. (1.23)), a non-zero phase δ implies T violation, and hence, if CPT invariance is assumed, CP violation. This is in contrast to the general form of the mixing matrix (6.39) for two generations only, which is purely real, so that CP violation could not arise from quark mixing if there were only two generations.

The mixing angles and the CP -violating phase must be determined from experiment; and the magnitude of a particular CP -violating effect depends not only on the magnitude of δ but also the mixing angles. Experimentally, $s_{12} \gg s_{23} \gg s_{13}$, and it is instructive to approximate the resulting matrix by the so-called *Wolfenstein parameterisation*:

$$V = \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 - \frac{1}{8}\lambda^4 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda + \frac{1}{2}A^2\lambda^5[1 - 2(\rho + i\eta)] & 1 - \frac{1}{2}\lambda^2 - \frac{1}{8}\lambda^4(1 + 4A^2) & A\lambda^2 \\ A\lambda^3[1 - (1 - \frac{1}{2}\lambda^2)(\rho + i\eta)] & -A\lambda^2 + \frac{1}{2}A\lambda^4[1 - 2(\rho + i\eta)] & 1 - \frac{1}{2}A^2\lambda^4 \end{pmatrix} + O(\lambda^6), \quad (7.84)$$

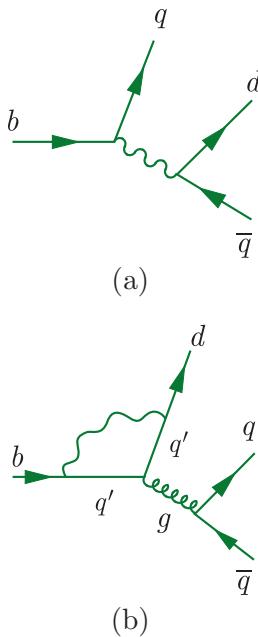


Figure 7.19 The tree (a) and penguin (b) diagrams for the decays $b \rightarrow dq\bar{q}$, where in (a) $q = u, c$ and in (b) $q = u, c, t$.

with parameters A , λ , ρ , and η , where a nonzero value of η is indicative of CP violation. The quantity $\lambda = |V_{ub}| \approx 0.225$ plays the role of an expansion parameter and from (7.84) one sees that the only elements that can give rise to CP violation, up to and including terms of order $\lambda^5 \approx 6 \times 10^{-4}$, are:

$$\begin{aligned} V_{ub} &= A\lambda^3(\rho - i\eta), & V_{cd} &= -\lambda + \frac{1}{2}A^2\lambda^5[1 - 2(\rho + i\eta)], \\ V_{td} &= A\lambda^3[1 - (1 - \frac{1}{2}\lambda^2)(\rho + i\eta)], & V_{ts} &= -A\lambda^2 + \frac{1}{2}A\lambda^4[1 - 2(\rho + i\eta)]. \end{aligned} \quad (7.85)$$

The other parameters are obtained from a detailed analysis of quark mixing data¹⁷ and the current values are:

$$A \approx 0.81, \quad \rho \approx 0.12, \quad \eta \approx 0.36.$$

From this one sees immediately that direct CP violation is predicted to be largest in those B decays in which V_{ub} plays an important role. This is illustrated in Figure 7.19 for the decays $b \rightarrow dq\bar{q}$, where $q = u, c$. Diagrams of the type shown in (a) are examples of *tree diagrams*, so called because they contain no closed loops. Diagrams of the type shown in (b) are called *penguin diagrams* because, suitably orientated, they remind some people of penguins. The loop in the penguin diagrams produces a suppression factor, which depends on the particular process, but is typically in the range 0.2 – 0.3. In addition, to estimate the relative importance of tree and penguin diagrams one must also take into account the powers of the expansion parameter λ in the vertices involved. For example, in the case $b \rightarrow du\bar{u}$, the tree diagram has vertices ubW and udW , which are of order unity and λ^3 , respectively, as can be seen from the entries in (7.84), giving a factor of λ^3 overall, whereas the leading penguin diagram has $q' = t$, so that the weak vertices are tbW and tdW , which also give a factor of order λ^3 overall. Hence the tree diagram is the most important in this case, but the penguin diagram is only suppressed by a loop factor of order 0.2–0.3, and so also makes a significant contribution. Examples of decays involving the transition $b \rightarrow du\bar{u}$ are $\bar{B}_s^0 \rightarrow K^+ + \pi^+$ and $\bar{B}^0 \rightarrow \pi^+ + \pi^-$, for both of which large direct CP -violating effects have been observed, as noted in Sections 7.4.1 and 7.4.3 respectively.

On the other hand, the leading tree and penguin diagrams for the decays $\bar{B}^0 \rightarrow f = J/\psi K_S^0$ and $J/\psi K_L^0$ involve the process $b \rightarrow sc\bar{c}$, as shown in Figure 7.20, where the superposition of K_S^0 and K_L^0 produced is initially a \bar{K}^0 . In this case, the tree diagram is of order λ^2 and if it dominates, there is no direct CP violation in these decays. The penguin contributions can give rise to CP -violating terms, i.e. terms proportional to $(\rho - i\eta)$, but one easily checks that they are of order λ^4 . Together with the loop factor, this gives a suppression factor of order 0.01 for the penguin diagrams relative to the tree diagram. Hence direct CP violation is

¹⁷The details are given in Ceccucci, Ligeti, and Sakai (2018).

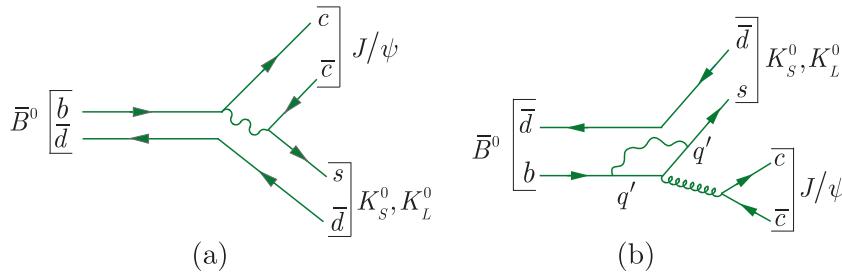


Figure 7.20 The leading tree and penguin diagrams for the decays $\bar{B}^0 \rightarrow f = J/\psi K_S^0$ and $J/\psi K_L^0$.

expected to be extremely small in these decay modes, as confirmed by the data on CP violation in interference for these decays, discussed in Section 7.4.3.

For K and D decays, however, the main contributions to CP violation arise from V_{cd} , and are of order λ^5 , whereas V_{cd} itself is of order λ , so that direct CP -violating effects are predicted to be much smaller in these cases. This is consistent with the data: as we saw in Section 7.3.2, direct CP violation in neutral kaon decays is very small compared to CP violation by mixing, while CP violation in D decays has yet to be detected.

Precise quantitative predictions are more complicated, since one must not only take into account a variety of diagrams like those shown in Figure 7.19 for the decay $b \rightarrow d u \bar{u}$, but, more problematically, the effects of the quarks being bound in mesons. We will not pursue this, but merely note that the data and predictions are consistent and the success of the mixing model in accounting for all CP -violating data in terms of a single CP -violating phase is a major triumph of the standard model. However, this is probably not the complete story on CP violation, because we will see in Chapter 10 that suggested explanations of the matter–antimatter asymmetry observed in the Universe require CP violation at a level several orders-of-magnitude larger than that needed to account for meson decays.

Finally, in Section 3.1.6, we noted that a complete discussion of neutrino oscillations must incorporate mixing between all three neutrino states, rather than just two. The most general form of this mixing can again be written in the standard form (7.83), i.e.

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{13}s_{23}e^{i\delta} & c_{12}c_{23} - s_{12}s_{13}s_{23}e^{i\delta} & c_{13}s_{23} \\ s_{12}s_{23} - c_{12}s_{13}c_{23}e^{i\delta} & -c_{12}s_{23} - s_{12}s_{13}c_{23}e^{i\delta} & c_{13}c_{23} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}, \quad (7.86)$$

where a value of the phase δ other than 0 or π will give rise to CP violation and differences in the behaviour of neutrinos and antineutrinos. The experimental values of the mixing angles θ_{12} , θ_{13} , and θ_{23} are now reasonably well determined by the neutrino oscillation data discussed in Section 3.1.5 and are given by (3.42), but the value of the phase δ is still

undetermined. It will hopefully be measured in future long-baseline oscillation experiments using beams of neutrinos and antineutrinos produced in accelerators, and a value other than 0 or π will imply CP violation in the neutrino sector. However, such effects, if they exist, are unlikely to be detected in leptonic decays, since their contributions will be proportional to the very small neutrino masses.

Problems 7

- 7.1** The intensity of the electrons emitted in the decay (7.1) of polarised cobalt-60 nuclei is found to be consistent with the form

$$I(v, \theta) = 1 + \alpha \frac{v}{c} \cos \theta,$$

where v is the magnitude of the electron velocity and θ is the angle between its direction and the direction of the ^{60}Co spin. Deduce the value of the coefficient α by considering events in which the electron is emitted in the direction of the spin of the decaying nuclei. The spins of the ^{60}Co and ^{60}Ni nuclei are $J = 5$ and $J = 4$, respectively, and the energies of the emitted particles are sufficiently small that orbital angular momenta may be neglected.

- 7.2** Neglecting the electron mass, the energy spectrum for the electrons emitted in muon decay is given by

$$\frac{d\Gamma}{dE_e} = \frac{2G_F^2 (m_\mu c^2)^2 E_e^2}{(2\pi)^3 (\hbar c)^6} \left(1 - \frac{4E_e}{3m_\mu c^2} \right).$$

What is the most probable energy for the electron? Draw a diagram showing the orientation of the momenta of the three outgoing particles and their helicities for the case when $E_e \approx m_\mu c^2/2$. Show also the helicity of the muon. Integrate the energy spectrum to obtain an expression for the total decay width of the muon. Hence calculate the muon lifetime in seconds. ($G_F/(\hbar c)^3 = 1.166 \times 10^{-5} \text{ GeV}^{-2}$.)

- 7.3** Derive (7.14) for the squared momentum transfer in Möller scattering. What is the maximum value of $|q^2|$ if the laboratory beam energy is $E_L = 50 \text{ GeV}$? Estimate the resulting error on $\sin^2 \theta_W$ if the parity-violating asymmetry (7.13) is measured with an error of about 10%.
- 7.4** The reaction $e^+ e^- \rightarrow \mu^+ \mu^-$ is studied in the centre-of-mass system using colliding beams of energy $E \gg m_e c^2$. If the beams are unpolarised and the final spins are not measured, the differential cross-section is of the form

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2 (\hbar c)^2}{4s} [F(s)(1 + \cos^2 \theta) + G(s) \cos \theta],$$

where $s = 4E^2$ and θ is the centre-of-mass scattering angle. Find an expression for the forward–backward asymmetry parameter

$$A_{FB} \equiv \frac{\sigma_F - \sigma_B}{\sigma_F + \sigma_B}$$

in terms of $F(s)$ and $G(s)$, where $\sigma_F(\sigma_B)$ is the total cross-section in the forward (backward) hemisphere, i.e. $0 \leq \cos \theta \leq 1$ ($-1 \leq \cos \theta \leq 0$). Does a non-zero value of A_{FB} imply parity non-conservation?

- 7.5** The couplings of the Z^0 to right-handed (R) and left-handed (L) fermions are given by

$$g_R(f) = -q_f \sin^2 \theta_W, \quad g_L(f) = \pm 1/2 - q_f \sin^2 \theta_W,$$

where q_f is the electric charge of the fermion f in units of e and θ_W is the weak mixing angle. The positive sign in g_L is used for neutrinos and the $q = u, c, t$ quarks; the negative sign is used for charged leptons and the $q = d, s, b$ quarks. If the partial width for $Z^0 \rightarrow f \bar{f}$ is given by

$$\Gamma_f = \frac{G_F M_Z^3 c^6}{3\pi \sqrt{2} (\hbar c)^3} [g_R^2(f) + g_L^2(f)],$$

calculate the partial widths to neutrinos Γ_ν and to $q\bar{q}$ pairs Γ_q and explain the relation of Γ_q to the partial width to hadrons Γ_{hadron} . Assume $\sin^2 \theta_W = 1/4$.

The widths to hadrons and to charged leptons are measured to be $\Gamma_{\text{hadron}} = (1744 \pm 2)$ MeV and $\Gamma_{\text{lepton}} = (251.9 \pm 0.3)$ MeV, and the total width to all final states is measured to be $\Gamma_{\text{total}} = (2495 \pm 2)$ MeV. Use these experimental results and your calculated value for the decay width to neutrinos to show that there are only three generations of neutrinos with masses $M_\nu < M_Z/2$.

- 7.6** Show that the existence of the decays $K^+ \rightarrow \pi^+ \pi^0$ and $K^+ \rightarrow \pi^+ \pi^+ \pi^-$ implies that parity is violated if the kaon is assumed to have spin-zero.
7.7 Show that the total decay rates for the reactions $K^0 \rightarrow \pi^- e^+ \nu_e$ and $\bar{K}^0 \rightarrow \pi^+ e^- \bar{\nu}_e$ are equal if CP is conserved.
7.8 By decomposing the $\pi^0 \pi^0$ state into components of definite isospin I , using the methods given in Appendix A.4, it can be shown that

$$\mathcal{M}(K_{S,L}^0 \rightarrow \pi^0 \pi^0) = \sqrt{\frac{2}{3}} \exp(i\delta_2) \mathcal{M}_{S,L}^{(2)} - \sqrt{\frac{1}{3}} \exp(i\delta_0) \mathcal{M}_{S,L}^0, \quad (\text{A})$$

where the phase factors are due to the strong interaction between the final-state pions and the superscripts (0, 2) refer to the isospin of the $\pi\pi$ state. ($I = 1$ is forbidden by the Bose statistic.) Use this result, together with expansion (7.34) to show that the ratio η_{00} defined in (7.39) may be written in the form (7.42), with

$$\varepsilon' = -i\sqrt{2} \frac{\text{Im} A_2}{A_0} \exp[i(\delta_2 - \delta_0)],$$

where

$$\mathcal{M}[K^0 \rightarrow (\pi\pi)_{0,2}] \equiv A_{0,2}, \quad (\text{B})$$

and by CPT invariance,

$$\mathcal{M}[\bar{K}^0 \rightarrow (\pi\pi)_{0,2}] \equiv A_{0,2}^*, \quad (\text{C})$$

with $A_2 \ll A_0$, and where A_0 may be taken as real. In deriving this result, neglect second-order terms in the small quantities ε and A_2 .

- 7.9** At SuperKEKB it is planned to collide 4 GeV positrons with 7 GeV electrons to form $\Upsilon(4S)$ particles. If the latter decay to B^\pm particles with

equal energies, how far will they travel on average before decaying? (The B^\pm particles have masses of $5.28 \text{ GeV}/c^2$ and lifetimes of $1.64 \times 10^{-12} \text{ s}$.)

- 7.10** Four mesons each of mass $5.28 \text{ GeV}/c^2$ are produced in a B factory and observed to decay to

$$(a) \bar{D}^0\pi^-\mu^+\nu_\mu, (b) \rho^+K^-, (c) \rho^+\pi^-, (d) D^-D_s^+,$$

where ρ^+ is the resonance with spin-parity $J^P = 1^-$ belonging to the meson supermultiplet Figure 3.15b. Is it possible to distinguish between a B^0 and a \bar{B}^0 in each case, and if so why?

- 7.11** Derive the Formulas (7.78a) and (7.78b) for CP violation in mixing.
7.12 Compare the expansions (7.62a) for mesons M_a and M_b with expansions of the form (7.34a) and (7.34b) for K_S^0 and K_L^0 , respectively, and show that in the case of no violation in mixing, (7.64c) implies $\varepsilon = i\tan\beta$.
7.13 Estimate the relative importance of the contributions of Figure 7.14a and b to the amplitude for the decay $B^0 \rightarrow K^+\pi^-$.
7.14 Explain why you might expect significant direct CP violation in $\bar{B}^0 \rightarrow \pi^+\pi^-$ decays and discuss its possible importance in $\bar{B}^0 \rightarrow D^+D^-$ decays.

8

Models and theories of nuclear physics

Nuclei are held together by the strong nuclear force between nucleons, so we start this chapter by looking at the form of this, which is more complicated than that generated by simple one-particle exchange. Much of the phenomenological evidence comes from low-energy nucleon–nucleon scattering experiments that we will simply quote, but we will interpret the results qualitatively in terms of the fundamental strong interaction between quarks. The rest of the chapter is devoted to various models and theories that are constructed to explain nuclear data in particular domains.

8.1 The nucleon–nucleon potential

The existence of stable nuclei implies that overall the net nucleon–nucleon force must be *attractive* and much stronger than the Coulomb force, although it cannot be attractive for all separations or otherwise nuclei would collapse in on themselves. So at very short ranges there must be a repulsive core. However, the repulsive core can be ignored in low-energy nuclear structure problems because low-energy particles cannot probe the short-distance behaviour of the potential. In lowest order, the potential may be represented dominantly by a central term (i.e. one that is a function only of the radial separation of the particles), although there is also a smaller noncentral part. We know from nucleon–nucleon scattering experiments¹ that the nucleon–nucleon force is *short-range*, of the same order as the size of the nucleus, and thus does not correspond to the exchange of gluons, as in the fundamental strong interaction. A schematic diagram of the resulting potential is shown in Figure 8.1. In practice,

¹There are useful reviews in, for example, Chapter 4 of Krane (1988) and Chapter 14 of Hodgson et al. (1997).

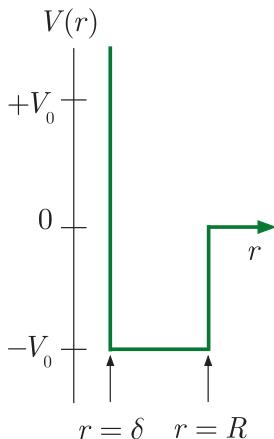


Figure 8.1 Idealised square well representation of the strong interaction nucleon–nucleon potential. The distance R is the range of the nuclear force and $\delta \ll R$ is the distance at which the short-range repulsion becomes important. The depth V_0 is approximately 40 MeV.

of course, the potential is smooth at the boundaries, and in the case of protons the strong interaction potential must be combined with the Coulomb potential.

A comparison of nn and pp scattering data (after allowing for the Coulomb interaction) shows that the nuclear force is *charge-symmetric* ($pp = nn$) and almost *charge-independent* ($pp = nn = pn$).² As we have commented in Section 3.3.1, charge-symmetry is also seen in comparisons of the energy levels of mirror nuclei (see, for example, Figure 3.13) and evidence for charge-independence comes from the energy levels of triplets of related nuclei with the same A values. Nucleon–nucleon forces are, however, *spin-dependent*. The force between a proton and neutron in an overall spin-1 state (i.e. with spins parallel) is strong enough to support a weakly bound state (called the *deuteron*), whereas the potential corresponding to the spin-0 state (i.e. spins antiparallel) has no bound states. Finally, nuclear forces *saturate*. This describes the fact that a nucleon in a typical nucleus experiences attractive interactions only with a limited number of the many other nucleons, and is a consequence of the short-range nature of the force. The evidence for this is the form of the nuclear binding energy curve in Figure 2.9, which shows that, other than for small nuclei with few nucleons present, the binding energy per nucleon is approximately independent of the number of nucleons present, giving rise to the volume term (2.51) in the semi-empirical mass formula (SEMF) (2.49).

Ideally one would like to be able to interpret the nucleon–nucleon potential in terms of the fundamental strong quark–quark interactions. It is not yet possible to give a complete explanation along these lines, but it is possible to go some way in this direction. If we draw an analogy with atomic and molecular structures, with quarks playing the role of electrons, then possibilities are: an ionic-type bond, a van der Waals type of force, or a covalent bond.³ The first can be ruled out because the confining forces are too strong to permit a quark to be ‘lent’ from one nucleon to another and the second can also be ruled out because the resulting two-gluon exchange is too weak. This leaves a covalent bond due to the sharing of single quarks between the nucleons, analogous to the covalent bond that binds the hydrogen molecule. However, nucleons have to remain ‘colourless’ during this process and so the shared quark from one nucleon has to have the same colour as the shared quark from the other nucleon. The effect of this is to reduce the effective force (because there are three possible colour states) and by itself is unable to explain the depth of the observed potential. In addition to the three (valence) quarks within

²For a discussion of these data see, for example, the references in Footnote 1 and Chapters 2 and 3 of Bertulani (2007).

³Recall from chemistry that in ionic bonding, electrons are permanently transferred between constituents to form positive and negative ions that then bind by electrostatic attraction; the van der Waals force is generated by the attraction between temporary charges induced on the constituents by virtue of slight movements of the electrons and in covalent bonding the constituents share electrons.

the nucleon, there are also present quark–antiquark pairs due to vacuum fluctuations.⁴ Hence a colourless quark–antiquark pair can also be shared between the nucleons. These pairs actually play a greater role in generating the nuclear strong interaction than single quarks. The lightest such diquarks are pions and their exchange gives the largest contribution to the attractive part of the nucleon–nucleon force and dominates at ranges above 1 fm. (See, for example, the Feynman diagrams in Figure 1.5b and c and the comments in the paragraph following (1.46).)

In principle, the short-range repulsion could be due to the exchange of heavier diquarks (i.e. heavier mesons), possibly also in different overall spin states, such as the spin-1 ρ meson and others discussed in Chapter 3. Experiment provides many suitable meson candidates, in agreement with the predictions of the quark model, and each exchange would give rise to a specific contribution to the overall nucleon–nucleon potential by analogy with the Yukawa potential resulting from the exchange of a spin-0 meson, as discussed in Chapter 1. It is indeed possible to obtain excellent fits to nucleon–nucleon scattering data in a model with several such exchanges.⁵ Thus this approach can yield a satisfactory potential model, but is semi-phenomenological only, as it requires the couplings of each of the exchanged particles to be found by fitting nucleon–nucleon scattering data. (The couplings that result broadly agree with values found from other sources.) Boson exchange models therefore cannot give a fundamental explanation of the repulsion.

The reason for the repulsion at small separations in the quark model lies in the spin dependence of the quark–quark strong interaction. We have discussed this in the context of calculating hadron masses in Section 3.3.4. When the two nucleons are very close, the wavefunction is effectively that for a 6-quark system with zero angular momentum between the quarks, i.e. a symmetric spatial wavefunction. Since the colour wavefunction is antisymmetric (recall the discussion of Chapter 5), it follows that the spin wavefunction is symmetric. However, the potential energy increases if all the quarks remain in the $L = 0$ state with spins aligned.⁶ The two-nucleon system will try to minimise its ‘chromomagnetic’ energy, but this will compete with the need to have a symmetric spin wavefunction. The optimum configuration at small separations is when one pair of quarks is in an $L = 1$ state, although the excitation energy is comparable to the decrease in chromomagnetic energy, so there will still be a net increase in energy at small separations.

Some tantalising clues exist about the role of the quark–gluon interaction in nuclear interactions, such as the small nuclear effects in deep

⁴These are the ‘sea’ quarks, which are probed in deep inelastic lepton scattering as discussed in Chapter 5.

⁵This approach is discussed in, for example, Chapter 3 of Cottingham and Greenwood (2001), Chapter 3 of Bertulani (2007), and also in the references quoted in Footnote 1.

⁶Compare the mass of the $\Delta(1232)$ resonance, where all three quark spins are aligned, to that of the lighter nucleon, where one pair of quark spins is antialigned to give a total spin of zero. This was discussed in detail in Section 3.3.4.

inelastic lepton scattering mentioned at the end of Section 5.5.3. (We will return to this question in Section 10.2.1.) There is also a considerable experimental programme in existence (for example at CEBAF, the superconducting accelerator facility at the Jefferson Laboratory, Virginia, USA, mentioned in Section 4.2.2(a)) to learn more about the nature of the strong nucleon–nucleon force in terms of the fundamental quark–gluon strong interaction, and further progress in this area may well result in the near future. Meanwhile, in the absence of a fundamental theory to describe the nuclear force, specific models and theories are used to interpret the phenomena in different areas of nuclear physics. In the remainder of this chapter we will discuss a number of such approaches.

8.2 Fermi gas model

In this model, the protons and neutrons that make up the nucleus are assumed to comprise two independent systems of nucleons, each freely moving inside the nuclear volume subject to the constraints of the Pauli principle. The potential felt by every nucleon is the superposition of the potentials due to all the other nucleons. This is assumed to be a finite-depth square well, modified by the Coulomb potential in the case of protons. A sketch of the potential wells for neutrons and protons is shown in Figure 8.2.

For a given ground-state nucleus, the energy levels will fill up from the bottom of the well. The energy of the highest level that is completely filled is called the *Fermi level* of energy E_F and has a momentum $p_F = (2ME_F)^{1/2}$, where M is the mass of the nucleon. Within the volume V , the number of states with a momentum between p and $p + dp$ is given by the *density of states factor*

$$\rho(p)dp = \frac{4\pi V}{(2\pi\hbar)^3} p^2 dp, \quad (8.1)$$

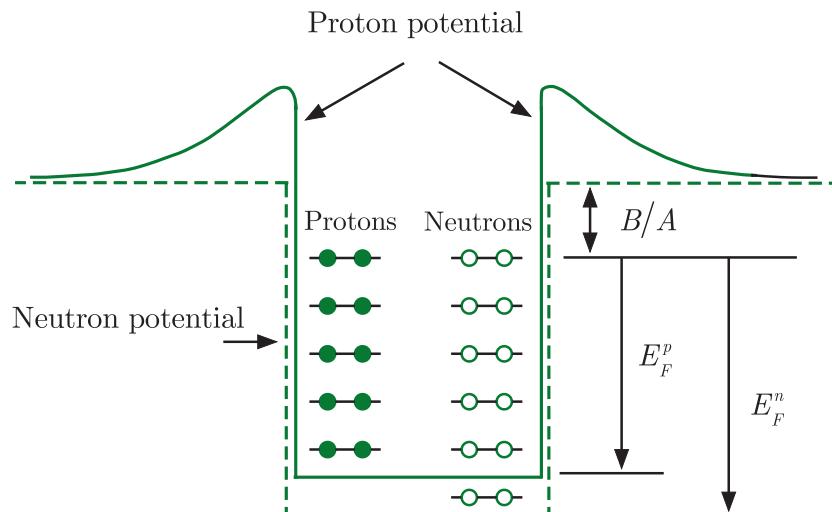


Figure 8.2 Proton and neutron potentials and states in the Fermi gas model.

which is derived in Appendix A. Since every state can contain two fermions of the same species,

$$n = 2 \int_0^{p_F} \rho(p) dp \quad (8.2a)$$

and so the numbers of neutrons and protons are given by

$$N = \frac{V(p_F^n)^3}{3\pi^2\hbar^3} \quad \text{and} \quad Z = \frac{V(p_F^p)^3}{3\pi^2\hbar^3}, \quad (8.2b)$$

respectively, with a nuclear volume

$$V = \frac{4}{3}\pi R^3 = \frac{4}{3}\pi R_0^3 A, \quad (8.3)$$

where experimentally $R_0 = 1.21$ fm, as we have seen from electron and hadron scattering experiments discussed in Section 2.2. Assuming for the moment that the depths of the neutron and proton wells are the same, we find for a nucleus with $Z = N = A/2$, the Fermi momentum

$$p_F = p_F^n = p_F^p = \frac{\hbar}{R_0} \left(\frac{9\pi}{8} \right)^{1/3} \approx 250 \text{ MeV}/c. \quad (8.4)$$

Thus the nucleons move freely within the nucleus with quite large momenta. The Fermi energy is

$$E_F = \frac{p_F^2}{2M} \approx 33 \text{ MeV}. \quad (8.5)$$

The difference between the top of the well and the Fermi level is constant for most heavy nuclei and is just the average binding energy per nucleon $\tilde{B} \equiv B/A = 7 - 8$ MeV. The depth of the potential and the Fermi energy are to a good approximation independent of the mass number A :

$$V_0 = E_F + \tilde{B} \approx 40 \text{ MeV}. \quad (8.6)$$

Heavy nuclei generally have a surplus of neutrons. Since the Fermi levels of the protons and neutrons in a stable nucleus have to be equal (otherwise the nucleus can become more stable by β decay), this implies that the depth of the potential well for the neutron gas has to be deeper than for the proton gas, as shown in Figure 8.2. Protons are therefore on average less tightly bound in nuclei than are neutrons.

We can use the Fermi gas model to give a theoretical expression for some of the dependence of the binding energy on the surplus of neutrons, as follows. Firstly, we define the average kinetic energy per nucleon as

$$\langle E_{\text{kin}} \rangle \equiv \left[\int_0^{p_F} E_{\text{kin}} p^2 dp \right] \left[\int_0^{p_F} p^2 dp \right]^{-1}. \quad (8.7)$$

Evaluating the integrals gives

$$\langle E_{\text{kin}} \rangle = \frac{3}{5} \frac{p_F^2}{2M} \approx 20 \text{ MeV}. \quad (8.8)$$

The total kinetic energy of the nucleus is then

$$E_{\text{kin}}(N, Z) = N \langle E_n \rangle + Z \langle E_p \rangle = \frac{3}{10M} [N(p_F^n)^2 + Z(p_F^p)^2], \quad (8.9)$$

which may be re-expressed as

$$E_{\text{kin}}(N, Z) = \frac{3}{10M} \frac{\hbar^2}{R_0^2} \left(\frac{9\pi}{4} \right)^{2/3} \left[\frac{N^{5/3} + Z^{5/3}}{A^{2/3}} \right], \quad (8.10)$$

where again we have taken the radii of the proton and neutron wells to be equal. This expression is for fixed A but varying N , and has a minimum at $N = Z$. Hence the binding energy gets smaller for $N \neq Z$. If we set

$$N = (A + \Delta)/2 \quad \text{and} \quad Z = (A - \Delta)/2,$$

where $\Delta \equiv N - Z$, and expand (8.10) as a power series in Δ/A , we obtain

$$E_{\text{kin}}(N, Z) = \frac{3}{10M} \frac{\hbar^2}{R_0^2} \left(\frac{9\pi}{8} \right)^{2/3} \left[A + \frac{5}{9} \frac{(N - Z)^2}{A} + \dots \right], \quad (8.11)$$

which gives the dependence on the neutron excess. The first term contributes to the volume term in the SEMF discussed in Chapter 2, while the second term describes the correction that results from having $N \neq Z$. This is a contribution to the asymmetry term we have met before in the SEMF and grows as the square of the neutron excess. Evaluating this term from (8.11) shows that its contribution to the asymmetry coefficient defined in (2.54) is about 44 MeV/c², compared to the empirical value of about 93 MeV/c² given in (2.57). In practice, to reproduce the actual term in the SEMF accurately we would have to take into account the change in the potential energy for $N \neq Z$.

8.3 Shell model

The nuclear shell model is based on the analogous model for the orbital structure of atomic electrons in atoms. In some areas it gives more detailed predictions than the Fermi gas model and it can also address questions that the latter cannot. Firstly, we recap the main features of the atomic case.

8.3.1 Shell structure of atoms

The binding energy of electrons in atoms is due primarily to the central Coulomb potential. This is a complicated problem to solve in general

because in a multi-electron atom we have to take account of not only the Coulomb field of the nucleus but also the fields of all the other electrons. Analytic solutions are not usually possible. However, many of the general features of the simplest case of hydrogen carry over to more complicated cases, so it is worth recalling the former.

Atomic energy levels are characterised by a number $n = 1, 2, 3, 4, \dots$ called the *principal quantum number*. This is defined so that it determines the energy of the system.⁷ For any n there are energy-degenerate levels with *orbital angular momentum quantum numbers* given by

$$l = 0, 1, 2, 3, \dots, (n - 1), \quad (8.12)$$

which follows from the form of the Coulomb potential, and for any value of l there are $(2l + 1)$ substates with different values of the projection of orbital angular momentum along any chosen axis (the *magnetic quantum number*):

$$m_l = -l, -l + 1, \dots, 0, 1, \dots, l - 1, l. \quad (8.13)$$

Due to the rotational symmetry of the Coulomb potential, all such substates are degenerate in energy. Furthermore, since electrons have spin $1/2$, each of the above states can be occupied by an electron with spin ‘up’ or ‘down’, corresponding to the *spin-projection quantum number*

$$m_s = \pm 1/2. \quad (8.14)$$

Again, both these states will have the same energy, so, finally, any energy eigenstate in the hydrogen atom is labelled by the quantum numbers (n, l, m_l, m_s) and, for any n , there will be n_d degenerate energy states, where

$$n_d = 2 \sum_{l=0}^{n-1} (2l + 1) = 2n^2. \quad (8.15)$$

The high degree of degeneracy can be broken if there is a preferred direction in space, such as that supplied by a magnetic field, in which case the energy levels could depend on m_l and m_s . Also, in addition to the Coulomb interaction, there is an interaction between the magnetic moment of the electron (due to its spin) and the magnetic field due to the motion of the nucleus (in the electron rest frame), which has been neglected in the above discussion. This is called the *spin-orbit interaction* and leads to corrections to the energy levels called *fine structure*, the size of which is determined by the electromagnetic fine structure constant α .

In atomic physics, the fine-structure corrections are small and so if we ignore them for the moment, in hydrogen we would have a system with electron orbits corresponding to shells of a given n , with each shell

⁷In nuclear physics we are not dealing with the same simple Coulomb potential, so it would be better to call n the *radial node quantum number*, as it still determines the form of the radial part of the wavefunction.

having degenerate subshells specified by the orbital angular momentum l . Going beyond hydrogen, we can introduce the electron–electron Coulomb interaction. This leads to a splitting in any energy level n according to the l value. The degeneracies in m_l and m_s are unchanged. It is straightforward to see that if a shell or subshell is filled, then we have

$$\sum m_s = 0 \quad \text{and} \quad \sum m_l = 0, \quad (8.16)$$

i.e. there is a strong pairing effect for closed shells. In these cases it can be shown that the Pauli principle implies

$$\mathbf{L} = \mathbf{S} = \mathbf{0} \quad \text{and} \quad \mathbf{J} = \mathbf{L} + \mathbf{S} = \mathbf{0}. \quad (8.17)$$

For any atom with a closed shell or a closed subshell structure, the electrons are paired off and thus no valence electrons are available. Such atoms are therefore chemically inert. It is straightforward to work out the atomic numbers at which this occurs. These are

$$Z = 2, 10, 18, 36, 54. \quad (8.18)$$

For example, the inert gas argon Ar($Z = 18$) has closed shells corresponding to $n = 1, 2$ and closed subshells corresponding to $n = 3, l = 0, 1$. These values of Z are called the *atomic magic numbers*.

8.3.2 Nuclear shell structure and magic numbers

In nuclear physics, there is also evidence for magic numbers, i.e. values of Z and N at which the nuclear binding is particularly strong. This can be seen from the B/A curve of Figure 8.3 (which only shows results for even values of the mass number A), where at certain values of A the data lie above the SEMF curve. The insert in Figure 8.3 shows this in more detail for small mass numbers.

The *nuclear magic numbers* are found from experiment to be

$$\begin{aligned} N &= 2, 8, 20, 28, 50, 82, 126, \\ Z &= 2, 8, 20, 28, 50, 82, \end{aligned} \quad (8.19)$$

and correspond to one or more closed shells, plus eight nucleons filling the s and p subshells of a nucleus with a particular value of n . Nuclei with both N and Z having one of these values are called *doubly magic*, and have even greater stability. They are clearly seen in Figure 8.3. An important example is the helium nucleus, the α particle.

Shell structure is also suggested by a number of other phenomena. For example, ‘magic’ nuclei have more stable isotopes than other nuclei; they have very small electric quadrupole moments, which means they are almost spherical, the most tightly bound shape; neutron capture cross-sections show sharp drops compared to neighbouring nuclei; and at magic numbers there are sharp changes in nucleon separation energies. However, to proceed further we need to know something about the effective potential.

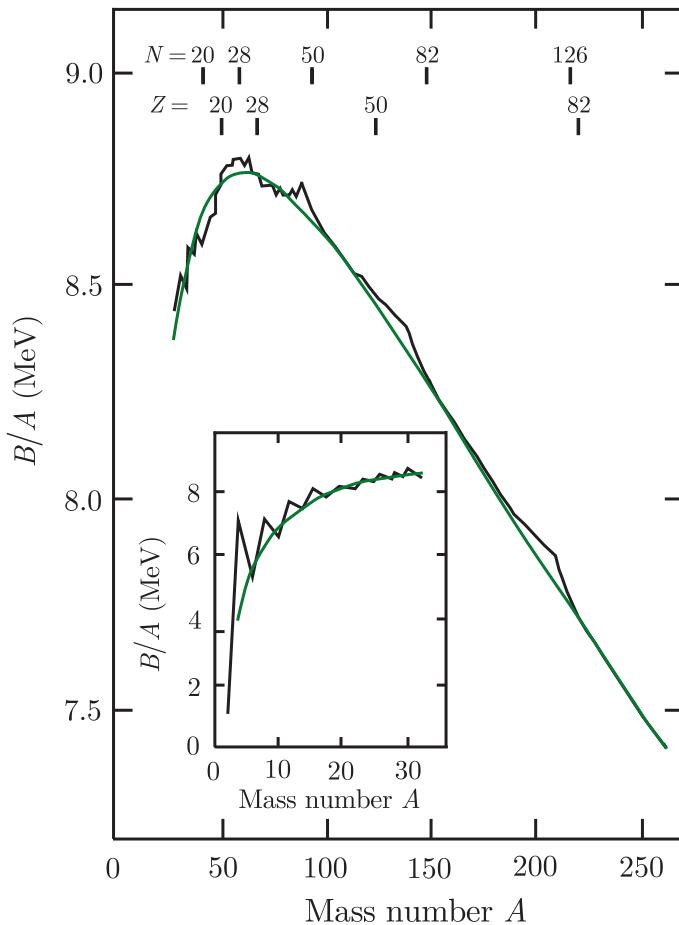


Figure 8.3 The black line shows the binding energy per nucleon for even values of A . The green curve is the fit from the semi-empirical mass formula. The inset shows the low- A region magnified.

A simple Coulomb potential is clearly not appropriate and we need some form that describes the effective potential of all the other nucleons. Since the strong nuclear force is short-ranged we would expect the potential to follow the form of the density distribution of nucleons in the nucleus. For medium and heavy nuclei, we have seen in Chapter 2 that the Fermi distribution fits the data and the corresponding potential is called the *Woods–Saxon* form

$$V_{\text{central}}(r) = \frac{-V_0}{1 + e^{(r-R)/a}}, \quad (8.20)$$

where V_0 , R , and a are constants. However, although these potentials can be shown to offer an explanation for the lowest magic numbers, they do not work for the higher ones. This is true of all purely central potentials.

The crucial step in understanding the origin of the magic numbers was taken in 1949 by Mayer and Jensen, who suggested that by analogy with atomic physics there should also be a spin-orbit term in the potential, so that the total potential is

$$V_{\text{total}} = V_{\text{central}}(r) + V_{ls}(r) (\mathbf{L} \cdot \mathbf{S}), \quad (8.21)$$

where \mathbf{L} and \mathbf{S} are the orbital and spin angular momentum operators for a single nucleon and $V_{ls}(r)$ is an arbitrary function of the radial coordinate.⁸ This form for the total potential is the same as used in atomic physics except for the presence of the function $V_{ls}(r)$. Once we have coupling between \mathbf{L} and \mathbf{S} then m_l and m_s are no longer ‘good’ quantum numbers and we have to work with eigenstates of the total angular momentum vector \mathbf{J} , defined by $\mathbf{J} = \mathbf{L} + \mathbf{S}$. Squaring this, we have

$$\mathbf{J}^2 = \mathbf{L}^2 + \mathbf{S}^2 + 2\mathbf{L} \cdot \mathbf{S}, \quad (8.22)$$

so that

$$\mathbf{L} \cdot \mathbf{S} = \frac{1}{2}(\mathbf{J}^2 - \mathbf{L}^2 - \mathbf{S}^2), \quad (8.23)$$

and hence the expectation value of $\mathbf{L} \cdot \mathbf{S}$, which we write as $\langle ls \rangle$, is

$$\langle ls \rangle = \frac{\hbar^2}{2}[j(j+1) - l(l+1) - s(s+1)] = \hbar^2 \begin{cases} l/2 & \text{for } j = l+1/2 \\ -(l+1)/2 & \text{for } j = l-1/2 \end{cases}. \quad (8.24)$$

(We are always dealing with a single nucleon, so that $s = 1/2$.) The splitting between the two levels is thus

$$\Delta E_{ls} = \frac{2l+1}{2}\hbar^2 \langle V_{ls} \rangle. \quad (8.25)$$

Experimentally, it is found that $V_{ls}(r)$ is negative, which means that the state with $j = l+1/2$ has a lower energy than the state with $j = l-1/2$. This is opposite to the situation in atoms. Also, the splittings are substantial and increase linearly with l . Hence for higher l , crossings between levels can occur. Namely, for large l , the splitting of any two neighbouring degenerate levels can shift the $j = l-1/2$ state of the initial lower level to lie above the $j = l+1/2$ level of the previously higher level.

An example of the resulting splittings up to the $1G$ state is shown in Figure 8.4, where the usual atomic spectroscopic notation has been used, i.e. levels are written nl_j , and following the historic notation used in atomic physics, the letters s, p, d, f, g, \dots are used for $l = 0, 1, 2, 3, 4, \dots$. Magic numbers occur when there are particularly large gaps between groups of levels. Note that there is no restriction on the values of l for a given n , because, unlike the atomic case, the strong nuclear potential is not Coulomb-like.

The *configuration* of a real nuclide (which of course has both neutrons and protons) describes the filling of its energy levels (subshells), for protons and for neutrons, in order, with the notation $(nl_j)^k$ for each subshell, where k is the occupancy of the given subshell. Sometimes, for brevity, the completely filled subshells are not listed, and if the highest subshell

⁸For their work on the shell structure of nuclei, Maria Goeppert-Mayer and J. Hans Jensen were awarded a half share of the 1963 Nobel Prize in Physics.

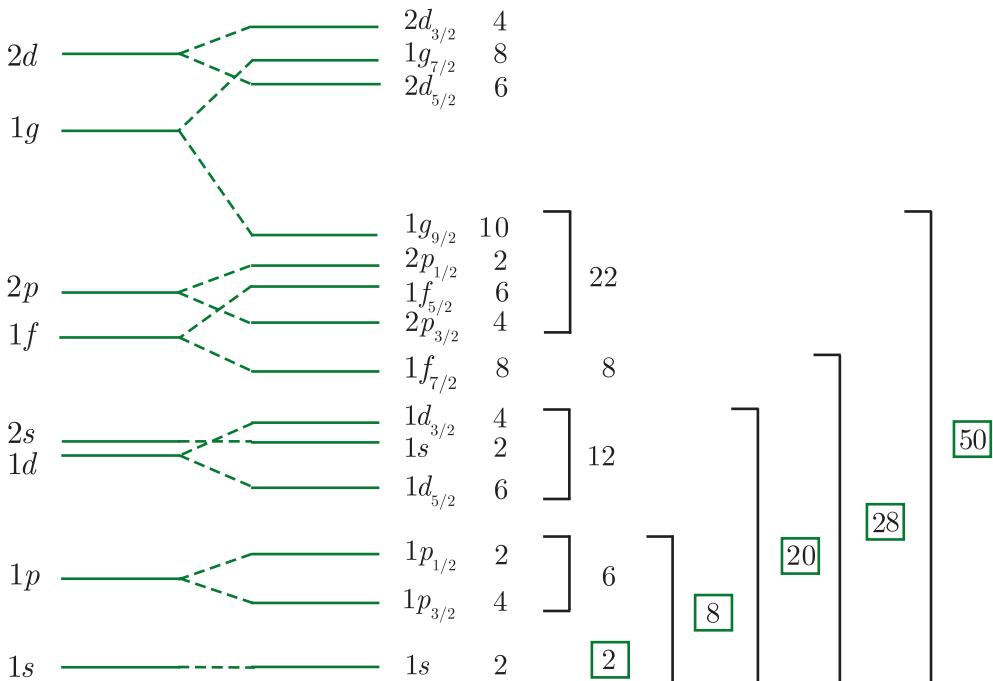


Figure 8.4 Low-lying energy levels in a single-particle shell model using a Woods–Saxon potential plus spin–orbit term. The integers in boxes correspond to nuclear magic numbers.

is nearly filled, k can be given as a negative number, indicating how far from being filled that subshell is. Using the ordering diagram above, and remembering that the maximum occupancy of each subshell is $2j + 1$, we predict, for example, the configuration for ${}^{17}_8\text{O}$ to be

$$(1s_{1/2})^2(1p_{3/2})^4(1p_{1/2})^2 \quad \text{for the protons} \quad (8.26\text{a})$$

and

$$(1s_{1/2})^2(1p_{3/2})^4(1p_{1/2})^2(1d_{5/2})^1 \quad \text{for the neutrons.} \quad (8.26\text{b})$$

Notice that all the proton subshells are filled and that all the neutrons are in filled subshells except for the last one, which is in a subshell on its own. Most of the ground-state properties of ${}^{17}_8\text{O}$ can therefore be found from just stating the neutron configuration as $(1d_{5/2})^1$.

8.3.3 Spins, parities, and magnetic dipole moments

The nuclear shell model can be used to make predictions about the *spins* of ground states. A filled subshell must have zero total angular momentum because j is always an integer-plus-a-half, so the occupancy of the subshell, $2j + 1$, is always an even number. This means that in a filled subshell, for each nucleon of a given m_j ($=j_z$) there is another having the opposite m_j . Thus the pair have a combined m_j of zero and so the complete subshell will also have zero m_j . Because this is true whatever axis we

choose for z , the total angular momentum must also be zero. Since magic-number nuclides have closed subshells, such nuclides are predicted to have a zero contribution to the nuclear spin from the neutrons or protons, or both, whichever are magic numbers. Hence magic- Z /magic- N nuclei are predicted to have zero nuclear spin. This is indeed found to be the case experimentally.

In fact, it is found that *all* even- Z /even- N nuclei have zero nuclear spin. We can therefore make the hypothesis that for ground-state nuclei, pairs of neutrons and pairs of protons in a given subshell always couple to give a combined angular momentum of zero, even when the subshell is not filled. This is called the *pairing hypothesis*. We can now see why it is the last proton and/or last neutron that determines the net nuclear spin, because these are the only ones that may not be paired up. In odd- A nuclides there is only one unpaired nucleon, so we can predict precisely what the nuclear spin will be by referring to the filling diagram, Figure 8.4. For even- A odd- Z /odd- N nuclides, however, we will have both an unpaired proton and an unpaired neutron. We cannot then make a precise prediction about the net spin because of the vectorial way that angular momenta combine; all we can say is that the nuclear spin will lie in the range $|j_p - j_n|$ to $(j_p + j_n)$.

Predictions can also be made about nuclear *parities*. Firstly, recall the following properties of parity from Chapter 1: (i) parity is the transformation $\mathbf{r} \rightarrow -\mathbf{r}$; (ii) the wavefunction of a single-particle quantum state will contain an angular part proportional to the spherical harmonic $Y_l^m(\theta, \phi)$, and under the parity transformation

$$P Y_l^m(\theta, \phi) = (-)^l Y_l^m(\theta, \phi); \quad (8.27)$$

(iii) a single-particle state will also have an *intrinsic parity*, which for nucleons is defined to be positive. Thus the parity of a single-particle nucleon state depends exclusively on the orbital angular momentum quantum number with $P = (-1)^l$. The total parity of a multiparticle state is the *product* of the parities of the individual particles. A pair of nucleons with the same l will therefore always have a combined parity of +1. The pairing hypothesis then tells us that the total parity of a nucleus is found from the product of the parities of the last proton and the last neutron. So we can predict the parity of *any* nuclide, including the odd/odd ones and these predictions are in agreement with experiment.

Unless the nuclear spin is zero, we expect nuclei to have *magnetic (dipole) moments*, since both the proton and the neutron have intrinsic magnetic moments, and the proton is electrically charged, so it can produce a magnetic moment when it has orbital motion. The shell model can make predictions about these moments. Using a notation similar to that used in atomic physics, we can write the nuclear magnetic moment as

$$\mu = g_j j \mu_N, \quad (8.28)$$

where μ_N is the *nuclear magneton* that was used in the discussion of hadron magnetic moments in Section 3.3.4, g_j is the *Landé g-factor*, and j is the nuclear spin quantum number. For brevity we can write simply

$\mu = g_j j$ nuclear magnetons. We will find that the shell model does not give very accurate predictions for magnetic moments, even for the even–odd nuclei where there is only a single unpaired nucleon in the ground state. We will not consider at all the much more problematic case of the odd–odd nuclei having an unpaired proton and an unpaired neutron.

For the even–odd nuclei, we would expect all the paired nucleons to contribute zero net magnetic moment, for the same reason that they do not contribute to the nuclear spin. Predicting the nuclear magnetic moment is then a matter of finding the correct way to combine the orbital and intrinsic components of the magnetic moment of the single unpaired nucleon. We need to combine the spin component of the moment, $g_s s$, with the orbital component, $g_l l$ (where g_s and g_l are the g -factors for spin and orbital angular momentum), to give the total moment $g_j j$. The general formula for doing this is⁹

$$g_j = \frac{j(j+1) + l(l+1) - s(s+1)}{2j(j+1)} g_l + \frac{j(j+1) - l(l+1) + s(s+1)}{2j(j+1)} g_s, \quad (8.29)$$

which simplifies considerably because we always have $j = l \pm 1/2$. Thus

$$jg_j = g_l l + g_s / 2 \quad \text{for } j = l + 1/2 \quad (8.30a)$$

and

$$jg_j = g_l j \left(1 + \frac{1}{2l+1} \right) - g_s j \left(\frac{1}{2l+1} \right) \quad \text{for } j = l - 1/2. \quad (8.30b)$$

Since $g_l = 1$ for a proton and 0 for a neutron, and g_s is approximately +5.6 for the proton and -3.8 for the neutron, (8.30a) and (8.30b) yield the results (where $g_{p,n}$ is the g -factor for nuclei with an odd proton (neutron), respectively):

$$\begin{aligned} jg_p &= l + \frac{1}{2} \times 5.6 = j + 2.3 \quad \text{for } j = l + 1/2, \\ jg_p &= j \left(1 + \frac{1}{2l+1} \right) - 5.6 \times j \left(\frac{1}{2l+1} \right) = j - \frac{2.3j}{j+1} \quad \text{for } j = l - 1/2, \\ jg_n &= -\frac{1}{2} \times 3.8 = -1.9 \quad \text{for } j = l + 1/2, \\ jg_n &= 3.8 \times j \left(\frac{1}{2l+1} \right) = \frac{1.9j}{j+1} \quad \text{for } j = l - 1/2. \end{aligned} \quad (8.31)$$

Accurate values of magnetic dipole moments are available for a wide range of nuclei, and plots of a sample of measured values for a range of odd- Z and odd- N nuclei across the whole periodic table are shown in Figure 8.5. It is seen that for a given j , the measured moments usually lie somewhere between the $j = l - 1/2$ and the $j = l + 1/2$ values (the so-called *Schmidt lines*), but beyond that the model does not predict the

⁹For a derivation, see, for example, Section 6.6 of Enge (1966).

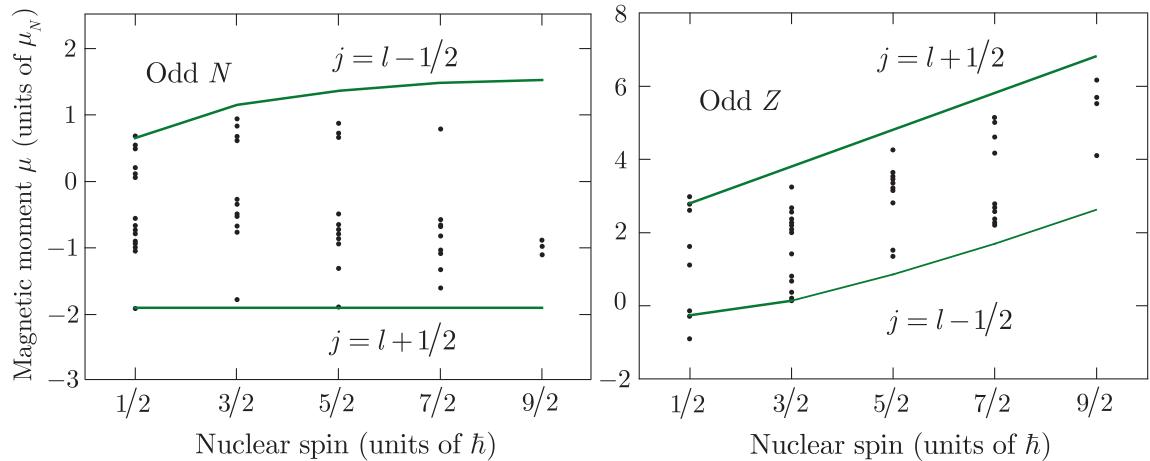


Figure 8.5 Magnetic moments for odd- N , even- Z nuclei (left diagram) and odd- Z , even- N (right diagram) as functions of nuclear spin compared to the predictions of the single-particle shell model (the Schmidt lines).

moments accurately. The only exceptions are a few low- A nuclei where the numbers of nucleons are close to magic values.

Why should the shell model work so well when predicting nuclear spins and parities, but be poor for magnetic moments? There are several likely problem areas, including the possibility that protons and neutrons inside nuclei may have effective intrinsic magnetic moments that are different to their free-particle values, because of their very close proximity to one another and the interactions that this induces.

8.3.4 Excited states

In principle, the shell model's energy level structure can be used to predict nuclear excited states. This works quite well for the first one or two excited states when there is only one possible configuration of the nucleus. However, for higher states the situation becomes very complicated because several nucleons can be excited simultaneously into a superposition of many different configurations to produce a given nuclear spin and parity. In some cases it may be necessary to take account of the quantum numbers of all nucleons in unfilled shells. When trying to predict the first one or two excited states using a filling diagram like Figure 8.4, we are looking for the configuration that is nearest to the ground-state configuration. This will normally involve *either* moving an unpaired nucleon to the next highest level *or* moving a nucleon from the subshell below the unpaired nucleon up one level to pair with it. Thus it is necessary to consider levels just above and below the last nucleons (protons and neutrons).

As an example, consider the case of $^{17}_8\text{O}$. Its ground-state configuration is given in (8.26a) and (8.26b). All the proton subshells are filled and all the neutrons are in filled subshells except for the last one, which is in a

subshell on its own. There are three possibilities to consider for the first excited state:

1. Promote one of the $1p_{1/2}$ protons to $1d_{5/2}$, giving a configuration of $(1p_{1/2})^{-1}(1d_{5/2})^1$, where the superscript -1 means that the shell is one particle short of being filled.
2. Promote one of the $1p_{1/2}$ neutrons to $1d_{5/2}$, giving a configuration of $(1p_{1/2})^{-1}(1d_{5/2})^2$.
3. Promote the $1d_{5/2}$ neutron to the next level, which is probably $2s_{1/2}$ (or the nearby $1d_{3/2}$), giving a configuration of $(1s_{1/2})^1$ or $(1d_{3/2})^1$.

Following the diagram of Figure 8.4, the third of these possibilities would correspond to the smallest energy shift, so it should be favoured over the others. The next excited state might involve moving the last neutron up a further level to $1d_{3/2}$, or putting it back where it was and adopting configurations (1) or (2). Option (2) is favoured over (1) because it keeps the excited neutron paired with another, which should have a slightly lower energy than creating two unpaired protons. Comparison of these predictions with the observed excited levels shows that the expected excited states do exist, but not necessarily in precisely the order predicted.

The shell model has many limitations, most of which can be traced to its fundamental assumption that the nucleons move independently of one another in a spherically symmetric potential. The latter, for example, is only true for nuclei that are close to having doubly filled magnetic shells and predicts zero electric quadrupole moments, whereas in practice many nuclei are deformed and quadrupole moments are often substantial. We discuss this important observation in the next section.

8.4 Nonspherical nuclei

So far we have discussed only spherical nuclei, but with nonsphericity new phenomena are allowed, including additional modes of excitation and the possibility of nonzero electric quadrupole moments.

8.4.1 Electric quadrupole moments

The charge distribution in a nucleus is described in terms of electric multipole moments and follows from the ideas of classical electrostatics. If we have a localised classical charge distribution with charge density $\rho(\mathbf{r})$, then the first moment that can be nonzero is the electric quadrupole Q ,¹⁰ defined by

$$eQ \equiv \int \rho(\mathbf{r}) (3z^2 - r^2) d^3\mathbf{r}, \quad (8.32)$$

¹⁰The electric dipole moment (EDM) $d_z = \frac{1}{e} \sum_i \int \psi^* q_i z_i \psi d^3\mathbf{r}$ will effectively vanish because it contains a sum of terms of the form $\langle \psi_i | z_i | \psi_i \rangle$, all of which are zero by parity conservation. This ignores the very small weak interaction component of nucleon interactions that violates parity. EDMs are discussed further in Section 10.5.3(a).

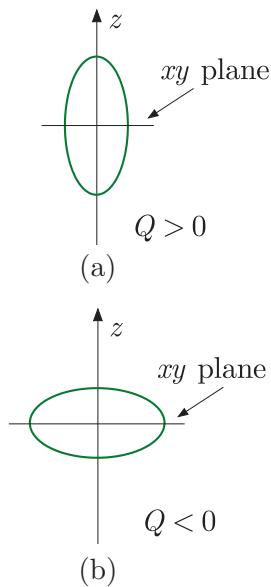


Figure 8.6 Shapes of nuclei leading to (a) $Q > 0$ (prolate) and (b) $Q < 0$ (oblate).

where we have taken the axis of symmetry to be the z axis. The analogous definition in quantum theory is

$$eQ = \sum_i \int \psi^* q_i (3z_i^2 - r^2) \psi d^3r, \quad (8.33)$$

where ψ is the nuclear wavefunction and the sum is over all relevant nucleons, each with charge q_i . The quadrupole moment is zero if the charge distribution, or $|\psi|^2$, is spherically symmetric and so a nonzero value of Q would be indicative of a nonspherical nuclear charge distribution. For example, evaluation of (8.32) using a classical ellipsoidal charge distribution with semi-axes defined as in Figure 2.17, leads to the result¹¹

$$Q_{\text{intrinsic}} = \frac{2}{5} Z (a^2 - b^2), \quad (8.34)$$

where $Q_{\text{intrinsic}}$ is the value of the quadrupole moment for an ellipsoid at rest and Ze is its total charge. For small deformations,

$$Q_{\text{intrinsic}} \approx \frac{6}{5} Z R^2 \varepsilon, \quad (8.35)$$

where ε is defined in (2.83) and R is the nuclear radius. Thus, for a prolate distribution ($a > b$), we have $Q > 0$ and for an oblate distribution ($a < b$), we have $Q < 0$, as illustrated in Figure 8.6. The same results hold in the quantum case.

If the nucleus has spin J and magnetic quantum number M , then Q will depend on M because it depends on the shape and hence the orientation of the charge distribution. The quadrupole moment is then defined as the value of Q for which M has its maximum value projected along the z axis. This may be evaluated from (8.33) in the single-particle shell model and without proof we state the resulting prediction that for odd- A , odd- Z nuclei with a single proton having a total angular momentum j outside closed subshells, the value of Q is given by

$$Q \approx -R^2 \frac{(2j-1)}{2(j+1)}. \quad (8.36)$$

Thus, $Q = 0$ for $j = 1/2$. For odd- A , odd- N nuclei with a single neutron outside closed subshells, Q is predicted to be zero because the neutron has zero electric charge, as will all even- Z , odd- N nuclei because of the pairing effect.

Unlike magnetic dipole moments, electric quadrupole moments are not always well measured and the quoted experimental errors are often larger than the differences between the values obtained in different experiments. Significant corrections also need to be made to the data to extract reliable values for the moments and this is not always done. The compilation of electric quadrupole moment data shown in Figure 8.7 is therefore

¹¹See Problem 8.5.

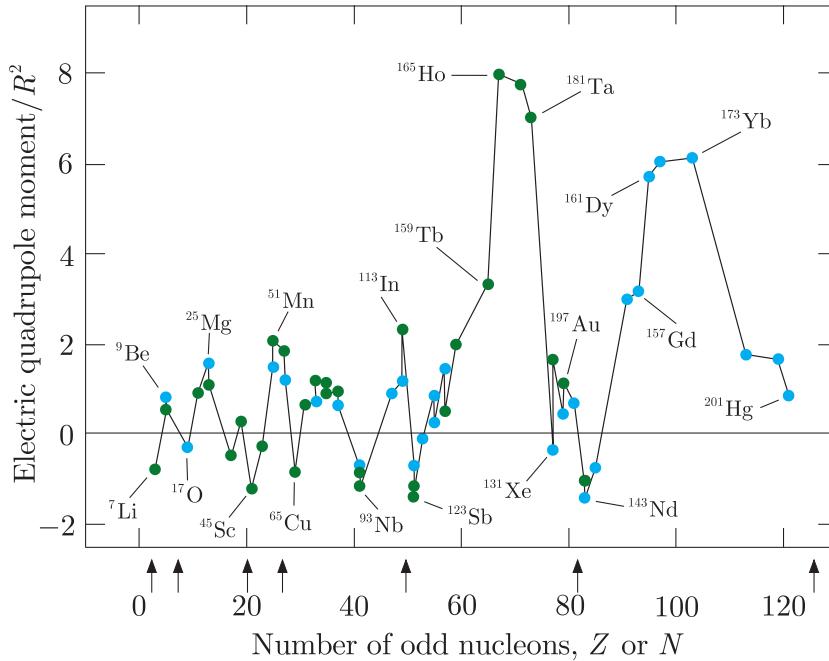


Figure 8.7 Some measured electric quadrupole moments for odd- A nuclei, normalised by dividing by R^2 , the squared nuclear radius. Blue circles denote odd- N nuclei and green circles odd- Z nuclei. The solid lines have no theoretical significance. The arrows denote the position of closed shells.

representative. The solid lines are simply to guide the eye and have no theoretical significance. The arrows indicate the positions of major closed shells. A change in sign of Q at these points is expected because a nucleus with one proton less than a closed shell behaves like a closed shell nucleus with a negatively charged proton (a proton hole) and there is some evidence for this in the data.

Two features emerge from the comparison of Figure 8.7 with the predictions of the shell model. Firstly, while odd- A , odd- Z nuclei with only a few nucleons outside a closed shell do have moments of order $-R^2$, in general the measured moments are larger by factors of 2–3 and for some nuclei the discrepancy can be as large as a factor of 10. Secondly, odd- A , odd- N nuclei also have nonzero moments, contrary to expectations, and moreover there is little difference between these and the moments for odd- A , odd- Z nuclei, except that the former tend to be somewhat smaller. These results strongly suggest that for some nuclei it is not a good approximation to assume spherical symmetry and that these nuclei must be considered to have non-spherical mass distributions.

The first attempt to explain the measured electric quadrupole moments in terms of non-spherical nuclei was due to Rainwater. His approach can be understood using the model we discussed in Section 2.7 when considering fission and used above to derive the results (8.34) and (8.35). There the sphere was deformed into an ellipsoid (see Figure 2.17) with axes parameterised in terms of a small parameter ε via (2.83). The resulting change in the binding energy ΔE_B was found to be

$$\Delta E_B = -\alpha \varepsilon^2, \quad (8.37)$$

where

$$\alpha = \frac{1}{5}(2a_s A^{2/3} - a_c Z^2 A^{-1/3})c^2 \quad (8.38)$$

and the coefficients a_s and a_c are those of the SEMF with numerical values given in (2.57). Rainwater assumed that this expression only held for closed-shell nuclei, but not for nuclei with nucleons in unfilled shells. In the latter cases he showed that the distortion gives rise to an additional term in ΔE_B that is linear in ε , so that the total change in binding energy is

$$\Delta E_B = -\alpha \varepsilon^2 - \beta \varepsilon, \quad (8.39)$$

where β is a parameter that could be calculated from the Fermi energy of the nucleus. This form has a minimum value $\beta^2/4\alpha$ when $\varepsilon = -\beta/2\alpha$. The ground state would therefore be deformed and not spherical.

The quantity Q above is the observed quadrupole moment, relative to some z axis in space. We can also define an intrinsic moment, relative to the axis of symmetry of the deformed nucleus, such as given by (8.35). A quantum calculation shows that in the ground state these are related by

$$Q = \frac{J(2J-1)}{(J+1)(2J+3)} Q_{\text{intrinsic}}, \quad (8.40)$$

where J is the total nuclear spin. This model gives values for Q that are of the correct sign, but overestimates them by typically a factor of 2. Refined variants of the model are capable of bringing the predictions into agreement with the data by making better estimates of the parameter β .

8.4.2 Collective model

The Rainwater model is equivalent to assuming an *aspherical* liquid drop, and Åge Bohr (the son of Niels Bohr) and Mottelson showed that many properties of heavy nuclei could be ascribed to the surface motion of such a drop. However, the single-particle shell model cannot be abandoned, because it explains many general features of nuclear structure. The problem was therefore to reconcile the shell model with the liquid drop model. The outcome is the *collective model*.¹²

This model views the nucleus as having a hard core of nucleons in filled shells, as in the shell model, with outer valence nucleons that behave like the surface molecules of a liquid drop. The motions of the latter introduce nonsphericity in the core that in turn causes the quantum states of the valence nucleons to change from the unperturbed states of the shell model. Such a nucleus can both rotate and vibrate and these new degrees of freedom give rise to rotational and vibrational energy levels. For example, the rotational levels are given by $E_J = J(J+1)\hbar^2/2I$, where I is the moment of inertia and J is the spin of the nucleus. The predictions of this

¹²For their development of the collective model, Åge Bohr, Ben Mottelson, and James Rainwater shared the 1975 Nobel Prize in Physics.

simple model are quite good for small J , but overestimate the energies for larger J . Vibrational modes are due predominantly to *shape oscillations*, where the nucleus oscillates between prolate and oblate ellipsoids. Radial oscillations are much rarer because nuclear matter is relatively incompressible. The energy levels are well approximated by a simple harmonic oscillator potential with spacing $\Delta E = \hbar\omega$, where ω is the oscillator frequency. In practice, the energy levels of deformed nuclei are very complicated, because there is often coupling between the various modes of excitation, but nevertheless many predictions of the collective model are confirmed experimentally.¹³

8.5 Summary of nuclear structure models

The shell model is based upon the idea that the constituent parts of a nucleus move independently. The liquid drop model implies just the opposite, since in a drop of incompressible liquid, the motion of any constituent part is correlated with the motion of all the neighbouring pairs. This emphasises the fact that models in physics have a limited domain of applicability and may be unsuitable if applied to a different set of phenomena. As knowledge evolves, it is natural to try and incorporate more phenomena by modifying the model to become more general, until we have a model with firm theoretical underpinning that describes a very wide range of phenomena, i.e. a theory. The collective model, which uses the ideas of both the shell and liquid drop models, is a step in this direction.

We will conclude this part of the chapter with a very brief summary of the assumptions of each of the nuclear models we have discussed and what each can tell us about nuclear structure.

Liquid drop model. This model assumes that all nuclei have similar mass densities, with binding energies approximately proportional to their masses, just as in a classical charged liquid drop. The model leads to the SEMF, which gives a good description of the average masses and binding energies of ground-state nuclei. It is largely classical, with some quantum mechanical terms (the asymmetry and pairing terms) inserted in an ad hoc way. Input from experiment is needed to determine the coefficients of the SEMF.

Fermi gas model. The assumption here is that nucleons move independently in a net nuclear potential. The model uses quantum statistics of a Fermi gas to predict the depth of the potential and the asymmetry term of the SEMF.

Shell model. This is a fully quantum mechanical model that solves the Schrödinger equation with a specific spherical nuclear potential. It makes

¹³The collective model is discussed in more detail, for example, in Section 2.5 of Lilley (2001) and Chapter 17 of Hodgson et al. (1997).

the same assumptions as the Fermi gas model about the potential, but with the addition of a strong spin-orbit term. It is able to successfully predict nuclear magic numbers, spins, and parities of ground-state nuclei and the pairing term of the SEMF. It is less successful in predicting magnetic moments.

Collective model. This is also a fully quantum mechanical model, but in this case the potential is allowed to undergo deformations from the strictly spherical form used in the shell model. The result is that the model can predict magnetic dipole and electric quadrupole magnetic moments with some success. Additional modes of excitation, both vibrational and rotational, are possible and lead to predictions for energy levels that are generally confirmed by experiment.

It is clear from the above that there is at present no universal nuclear model. What we currently have is a number of models and theories that have limited domains of applicability and, even within these, they are not always able to explain all the observations. For example, the shell model, while able to give a convincing account of the spins and parities of the ground states of nuclei, is unable to predict the spins of excited states with any real confidence; and of course the shell model has absolutely nothing to say about whole areas of nuclear physics phenomena. Some attempt has been made to combine features of different models, such as is done in the collective model, with some success. A more fundamental theory will require the full apparatus of many-body theory applied to interacting nucleons and some progress has been made in this direction for light nuclei, as we will mention in Chapter 10. A theory based on interacting quarks is a more distant goal.

8.6 α decay

As we saw in Section 2.4, because the ${}^4\text{He}$ nucleus is so tightly bound, many nuclei with $A > 100$ can decay by emitting an α particle in the reaction

$$(Z, A) \rightarrow (Z - 2, A - 2) + {}^4\text{He}. \quad (8.41)$$

Because this is a two-body reaction, the kinetic energy E_α of the α particle is uniquely determined by energy and momentum conservation and is given in the rest frame of the parent nucleus (A, Z) by

$$E_\alpha = \frac{1}{2}m_\alpha v_\alpha^2 = M_D \frac{Q_\alpha}{m_\alpha + M_D}, \quad (8.42)$$

where M_D is the mass of the ‘daughter’ nucleus $(Z - 2, A - 4)$, m_α and v_α are the mass and velocity of the emitted α particle, and Q_α is the Q-value (2.60). Values of E_α are typically of order a few MeV. On the other hand, lifetimes of α emitters span an enormous range, and examples are known from 10 ns to 10^{17} years. For different isotopes of a given

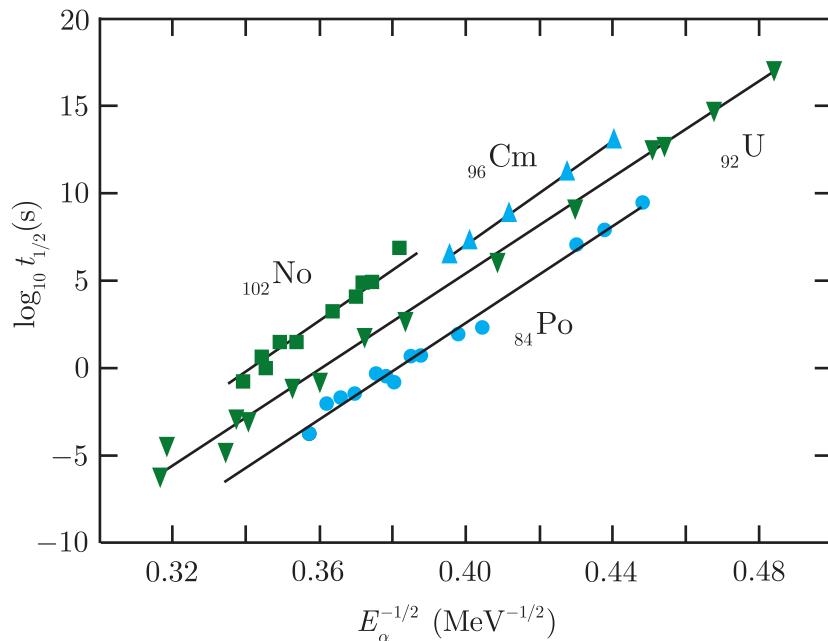


Figure 8.8 Comparison of the Geiger–Nuttall relation with experimental data for some α emitters. The lines connect sequences of isotopes.

element Z , the lifetimes are very sensitive to the value of E_α and, to a good approximation, the half-lives $t_{1/2}$ obey the *Geiger–Nuttall relation*

$$\log_{10} t_{1/2} = a + bZ E_\alpha^{-1/2}, \quad (8.43)$$

where the constant a depends on Z but b does not. This relation was found empirically by Geiger and Nuttall in 1911, long before its theoretical derivation in 1928, and is compared to experiment for four radioactive series in Figure 8.8. The Geiger–Nuttall relation is also sometimes written in the equivalent form

$$\log_{10} \lambda = c + d \log_{10} r, \quad (8.44)$$

where the decay rate $\lambda = \log_e 2 / t_{1/2}$, r is the range of the alpha particle in air, and c and d are constants for a given radioactive series.

The origin of this behaviour lies in the quantum mechanical phenomenon of *tunnelling*. Figure 8.9 shows the potential energy of an α particle as a function of r , its distance from the centre of the nucleus. Beyond the range of the nuclear force, $r > R$, the α particle feels only the Coulomb potential

$$V_C(r) = 2Z \alpha \hbar c / r, \quad (8.45)$$

where we now use Z to be the atomic number of the daughter nucleus. Within the range of the nuclear force, $r < R$, the strong nuclear potential prevails, with its strength characterised by the depth of the well. Since the α particle can escape from the nuclear potential, this implies that $E_\alpha > 0$. It is this energy that is released in the decay. Unless E_α is larger than the Coulomb barrier (in which case the decay would be so fast as to

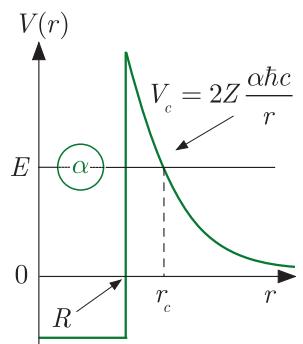


Figure 8.9 Schematic diagram of the potential energy of an α particle as a function of its distance r from the centre of the nucleus.

be unobservable) the only way the α particle can escape is by quantum mechanical tunnelling through the barrier.¹⁴

The probability T for transmission through a barrier of height V and thickness Δr by a particle of mass m_α with energy E_α is given approximately by

$$T \approx e^{-2\kappa\Delta r}, \quad (8.46)$$

where $\hbar\kappa = (2m|V_C - E_\alpha|)^{1/2}$ and m is the reduced mass of the α particle and the daughter nucleus, i.e. $m = m_\alpha m_D / (m_\alpha + m_D) \approx m_\alpha$. Using this result, we can model the Coulomb barrier as a succession of thin barriers of varying height (cf. Figure A.1). The overall transmission probability is then

$$T = e^{-G}, \quad (8.47)$$

where the *Gamow factor* G is

$$G = \frac{2}{\hbar} \int_R^{r_c} [2m|V_C(r) - E_\alpha|]^{1/2} dr. \quad (8.48)$$

This assumes that the orbital angular momentum of the α particle is zero, i.e. we ignore possible centrifugal barrier corrections.¹⁵ Since r_c is the value of r where $E_\alpha = V_C(r_c)$,

$$r_c = 2Ze^2 / 4\pi\varepsilon_0 E_\alpha \quad (8.49)$$

and hence

$$V_C(r) = 2Ze^2 / 4\pi\varepsilon_0 r = r_c E_\alpha / r. \quad (8.50)$$

Then substituting into (8.48) gives

$$G = \frac{2(2mE_\alpha)^{1/2}}{\hbar} \int_R^{r_c} \left(\frac{r_c}{r} - 1 \right)^{1/2} dr \quad (8.51)$$

and evaluating the integral (8.51) gives

$$G = 4Z\alpha \left(\frac{2mc^2}{E_\alpha} \right)^{1/2} \left[\left(\cos^{-1} \sqrt{\frac{R}{r_c}} \right) - \sqrt{\frac{R}{r_c} \left(1 - \frac{R}{r_c} \right)} \right]. \quad (8.52)$$

Finally, since E_α is typically 5 MeV and the height of the barrier is typically 40 MeV, $r_c \gg R$ and, from (8.52),

$$G \approx 4\pi\alpha Z/\beta, \quad (8.53)$$

where $\beta = v_\alpha/c$.

¹⁴Quantum mechanical tunnelling is discussed in Appendix A and the results (8.46) to (8.48) that follow are derived in Section A.1.

¹⁵The existence of an angular momentum barrier will suppress the decay rate (i.e. increase the lifetime) compared to a similar nucleus without such a barrier. Numerical estimates of the suppression factors, which increase rapidly with angular momentum, are given in Blatt and Weisskopf (1952).

The probability per unit time λ of the α particle escaping from the nucleus is proportional to the product of: (i) the probability $w(\alpha)$ of finding the α particle in the nucleus; (ii) the frequency of collisions of the α particle with the barrier, which is $v/2R$, where v is the velocity of the α particle within the nucleus¹⁶; and (iii) the transition probability. Combining these factors, λ is given by

$$\lambda = w(\alpha) \frac{v}{2R} e^{-G} \quad (8.54)$$

and since

$$G \propto \frac{Z}{\beta} \propto \frac{Z}{\sqrt{E_\alpha}}, \quad (8.55)$$

small differences in E_α have large effects on the lifetime. Taking logarithms of (8.54) and using (8.55) together with

$$\lambda = \log_e 2/t_{1/2} \quad (8.56)$$

then leads to the Geiger–Nuttall relation (8.43), where the quantity a depends on the probability $w(\alpha)$ and so is a function of the nucleus, whereas b is a constant that may be estimated from the above equations to be about 1.8. The predicted very strong variation with alpha particle energy is evident in the examples shown in Figure 8.8; changing E_α by a factor of about 2.5 changes the lifetime by 20 orders of magnitude. In all cases the agreement with the Geiger–Nuttall relation is very reasonable and the slopes are compatible with the estimate for b above. Thus the simple barrier penetration model is capable of explaining the very wide range of lifetimes of nuclei decaying by α emission.

8.7 β decay

In Section 2.6, we used the SEMF together with energy conservation to determine which isotopes of a given element were stable against β decay and to identify energetically allowed decay chains among the different isotopes, as illustrated in Figure 2.15. Here we discuss the mechanism of β decay, starting with a qualitative account of the form of the fundamental interaction.

8.7.1 $V - A$ theory

In Chapter 2, we interpreted electron β decay as the decay of a bound neutron. In general, it is possible for the internal state of the nucleus to change in other ways during the transition, but we will simplify matters by initially considering just the basic neutron decay process



A similar formalism can be used to describe electron capture.

¹⁶Note that v is not the same as v_α defined earlier.

In Figure 3.11 we gave the quark interpretation of this decay, in which the fundamental process is

$$d \rightarrow u + e^- + \bar{\nu}_e. \quad (8.58)$$

This is a low-energy process mediated by W -exchange, as shown in Figure 3.11. Hence it can be treated in the zero-range approximation, in which case the fundamental interaction Hamiltonian density, i.e. the analogue of $\psi^* V \psi$ in potential theory, can be shown to take the form

$$\frac{G_F}{\sqrt{2}} J_w(t, x) J_w^*(t, x). \quad (8.59)$$

Here $J_w(t, x)$ is the weak analogue of the familiar charge-current four-vector $J = (c\rho, \mathbf{j})$ of electromagnetism, where ρ and \mathbf{j} are the electric charge and current densities, respectively, and $J_w J_w^*$ indicates a scalar product of the two current four vectors, which are evaluated at the same point, because we are using the zero-range approximation. The current $J_w(t, x)$ is the sum of two terms, corresponding to the leptons and quarks, respectively, while G_F is the universal Fermi coupling constant (1.54), which has the same value (1.55) for leptons and quarks, provided quark mixing is taken into account.

When Fermi first formulated a theory of this type in 1934, he assumed the weak current transformed as a proper four-vector current V , like the electromagnetic interaction, giving rise to a parity conserving VV interaction in (8.59). However, with the discovery of parity violation in 1957, it was clear that this was not tenable, and it eventually became clear that both the lepton and quark components of the current were of the form $J_w = V - A$, where A indicates an axial current, which transforms in the opposite way to V under a parity transformation. Both terms have equal strengths and on substituting into (8.59) one obtains VV and AA interactions, which are parity conserving, and cross-terms VA , which give rise to parity violation. The $V - A$ form can also be shown to guarantee that if highly relativistic leptons are emitted they are in left-handed polarisation states, as we have seen in Chapter 7.

Finally, we need to take account of the fact that the quarks are bound within nucleons, which are extended objects. This can be done in a semi-empirical manner by replacing the quark component of the weak current by a nucleon component with the same transformation properties, i.e. it again has vector and axial components, and it can be shown on general principles that the coupling of the vector component V remains unchanged.¹⁷ However, the coupling of the axial current A is affected by the binding, and cannot in practice be accurately calculated, but must be determined from experiment. Its sign is unchanged, but its magnitude

¹⁷This follows from the fact that the vector current V , like the electromagnetic current, is exactly conserved, so that the associated charge is unaffected by the binding. The axial current A , in contrast, is not exactly conserved, so the details of nucleon structure would be required to calculate it theoretically.

is found to be enhanced by a factor of 1.273 ± 0.023 relative to the vector coupling, so that we no longer have a pure $V - A$ form for the nucleon current. The lepton current, of course, remains unchanged.

8.7.2 Electron and positron momentum distributions

In Appendix A.3 we derive the Second Golden Rule (A.39), which enables transition rates to be calculated provided the interaction is relatively weak. We will write the Golden Rule as

$$\omega = \frac{2\pi}{\hbar} |\mathcal{M}_{fi}|^2 \rho(E), \quad (8.60)$$

where ω is the *transition rate* (number of decays per unit time) and $\rho(E)$ is the *density of states*, i.e. the number of quantum states available to the final system per unit interval of total energy. The density-of-states factor $\rho(E)$ can be calculated from purely kinematical quantities, such as energies, momenta, masses, and spins where appropriate. The dynamics of the process is contained in the transition amplitude¹⁸ $\mathcal{M}_{fi} = \langle f | H_I | i \rangle$, where i and f denote the initial and final states. In β decay, one of the W -boson couplings is to a ud quark pair, so that its coupling strength is reduced by a factor $V_{ud} = \cos \theta_C$ when quark mixing is taken into account, as discussed in Section 6.2.2. Because of this, in what follows, it will be convenient to rewrite the transition amplitude in the form

$$\mathcal{M}_{fi} \equiv \frac{G}{V} M_{fi}, \quad (8.61a)$$

where the modified Fermi coupling

$$G = G_F \cos \theta_C \quad (8.61b)$$

and we have also abstracted an arbitrary volume V , which is used to normalise the wave functions. (It will eventually cancel out with a factor in the density-of-states term.) We will use the term *matrix element* to refer to the amplitude M_{fi} .

We see from (8.60) that the transition rate depends on kinematical factors arising through the density-of-states factor $\rho(E)$. To simplify the evaluation of this factor, we consider the neutron and proton to be ‘heavy’, so that they have negligible kinetic energy, and all the energy released in the decay process goes into creating the electron and neutrino and in giving them kinetic energy. Thus we write

$$E_0 = E_e + E_\nu, \quad (8.62)$$

where E_e is the total (relativistic) energy of the electron, E_ν is the total energy of the neutrino, and E_0 is the total energy released. The latter

¹⁸In (A.39), this is denoted by V_{fi} because the interaction Hamiltonian is denoted by a potential V .

equals $(\Delta m)c^2$, if Δm is the neutron–proton mass difference or the change in mass of the decaying nucleus.

The transition rate ω can be measured as a function of the electron momentum, so we need to obtain an expression for the spectrum of β decay electrons. Thus we will fix E_e and find the differential transition rate for decays where the electron has energy in the range E_e to $E_e + dE_e$. From the Golden Rule, this is

$$d\omega = \frac{2\pi}{\hbar} |\mathcal{M}_{fi}|^2 \rho_\nu(E_0 - E_e) \rho_e(E_e) dE_e, \quad (8.63)$$

where $\rho_e(E_e)$ and $\rho_\nu(E_0 - E_e)$ now are the density-of-states factors for the electron and neutrino, respectively. Their explicit form is derived in Appendix A.2. Integrating (A.24) over all solid angles gives

$$\rho(p_e) dp_e = \frac{V}{(2\pi\hbar)^3} 4\pi p_e^2 dp_e, \quad (8.64)$$

and changing variables using

$$\frac{dp}{dE} = \frac{E}{pc^2} \quad (8.65)$$

gives

$$\rho(E_e) dE_e = \frac{4\pi V}{(2\pi\hbar)^3 c^2} p_e E_e dE_e, \quad (8.66)$$

with a similar expression for $\rho(E_\nu)$. Using these in (8.63) and setting $\mathcal{M}_{fi} = (G/V)M_{fi}$ gives

$$\frac{d\omega}{dE_e} = \frac{G^2 |M_{fi}|^2}{2\pi^3 \hbar^7 c^4} p_e E_e p_\nu E_\nu, \quad (8.67)$$

where in general

$$p_\nu c = (E_\nu^2 - m_\nu^2 c^4)^{1/2} = [(E_0 - E_e)^2 - m_\nu^2 c^4]^{1/2}. \quad (8.68)$$

Finally, it is useful to change the variable to p_e by writing

$$\frac{d\omega}{dp_e} = \frac{dE_e}{dp_e} \frac{d\omega}{dE_e} = \frac{G^2 |M_{fi}|^2}{2\pi^3 \hbar^7 c^2} p_e^2 p_\nu E_\nu. \quad (8.69)$$

If we take the antineutrino to be precisely massless, then $p_\nu = E_\nu/c$ and (8.69) reduces to

$$\frac{d\omega}{dp_e} = G^2 |M_{fi}|^2 \frac{p_e^2 E_\nu^2}{2\pi^3 \hbar^7 c^3} = G^2 |M_{fi}|^2 \frac{p_e^2 (E_0 - E_e)^2}{2\pi^3 \hbar^7 c^3}, \quad (8.70)$$

where E_0 is the endpoint of the spectrum.

8.7.3 Selection rules

The matrix element may be written as

$$M_{fi} = \int [\psi_f^* \psi_e^* \psi_\nu^*] H \psi_i dV, \quad (8.71)$$

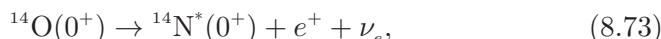
where ψ_f and ψ_i are the nuclear wavefunctions for the final and initial states, respectively, and H is the Hamiltonian. If we take the lepton wavefunctions to be free-particle plane waves, then the integral will contain a term of the form

$$\exp[-i(\mathbf{p}_e + \mathbf{p}_\nu) \cdot \mathbf{r}/\hbar]. \quad (8.72)$$

For an electron with a typical energy of 1 MeV, $p = 1.4 \text{ MeV}/c$ and $p/\hbar = 0.007 \text{ fm}^{-1}$. Thus over the nuclear volume, $pr/\hbar \ll 1$ and we can expand the exponential keeping only the first term, unity. Thus, in this approximation, M_{fi} does not depend on the electron momentum p_e and the differential transition rate therefore depends only on the kinematical variables via the density-of-states factor. Decays for which this is a good approximation are called *allowed transitions*. We consider these further below.

We need also to consider spin effects, which are absent in the simple Fermi theory described above. Since the lepton wavefunctions are constants, there can be no orbital angular momentum, i.e. the total orbital angular momentum of the leptons $\mathbf{L} = \mathbf{0}$. However, both the neutrino and the charged lepton (electron or positron) have intrinsic spins of $1/2$ (in units of \hbar). Thus their total spin \mathbf{S} can be either $\mathbf{0}$ (antiparallel) or $\mathbf{1}$ (parallel). In the antiparallel case it follows that there can be no change in the nuclear spin, i.e. $\Delta J \equiv |J_i - J_f| = 0$, and neither can there be any nuclear parity change, i.e. $\Delta P = 0$. These transitions are called *Fermi transitions* and the modern interpretation would be that they are due to the Coulomb-like part of the field generated by the exchanged W meson. In the case of parallel lepton spins, the nonzero total lepton spin implies that $\Delta J \equiv |J_i - J_f| = 0$ or 1 , with $\Delta P = 0$ still. However, the transition $0^+ \rightarrow 0^+$ is forbidden since the lepton pair carries away a unit of angular momentum. These are called *Gamow–Teller transitions* and in modern theory they would correspond to the magnetic-like part of the field due to the exchanged W meson.

An example of a pure Fermi transition is



where ${}^{14}\text{N}^*(0^+)$ is an excited state of nitrogen. An example of a pure Gamow–Teller transition is:



Mixed transitions are also possible, such as neutron decay itself:



In the latter cases, the matrix element M_{fi} is replaced by the weighted spin-averaged Fermi and Gamow–Teller matrix elements.

Because in an allowed Fermi transition, a neutron becomes a proton (or vice versa), without having an effect on the rest of the wavefunction, the selection rule for isospin I is $\Delta I = 0$. However, in a Gamow–Teller

transition, the nuclear wavefunction can change, so the isospin change is that due to a simple nucleon transition, and since the isospin of the nucleon is $1/2$ and isospin is a vector, the corresponding selection rule is $\Delta I = 0$ or 1 .

Situations can occur where the selection rules may yield a zero matrix element if only the first term (unity) in the expansion of the exponential of (8.72) is kept. In these cases, higher terms in the expansion of the exponential will be required. Each of these will contain functions of the polar angles of the coordinate \mathbf{r} and will thus correspond to the lepton pair having a nonzero orbital angular momentum. In these cases the nucleus may change its spin by more than one unit and there may also be a change of parity between the initial and final nuclear states. These decays are traditionally known as *forbidden transitions*, although they are not in fact completely forbidden. The decay rates in these cases are generally suppressed with respect to allowed decays. The nomenclature ‘first-forbidden’, ‘second forbidden’, etc. is used to correspond to $L = 1$, $L = 2$, etc. for the lepton pair.¹⁹

8.7.4 Applications of Fermi theory

We conclude the discussion of Fermi theory by considering a few of its other applications.

8.7.4(a) Kurie plots

Because the matrix element for allowed transitions does not depend on the electron kinematic variables, expression (8.70) gives rise to a bell-shaped electron momentum distribution, which rises from zero at zero momentum, reaches a peak, and falls to zero again at an electron energy equal to E_0 , as illustrated in the curve labelled $Z = 0$ in Figure 8.10. The curve ignores the effect of the electromagnetic force between the positive nucleus and the outgoing charged lepton. In the case of an emitted positron, the spectrum will be shifted to the right and conversely for an emitted electron it will be shifted to the left, with the greatest effect occurring at low lepton momenta (see Figure 8.10). The precise form of these effects is complicated to calculate, and requires quantum mechanics, but the results are published in tables of a factor $F(Z, E_e)$, called the *Fermi screening factor*, to be applied to the basic beta spectrum.

The usual way of experimentally testing the form of the electron momentum spectrum given by the Fermi theory is by means of a *Kurie plot*. From (8.70), with the Fermi screening factor included, we have

$$\frac{d\omega}{dp_e} = \frac{F(Z, E_e) G^2 |M_{fi}|^2 p_e^2 (E_0 - E_e)^2}{2\pi^3 \hbar^7 c^3}, \quad (8.76)$$

¹⁹For a discussion of forbidden transitions see, for example, Cottingham and Greenwood (2001) and Krane (1988).

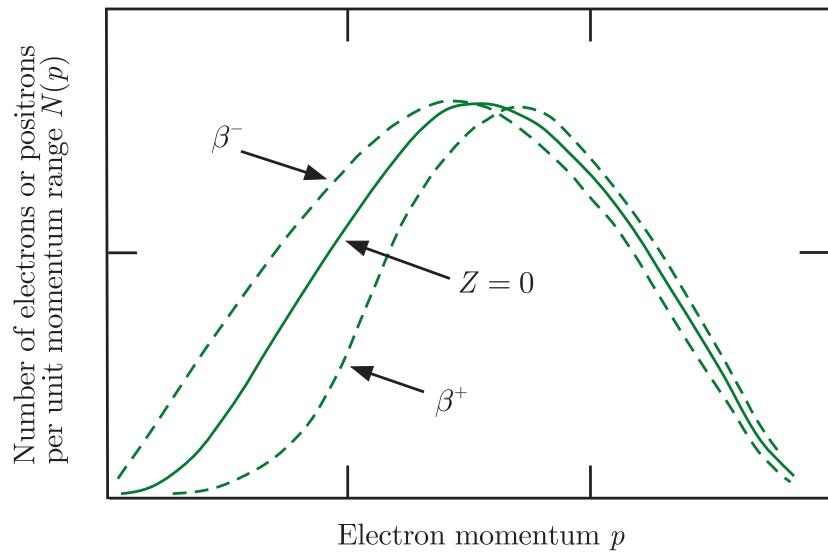


Figure 8.10 Predicted electron spectra: $Z = 0$, without the Fermi screening factor; β^\pm , with the Fermi screening factor.

which, for allowed transitions, may be written as

$$H(E_e) \equiv \left[\left(\frac{d\omega}{dp_e} \right) \frac{1}{p_e^2 F(Z, p_e)} \right]^{1/2} \propto E_0 - E_e, \quad (8.77)$$

because M_{fi} does not depend on the electron kinematic variables. A plot of the quantity $H(E_e)$ against the electron energy E_e , obtained using the measured $d\omega/dp_e$ and p_e , together with the calculated value of $F(Z, p_e)$, should then give a straight line with an intercept of E_0 . It is usual to make the plot as a function of the electron's kinetic energy $T_e = E_e - m_e c^2$ and two examples are shown in Figure 8.11. In the case of a forbidden

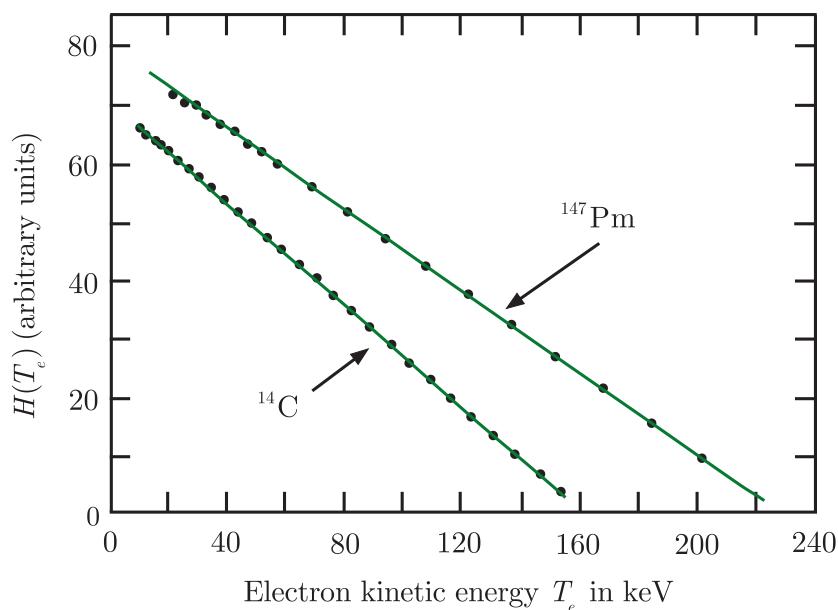


Figure 8.11 Kurie plots for the β decays of ^{14}C and ^{147}Pm . Source: Adapted from Pohl et al. (1955). Copyright (1955) the American Physical Society, reprinted with permission.

transition, the Kurie plot is not a straight line because the conditions under which (8.77) was derived no longer hold.

8.7.4(b) Mass of the electron neutrino

Studying the precise shape of the momentum distribution near its upper endpoint is one way in principle of finding the value of the electron neutrino mass. It is simplest in practice to study this via the Kurie plot. If the neutrino mass is not exactly zero then it is straightforward to repeat the derivation of (8.77) and show that the left-hand side of the Kurie plot is proportional to

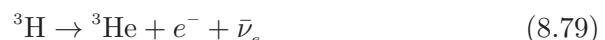
$$\{(E_0 - E_e)[(E_0 - E_e)^2 - m_\nu^2 c^4]^{1/2}\}^{1/2}, \quad (8.78a)$$

or, equivalently,

$$\{(T_0 - T_e)[(T_0 - T_e)^2 - m_\nu^2 c^4]^{1/2}\}^{1/2}, \quad (8.78b)$$

where $T_0 \equiv E_0 - m_e c^2$. If the neutrino has zero mass, then the plot of $H(T_e)$ falls to zero at the endpoint linearly, whereas (8.78b) shows that any nonzero value of the electron mass results in a very small deviation from linearity extremely close to the endpoint of the spectrum as a function of T_e , and the straight line will curve near the endpoint and cut the axis vertically at $T'_0 = T_0 - m_\nu c^2$.

In order to have the best conditions for measuring the neutrino mass, it is necessary to use a nucleus where a nonzero neutrino mass would have a maximum effect, i.e. the maximum kinetic energy of the electron should only be a few keV. Also at such low energies atomic effects have to be taken into account, so the initial and final atomic states must be very well understood. The most suitable case is the decay of tritium,



where T_0 is only 18.6 keV. The predicted Kurie plot very close to the endpoint is shown in Figure 8.12.

Since the counting rate near T_0 is vanishingly small, the experiment is extremely difficult. In practice, the expression (8.78b) is fitted to data close to the endpoint of the spectrum and extrapolated to T_0 . The best experiments are consistent with a zero neutrino mass, but when experimental and theoretical uncertainties are taken into account, an upper limit of about $2 \text{ eV}/c^2$ results, as we have remarked in Section 3.1.3. To reduce this bound further would require a better knowledge of the nuclear wavefunctions and, crucially, the value of T_0 , or equivalently the (${}^3\text{H} - {}^3\text{He}$) mass difference. For example, to reduce the bound to $0.2 \text{ eV}/c^2$ would require the mass difference to be known with a precision of 1 part in 10^{11} , which is close to the limit that current Penning trap measurements can achieve.

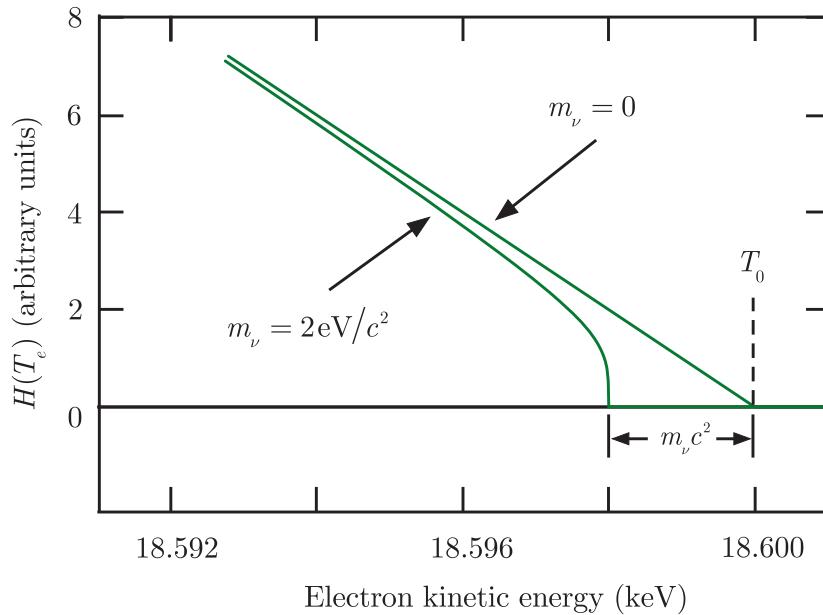


Figure 8.12 Predicted Kurie plot for tritium decay very close to the end point of the electron energy spectrum for two cases: $m_\nu = 0$ and $m_\nu = 2 \text{ eV}/c^2$.

8.74(c) Total decay rates

The total decay rate follows from integrating expression (8.76). For allowed transitions, the matrix element may be taken outside the integral and the total decay rate λ is given by

$$\lambda = \frac{G^2 |M_{fi}|^2}{2\pi^3 \hbar^7 c^3} \int_0^{p_{\max}} F(Z, p_e) p_e^2 (E_0 - E_e)^2 dp_e, \quad (8.80)$$

where p_{\max} is the maximum value of p_e . The integral depends only on E because $cp_{\max} = \sqrt{(E_0^2 - m_e^2 c^4)}$. Therefore, if we define the dimensionless integral

$$f(Z, E) \equiv \frac{1}{(m_e c)^3 (m_e c^2)^2} \int_0^{p_{\max}} F(Z, p_e) p_e^2 (E_0 - E_e)^2 dp_e, \quad (8.81a)$$

we have

$$\lambda = \frac{G^2 m_e^5 c^4}{2\pi^3 \hbar^7} |M_{fi}|^2 f(Z, E_0), \quad (8.81b)$$

and since $\lambda = \ln 2/t_{1/2}$, where $t_{1/2}$ is the half-life, finally

$$ft_{1/2} = \frac{(2 \ln 2) \pi^3 \hbar^7}{G^2 m_e^5 c^4} \cdot \frac{1}{|M_{fi}|^2}. \quad (8.82)$$

The value of $ft_{1/2}$, called the *comparative half-life*, is thus a direct measure of the matrix element $|M_{fi}|$. In practice $ft_{1/2}$ values vary enormously

and so it is usual to compare values of $\log_{10}(ft_{1/2})$, where $t_{1/2}$ is measured in seconds. There is considerable overlap in the comparative half-lives corresponding to the various transitions. Thus $\log_{10}(ft_{1/2})$ spans the approximate ranges: 5.5 ± 1.5 for allowed transitions and 7.5 ± 1.5 for first forbidden transitions. In addition, there are some decays with $\log_{10}(ft_{1/2})$ values in the range 3–4. These are called *superallowed transitions*.

Conversely, for decays where the matrix element can be reliably calculated, the measured transition rate can be used to determine the coupling constant G . For allowed transitions, $|M_{fi}|$ is independent of the electron momentum, and for allowed Fermi transitions they are also spin-independent. Finally, while $|M_{fi}|$ in general depends on nuclear wavefunctions, which are usually difficult to calculate precisely, for superallowed Fermi transitions the parent and daughter belong to the same isospin multiplet, and to a very good approximation the neutron converts to a proton with no change in the nuclear wavefunction, giving $|M_{fi}| = 1$. A precise study of 20 superallowed Fermi decays including (8.73) (Hardy and Towner 2009), when combined with the known value of G_F^{20} using (8.61b), yields a value

$$V_{ud} = \cos \theta_C = 0.97425 \pm 0.00022$$

for the ud element of the CKM mixing matrix (6.40).

Finally, we consider the dependence of decay rates of allowed transitions on the maximum energy E_0 of the emitted electrons. In order to obtain a simple approximate result, we shall ignore the Fermi screening factor $F(Z, E_e)$ and use the relativistic approximation $p_e c \approx E_e$, which is a reasonable first approximation for many, but not all decays. Equation (8.80) then reduces to

$$\lambda = \frac{G^2 |M_{fi}|^2 E_0^5}{60\pi^3 \hbar^7 c^6}, \quad (8.83)$$

so that, if we ignore the variation of the matrix element for allowed decays, λ is approximately proportional to the fifth power of the disintegration energy. This is often referred to as *Sargent's Rule*²¹ and is similar to the much more precise results obtained for the β decays (3.3a) and (3.3b) of the muon and tauon, respectively, in Section 3.1.2. The above argument, however, is at best qualitative, and in practice better agreement is obtained if E_0 is replaced by the maximum kinetic energy $T_0 \equiv E_0 - m_e c^2$. This is illustrated in Figure 8.13, which shows the decay rates against T_0 on a log–log plot for a number of allowed decays with $T_0 > 1$ MeV, together with the corresponding results for some strangeness conserving hadronic decays, like $\pi^- \rightarrow \pi^0 e^- \bar{\nu}_e$, $\Sigma^- \rightarrow \Lambda e^- \bar{\nu}_e$, and $\Sigma^+ \rightarrow \Lambda e^+ \nu_e$.

²⁰This is most precisely measured in muon decay (3.3a), which is a purely leptonic decay and so free of hadronic and nuclear complications.

²¹In 1930, Sargent observed a number of β decays and found that their rates were approximately proportional to E_0^n , where n was approximately 5.

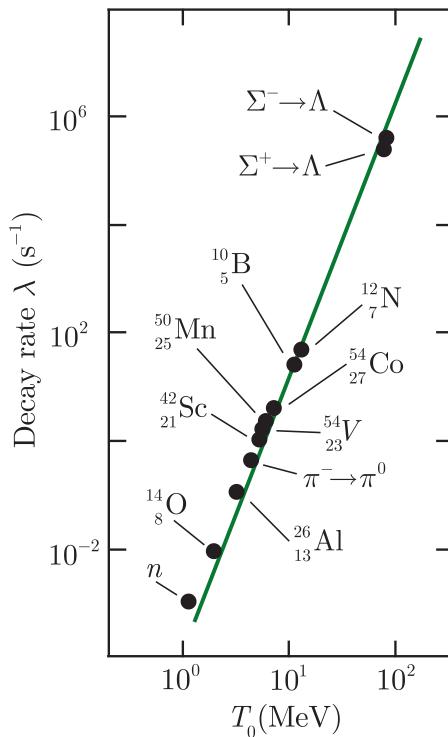


Figure 8.13 Decay rates for various β decays as a function of T_0 . The line is the prediction of Sargent's Rule (8.85), which predicts a slope of 5. Source: Adapted from Williams (1991), by permission of Oxford University Press.

8.8 γ decay

In Chapter 2 we mentioned that excited states of nuclei frequently decay to lower states (usually the ground state) by the emission of photons in the energy range appropriate to γ rays and that in addition it is possible for the nucleus to de-excite by ejecting an electron from a low-lying atomic orbit. We shall discuss this only briefly because a proper treatment requires using a quantised electromagnetic radiation field and is beyond the scope of this book. Instead, we will outline the results, without proof.

8.8.1 Selection rules

Gamma emission is a form of electromagnetic radiation and, like all such radiation, is caused by a changing electric field inducing a magnetic field. There are two possibilities, called electric (E) radiation and magnetic (M) radiation. These names derive from the semiclassical theory of radiation, in which the radiation field arises because of the time variation of charge and current distributions. The classification of the resulting radiation is based on the fact that total angular momentum and parity are conserved in the overall reaction, the latter because it is an electromagnetic process.

The photon carries away a total angular momentum, given by a quantum number L , which must include the fact that the photon is a spin-1 vector meson. The minimum value is $L = 1$. This is because a real photon has two possible polarisation states corresponding, for example, to

Table 8.1 Selection rules for γ emission

Multipolarity	Dipole		Quadrupole		Octupole	
Type of radiation	E1	M1	E2	M2	E3	M3
L	1	1	2	2	3	3
ΔP	Yes	No	No	Yes	Yes	No

$L_z = \pm 1$. Thus in the transition there must be a change in L_z of ± 1 for the emitting nucleus and this cannot happen if $L = 0$. Hence, if the spins of the initial and final nuclei states are denoted by \mathbf{J}_i and \mathbf{J}_f , respectively, the transition $\mathbf{J}_i = \mathbf{0} \rightarrow \mathbf{J}_f = \mathbf{0}$ is strictly forbidden. In general, the photons are said to have a multipolarity L and we refer to ‘multipole radiation’; transitions are called dipole ($L = 1$), quadrupole ($L = 2$), octupole ($L = 3$), etc. Thus, for example, M2 stands for magnetic quadrupole radiation. The allowed values of L are restricted by the conservation equation relating the photon total angular momentum \mathbf{L} and the spins of the initial and final nuclear states, i.e.

$$\mathbf{J}_i + \mathbf{J}_f + \mathbf{L}. \quad (8.84)$$

Thus, L may lie in the range

$$J_i + J_f \geq L \geq |J_i - J_f|. \quad (8.85)$$

It is also necessary to take account of parity. In classical physics, an electric dipole $q\mathbf{r}$ is formed by having two equal and opposite charges q separated by a distance \mathbf{r} . It therefore has negative parity under $\mathbf{r} \rightarrow -\mathbf{r}$. Similarly, a magnetic dipole is equivalent to a charge circulating with velocity \mathbf{v} to form a current loop of radius \mathbf{r} . The magnetic dipole is then of the form $q\mathbf{r} \times \mathbf{v}$, which does not change sign under a parity inversion and thus has positive parity. The general result, which we state without proof, is that electric multipole radiation has parity $(-1)^L$, whereas magnetic multipole radiation has parity $(-1)^{L+1}$. We thus are led to the selection rules for γ emission shown in Table 8.1.

Using this table we can determine which radiation types are allowed for any specific transition. Some examples are shown in Table 8.2.

Table 8.2 Examples of nuclear electromagnetic transitions

$J_i^{P_i}$	$J_f^{P_f}$	ΔP	L	Allowed transitions
0^+	0^+	No	—	None
$1/2^+$	$1/2^-$	Yes	1	E1
1^+	0^+	No	1	M1
2^+	0^+	No	2	E2
$3/2^-$	$1/2^+$	Yes	1, 2	E1, M2
2^+	1^+	No	1, 2, 3	M1, E2, M3
$3/2^-$	$5/2^+$	Yes	1, 2, 3, 4	E1, M2, E3, M4

Although transitions $\mathbf{J}_i = \mathbf{0} \rightarrow \mathbf{J}_f = \mathbf{0}$ are forbidden if the photon is a real particle, such transitions could occur if a virtual photon is emitted and absorbed, provided parity does not change. This is because a virtual photon does not have the restriction on its states of polarisation that a real photon does. In practice, the energy of the virtual photon can be transferred to an orbital atomic electron that can thereby be ejected. This is the process of *internal conversion*. There is another possibility whereby the virtual photon can create an internal e^+e^- pair. This is referred to as *internal pair production*. We will only discuss gamma radiation.

8.8.2 Transition rates

In semi-classical radiation theory, the transition probability per unit time, i.e. the emission rate, is given by²²

$$T_{fi}^{E,M}(L) = \frac{1}{4\pi\varepsilon_0} \frac{8\pi(L+1)}{L[(2L+1)!!]^2} \frac{1}{\hbar c} \left(\frac{E_\gamma}{\hbar c} \right)^{2L+1} B_{fi}^{E,M}(L), \quad (8.86)$$

where E_γ is the photon energy, E and M refer to electric and magnetic radiation, and for odd- n , $n!! \equiv n(n-2)(n-4)\dots 3.1$. The function $B_{fi}^{E,M}(L)$ is called the *reduced transition probability* and contains all the nuclear information. It is essentially the square of the matrix element of the appropriate operator causing the transition and producing photons with multipolarity L , taken between the initial and final nuclear wavefunctions. For electric transitions, B is measured in units of $e^2 \text{ fm}^{2L}$ and for magnetic transitions in units of $(\mu_N/c)^2 \text{ fm}^{2L-2}$, where μ_N is the nuclear magneton.

To go further requires knowledge of the nuclear wavefunctions. An approximation due to Weisskopf²³ is based on the single-particle shell model. This approach assumes that the radiation results from the transition of a single proton from an initial orbital state of the shell model to a final state of zero angular momentum. In this model the general formulas reduce to

$$B^E(L) = \frac{e^2}{4\pi} \left(\frac{3R^L}{L+3} \right)^2 \quad \text{for electric radiation} \quad (8.87a)$$

and

$$B^M(L) = 10 \left(\frac{\hbar c}{m_p c^2 R} \right)^2 B^E(L) \quad \text{for magnetic radiation}, \quad (8.87b)$$

where R is the nuclear radius and m_p is the mass of the proton. Finally, from the work in Chapter 2 on nuclear sizes, we can substitute $R = R_0 A^{1/3}$, with $R_0 = 1.21 \text{ fm}$, to give

$$B^E(L) = \frac{e^2}{4\pi} \left(\frac{3}{L+3} \right)^2 (R_0)^{2L} A^{2L/3} \quad (8.88a)$$

²²See, for example, Chapter 16 of Jackson (1975) and Chapter 5 of Burcham and Jobes (1995).

²³See, for example, Blatt and Weisskopf (1952).

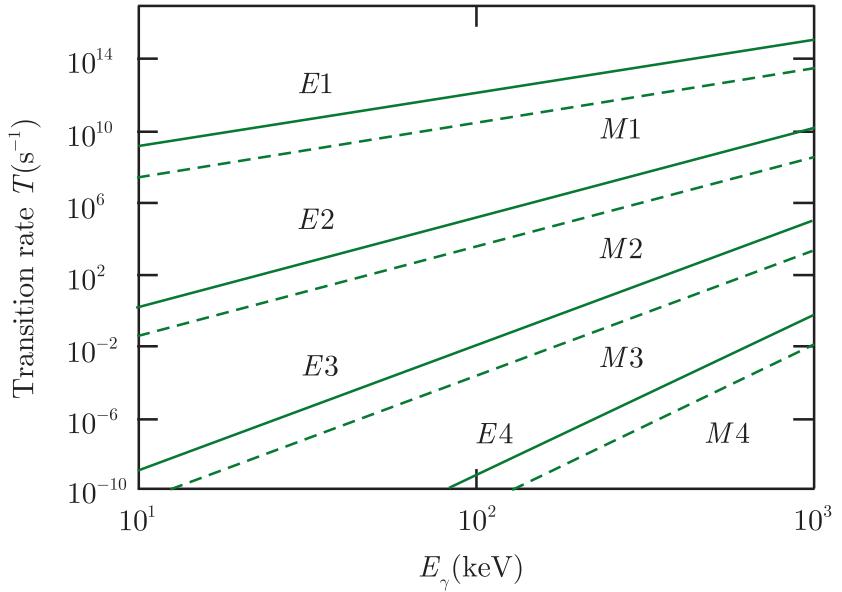


Figure 8.14 Transition rates using the single-particle shell model formulas of Weisskopf as a function of photon energy for a nucleus of mass number $A = 60$.

and

$$B^M(L) = \frac{10}{\pi} \left(\frac{e\hbar c}{2m_p c^2} \right)^2 \left(\frac{3}{L+3} \right)^2 (R_0)^{2L-2} A^{(2L-2)/3}. \quad (8.88b)$$

Figure 8.14 shows an example of the transition rates $T^{E,M}$ calculated from (8.86) using the approximations (8.88a) and (8.88b). Although these are only approximate predictions, they do confirm what is observed experimentally: that for a given transition there is a very substantial decrease in decay rates with increasing L and that electric transitions have decay rates about two orders of magnitude larger than the corresponding magnetic transitions.

Finally, it is often useful to have simple formulas for *radiative widths* Γ_γ . These follow from (8.86), (8.87a), (8.87b), (8.88a) and (8.88b) and for the lowest multipole transitions may be written as

$$\begin{aligned} \Gamma_\gamma(E1) &= 0.068 E_\gamma^3 A^{2/3}; \quad \Gamma_\gamma(M1) = 0.021 E_\gamma^3; \\ \Gamma_\gamma(E2) &= (4.9 \times 10^{-8}) E_\gamma^5 A^{4/3}, \end{aligned} \quad (8.89)$$

where Γ_γ is measured in eV, the transition energy E_γ is measured in MeV, and A is the mass number of the nucleus. These formulas are based on the single-particle approximation and in practice collective effects often give values that are greater than those predicted by (8.89).

Problems 8

- 8.1** Assume that in the shell model the nucleon energy levels are ordered as shown in Figure 8.4. Write down the shell model configuration of the nucleus ${}^7\text{Li}$ and hence find its spin, parity, and magnetic moment

- (in nuclear magnetons). Give the two most likely configurations for the first excited state, assuming that only protons are excited.
- 8.2** A certain odd-parity shell-model state can hold up to a maximum of 16 nucleons. What are its values of j and l ?
- 8.3** The ground state of the radioisotope $^{17}_9\text{F}$ has spin-parity $j^P = 5/2^+$ and the first excited state has $j^P = 1/2^-$. By reference to Figure 8.4, suggest two possible configurations for the latter state.
- 8.4** What are the configurations of the ground states of the nuclei $^{93}_{41}\text{Nb}$ and $^{33}_{16}\text{S}$ and what values are predicted in the single-particle shell model for their spins, parities, and magnetic dipole moments?
- 8.5** Show explicitly that a uniformly charged ellipsoid at rest with total charge Ze and semi-axes defined in Figure 2.17 has a quadrupole moment $Q = 2Z(a^2 - b^2)/5$.
- 8.6** The ground state of the nucleus $^{165}_{67}\text{Ho}$ has an electric quadrupole moment $Q \approx 3.5$ b. If this is due to the fact that the nucleus is a deformed ellipsoid, use the result of Question 8.5 to estimate the sizes of its semi-major and semi-minor axes.
- 8.7** The decay $^{226}_{90}\text{Th}(0^+) \rightarrow ^{222}_{88}\text{Ra}(0^+) + \alpha$ has a Q -value of 6.451 MeV and a half-life of 30.57 min. If the frequency and probability of forming alpha particles (see (8.53)) for this decay are the same as those for the decay $^{228}_{90}\text{Th}(0^+) \rightarrow ^{224}_{88}\text{Ra}(0^+) + \alpha$, estimate the half-life for the α decay of $^{228}_{90}\text{Th}$, given that its Q -value is 5.520 MeV.
- 8.8** The $^{238}_{92}\text{U}$ is unstable and decays via alpha emission. (a) Calculate the height of the Coulomb barrier shown in Figure 8.2 for this decay. (b) What is the distance from the centre of the nuclear potential beyond which the kinetic energy of the α -particle is positive?
- 8.9** Three nuclei A, B, C are radioactive isotopes of the same element and all decay via α -emission. The half-lives $t_{1/2}$ and ranges r of the alpha particles for A and B are:

$$\text{A } (t_{1/2} = 10^3 \text{ yr}, r = 3 \text{ cm}), \quad \text{B } (t_{1/2} = 10^2 \text{ days}, r = 4 \text{ cm}).$$

If the range of the α particles from the decay of C is $r = 6 \text{ cm}$, estimate its half-life.

- 8.10** The reaction $^{34}\text{S}(p,n)^{34}\text{Cl}$ has a threshold proton laboratory energy of 6.45 MeV. Calculate nonrelativistically the upper limit of the positron energy in the β decay of ^{34}Cl , given that the mass difference between the neutron and the hydrogen atom is 0.78 MeV.
- 8.11** To determine the mass of the electron neutrino from the beta decay of tritium requires measurements of the electron energy spectrum very close to the endpoint, where there is a paucity of events (see Figure 8.12). Obtain a rough estimate of the fraction of electrons with kinetic energies within 5 eV of the endpoint by ignoring the Fermi screening factor and approximating the spectrum as a function of kinetic energy T by $d\omega/dT = T^{1/2}(T_0 - T)^2$, where T_0 is the endpoint. You may assume the integral

$$\int_{a-\varepsilon}^a x^{1/2}(a-x)^2 dx \approx \frac{1}{3}a^{1/2}\varepsilon^3, \quad \varepsilon \ll a.$$

- 8.12** Use the approximation given in Problem 8.11 for the kinetic energy spectra of β decays with very low-energy endpoints T_0 to show that in these cases the mean kinetic energy is $T_0/3$.

- 8.13** The ground state of $^{73}_{35}\text{Br}$ has $J^P = 1/2^-$ and the first two excited states have $J^P = 5/2^-$ (26.92 keV) and $J^P = 3/2^-$ (178.1 keV). List the possible γ -transitions between these levels and estimate the half-life of the $3/2^-$ state.
- 8.14** The hadrons Σ^0 and Δ^0 can both decay via photon emission: $\Sigma^0(1193) \rightarrow \Lambda(1116) + \gamma$ (branching ratio $\sim 100\%$); $\Delta^0(1232) \rightarrow n + \gamma$ (branching ratio 0.56%). If the lifetime of the Δ^0 is 0.6×10^{-23} s, estimate the lifetime of the Σ^0 . The J^P values of the hadrons are given in Figure 3.16.
- 8.15** Use the Weisskopf formulas (8.88a) and (8.88b) to calculate the radiative width $\Gamma_\gamma(E3)$ expressed in a form analogous to (8.89).

9

Applications of nuclear and particle physics

Nuclear and particle physics impinge on our everyday lives in many ways.¹ Some of these interactions directly exploit detailed knowledge of nuclear and particle physics to produce practical devices, an important example being the production of power using nuclear fission reactors. A second type of interaction uses knowledge of the subjects to better understand other related scientific questions, for example the origin of the Sun's energy and how stars are formed. A third type exploits techniques developed within nuclear and particle physics and adapts them for use in unrelated area, such as medicine. A well-known example of this type is the World Wide Web, which originated at CERN from a need to connect many physicists working on a common experiment. The resulting Internet is now ubiquitous and an essential tool of modern life. For convenience, we will refer to all these interactions as 'applications', although some of the third type are better described as 'spin-offs'. For reasons of space and relevance, we will discuss in detail only a few examples of these various types of application, specifically: fission, fusion, and biomedicine, concentrating in the latter on diagnostic imaging and the therapeutic uses of radiation. In the final section we will briefly mention some other examples, particularly of the third type, to give a flavour of the breath of applications having their origins in technology originally developed in nuclear and particle physics.

9.1 Fission

Fission was discussed in Section 2.7 in the context of the semi-empirical mass formula (SEMF), and there we showed that spontaneous fission only

¹This is literally true for nuclear physics, because the energy of the Sun has its origins in nuclear reactions. This is discussed in Section 9.2.3(b).

occurs for very heavy nuclei. In this section, we firstly extend that discussion to include induced fission and the idea of chain reactions, and then show how the latter can be used to provide controlled production of energy in fission reactors (The explosive production of energy is discussed in Section 9.3.)

9.1.1 Induced fission and chain reactions

9.1.1(a) Fissile materials

For nuclei with $A \approx 240$, the Coulomb barrier, which inhibits spontaneous fission, is between 5 and 6 MeV. If a neutron with zero kinetic energy enters a nucleus to form a compound nucleus, the latter will have an excitation energy above its ground state equal to the neutron's binding energy in that state. For example, a zero-energy neutron entering a nucleus of ^{235}U forms a state of ^{236}U with an excitation energy of 6.5 MeV. This energy is well above the fission barrier and the compound nucleus quickly undergoes fission, with decay products similar to those found in the spontaneous decay of ^{236}U . On the other hand, to induce fission in ^{238}U requires a neutron with kinetic energy of at least 1.2 MeV. The binding energy of the last neutron in ^{239}U is only 4.8 MeV and an excitation energy of this size is below the fission threshold of ^{239}U .

The difference in the binding energies of the last neutron in even- A and odd- A nuclei is given by the pairing term in the SEMF. Examination of the value of this term (see (2.55)) leads to the explanation of why the odd- A nuclei

$$^{233}_{92}\text{U}, \quad ^{235}_{92}\text{U}, \quad ^{239}_{94}\text{Pu}, \quad ^{241}_{94}\text{Pu} \quad (9.1)$$

are ‘fissile’, i.e. fission may be induced by even-zero energy neutrons, whereas the even- A (even- Z /even- N) nuclei

$$^{232}_{90}\text{Th}, \quad ^{238}_{92}\text{U}, \quad ^{240}_{94}\text{Pu}, \quad ^{242}_{94}\text{Pu} \quad (9.2)$$

require an energetic neutron to induce fission. In what follows, we will focus on uranium, since this element is the most commonly used fuel in nuclear fission reactors and is also frequently used in fission weapons.

Natural uranium consists of 99.3% ^{238}U and only 0.7% ^{235}U . The total and fission cross-sections, σ_{tot} and σ_{fission} , for neutrons incident on ^{235}U and ^{238}U are shown in Figure 9.1.

The most important features of these figures are (cf. the discussion of nuclear reactions in Section 2.9):

1. At energies below 0.1 eV, σ_{tot} for ^{235}U is much larger than that for ^{238}U and the fission fraction is large ($\sim 84\%$). (The other 16% is mainly radiative capture with the formation of an excited state of ^{236}U , plus one or more photons.)
2. In the region between 1 eV and 1 keV, the cross-sections for both isotopes show prominent peaks corresponding to neutron capture into

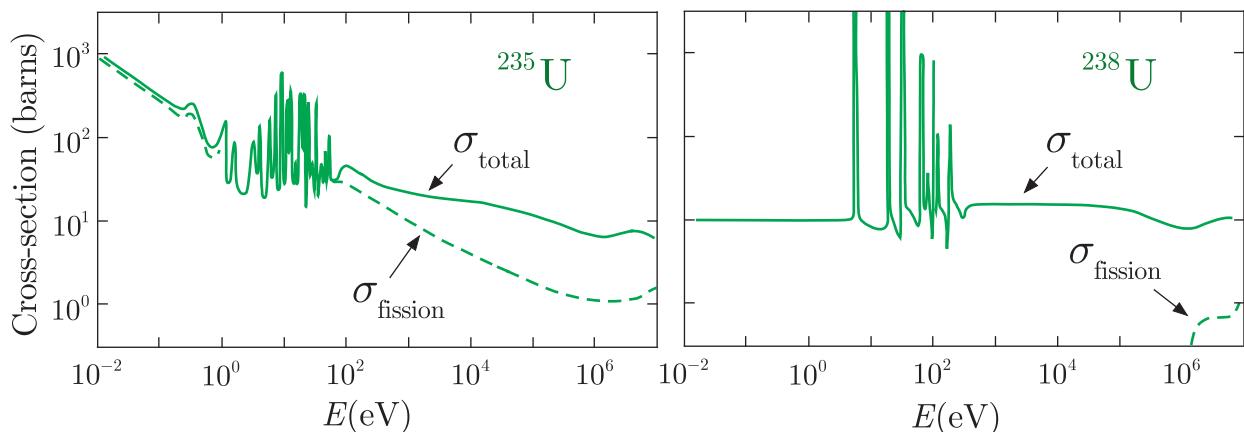


Figure 9.1 Total cross-section σ_{tot} and fission cross-section σ_{fission} as functions of energy for neutrons incident on ^{235}U (left-hand figure) and ^{238}U (right-hand figure). Source: Adapted from Garber and Kinsey (1976). Courtesy of Brookhaven National Laboratory.

resonances. The widths of these states are $\approx 0.1 \text{ eV}$ and thus their lifetimes are of order $\tau_f \approx \hbar/\Gamma_f \approx 10^{-14} \text{ s}$. In the case of ^{235}U , these compound nuclei lead to fission, whereas in the case of ^{238}U , neutron capture leads predominantly to radiative decay of the excited state.

3. Above 1 keV, the ratio $\sigma_{\text{fission}}/\sigma_{\text{tot}}$ for ^{235}U is still significant, although smaller than at very low energies. In both isotopes, σ_{tot} is mainly due to contributions from elastic scattering and inelastic excitation of the nucleus.

The fission fragments (which are not unique – several final states are possible) carry away about 180 MeV per fission, which is about 80–85% of the total energy released in the primary fission reaction. The accompanying neutrons (referred to as *prompt* neutrons) carry away only about 2.5% of the energy. For ^{235}U , the number of prompt neutrons per fission varies from 0 to 6 or more, with an average $n \approx 2.5$, the value depending a little on the incident neutron energy. Their energy varies from 0.5 MeV to more than 4 MeV, with an average of about 2 MeV.

In addition to the neutrons produced in the primary fission, the decay products will themselves decay by chains of β decays and some of the resulting nuclei will themselves give off further neutrons. This *delayed* component constitutes about 13% of the energy release in the fission of ^{235}U , about half of which is ‘lost’ through neutrino emission. Although the mean delay is about 13 seconds, some components have very long lifetimes and may not decay until many years later. This has no consequences for an explosive release of energy. However, in a reactor, one consequence is that heat will still be produced even after it has ceased to be used for power production, and another is that the delayed component may be emitted after the fuel has been used and removed from the reactor, leading to the

biological hazard of radioactive waste. We will return to these points in later sections of this chapter.²

9.1.1(b) Chain reactions

The fact that in each fission reaction a large amount of energy is produced is of course what is needed for power production. However, just as important is the fact that the fission decay products contain other neutrons. For example, we have said above that in the case of fission of ^{235}U , on average 2.5 neutrons per fission are produced. Since neutrons can induce fission, the potential exists for a sustained chain reaction, although a number of conditions have to be fulfilled for this to happen in practice. If we define the multiplication factor

$$k \equiv \frac{\text{number of neutrons produced in the } (n+1)\text{ th stage of fission}}{\text{number of neutrons produced in the } n\text{ th stage of fission}}, \quad (9.3)$$

then if $k = 1$ the process is said to be *critical* and a sustained reaction can occur. This is the ideal situation for the operation of a power plant based on nuclear fission. If $k < 1$, the process is said to be *subcritical* and the reaction will die out; if $k > 1$, the process is *supercritical* and the energy will grow very rapidly, leading to an uncontrollable explosion, i.e. a nuclear fission bomb.³

Again we will focus on uranium as the fissile material and consider the length and timescales for a chain reaction to occur. Consider firstly the length scale. If we assume that the uranium is a mixture of the two isotopes, ^{235}U and ^{238}U , with an average neutron total cross-section $\bar{\sigma}_{\text{tot}}$, then the mean free path, i.e. the mean distance the neutron travels between interactions (see Section 4.3.1), is given by

$$\bar{l} = 1/(\rho_{\text{nucl}} \bar{\sigma}_{\text{tot}}), \quad (9.4)$$

where $\rho_{\text{nucl}} = 4.8 \times 10^{28}$ nuclei/m³ is the nuclear number density of uranium metal. For example, at the average energy of a prompt neutron from fission of 2 MeV, from Figure 9.1 $\bar{\sigma}_{\text{tot}} \approx 7$ barns, so that $\bar{l} \approx 3$ cm. A 2 MeV neutron will travel this distance in about 1.5×10^{-9} s.

Next we consider the timescale, and for simplicity we will assume a sample of 100% ^{235}U . From Figure 9.1, we see that a neutron with energy of 2 MeV has a probability of about 18% to induce fission in an interaction with a ^{235}U nucleus. Otherwise it will scatter and lose energy, so that the probability for a further interaction will be somewhat increased (because the cross-section increases with decreasing energy). If the probability of inducing fission in a collision is p , the probability that a neutron has

²The production of power from nuclear reactors is discussed in Section 9.1.2, and the effect of radiation on living tissue is discussed in Section 9.4.1.

³In practice, in a fission weapon $k \geq 2$.

induced fission after n collisions is $p(1-p)^{n-1}$ and the mean number of collisions to induce fission will be

$$\bar{n} = \sum_{n=1}^{\infty} np(1-p)^{n-1} = p^{-1}, \quad (9.5)$$

provided the neutron does not escape outside the target. The value of \bar{n} can be estimated using the measured cross-sections and is about 6. Thus the neutron will move a linear (net) distance of $3\sqrt{6}$ cm \approx 7 cm in a time $t_p \approx 10^{-8}$ s before inducing a further fission and being replaced on average by 2.5 new neutrons with an average energy of 2 MeV.⁴

The above argument suggests that the critical mass of uranium ^{235}U that would be necessary to produce a self-sustaining chain reaction is a sphere of radius about 7 cm. However, not all neutrons will be available to induce fission. Some will escape from the surface and some will undergo radiative capture. If the probability that a newly created neutron induces fission is q , then each neutron will on average lead to the creation of $(nq-1)$ additional neutrons in the time t_p . If there are $N(t)$ neutrons present at time t , then at time $t+\delta t$ there will be

$$N(t + \delta t) = N(t) [1 + (nq - 1)(\delta t/t_p)] \quad (9.6)$$

neutrons and hence

$$\frac{N(t + \delta t) - N(t)}{\delta t} = \frac{N(t)(nq - 1)}{t_p}. \quad (9.7)$$

In the limit as $\delta t \rightarrow 0$, this gives

$$\frac{dN(t)}{dt} = \frac{(nq - 1)}{t_p} N(t), \quad (9.8)$$

and hence by integrating (9.8)

$$N(t) = N(0)\exp[(nq - 1)t/t_p]. \quad (9.9)$$

Thus the number increases or decreases exponentially, depending on whether $nq > 1$ or $nq < 1$. For ^{235}U , the number increases if $q > 1/n \approx 0.4$ (recall that $n \approx 2.5$). Clearly, if the dimensions of the metal are

⁴The square root appears because we are assuming that at each collision the direction changes randomly, i.e. the neutron executes a *random walk*. Thus if the distance travelled in the i th collision is \mathbf{l}_i , the displacement vector \mathbf{d} after n collisions will be $\mathbf{d} = \sum_{i=1}^n \mathbf{l}_i$ and the net distance d travelled will be given by

$$d^2 = \sum_{i=1}^n \sum_{j=1}^n (\mathbf{l}_i \cdot \mathbf{l}_j) = l_1^2 + l_2^2 + l_3^2 + \dots + l_n^2 + 2(\mathbf{l}_1 \cdot \mathbf{l}_2 + \mathbf{l}_1 \cdot \mathbf{l}_3 + \dots).$$

When the average is taken over several collisions, the scalar products will cancel because the direction of each step is random. Finally, setting $l_i = \bar{l}$, the mean distance travelled per collision gives $d = \bar{l}\sqrt{n}$.

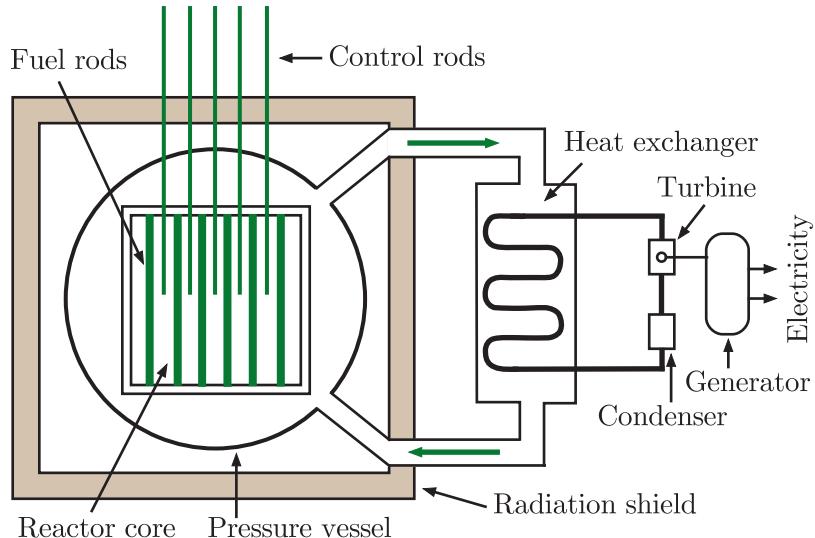


Figure 9.2 Sketch of the main elements of a thermal reactor. The components are not to scale. Source: Adapted from Lilley (2001). Copyright (2001) John Wiley & Sons, reprinted with permission.

substantially less than 7 cm, q will be small and the chain reaction will die out exponentially. However, a sufficiently large mass brought together at $t = 0$ will have $q > 0.4$. There will always be some neutrons present at $t = 0$ arising from spontaneous fission and, since $t_p \approx 10^{-8}$ s, *in principle* an explosion will occur very rapidly.⁵ For a simple sphere of ^{235}U the critical radius at which $nq = 1$ is actually close to 9 cm and the critical mass is about 52 kg. For ^{239}Pu , the critical mass is much smaller, about 10 kg.

9.1.2 Thermal fission reactors

The production of power in a controlled way for peaceful use via fission is done in a *nuclear fission reactor*. There are several distinct types of reactor available. We will concentrate on just one of these, the *thermal reactor*, which uses uranium as the fuel and low-energy neutrons to establish a chain reaction. The discussion will concentrate on the principles of operating such a reactor and not on practical details.

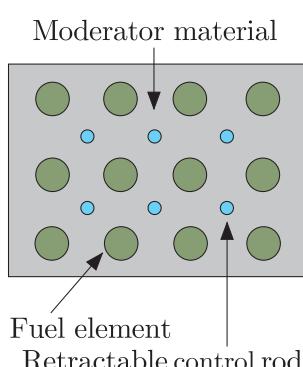


Figure 9.3 Sketch of the elements of the core of a reactor.

⁵The substantial practical difficulties of creating such an explosion are discussed in Section 9.3.1.

of the neutron will be below the threshold of 1.2 MeV for inducing fission in ^{238}U . A neutron with its energy so reduced will have to find a nucleus of ^{235}U if it is to induce fission, but its chances of doing this are very small unless its energy has been reduced to a very low value below 0.1 eV, where the cross-section is large (see Figure 9.1). Before that happens it is likely to have been captured into one of the ^{238}U resonances with the emission of photons. Thus, to sustain a chain reaction, either the fuel must be enriched with a greater fraction of ^{235}U (2–3% is common in some types of commercial reactor) or if natural uranium is to be used, some method must be devised to overcome this problem.

This is where the *moderator* comes in. This surrounds the fuel elements and its volume is much greater than that of the latter. Its main purpose is to slow down fast neutrons produced in the fission process. Fast neutrons will escape from the fuel rods into the moderator and are reduced to very low energies by elastic collisions. In this way the absorption into resonances of ^{238}U is avoided. The moderator must therefore be a material with a negligible cross-section for absorption and ideally should also be inexpensive. In practice, heavy water (a form of water where the hydrogen atoms are replaced by atoms of deuterium) or carbon (in the form of graphite) are the moderators of choice in many thermal reactors using natural uranium. For enriched reactors, ordinary water may be used.

Consider now the stability of the chain reaction. This is where the control rods play their part. They are usually made of cadmium, which has a very high absorption cross-section for neutrons. By mechanically manipulating the control rods, i.e. by retracting or inserting them, the number of neutrons available to induce fission can be regulated. This mechanism is the key to maintaining a constant k value of unity and therefore a constant power output. However, safe working of the reactor is not possible with prompt neutrons alone. To see this, we return to (9.9) and set $nq = k$, so that

$$N(t) = N(0)\exp \left[(k - 1)t/t_p \right]. \quad (9.10)$$

The value of t_p is determined by the mean free path for neutron absorption, and unlike the case of pure ^{235}U we considered in Section 9.1.1, is given approximately by $t_p \sim 10^{-3}$ s. Thus, for example, if we take $k = 1.001$, i.e. an increase of only 0.1%, the reactor flux would increase by $e^{60} \approx 10^{26}$ in only one minute. Clearly a much smaller rate of increase has to be achieved for safe manipulation of the control rods if a disaster is to be averted. This is where the delayed neutrons play a crucial role.

In an explosive reaction, the delayed neutrons are of no consequence, because the explosion will have taken place long before they would have been emitted, but in a power reactor they are vital for reactor safety. Taking account of delayed neutrons, each fission leads to $[(n + \delta n)q - 1]$ additional neutrons, where we have defined δn as the number of delayed neutrons per fission. In practice, $\delta n \sim 0.02$. In the steady-state operation,

with constant energy output, the neutron density must remain constant, i.e. $k = 1$ in (9.3). Thus q must satisfy the critical condition

$$(n + \delta n)q - 1 = 0. \quad (9.11)$$

Equation (9.10) is now modified to have an additional term that depends on the mean time τ_d of the delayed neutrons, which is about 13 s. Provided $n(k-1) \ll \delta n$, it is the latter term that dominates and, without proof, the modified form of (9.10) is given approximately by

$$N(t) \approx N(0)\exp\left\{\frac{n(k-1)t}{[\delta n - n(k-1)]t_d}\right\}. \quad (9.12)$$

Thus the timescale to manipulate the control rods is determined by that of the delayed neutrons. For example, using $n = 2.5$, $\delta n = 0.02$, $k = 1.001$, and $t_d = 13$ s in (9.12) gives an increase in reactor flux of about a factor 2 in one minute. Clearly, the precise increase is sensitive to the parameters chosen, but factors of this size are manageable. The reactor design therefore ensures that $nq - 1 < 0$ always, so that the reactor can only become critical in the presence of delayed neutrons.

This simple discussion has ignored many practical details that will modify the real formulas used in reactor dynamics, such as the fact that the fuel and moderator are not uniformly distributed throughout the core and that some of the fission products themselves have appreciable cross-sections for neutron absorption and will therefore reduce the flux of neutrons available to sustain the chain reaction.⁶

Returning to Figure 9.2, the core is surrounded by a coolant (often water), which removes the heat generated in the core from the energy deposited by the fission fragments. A thick concrete shield to prevent radiation leaks surrounds the entire setup. At start-up, the value of k is set slightly higher than unity and is kept at that value until the desired power output is achieved and the operating temperature is reached, after which the value of k is lowered by adjusting the control rods. It is very important for safety reasons that $dq/dT < 0$, so that an increase in temperature T leads to a fall in reaction rate. The rest of the plant is conventional engineering. Thus, the heated coolant gives up its heat in a heat exchanger and is used to boil water and drive a steam turbine, which in turn produces electricity.

It is worth calculating the efficiency with which one can expect to produce energy in a nuclear reactor. We can use the SEMF to calculate the energy released during fission by finding the binding energies of the two fission products and comparing their sum to the binding energy of the decaying nucleus. For the fission of a single ^{235}U nucleus this is ~ 200 MeV or 3.2×10^{-11} joules. (As we have mentioned above, about 90% of this is

⁶More details of reaction dynamics are discussed in, for example, Section 10.3 of Lilley (2001). In Section 10.6 of this reference there is also a discussion of several other types of commercial reactor.

in the form of ‘prompt’ energy.) We also know that 1 g of any element contains N_A/A atoms, where N_A is Avogardo’s number. Thus 1 g of ^{235}U has about 3×10^{21} atoms and if fission were complete would yield a total energy of about 10^{11} joules, or 1 MW-day. This is about 3×10^6 times greater than the yield obtained by burning (chemical combustion) 1 g of coal. In practice, only about 1% of the energy content of natural uranium can be extracted (the overall efficiency is also greatly reduced by the conventional engineering required to produce electricity via steam turbines), but this can be significantly increased in another type of reactor, called a *fast breeder*, discussed below.

We can also calculate the power output from an ideal thermal reactor for a given mass of uranium. From (1.57a), the reaction rate for fission W_f is given by

$$W_f = J N \sigma_{\text{fission}}, \quad (9.13)$$

where J is the flux, N is the number of nuclei undergoing fission and σ_{fission} is the fission cross-section. Consider, for example, a reactor containing 100 t of natural uranium, generating a neutron flux of $10^{13} \text{ cm}^{-2} \text{ s}^{-1}$ and with a fission cross-section for ^{235}U of 580 b at the appropriate energy (see Figure 9.1). Since the fraction of ^{235}U in natural uranium is 0.72%, the number of ^{235}U nuclei undergoing fission is given by

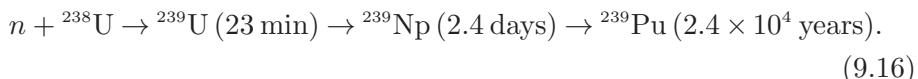
$$N = \frac{100 \times 10^3 \times 0.0072 \times N_A}{A} = 1.82 \times 10^{27}, \quad (9.14)$$

where $A = 238.03$ is the mass number of natural uranium. The power generated is thus

$$P = W_f E, \quad (9.15)$$

where $E = 200 \text{ MeV}$ is the total energy released per fission (see above). Evaluating (9.15) gives $P \approx 340 \text{ MW}$. In addition to causing fission, neutrons will be absorbed by ^{235}U without causing fission. If the total absorption cross-section σ_a is 680 b, then the number of ^{235}U nuclei that will be consumed per second will be $N J \sigma_a$, i.e. $1.24 \times 10^{19} \text{ s}^{-1}$. Since we started with 1.82×10^{27} nuclei, the fuel will be used at the rate of about 1.8% per month.

Finally, we turn briefly to the fast breeder reactor mentioned above. In this reactor there is no large volume of moderator and no large density of thermal neutrons is established. In such a reactor, the proportion of fissile material is increased to about 20% and fast neutrons are used to induce fission. The fuel used is ^{239}Pu rather than ^{235}U , the plutonium being obtained by chemical separation from the spent fuel rods of a thermal reactor. This is possible because some ^{238}U nuclei in the latter will have captured neutrons to produce ^{239}U , which subsequently decays via a chain of β -decays to plutonium. The whole sequence is, with half-lives shown in brackets,



The mean number of neutrons produced in the fission of ^{239}Pu is 2.96, so this nucleus is very suitable for use in a fast reactor. In practice, the core is a mixture of 20% ^{239}Pu and 80% ^{238}U surrounded by a blanket of ^{238}U (obtained as a byproduct of enriching uranium for use in thermal reactors and called *depleted uranium*) where more plutonium is made. Such a reactor can produce more fissile ^{239}Pu than it consumes, hence the name ‘breeder’. In principle, such a reactor can consume all the energy content of natural uranium, rather than the 1% used in thermal reactors, although in practice there are limits to its efficiency.

9.1.3 Radioactive waste

Although the production of power from nuclear fusion (to be discussed in Section 9.2.4) is being actively pursued, based on experience with fission reactors it could be several decades before a practical fusion power plant is achieved. In the shorter term the world faces the difficult problem of producing sufficient energy to sustain economic growth, particularly in the developing countries, in a way that is not harmful to the global environment. Assuming that renewable sources of energy are insufficient by themselves to fulfil the world’s increasing energy needs, power plants based on fission reactions will remain an important source of energy and a significant contribution to replacing fossil fuels in the medium term. This raises several problems, including the major one of radioactive waste.

Whatever type of conventional uranium reactor is used, radioactive waste will result, including transuranic elements and long-lived fission fragments, which in some cases may have to be stored safely for hundreds of years.⁷ Much effort has been expended on this problem, but a totally satisfactory solution is still not available. Short-lived waste with very low activity (for example, consumables such as protective clothing) is simply buried in the ground, but waste with longer-lived components cannot be treated in this way.

The waste from light water reactors, the most common type of power reactor, has two major components: the actinides, i.e. any of the series of radioactive elements with atomic numbers between 89 and 103 (mainly uranium but also smaller amounts of heavier elements, the transuranic elements like plutonium, and the minor actinides such as neptunium, americium, and curium); and fission products, which are medium-weight elements from fission processes in the nuclear fuel. The transuranic elements constitute about 1% of spent nuclear fuel. Their very long lifetimes mean that they dominate the long-term toxicity of nuclear waste and they can only be destroyed by fission. The fission fragments constitute about 4% of nuclear waste and can only be destroyed by neutron capture. While it is generally agreed that radioactive nuclei with relatively short lifetimes

⁷In principle, there would be no such problem with fast breeder reactors, but in practice the ideal is not realised.

(<100 years) can be safely stored in deep geological disposal facilities, possibly after vitrification to reduce them to physical stable forms, the same is not true of waste with very long lifetimes, some of which are water-soluble and so have the potential to contaminate ground water over very long time periods.

An ingenious idea is to ‘neutralise’ long-lived fission fragments by using the resonance capture of neutrons to convert them to short-lived, or even stable, nuclei. For example, ^{99}Tc (technetium), which concentrates in several organs of the body and also in the blood, has a very long half-life. However, it has a large resonant cross-section for neutron capture to a stable isotope ^{100}Ru (ruthenium) and in principle this reaction could be used to render ^{99}Tc harmless. Needless to say, the problems to be overcome are far from trivial. Firstly, the amount of radioactive waste is very large, so one problem is to find a source of neutrons capable of handling it. Secondly, the neutron energy has to be matched to the particular waste material, which therefore would have to be separated and prepared before being bombarded by the neutrons. All this would take energy and would increase the overall cost of energy production by nuclear power, which is already more expensive than conventional burning of fossil fuels. Nevertheless, there is considerable interest in the principle of this method and proposals have been made to exploit it without the attendant drawbacks above.

One possibility that does not use reactors themselves as the source of neutrons to effect the transmutation of long-lived waste is to use an accelerator to ‘incinerate’ nuclear waste with the aim of reducing the waste lifetimes to less than 100 years. This is referred to as an Accelerator Driven System (ADS). It would also address the two other common objections to conventional fission reactors: (i) fear of accidents at nuclear power plants, such as the one at Chernobyl in Russia in 1986, which resulted in the widespread dispersal of radioactive material into the atmosphere, and (ii) the danger of fissile material, i.e. ^{239}Pu and ^{235}U , being diverted to the production of nuclear weapons.

In one proposed scheme, uranium and most of the plutonium would be separated prior to proton irradiation and used again as reactor fuel. The most important long-lived components of the remaining waste would be isotopes of neptunium, americium, curium, and iodine, some with half-lives of 10 000 years or more. The approach would be to irradiate this material with a new source of fast neutrons produced by spallation reactions⁸ initiated using protons from a high-current accelerator. The function of the spallation process is to convert incident high-energy protons to a copious supply of low-energy neutrons. It is a complex process that consists of a series of very rapid direct reactions where a single nucleon, or small group of nucleons, is ejected from a nucleus that then interacts with nearby nuclei, leading to a cascade. After this phase, the

⁸The spallation process was mentioned in Section 4.2.3 in the context of producing neutron beams and will be again in Section 9.3 in the context of nuclear weapon design.

nucleus is left in an excited state and relaxes to its ground state by ‘boiling-off’ more nucleons, mainly neutrons. The neutron yield can be about 15 per incident nucleon. Nevertheless, it is not practical to use spallation neutrons alone because the proton current that would be required (~ 300 A) is far greater than can be realised. In practice, therefore, a subcritical mass of fissile material surrounds the spallation target to multiply the production of neutrons. This can boost the yield of neutrons to as high as 30–40 per incident proton. The contribution of the spallation production must of course contribute very little to the nuclear waste output. The optimal choice of spallation target is believed to be lead, although it does need high operating temperatures. A crucial question is the reliability of the spallation target at high proton fluxes and experiment has demonstrated this for a liquid lead–bismuth target exposed to a 1.4 mA proton beam.

In this way, the capacity to ‘burn’ long-lived fission products and actinides is greatly increased, leaving waste with much shorter lifetimes that can be disposed of by conventional means. The accelerator would deliver a high-current (10–20 mA) proton beam of about 1 GeV energy to a heavy metal (spallation) target surrounded by the nuclear waste to be incinerated. The accelerator–waste combination would always be operated at a subcritical level – by itself it could not sustain a chain reaction – so that no reactor-core meltdown accident could occur. The accelerator itself provides a control mechanism for a subcritical assembly that is more convenient than the control rods used in the critical assembly of a conventional nuclear power plant.

So far, research has shown that initial partitioning of the radioactive waste at the level of 95–99% is possible depending on the actinide species, and a number of successful reactor transmutation and spallation studies have been made on a full ADS system. The latter device (TRADE) consisted of coupling a cyclotron delivering a 140 MeV, 0.5–1.0 mA proton beam to an existing 1 MW water-cooled reactor sited in Italy, and used a spallation target of tantalum. Later research (for example MYRRHA, discussed in Section 9.1.4 below) has also shown that some long-lived isotopes can be successfully incinerated.

9.1.4 Power from ADS systems

Suggestions that the concept of ADS might be carried one step further, and a particle beam might be used to produce additional neutrons directly in a nuclear-reactor-like core, has a long history. Building on these, the first proposal for a practical device, called the Energy Amplifier, was made in 1995 by Rubbia,⁹ and versions of it have been studied in America and Europe. In this scheme, the core of the reactor would again be subcritical and the accelerator beams would provide sufficient additional neutrons

⁹Rubbia et al. (1995), the same man who shared the 1984 Nobel Prize in Physics for the discovery of the W and Z bosons.

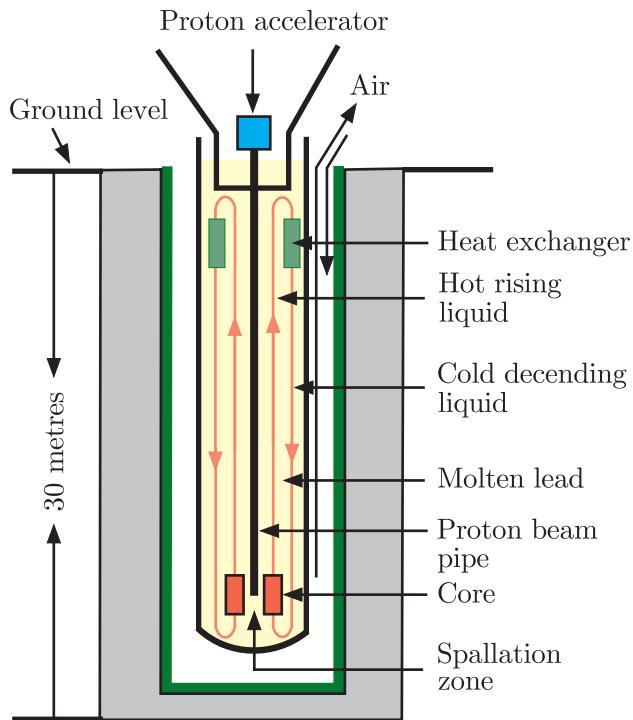


Figure 9.4 Schematic diagram of a possible configuration of an energy amplifier. In this design the coolant and spallation metal is molten lead. Source: Adapted from CERN report CERN/AT/95-44(ET).

via the spallation reaction to run the reactor. An idealised possible setup is shown in Figure 9.4.

The device shown is about 30 m deep and 6 m in diameter and contains liquid lead. Protons from an accelerator above ground are directed via a beam pipe to the bottom of the assembly, where the reactor core is situated, and initiate the production of neutrons via the spallation process in the lead. Because the spallation neutrons would have high energies, a less enriched element, such as natural thorium, could serve as the fuel. Thorium has the great advantage over uranium of being an abundant element that does not require costly isotope separation and does not produce plutonium.¹⁰ The thorium fuel would not require enrichment, although it would need to be recharged every five years or so.¹¹ The heat produced in the core flows upwards carried by the liquid lead and is extracted via

¹⁰Thorium might be thought suitable for power production in a conventional plant because of the high yield of neutrons from ^{233}U via the chain $n + ^{232}\text{Th} \rightarrow ^{233}\text{Th} \rightarrow ^{233}\text{Pa} \rightarrow ^{233}\text{U}$. However, the intermediate nucleus ^{233}Pa has a very large neutron capture cross-section that must be compensated by enrichment of the fissile material. In addition, the fraction of delayed neutrons from the fission of ^{233}U (which we have seen in Section 9.1.2 are essential for the safe operation of a reactor) is less than half that from the fission of ^{235}U , leading to smaller safety margins. Also, ^{233}U is weapon grade unless denatured and waste processing in this reaction has not yet been developed. Only a few experimental reactors have been built using thorium.

¹¹When used for ‘burning’ radioactive waste the core would consist of a range of elements.

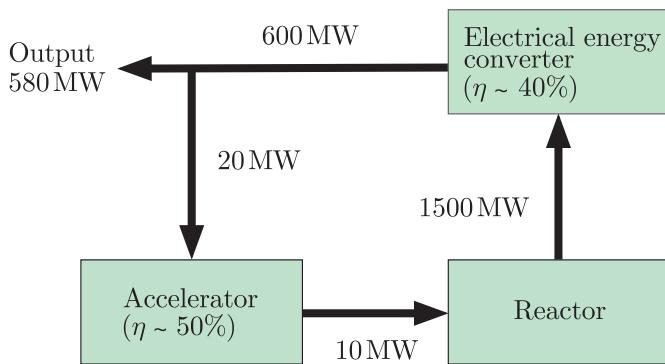


Figure 9.5 Possible energy flows in an energy amplifier system of the Rubbia design. The conversion efficiencies are denoted by η .

a heat exchanger. The cooled liquid lead then flows downwards and completes the circulation.

The possible energy flow in a commercial system is shown in Figure 9.5. This assumes a 1 GeV, 20 mA proton beam requiring about 20 MW of input power. The latter is taken from the output of the reactor, leaving a net electrical output of 580 MW, i.e. a gain factor of about 30. The actual gain factor depends on quantities such as the type of spallation material and the multiplication factor defined in (9.3) (see Problem 9.4). Whether the Energy Amplifier would be economically competitive is an open question.

Practical realisations of ADS systems are being pursued at a number of sites. For example, the Belgium research centre SCK•EN has been developing an ADS system (the MYRRHA project) since 1998, using a linac producing a 600 MeV proton beam developing an intensity of 4 mA, and a liquid lead–bismuth mixture as coolant and spallation target, coupled to a 100 MW subcritical reactor. As mentioned above, the MYRRHA experiment has already shown that some long-lived isotopes can be successfully incinerated. Another advanced project is the Kumatori Accelerator-Driven Reactor Test Facility (KART) at Kyoto University, which in 2006 coupled a purpose-built 150 MeV proton accelerator to an existing test reactor facility using a core of thorium and enriched uranium.

To summarise, ADS has a number of important advantages over a conventional power reactor. These include: it is subcritical without the spallation neutrons and so is inherently safe as a meltdown or explosion is not possible; radioactive waste is consumed in the reactor and no long-lived waste is produced; and there is no overlap with the nuclear weapons fuel cycle and so the energy amplifier cannot be used as the basis for producing materials for nuclear weapons, making installations politically acceptable worldwide. Thus all three objections listed earlier are in principle addressed. However, there are still a great many problems to be overcome and questions to be answered. For example, the estimated time for the MYRRHA project to be fully operational is 2033 and time for completion of research and development work and commencement of an industrial plant based on ADS could be as long as several decades.

9.2 Fusion

We have seen that the plot of binding energy per nucleon (Figure 2.9) has a maximum at $A \approx 56$ and slowly decreases for heavier nuclei. For lighter nuclei, the decrease is much quicker, so that with the exception of magic nuclei, lighter nuclei are less tightly bound than medium-size nuclei. Thus, in principle, energy could be produced by two light nuclei fusing to produce a heavier and more tightly bound nucleus – the inverse process to fission. Just as for fission, the energy released comes from the difference in the binding energies of the initial and final states. This process is called *nuclear fusion* and is very attractive as a potential source of power, because of the far greater abundance of stable light nuclei in nature than very heavy nuclei. Thus fusion would offer enormous potential for power generation, if the huge practical problems could be overcome. Fusion processes also explain how stars are formed. We will discuss both of these applications in what follows, but first we start with some basic concepts.

9.2.1 Coulomb barrier

The practical problem to obtaining fusion, whether in power production or more generally, has its origin in Coulomb repulsion, which inhibits two nuclei getting close enough together to fuse. This is given by the Coulomb potential

$$V_C = \frac{ZZ'e^2}{4\pi_0(R + R')}, \quad (9.17)$$

where Z and Z' are the atomic numbers of the two nuclei and R and R' are their effective radii. The quantity $(R + R')$ is therefore classically the distance of closest approach. Recalling, from the work on nuclear structure in Chapter 2, that for medium and heavy nuclei $R = 1.2A^{1/3}\text{fm}$, we have

$$V_C = \left(\frac{e^2}{4\pi e_0 \hbar c} \right) \frac{\hbar c ZZ'}{1.2 \left[A^{1/3} + (A')^{1/3} \right] \text{fm}} = 1.198 \frac{ZZ'}{A^{1/3} + (A')^{1/3}} \text{MeV}. \quad (9.18)$$

If for illustration we set $A \approx A' \approx 2Z \approx 2Z'$, then

$$V_C \approx 0.15A^{5/3} \text{ MeV}. \quad (9.19)$$

Thus, for example, with $A \approx 8$, $V_C \approx 4.8 \text{ MeV}$ and this energy has to be supplied to overcome the Coulomb barrier. This is a relatively small amount of energy and it might be thought that it could be achieved by simply colliding two accelerated beams of light nuclei, but in practice nearly all the particles would be elastically scattered. The only practical way is to heat a confined mixture of the nuclei to supply enough thermal energy to overcome the Coulomb barrier. The temperature necessary may

be estimated from the relation $E = kT$, where $k = k_B$ and k_B is Boltzmann's constant. For an energy of 4.8 MeV, this implies a temperature of 5.6×10^{10} K. This is well above even the typical temperature of 10⁷ K found in stellar interiors.¹² It is also the major hurdle to be overcome in achieving a controlled fusion reaction in a reactor, as we shall see later.

Fusion actually occurs at a lower temperature than this estimate due to a combination of two reasons. The first and most important is the phenomenon of quantum tunnelling, which means that the full height of the Coulomb barrier does not have to be overcome. In Section 8.6 we discussed a similar problem in the context of α decay, and we can draw on that analysis here. The probability of barrier penetration depends on a number of factors, but the most important is the Gamow factor, which is a function of the relative velocities and the charges of the reaction products. In particular, the probability is proportional to $\exp[-G(E)]$, where $G(E)$ is a generalisation of the Gamow factor of (8.48). This may be written as $G = \sqrt{E_G/E}$, where again, generalising the equations in Chapter 8,

$$E_G = 2mc^2(\pi\alpha Z_1 Z_2)^2. \quad (9.20)$$

Here m is the reduced mass of the two fusing nuclei that have electric charges $Z_1 e$ and $Z_2 e$. Thus the probability of barrier penetration increases as E increases. Nevertheless, the probability of fusion is still extremely small. For example, if we consider the fusion of two protons (which we will see below is an important ingredient of the reactions that power the Sun), at a typical stellar temperature of 10⁷ K, we find $E_G \approx 490$ keV and $E \approx 1$ keV. Hence the probability of fusion is proportional to $\exp[-(E_G/E)^{1/2}] \approx \exp(-22) \approx 10^{-9.6}$, which is a very large suppression factor, and so the actual fusion rate is still extremely slow.

The other reason that fusion occurs at a lower temperature than expected is that a collection of nuclei at a given mean temperature, whether in stars or elsewhere, will have a Maxwellian distribution of energies about the mean and so there will be some with energies substantially higher than the mean energy. Nevertheless, even a stellar temperature of 10⁸ K corresponds to an energy of only about 10 keV, so the fraction of nuclei with energies of order 1 MeV in such a star would only be of order $\exp(-E/kT) \sim \exp(-100) \sim 10^{-43}$, a minute amount. We now examine the interplay of these two factors.

9.2.2 Fusion reaction rates

We have discussed in Section 9.2.1 how quantum tunnelling and the Maxwellian distribution of energies combine to enable fusion to occur at a lower temperature than might at first be expected. The product of the increasing barrier penetration factor with energy and the Maxwellian

¹²For this reason, many scientists initially refused to accept that fusion occurred in stars when the suggestion was first made by Arthur Eddington.

decreasing exponential means that in practice fusion takes place over a rather narrow range of energies. To see this we will consider the fusion between two types of nuclei, a and b , having number densities n_a and n_b (i.e. the number of particles per unit volume) and at a temperature T . We assume that the temperature is high enough so that the nuclei form a plasma, i.e. are fully ionised, with uniform values of number densities and temperature. We also assume that the velocities of the two nuclei are given by the Maxwell–Boltzmann distribution, so that the probability of having two nuclei with a relative speed v in the range v to $v + dv$ is

$$P(v) dv = \left(\frac{2}{\pi}\right)^{1/2} \left(\frac{m}{kT}\right)^{3/2} \exp\left(-\frac{mv^2}{2kT}\right) v^2 dv, \quad (9.21)$$

where m is the reduced mass of the pair. The fusion reaction rate per unit volume is then

$$R_{ab} = n_a n_b \langle \sigma_{ab} v \rangle, \quad (9.22)$$

where σ_{ab} is the fusion cross-section¹³ and the brackets denote an average, i.e.

$$\langle \sigma_{ab} v \rangle \equiv \int_0^\infty \sigma_{ab} v P(v) dv. \quad (9.23)$$

The fusion cross-section may be written

$$\sigma_{ab}(E) = \frac{S(E)}{E} \exp\left[-\left(\frac{E_G}{E}\right)^{1/2}\right], \quad (9.24)$$

where the exponential follows from the previous discussion of quantum tunnelling and $S(E)$ contains the details of the nuclear physics. The term $1/E$ is conveniently factored out because many nuclear cross-sections have this behaviour at low energies. Using (9.21) and (9.24) in (9.23) gives, from (9.22),

$$R_{ab} = n_a n_b \left(\frac{8}{\pi m}\right)^{1/2} \left(\frac{1}{kT}\right)^{3/2} \int_0^\infty S(E) \exp\left[-\frac{E}{kT} - \left(\frac{E_G}{E}\right)^{1/2}\right] dE. \quad (9.25)$$

Because the factor $1/E$ has been taken out of the expression for $\sigma(E)$, the quantity $S(E)$ is slowly varying and the behaviour of the integrand is dominated by the behaviour of the exponential term. The falling exponential of the Maxwellian energy distribution combines with the rising

¹³The product $n_a n_b$ is the number of pairs of nuclei that can fuse. If the two nuclei are of the same type, with $n_a = n_b = n$, then the product must be replaced by $n(n-1)/2 \approx n^2/2$, because in quantum theory such nuclei are indistinguishable.

exponential of the quantum tunnelling effect to produce a maximum in the integrand at $E = E_0$, where

$$E_0 = \left[\frac{1}{4} E_G (kT)^2 \right]^{1/3} \quad (9.26)$$

and fusion takes place over a relatively narrow range of energies $E_0 \pm \frac{1}{2} \Delta E_0$, where

$$\Delta E_0 = \frac{4}{3^{1/2} 2^{1/3}} E_G^{1/6} (kT)^{5/6}. \quad (9.27)$$

The importance of the temperature and the Gamow energy $E_G = 2mc^2(\pi\alpha Z_a Z_b)^2$ is clear. A schematic illustration of the interplay between these two effects is shown in Figure 9.6a. Figure 9.6b shows the function $\exp[-E/kT - (E_G/E)^{1/2}]$ for the real example of the fusion of two protons at a temperature of 2×10^7 K. We have $E_G = 496$ keV and $kT = 1.7$ keV, so that fusion is most likely at $E_0 = 7.2$ keV and the width of the distribution is $\Delta E_0 = 8.2$ keV.

In the approximation where we take $S(E)$ as a constant $S(E_0)$, the integral in (9.25) may be done and gives

$$\langle \sigma_{ab} v \rangle \approx \frac{8}{9} S(E_0) \left(\frac{2}{3mE_G} \right)^{1/2} \tau^2 \exp(-\tau), \quad (9.28)$$

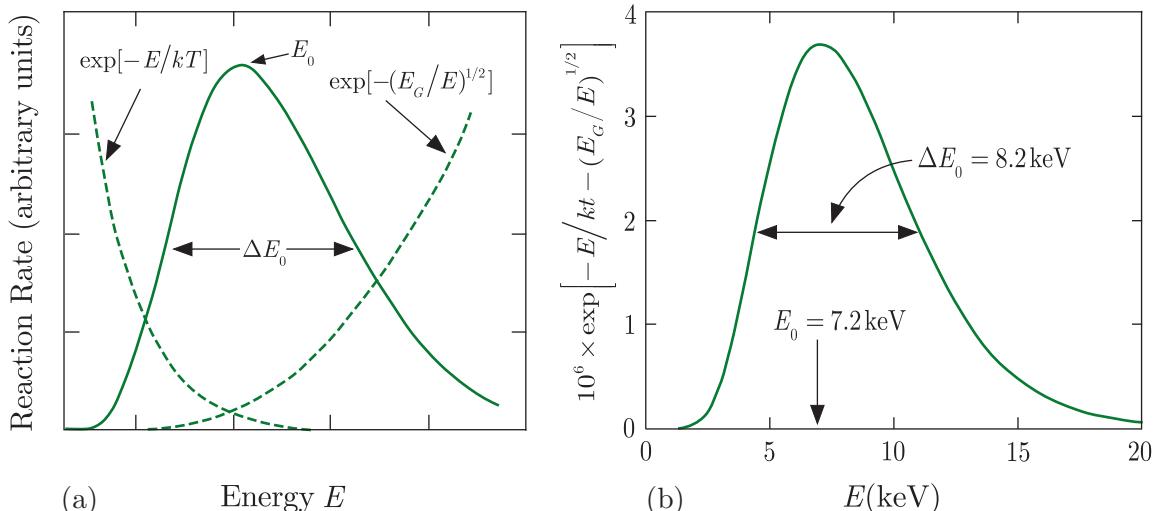


Figure 9.6 (a) The dashed curve increasing with energy is proportional to the barrier penetration factor and the dashed curve decreasing with energy is proportional to the Maxwell distribution. The solid curve is the combined effect and is proportional to the overall probability of fusion with a peak at E_0 and a width at half-maximum height of ΔE_0 . (b) The exponential part of the integrand in (9.25) for the case of $p-p$ fusion at a temperature of 2×10^7 K.

where $\tau = 3(\frac{1}{2})^{2/3}(E_G/kT)^{1/3}$. If we take the masses to be $A_{a,b}$ in atomic mass units we can use this result to evaluate (9.25), using the expression (9.20) for E_G , to give

$$R_{ab} = \frac{7.21 \times 10^{-22}}{Z_a Z_b} n_a n_b \frac{(A_a + A_b)}{A_a A_b} \left(\frac{S(E_0)}{1 \text{ MeV b}} \right) \tau^2 \exp(-\tau) \text{ m}^3 \text{ s}^{-1}, \quad (9.29)$$

with

$$\tau = 18.8(Z_a Z_b)^{2/3} \left(\frac{A_a A_b}{A_a + A_b} \right)^{1/3} \left(\frac{1 \text{ keV}}{kT} \right)^{1/3}. \quad (9.30)$$

The rate depends very strongly on both the temperature and the nuclear species because of the factor $\tau^2 \exp(-\tau)$. This is illustrated in Figure 9.7 for $p-p$ and $p-^{12}\text{C}$ fusion, both of which are important reactions in stellar processes, as we shall see in the following section.

9.2.3 Nucleosynthesis and stellar evolution

Nuclear fusion plays a vital role in the evolution of astrophysical objects of many types, because it determines how heavier elements are formed from lighter ones. We will start by considering how the key elements of hydrogen and helium were formed in the early universe and their role in the initial stages of evolution. Then we discuss how heavier nuclei are formed in stars.

9.2.3(a) Nucleosynthesis in the early universe

The current theory of the origin of the universe – the Big Bang theory – starts with the creation of the fundamental particles, quarks, leptons, and others, each with their own antiparticle. How this particle–antiparticle symmetry evolved into what we see today, where matter is overwhelmingly made of particles, is not yet fully understood (we will return to this in Chapter 10), but within 10^{-2} s after the Big Bang, quarks in the rapidly cooling primordial plasma coalesced to form protons and neutrons, and the reactions



were possible. Initially these reactions could proceed in either direction with equal rates, as indicated by the two-way arrows, despite the neutron–proton mass difference, because the particles had high kinetic energies derived from the extreme temperature. However, as the universe continued to expand and cool, within one second the particles had insufficient energy to maintain equilibrium and only the decay $n \rightarrow p + e^- + \bar{\nu}_e$ was possible, with a half-life of about 10 minutes. The neutrinos escaped to become ‘relics’ of the Big Bang. However, before the neutrons could decay, within a few minutes most of them were captured by protons to form the bound state deuterium, $d = np$, via the reactions

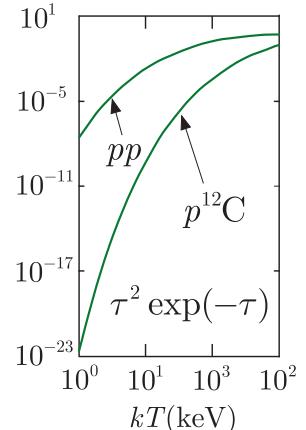
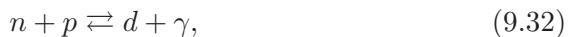


Figure 9.7 The function $\tau^2 \exp(-\tau)$ of (9.30) for the $p-p$ and $p-^{12}\text{C}$ fusion reactions.

where again initially the two reactions were in equilibrium. However, by analogy with reaction (9.31), as the universe cooled further the fraction of photons having sufficient energy to photodissociate deuterium was reduced, so deuterium production started to dominate. Production of deuterium in sufficient quantities marked the start of nucleosynthesis.

If we use our usual notation for nuclei, with $p = {}_1^1\text{H}$, $n = {}_0^1\text{H}$, and $d = {}_1^2\text{H}$, then helium-3 can be produced via deuterium reactions with hydrogen,



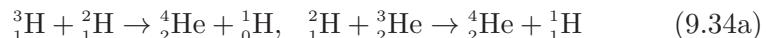
or in deuteron-deuteron collisions,¹⁴



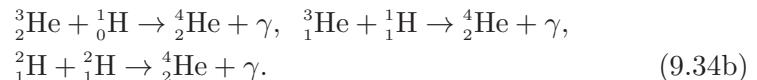
The latter interaction can also produce tritium ${}_{\text{1}}^3\text{H}$, with a lifetime of 12 years, i.e.



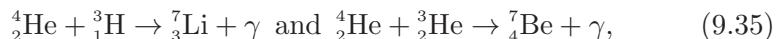
Next, helium-4 can be produced via a number of reactions involving protons, neutrons, deuterium, tritium, and helium-3, including



and



In principle, successive sequences analogous to those above could synthesise the next highest nuclei. However, there is no mass-5 nucleus and ${}_{\text{4}}^8\text{Be}$ is highly unstable. Fusion can produce mass-7 nuclei via the reactions



but the Coulomb barriers to be overcome are about 1 MeV, whereas at the temperature prevalent at this stage ($T \approx 10^9$ K) the nuclei energies are less than 0.1 MeV, so these reactions have very low probabilities. Moreover, if the lithium or beryllium nuclei collide with a free nucleon they will immediately fragment back to isotopes of helium. Thus from the very small quantities of mass-7 nuclei initially produced, only traces survived. Three minutes after the Big Bang, the hadron constituents of the universe were: protons 75%, helium 24%, and in addition there were small quantities of deuterium, with traces of lithium and beryllium nuclei.

At first sight it might seem plausible that heavier nuclei could be formed by fusion reactions involving lighter nuclei analogous to those described above, but this idea is frustrated by the very strong binding

¹⁴In principle, two protons could form ${}_{\text{2}}^2\text{He}$, but in practice it is too unstable.

energy of helium-4 (recall it is a doubly magic nucleus). Thus in the early universe scenario, the nucleus ${}^8\text{Be}$ formed in the reaction

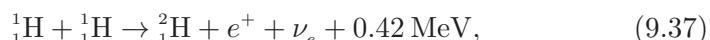


disintegrates back to two helium-4 nuclei very rapidly before a third helium-4 nucleus has time to react with it to form stable ${}^{12}_6\text{C}$, and the probability of three helium-4 nuclei colliding simultaneously is negligible. However, the carbon nucleus can be formed in stars at lower temperatures, and so we next turn to considering stellar nucleosynthesis.

9.2.3(b) Stellar nucleosynthesis

Stellar evolution involves the release of gravitational potential energy through contraction of the nuclear material of the star. This increases its temperature to a point where a particular fusion reaction can occur, producing energy that flows to the surface of the star and temporarily halts the gravitational contraction. When all the nuclear ‘fuel’ for the reaction is ‘burnt’, contraction restarts and, provided the mass of the star is great enough, the temperature of the star again rises to a point where the products of the first stage of burning are used as the fuel for the next set of fusion reactions, and so on. The full sequence for a typical very heavy star fulfilling this criterion are summarised in Table 9.1.

In the case of the Sun, for example, hydrogen is initially burned by the so-called *proton–proton cycle*. This has more than one branch, but one of these, the PPI cycle, is dominant. It starts with the fusion of hydrogen nuclei to produce nuclei of deuterium:



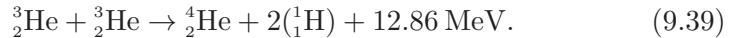
which is the primary source of the solar neutrinos discussed in Chapter 3. The deuterium then fuses with more hydrogen to produce ${}^3_2\text{He}$:



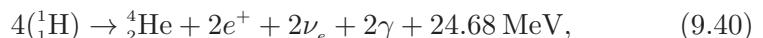
Table 9.1 The main stages of nuclear burning for a star of mass 25 solar masses.

Fuel	Products	Timescale	Approximate ignition temperature (K)
Hydrogen	Helium	1×10^7 yr	1×10^7
Helium	Carbon	5×10^5 yr	1×10^8
	Oxygen		
Carbon	Oxygen	600 yr	5×10^8
	Neon		
	Sodium		
	Magnesium		
Neon	Oxygen	1 yr	1×10^9
	Magnesium		
Oxygen	Magnesium to sulfur	6 months	2×10^9
Silicon	Iron and nearby nuclei	1 day	3×10^9

and finally, two ${}^3_2\text{He}$ nuclei fuse to form ${}^4_2\text{He}$:



The relatively large energy release in the last reaction is again because ${}^4_2\text{He}$ is a doubly magic nucleus and so is very tightly bound. The first of these reactions (9.37), being a weak interaction, proceeds at an extremely slow rate, and sets the scale for the long lifetime of the Sun. Combining these equations, we have overall



and the nuclear end product is helium-4. Because the temperature of the Sun is $\sim 10^7 \text{ K}$, all its material is in the form of a plasma. The positrons produced in reaction (9.37) will annihilate with electrons in the plasma to release a further 1.02 MeV of energy per positron and so the total energy released is 26.72 MeV. However, of this each neutrino will carry off 0.26 MeV on average, which is lost into space. Thus on average, 6.55 MeV of electromagnetic energy is radiated from the Sun for every proton consumed in the PPI chain. The PPI chain is not the only fusion cycle contributing to the energy output of the Sun, but it is the most important.

When hydrogen has been depleted as a fuel, helium burning starts. The difficulty of synthesising carbon in the environment of the early universe has been mentioned earlier, but at the lower temperatures in stars the slower dissociation of ${}^8_4\text{Be}$ in reaction (9.36) allows the rare reaction



to occur, where C^* is an excited state of carbon. A very small fraction of the latter will decay to the ground state, so that overall we have

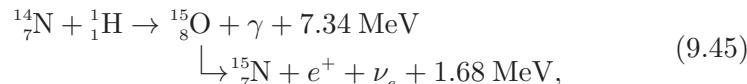
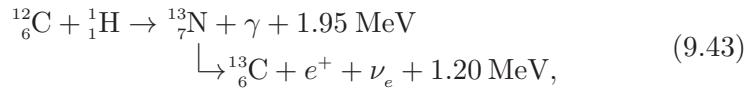


The occurrence of this crucial reaction depends critically on the existence of a particular excited state of ${}^{12}\text{C}$, which was predicted to exist by the astronomer Fred Hoyle to explain the ‘anomaly’ that ${}^{12}\text{C}$ was present in stars in greater amounts than expected. The existence of the state was confirmed by William Fowler.¹⁵

Once carbon is present, another cycle, called the CNO chain, can start. Although this contributes only about 3% of the Sun’s energy output, it plays an important role in the evolution of stellar objects of greater mass.

¹⁵William Fowler shared the 1983 Nobel Prize in Physics for his studies of nuclear reactions of importance in the formation of chemical elements in stars.

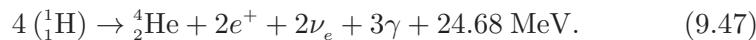
In the presence of any of the nuclei $^{12}_6\text{C}$, $^{13}_6\text{C}$, $^{14}_7\text{N}$, or $^{15}_7\text{N}$, hydrogen will catalyse burning via the reactions



and



Thus overall in the CNO cycle we have



These and other fusion chains all produce electron neutrinos as final-state products, and using detailed models of the Sun, the flux of such neutrinos at the surface of the Earth can be predicted.¹⁶ However, the actual count rate is far lower than the theoretical expectation. This is the solar neutrino problem that we met in Section 3.1.5 and we saw there that the solution lies in the phenomenon of neutrino oscillations, where some electron neutrinos are converted to neutrinos of other flavours in their passage from the Sun to Earth.

The process whereby heavier elements are produced by fusion of lighter ones can continue beyond the reactions above. Thus, for example, the presence of $^{12}_6\text{C}$ enables the production of $^{16}_8\text{O}$ via the reaction



and the production of neon, sodium, and magnesium are possible via the reactions



Fusion processes continue to synthesise heavier elements until the core of the stellar object is composed mainly of nuclei with $A \approx 56$ (see Table 9.1), i.e. the peak of the binding energy per nucleon curve (see Figure 2.9), at which point fusion ceases to be a way of creating heavier elements.

Heavier nuclei are produced mainly by the accretion of neutrons in collisions. These reactions are intrinsically implausible and the process is very slow – hence the name *s*-process. The majority of stars die quietly, accumulating more neutrons and expanding until they eventually collapse slowly under gravity. However, if part of a binary system, they

¹⁶The expectations are based on the detailed model of the Sun known as the ‘standard solar model’ that we met in Section 3.1.5.

may attract hydrogen and helium from their companion star and the accreted matter may ignite reactions in which protons are created to create unstable proton-rich nuclei. This is the so-called *rp*-process. Finally, a route of great importance, although very rare, is where a very massive star collapses very rapidly in just a few seconds, resulting in a spectacular explosion called a supernova, during which ultrarapid reactions (called the *r*-process) involving neutron-rich nuclei create elements as heavy as uranium, which are then distributed throughout the universe.¹⁷ The paths of these three processes are shown on the Segrè plot of Figure 2.13.

The synthesis of nuclei in the universe is one of the cornerstones of modern astrophysics, but even here there are still surprises. One example is the production of heavy elements via the rare reaction $3(^4\text{He}) \rightarrow ^{12}\text{C}$, which depends critically on the existence of a particular excited state of ^{12}C , as we have seen earlier. However, subsequently another excited state has been discovered at a somewhat higher energy that has the effect of significantly altering the energy dependence (or equivalently the temperature dependence) of this reaction from the values usually assumed. This could have consequences for models and theories of stellar evolution. Nuclear reactions are important in understanding other astrophysical processes. For example, the lifetime of the doubly-magic nucleus ^{78}Ni has been measured and found to be shorter than expected, implying that supernova explosions may produce gold and other heavier elements much faster than had previously been thought. This is important because the production of ^{78}Ni is believed to be the pathway that leads to the production of more than half the elements in the universe that are heavier than iron. Another example concerns ^{22}Na production in novas. Until recently it was thought that the production rate of this isotope would be such that its presence would be observable by γ -ray astronomy, but satellites have failed to detect it. However, measurement of the rate for $^{21}\text{Na}(p, \gamma)^{22}\text{Mg}$, which determines the abundance of galactic ^{22}Na , shows that it is higher than previously expected and so ^{22}Na is produced earlier in novas than thought, whence it is effectively removed by the reaction $^{22}\text{Na}(p, \gamma)^{23}\text{Mg}$, thus explaining the satellite data. These examples show that there is still much to be learned about nuclear synthesis.

9.2.4 Fusion reactors

For many years there has been an international large-scale effort to achieve controlled fusion in the laboratory, with the eventual aim of producing a fusion-powered reactor. For this, the *p-p* reactions are far too slow to be useful (see Figure 9.7). However, the Coulomb barrier for the deuteron ${}_1^2\text{H}$ is the same as for the proton, and the exothermic reactions



¹⁷It is a remarkable fact that some of the material of our bodies was once part of distant stars.

and



suggest that deuterium might be a suitable fuel for a fusion reactor. Deuterium is present in huge quantities in sea water and is easy to separate at low cost.

An even better reaction in terms of energy output is deuterium–tritium fusion:



The quantity $\langle\sigma v\rangle$ (called the *thermal reactivity*) for the $d-t$ reaction (9.51) and the combined $d-d$ reactions (9.50) is available as a function of temperature from several calculations.¹⁸ Typical values are shown in Figure 9.8. It can be seen that the deuterium–tritium ($d-t$) reaction has the advantage over the deuterium–deuterium ($d-d$) reaction of having a much higher cross-section. The heat of the reaction is also greater. The principal disadvantage is that tritium does not occur naturally (it has a mean life of only 17.7 years) and is expensive to manufacture, which increases the overall cost. From Figure 9.8 it can be seen that the rate for the $d-t$ reaction peaks at about $E = kT = 30 - 40$ keV, and a working energy where the cross-section is still considered reasonable is about 20 keV, i.e. 3×10^8 K.

The effective energy produced by the fusion process will be reduced by the heat radiated by the hot plasma. The mechanism for this is predominantly electron bremsstrahlung. The power loss per unit volume due to this process is proportional to $T^{1/2}Z^2$, where Z is the atomic number of the ionised atoms. Thus for a plasma with given constituents and at a fixed ion density, there will be a minimum temperature below which the radiation losses will exceed the power produced by fusion. For example, for the $d-t$ reaction with an ion density 10^{21} m^{-3} , $kT_{\min} \approx 4 \text{ keV}$. It would be 10 times larger for the $d-d$ reactions of (9.50a) and (9.50b) because of the form of $\langle\sigma v\rangle$ (see Figure 9.8), which is another reason for using the $d-t$ reaction. In practice the situation is worse than this because most of the neutrons in (9.51) will escape, so even at the theoretical ‘breakeven’ temperature, external energy would have to be supplied to sustain the fusion process. Only when the energy deposited in the plasma by the alpha particles exceeds the radiation losses would the reaction be self-sustaining. This is referred to as the ‘ignition point’.

A numerical expression that embodies these ideas is the *Lawson criterion*, which provides a measure of how close to practicality is a particular reactor design. We will assume a $d-t$ reaction. To achieve a temperature T in a deuterium–tritium plasma, there has to be an input of energy $4n_d(3kT/2)$ per unit volume. Here n_d is the number density of deuterium ions and the factor of 4 is because n_d is equal to the number density of tritium ions and the electron density is twice this, giving $4 n_d$ particles per

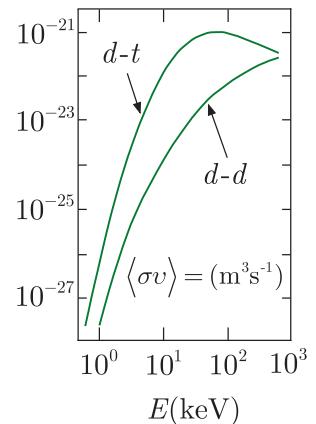


Figure 9.8 Typical values of the thermal reactivity $\langle\sigma v\rangle$ for the $d-t$ reaction (9.51) and the combined $d-d$ reactions (9.50a) and (9.50b).

¹⁸See, for example, Hively (1983) and Bosch and Hale (1992).

unit volume. The reaction rate in the plasma is $n_d^2 \langle \sigma_{dt} v \rangle$. If the plasma is confined for time t_c , then per unit volume of plasma,

$$L = \frac{\text{energy output}}{\text{energy input}} = \frac{n_d \langle s_{dt} u \rangle t_c Q}{6kT}, \quad (9.52)$$

where Q is the energy released in the fusion reaction. For a useful device, $L > 1$. For example, if we assume $kT = 20\text{ keV}$ and use the experimental value $\langle s_{dt} u \rangle \approx 10^{-22}\text{ m}^3\text{s}^{-1}$, then the Lawson criterion may be written as

$$n_d t_c > 7 \times 10^{19}\text{ m}^{-3}\text{s}. \quad (9.53)$$

Thus either a very high particle density or a long confinement time, or both, is required. At the temperatures required for fusion, any material container will vapourise, so the central problem is how to contain the plasma for a sufficiently long time for the reaction to take place. The two main methods are *magnetic confinement* and *inertial confinement*. Both techniques present enormous technical challenges.

In the inertial confinement method, small pellets of the deuterium–tritium ‘fuel’ mixture are bombarded with intense laser energy from several directions simultaneously. As material is ejected from the surface, other material interior to the surface implodes, compressing the core of the pellet to densities and temperatures where fusion can take place. The laser pulses are extremely short, typically $10^{-7} – 10^{-9}\text{ s}$, which is many orders of magnitude shorter than the times associated with the pulsed current in devices using magnetic confinement (which could be as long as a second), but this is compensated for by much higher plasma densities. Research is being actively pursued on inertial confinement, for example at the National Ignition Facility (NIF) at Lawrence Livermore Laboratory in California and the Laser Mégajoule in France. At NIF a system of 192 laser beams is directed at a small hollow target containing helium, with a tiny capsule less than 1 cm in size at its centre filled with the deuterium–tritium fuel. The major problem that has not yet been overcome is to produce a system where the laser beams impact the target perfectly symmetrically to ensure that the implosion produced is also symmetric, and thus capable of compressing the fuel to enable fusion to take place. The European Union is also funding the High Power Laser Energy Research Facility (HiPER). This is the first experiment designed specifically to study the so-called ‘fast ignition’ approach to generating nuclear fusion, which uses much smaller lasers than conventional designs yet produces fusion power outputs of about the same magnitude.

In practice, most work has been done on magnetic confinement. In this approach, the plasma is confined by magnetic fields and heated by electromagnetic fields. Firstly, we recall the behaviour of a particle of charge q in a uniform magnetic field \mathbf{B} . In the extreme case where the velocity \mathbf{v} of the particle is at right angles to \mathbf{B} , the particle traverses a circular orbit of fixed radius (compare the principle of the cyclotron discussed in Chapter 4). Otherwise, the velocity component in the field direction is

unaffected and the path is a helix of fixed pitch along the direction of the field (compare the motion of electrons in a time projection chamber, also discussed in Chapter 4). Two techniques have been proposed to stop particle losses: magnetic ‘mirrors’ and a geometry that would ensure a stable indefinite circulation. In the former, it is arranged that the field is greater at the boundaries of a region than in its interior. Then as the particle approaches the boundary, the force it experiences will develop a component that points into the interior where the field is weaker. Thus the particle is trapped and will oscillate between the interior and the boundaries.¹⁹ However, most practical work has been done on case (b) and for that reason we will restrict our discussion to this technique.

The simplest configuration is a toroidal field produced by passing a current through a doughnut-shaped solenoid. In principle, charged particles in such a field would circulate endlessly, following helical paths along the direction of the magnetic field. In practice the field would be weaker at the outer radius of the torus and the nonuniformity of the field would produce instabilities in the orbits of some particles and hence lead to particle loss. To prevent this a second field is added, called a poloidal field. This produces a current around the axis of the torus and under the combined effect of both fields charged particles in the plasma execute helical orbits about the mean axis of the torus. Most practical realisations of these ideas are devices called *tokamaks*, in which the poloidal field is generated along the axis of the torus through the plasma itself.

The largest and most powerful tokamak in existence is the Joint European Torus (JET), which is a European collaboration sited at the Culham Laboratory in Berkshire, UK. A cutaway view of JET is shown in Figure 9.9a and a schematic view of the arrangement of the fields in the device is shown in Figure 9.9b. This shows the external coils that generate the main toroidal field. The poloidal field is generated by transformer action on the plasma. The primary windings of the transformer are shown with the plasma itself forming the single-turn secondary winding. The current induced in the plasma not only generates the poloidal field but also supplies several megawatts of resistive heating to the plasma. However, even this is insufficient to ensure a high enough temperature for fusion and additional energy is input via other means, including rf sources.

Considerable progress has been made towards the goal of reaching the ignition point. For example, in 1997 JET produced 16 MW of fusion power from a total input of 24 MW, a ‘gain’ of about 0.7, although only for a few seconds, and in 2013 a small real gain, i.e. greater than 1, was fleetingly achieved, the first for any fusion facility. However, although appropriate values of n_d , t_c , and T have been obtained separately, to date no device has yet succeeded in achieving the Lawson criterion in a sustainable way.

¹⁹The Van Allen radiation belts that occur at high altitudes consist of charged particles from space that have become trapped by a magnetic mirror mechanism because the Earth’s magnetic field is stronger at the poles than at the Equator.

The best value is still well below the ignition point and gains of 10–30 will need to be achieved for fusion power to be practical.

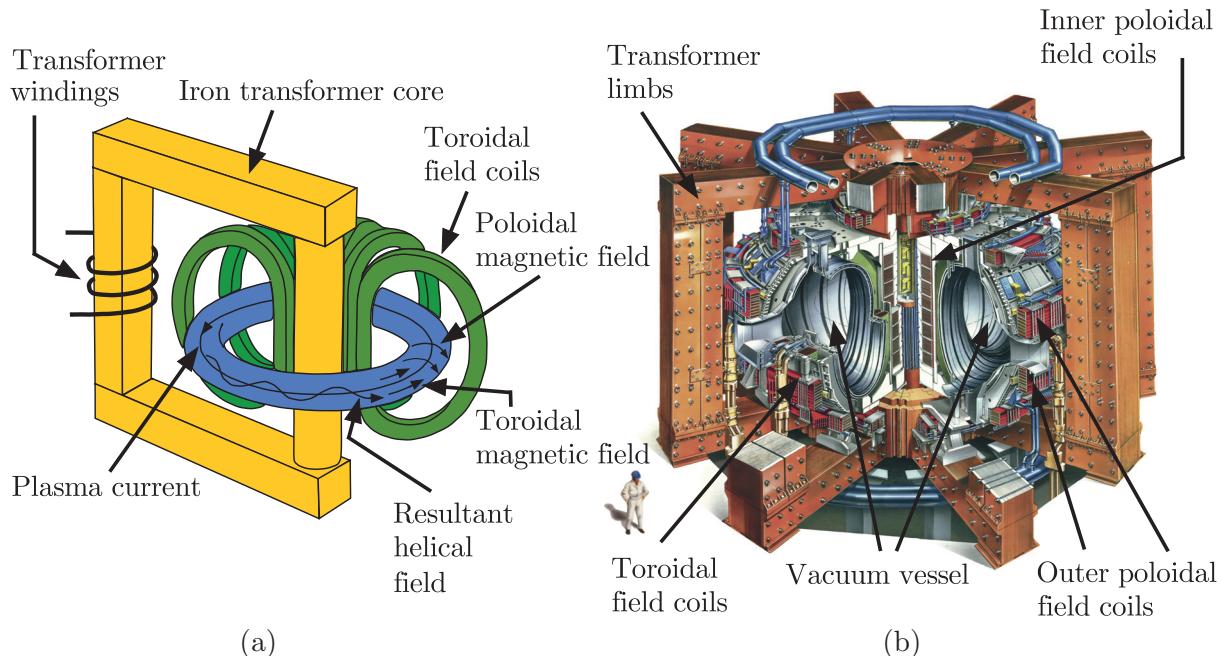


Figure 9.9 Schematic diagrams showing: (a) the main magnetic field components of the JET tokamak; (b) how these elements are incorporated into the JET device. Source: Adapted from images courtesy of EUROfusion.

Nuclear fusion holds the promise of unlimited power without the problem of radioactive waste, but the road to realisation of this goal is long and we are far from the end. In Section 9.2.4 we introduced the Lawson criterion as a measure of how close a design was to the ignition point, i.e. the point at which a fusion reaction becomes self-sustaining. To date no device has yet succeeded in achieving the Lawson criterion and much work remains to be done on this important problem. In recognition of this, a major new tokamak machine, the International Thermonuclear Experimental Reactor (ITER), is being built in France by a global collaboration, including Japan, USA, China, Russia, and the countries of the EU. ITER, although considerably larger than JET, has the same basic torus-like configuration. The project has experienced a number of problems, not always scientific, and the first plasma is now planned for 2025. Operation with tritium-deuterium fuel is not anticipated until 2035. The aim is to produce 500 MW of power sustained for up to 1000 seconds, compared to JET's peak of 16 MW for less than a second.

Although tokamak machines such as ITER and JET are likely to be the first to achieve a self-sustaining reaction, a commercial reactor would have to satisfy additional financial and engineering constraints and other configurations are not ruled out. One of these is the spherical tokamak, the limiting case of the torus type. Experiments on this type of configuration

started in the late 1980s and there is an on-going experimental programme at a small number of centres, an important one being Culham, the home of JET. Here a machine called MAST (Mega-Ampere Spherical Tokamak) is being used to carry out a range of studies to support the design work on ITER by studying plasma behaviour at the limit of conventional tokamak design. There are other groups in institutes around the world using small tokamaks to explore different approaches to producing a sustainable plasma.

9.3 Nuclear weapons

When nuclear reactions are used to produce power in a controlled way there is a clear distinction between fission and fusion processes. However, in the case of explosive power production, i.e. bombs, the distinction is not always so clear and some weapons use both fission and fusion in the same device. It is therefore appropriate to separate the discussion of nuclear weapons from the use of nuclear reactions for peaceful purposes.²⁰

9.3.1 Fission devices

Despite the simple analysis in Section 9.1.1, fortunately it is not easy to make a nuclear bomb! A major problem is that the thermal energy released as the assembly becomes critical will produce an outward pressure that is sufficient to blow apart the fissile material before criticality is achieved, unless special steps are taken to prevent this. Even before the chain reaction occurs, there are problems to be overcome, which can be summarised as follows:

- (a) keeping the fissile materials subcritical before detonation;
- (b) bringing the material into a supercritical state while keeping it relatively free of neutrons;
- (c) introducing neutrons to the critical mass at the optimum time;
- (d) keeping the fissile material together until a substantial fraction of the mass has fissioned.

Problems (a) and (b) are complicated by the fact that there will always be neutrons present from spontaneous fission of the fissile material. If we define the ‘insertion time’ as the time to reach a supercritical value of $k = 2$ starting from the critical value $k = 1$, then it follows that the insertion time must be much less than the average time between spontaneous fissions. The key to achieving this is the fact that the critical mass is inversely

²⁰Some of the material in this section is based on Sublette (2007). Our discussion is qualitative, and more details, including a comprehensive history of nuclear weapons, may be found in this reference.

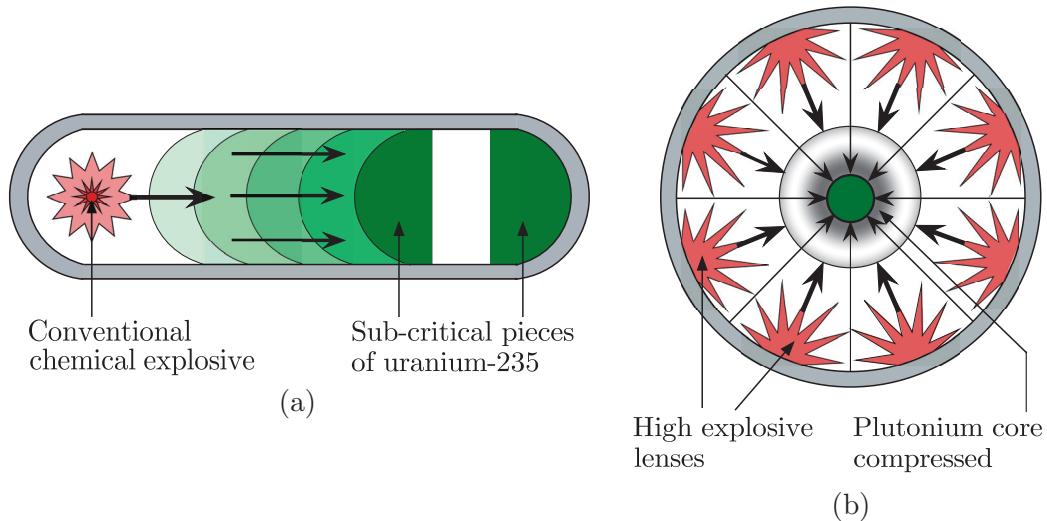


Figure 9.10 Schematic diagrams of: (a) gun assembly and (b) implosion assembly technique for an explosive fission device. In practice, in the gun assembly method the ‘missile’ was the smaller of the two pieces and was shaped to fit into a prepared hole in the ‘target’.

proportional to the density squared. Thus, if we can contain a subcritical assembly of fissile material in such a way that its average density can be rapidly increased, the necessary very short insertion time can be achieved. In practice, the density of a highly supercritical mass must be 2–3 times that of a mass of the same shape that is just critical. Two methods of achieving this have been used and we will describe each of them.

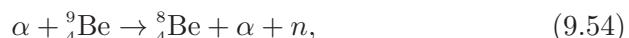
The first technique is the *gun assembly* method, illustrated schematically in Figure 9.10a. Here a subcritical mass is fired down a barrel to combine with another (just) subcritical mass, so that together the assembly is supercritical. In practice, the ‘missile’ is shaped to fit a hollow prepared in the ‘target’ material. This method was used in the bomb (called ‘Little Boy’) dropped on Hiroshima. No other weapon using this technique has ever been exploded. Although the gun assembly method has the advantage of simplicity, the insertion times are large, of order milliseconds, and this, plus the lack of compression, means that a large amount of fissile material must be used. The fissile material in Little Boy was uranium.²¹

The second technique is the *implosion assembly* method, illustrated schematically in Figure 9.10b for the case of a spherical assembly. This

²¹It was originally proposed to use ^{239}Pu , produced in reactors, as the fissile material, but due to neutron absorption the ^{239}Pu is always accompanied by ^{240}Pu , which is extremely difficult to separate from the lighter isotope. ^{240}Pu has a high rate for spontaneous fission and the neutrons produced would have caused a weak premature detonation before a supercritical mass could be assembled. There is not the same problem with uranium, because the fissile material was more than 90% ^{235}U .

uses specially designed conventional high explosives to compress a subcritical mass to high density and hence criticality. The explosives are ‘shaped’ so that on detonation an *imploding* shock wave is created that compresses the critical mass. There will also usually be a layer of dense material (a ‘tamper’), typically natural uranium or tungsten, surrounding the critical mass with a gap between this and the core. This allows the tamper to generate momentum before it impacts the core and aids the compression. In addition, the tamper can scatter neutrons back into the critical mass, thereby reducing the amount of fissile material needed for criticality. The pressure generated can be as high as several megabars, sufficient to increase the density at least twofold, even for a solid uranium or plutonium core. The compression is very rapid, resulting in very short insertion times, of order microseconds. Taken together, an efficient bomb can be made using relatively small amounts of fissile material compressed by a few kilotons of explosive. This method was used in the device (called ‘Fat Man’) that was dropped on Nagasaki in World War II. In this case the fissile material was plutonium.²²

Assembly techniques address the first two problems in the list above. Turning to (c), we have to make sure that fission occurs at the optimum time. In the gun assembly, this could in principle be done by holding the supercritical mass together until spontaneous neutrons start the chain reaction. However, this is not possible for the implosion method, because the compressed material will start to expand as soon as the shock wave ceases. It is therefore necessary for the design to incorporate a neutron generator, whose operation is precisely synchronised with the assembly process. There are a number of ways of doing this. In early bombs, the neutrons were obtained from the very rare reaction



which was initiated by a high flux of alpha particles from an α emitter such as ${}^{210}\text{Po}$. Both the α emitter and the beryllium sources were located at the centre of the assembly, but were kept apart until brought together by the implosion process. A drawback of this method is that strong α emitters, such as ${}^{210}\text{Po}$, have very short lifetimes measured in days, so in order to maintain a stockpile of weapons it was necessary to continually replace the neutron generator. Later weapons generated the neutrons from a fusion process. We will defer discussing the latter until we have discussed the use of fusion in weapons in Section 9.3.2 below.

Finally, we have to consider problem (d) – preventing the assembly disassembling before the chain reaction has been established. In practice, this means reducing the rate of expansion through better confinement of

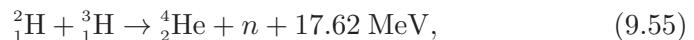
²²Plutonium is a very complex element, with several distinct solid crystalline forms that convert to each other relatively easily and not all of which are suitable for mechanical working. In practice, a stable form of the so-called δ -phase was achieved by alloying it with a small percentage of gallium. An interesting short history of plutonium is given in Bernstein (2007).

the critical mass. Here the tamper plays an important role, because the expanding material has to drive a shock wave through this dense medium, which greatly impedes the rate of expansion.

The typical energy released in a fission weapon is of order 10^{14} J, which is usually expressed in equivalent tons of the high explosive TNT using the conversion 1 ton of TNT $\approx 4 \times 10^9$ J, i.e. about 25 kt equivalent of TNT. The largest pure fission bomb probably produced an output of about 500 kt of TNT equivalent.

9.3.2 Fission/fusion devices

The rate at which a given fusion reaction proceeds is a strong function of temperature (see Figures 9.7 and 9.8), but even at the temperatures occurring in stellar interiors the rates of reactions of interest to astrophysicists are far too low for the useful production of controlled power in a fusion reactor. The same is true for the explosive production of power and the most important fusion reactions for weapons are those discussed in Section 9.2.4 for controlled power production. They are: the deuterium–tritium reaction



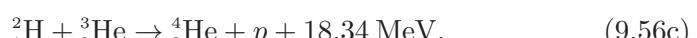
where the neutron carries off 14.06 MeV of energy; the two ${}_{1}^2\text{H} - {}_{1}^2\text{H}$ reactions (which are equally likely)



where the neutron carries off 2.45 MeV of energy, and



where the proton carries off 3.03 MeV; and the ${}_{1}^2\text{H} - {}_{2}^3\text{He}$ reaction



where the proton carries off 14.67 MeV.

Because high temperatures are required for fusion to occur, fusion weapons are sometimes also called *thermonuclear weapons*. The high temperature is produced by initially detonating a fission bomb, so a better name might be fission–fusion weapons. At the temperature produced in a fission bomb, $\sim 10^8$ K (i.e. ~ 10 keV),²³ reaction (9.55) has a rate about 100 times greater than that of (9.56a) and (9.56b) combined (see Figure 9.8) and is the only one that occurs to a sufficient extent. Since the neutron carries off most of the energy, this reaction can be used as an initiator to ‘boost’ fission bombs (as mentioned in Section 9.3.1), thus removing the need for an initiator of the $\text{Be}/{}^{210}\text{Po}$ type with its need for frequent renewal. This is a distinct advantage, but the disadvantage is that the implosion initiator is more difficult to engineer. Most modern weapons

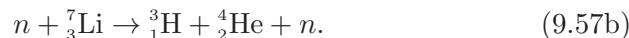
²³Compare this to the typical temperatures in stellar interiors of $\sim 1 \times 10^7$ K.

using fission, either alone or as a trigger for fusion reactions, incorporate a device called a pulse neutron tube that uses an electronic technique to generate neutrons in short pulses from $d-t$ and $d-d$ reactions. Because the technique generates copious quantities of neutrons, the device can be placed anywhere within the assembly.

The main problem with the $d-t$ reaction is the short lifetime of tritium. Since it does not occur naturally it has to be manufactured, and the process for doing this is expensive. Attention has therefore focussed on the $d-d$ reactions (9.56a) and (9.56b). Although deuterium is stable, it is difficult to store and must be either highly compressed or liquefied at very low temperatures. This problem can be circumvented by using the stable solid compound lithium deuteride, which has the added advantage that lithium can participate in the thermonuclear process via the reactions



and



Because the $d-d$ reactions have much slower reaction rates than the $d-t$ one (9.55) at the same temperature, this has to be offset by compressing the fuel, typically to densities $10^2 - 10^3$ times greater than those of normal conditions, because at a fixed temperature the rate is proportional to the square of the density.

The key to making a large fission/fusion bomb is to find a way of using the energy of a fission bomb to compress a mass of deuterium sufficiently for the $d-d$ reaction to become practical, followed by heating the mass to ignite the fusion process. The technique for doing this is a staged radiation implosion, called the Teller–Ulam configuration after its original inventors. This makes use of the fact that in a fission explosion about 80% or more of the energy is in the form of photons – X-rays and prompt gamma rays. Thus the transport of energy from the fission core is far faster than the rate of expansion of the core material. This energy can therefore be used to compress and ignite a physically separate mass of fusion fuel through radiation implosion before the ‘debris’ of the expanding trigger reaches the fusion capsule and disrupts it.

Although there is much information about thermonuclear weapons in the public domain, the full details of bomb designs are still classified. However, the principles can be illustrated by reference to Figure 9.11, which shows schematically an example of an idealised cylindrical shape. It consists of two sections: the primary one where the initial fission/fusion reaction occurs and the secondary one where the main fusion reaction takes place. The structure of the primary part is a core of plutonium surrounded by a layer of beryllium to reflect neutrons back into the assembly to increase the efficiency of the fission process, with an outer shell of conventional high explosives. There will also be a booster mechanism at the core of the assembly, or elsewhere (see the discussion above). The secondary section consists of a cylinder of fusion fuel, typically lithium

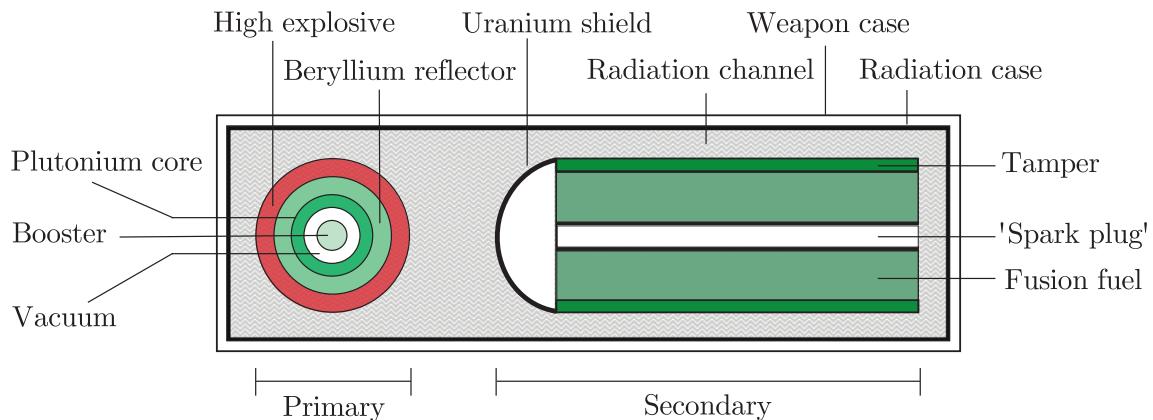


Figure 9.11 Schematic diagram of the Teller-Ulam staged explosion technique for an idealised cylindrical configuration.

deuteride, enriched using the ^6Li isotope, through which is a hollow core of plutonium about 2–3 cm in diameter, which acts as a ‘spark plug’ for fusion, as we shall see. The fusion fuel is contained by a tamper made of a heavy metal such as uranium or tungsten, and there is a shaped shield of uranium between the fuel capsule and the primary assembly. Both stages are enclosed in a radiation case made of a heavy metal, typically lead, to contain radiation. The space between this and the two stages, called the radiation channel, is filled with plastic foam such as polystyrene. Finally, there is an outer weapon case made of a light metal such as aluminium.

When the trigger explodes, most of the energy released is in the form of photons that very rapidly fill the radiation channel and fully ionise the plastic foam within it. One of the roles of the resulting plasma is to ensure that complete thermal equilibrium is quickly established and maintained. The inner casing and outer capsule surfaces are quickly raised to very high temperatures, but the uranium shield and tamper prevent the fusion fuel from being prematurely heated. As the surface temperature of the tamper rises it expands and ‘boils off’ from the surface of the fuel capsule (called ‘ablation’). This generates an enormous pressure and causes a rapid implosion of the fusion fuel, so that its density rises, to perhaps 1000 times its initial value.²⁴ Compression of the plutonium core (the ‘spark plug’) will also occur, and although not to anything like the extent of the fusion fuel, the compression is sufficient to make it supercritical and fission will be initiated by neutrons from the fusion fuel. This second fission explosion, plus the effect of the continuing implosion, is sufficient to raise the temperature and density to a point where fusion takes place and spreads outwards. Moreover, the temperature is raised still further (to as high as 3×10^8 K), thereby increasing the efficiency, by the remains of the tamper that traps emerging thermal radiation. Finally, the neutrons

²⁴This is similar to the inertial confinement method for controlled fusion mentioned in Section 9.2.4.

emerging from the fusion reactions (9.55) and (9.56a) are of sufficient energy to induce fission in the highly compressed fusion tamper (typically natural uranium) and can contribute substantially to the total output of a fission/fusion/fission bomb.

Many variations have been proposed on the above construction, some of which have been implemented. For example, fission of the uranium tamper during the final stages can in practice yield most of the energy output, but at the expense of producing large amounts of radioactive fallout. A ‘cleaner’ device can be built by using a nonfissile material for the tamper at the expense of reducing the yield. Devices have been made where as much as 90% of the yield comes from the pure fusion process. On the other hand, the nature of the contamination to the ground over which the bomb is detonated (the degree of ‘dirtiness’) can be varied by ‘salting’ the tamper with a variety of materials designed to produce radioactive isotopes of different lifetimes when exposed to neutrons from the fusion stage. Finally, the output of the fusion stage of a two-stage device can be used to compress and explode a third stage, with each stage increasing the output 10–100-fold. The largest device built to date was a three-stage weapon with an output equivalent to 50 megatons of TNT.

9.4 Biomedical applications

Biomedicine is a very large subject and for reasons of space we will therefore concentrate on just two topics: the therapeutic and diagnostic uses of radiation, with particular reference to medical imaging techniques and the treatment of cancers. These use accelerators, precision detectors, and techniques that were developed, and are still improving, in nuclear and particle physics. However, before that we briefly discuss the effect of radiation on living matter.

9.4.1 Radiation and living matter

Exposure of living tissue to radiation is a complex process. Immediate physical damage may be caused by the initial deposition of energy, but in addition there can be secondary damage due to the production of highly active chemicals. The latter may not be evident in full for several hours after exposure. For low levels of radiation this effect is the only one. High levels of damage may lead to the rapid death of living cells, but cells that survive in a damaged form may still have serious consequences. However caused, damage to the DNA of the nucleus of cells can result in long-term biological effects, such as cancer or genetic abnormalities, which may not reveal themselves for years, even decades, after the original exposure.²⁵

²⁵This has been known for a long time. For example, Hermann Muller was awarded the 1946 Nobel Prize in Physiology and Medicine for his discovery that mutations could be induced by X-rays.

To make statements like ‘low-level’ and ‘high-level’ used above meaningful needs a more detailed discussion, including the question of how dosages are defined. We will do this only briefly. Roughly speaking, the average absorbed dose D is the total energy deposited per unit mass of tissue. This is measured in ‘grays’, defined by $1 \text{ Gy} = 1 \text{ J kg}^{-1}$, which has largely replaced the older unit of the ‘rad’ ($1 \text{ Gy} = 100 \text{ rad}$). However, in practice, biological effects depend not only on the total dose, but also on other factors, including the type of radiation, the rate of deposition, and whether the whole organ is uniformly irradiated. These considerations lead to the definitions used in medical applications of *equivalent* and *effective doses*, where multiplicative weighting factors are included to take account of different types of radiation and different organs being irradiated. To distinguish these latter doses from the simple absorbed dose, the sievert (Sv) unit is used, also defined as 1 J kg^{-1} because the weighting factors are dimensionless. For example, the dose rate absorbed in tissue at a distance r from an external source of activity \mathcal{A} emitting gamma rays of energy E_γ is given approximately by

$$\frac{dD}{dt} (\mu\text{Sv h}^{-1}) \approx \frac{\mathcal{A}(\text{MBq}) \times E_\gamma(\text{MeV})}{6r^2(\text{m}^2)}, \quad (9.58\text{a})$$

and for an internal source emitting radiation of energy E_R , the effective dose rate for an organ of mass M is

$$\frac{dD}{dt} = \frac{\mathcal{A} E_R f}{M}, \quad (9.58\text{b})$$

where f is the fraction of the energy deposited in the organ.

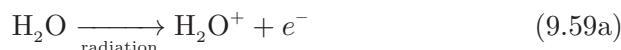
To get some idea of scale, the total annual effective dose to the UK population is approximately $2600 \mu\text{Sv}$, of which 85% is due to naturally occurring background radiation, although much higher doses can occur in specific cases, for example workers whose occupational activities expose them to radiation on a daily basis, or people who live in areas rich in granite rocks (which emit radon, the source of about half of the background). The recommended limit for additional whole-body exposure of the general population is 1 mSv yr^{-1} .²⁶

The primary deposition of energy is due, as in nonliving matter, to ionisation and excitation of atoms and molecules in the path of the radiation. This occurs on a timescale of 10^{-16} s or less and was described in Chapter 4. We can draw on that discussion here, bearing in mind that living tissue consists mainly of light elements, and in particular has a high proportion (about 80%) of water. For heavy particles, such as protons and alpha particles, the most important process is ionisation via interactions

²⁶For a discussion of (9.58) and quantitative issues of acceptable doses for various sections of the population and to different organs, see, for example, Chapter 7 of Lilley (2001) and Chapter 11 of Dendy and Heaton (1999).

with electrons, and the energy losses are given by the Bethe–Bloch formula (4.11). The rate of energy loss by a heavy particle is high, peaking near the end of its range, and so the penetrating power is low. For example, a 1 MeV alpha particle travels only a few tens of microns and is easily stopped by skin. However, considerable damage can be caused to sensitive internal organs if an alpha-emitting isotope is ingested. An exception to the above is neutron radiation, which being electrically neutral does not produce primary ionisation. Its primary interaction is via the nuclear strong force and it will mainly scatter from protons contained in the high percentage of water present. The scattered protons will, however, produce ionisation, as discussed above. The overall effect is that neutrons are more penetrating than other heavy particles and at MeV energies can deposit their energy to a depth of several centimetres. Electrons also lose energy by interaction with electrons, but the rate of energy loss is smaller than for heavy particles. Also, because they have small mass, they are subject to greater scatter and so their paths are not straight lines. In addition, electrons can in principle lose energy by bremsstrahlung, but this is not significant in the low- Z materials that make up living material. The overall result is that electrons are more penetrating than heavy particles and deposit their energy over a greater volume. Finally, photons lose energy via a variety of processes (see Section 4.3.4), the relative importance of which depends on the photon energy. Photons are very penetrating and deposition of their energy is not localised.

In addition to the physical damage that may be caused by the primary ionisation process, there is also the potential for chemical damage, as mentioned above. This comes about because most of the primary interactions result in the ionisation of simple molecules and the creation of neutral atoms and molecules with an unpaired electron, called *free radicals*. These reactions occur on much longer timescales of about 10^{-6} s. For example, ionisation of a water molecule produces a free electron and a positively charged molecule:



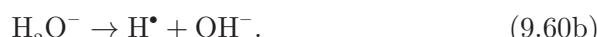
and the released electron is very likely to be captured by another water molecule producing a negative ion:



Both ions are unstable and dissociate to create free radicals (denoted by black circles):



and



Free radicals are chemically very active, because there is a strong tendency for their electrons to pair with one in another free radical. Thus the free radicals in (9.60a) and (9.60b) will interact with organic molecules

(denoted generically by RH, as the combination of the free radical R[•] with hydrogen) to produce organic free radicals:



and



The latter may then induce chemical changes in critical biological structures (e.g. chromosomes) some way from the site of the original radiation interaction that produced them. Alternatively, the radiation may interact directly with the molecule RH, again releasing a free radical R[•]:



Finally, if the irradiated material is rich in oxygen, yet another set of reactions is possible:



followed by



with the release of another free radical. This is the *oxygen effect* that complicates the treatment of tumours and is considered below.

Fortunately, for low-level radiation, living matter has the ability itself to repair much of the damage caused by radiation and it does not lead to permanent consequences. This ability is an important factor in determining the *relative biological effectiveness* (RBE) of different types of radiation. Indeed, if cells did not have this capacity, then life may not have evolved in the way it did, because we are all exposed to low levels of naturally occurring radiation throughout our lives (which may well have been far greater in the distant past) and the modern use of radiation for a wide range of industrial and medical purposes has undoubtedly increased that exposure. However, the repair mechanism is not effective for high levels of exposure.

9.4.2 Radiation therapy

In developed countries, approximately 40% of the population can expect to be diagnosed with cancer during their lifetime and its treatment is a major medical activity. Radiation therapy is a long-standing treatment for cancer, often combined with chemotherapy (the administration of cell-destroying chemicals) and/or surgery. Other treatments, particularly for localised cancers, involve the introduction of a radionuclide, either injected via a needle or by ingestion of a compound containing the radionuclide. Chemicals that preferentially target specific organs or bones are commonly used, and the production of suitable radionuclides using cyclotrons is an important biochemical industry using knowledge of nuclear physics. By damaging DNA, the ability of the cell to reproduce is inhibited and so

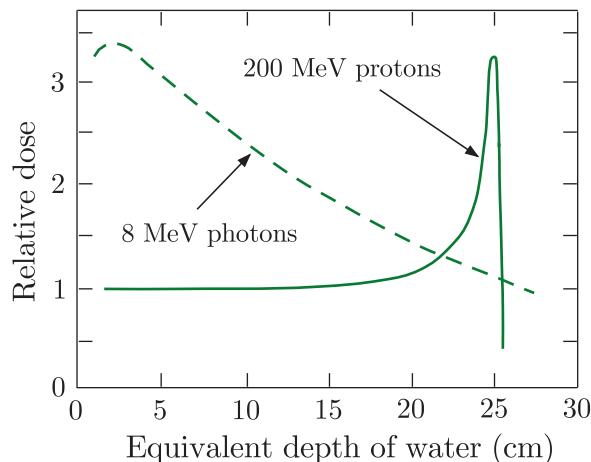


Figure 9.12 Relative absorption of photons and protons as functions of equivalent depth of water.

tumour tissue can in principle be destroyed. Of course the same applies to healthy tissue, so when using radiation in a medical environment, a balance has to be struck between the potential diagnostic and/or therapeutic benefits and the potential deleterious effects of damage done by the radiation. This is a particularly delicate balance for cancer treatment, because highly oxygenated tissue has a greater sensitivity to radiation and unfortunately many tumours are less oxygenated than healthy tissue and therefore more resistant to radiation.

In the context of radiation therapy, an important quantity is the *linear energy transfer* (LET) that measures the energy deposited per unit distance over the path of the radiation. Except for bremsstrahlung, LET is the same as dE/dx discussed in Chapter 4. Examples of high-LET particles are heavy ions and α particles, which lose their energy rapidly and have short ranges. LET values of the order 100 keV/mm and ranges 0.1–1.0 mm are typical. Low-LET particles are electrons and photons with LET values of order 1 keV/mm and ranges of order 1 cm.

Most current cancer therapy work uses low-LET particles. Thus around half of all cancer patients are treated with photon radiation²⁷ in multiple sessions, extending over several weeks. The photons are produced by electron linacs, of which there are about 10 000 worldwide. A major disadvantage of using photons is that their absorption in matter, after an initial increase, is roughly exponential (see Figure 9.12), so that most of the radiation is absorbed before the tumour is reached, with subsequent damage to healthy tissue. For example, a beam of 8 MeV photons delivers a maximum dose at about 2–3 cm, but at the depth of about 25 cm water equivalent, the depth of deep-seated tumours, the dose is only about one-third of its maximum. The exposure of healthy tissue can be reduced, while maintaining the total dose to the tumour, by directing a beam at a

²⁷In radiation treatments the photons typically have energies in the MeV range, but despite this, in a medical environment the photons are often still referred to as ‘X-rays’ rather than γ -rays.

cancer site from several directions using a rotating gantry and by modern developments such as the ‘gamma knife’ used for treating brain tumours, which can precisely focus an intersecting array of gamma beams. Other techniques include giving the dose in several stages, so that the outer regions of the tumour, which are relatively oxygen-rich, are successively destroyed as they become reoxygenated.

Nucleons are examples of high-LET particles. They overcome a cancer cell’s resistance to radiation damage more effectively than low-LET photon or electron radiation. Thus neutrons appear to be biologically more effective in killing cancers than are many other forms of radiation, especially in oxygen-poor cells, and are particularly effective against large tumours. However, neutron therapy is not widely used because of the problem of producing a strongly collimated beam, plus the difficulty of ensuring that the energy is deposited primarily at the tumour site. Neutrons also share with low-LET radiation the drawback that their attenuation in matter is exponential. Nevertheless, some facilities do exist, such as the Neutron Therapy Facility at Fermilab, which has one of the highest energy and deepest penetration beams available.

In contrast to photons, the rate of energy loss of protons and other charged particles increases with penetration depth, culminating at a maximum, the *Bragg peak*, close to the end of their range (see Figure 9.12). In principle, this means that a greater fraction of the energy would be deposited at the tumour site and less damage caused along the path length to the site.²⁸ The suggestion of using protons as the basis for cancer therapy was first made in 1946 by Robert Wilson, one of the pioneers of accelerator physics. However, it was not until the 1990s that its potential was recognised by radiation oncologists and eventually in 1993 the first facility dedicated to proton therapy was established. Fittingly, this used a synchrotron built by the Fermi Laboratory, the particle physics laboratory that Wilson had founded and where he had served as its first Director.

There have been important and continuing technical developments in proton therapy in the decade or so since the first dedicated facility was opened, aimed at focussing more of the beam energy on to the target site while reducing the energy deposited in healthy cells *en route* to the tumour. For example, to reach deep-seated tumours, protons with energies of about 200 MeV are necessary. If circumstances require, for example for large tumours, it is possible to widen and flatten the maximum in the proton curve shown in Figure 9.12 by using a system of scattering foils to spread the very sharp Bragg peak laterally across the dimensions of the tumour while maintaining a constant range (so-called ‘passive scanning’). Modern developments aim to produce a dose distribution tailored in three dimensions by using an ‘active scanning’ technique. Here the

²⁸For a review of the use of proton beams in cancer therapy see, for example, Paganetti (2017).

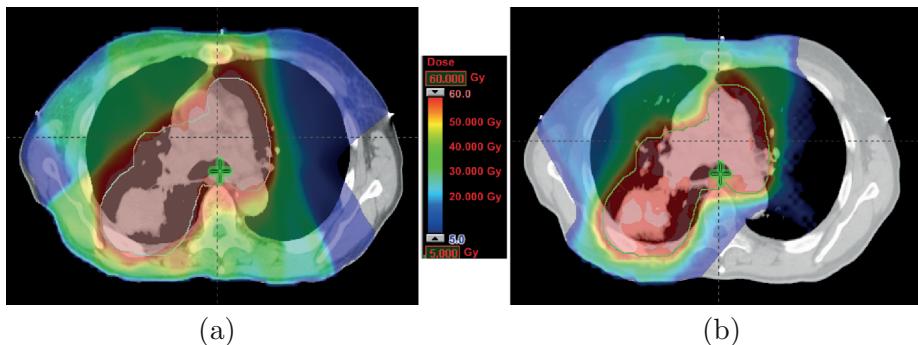


Figure 9.13 Treatment plans for an extensive non-small-cell lung cancer (shown in brown) using (a) an intensity-modulated photon beam and (b) an intensity-modulated proton beam. The dosages in Grays (Gy) are shown on the scale. There is a clear reduction of radiation in the tissues surrounding the tumour when using protons. Source: Zhang et al. (2010). Copyright (2010), Elsevier, reprinted with permission.

dose delivery is achieved by the sequential superposition of single pencil beams of protons, each of which deposits its energy in a local hot spot at the Bragg peak. Lateral scanning is performed either by using magnets or by moving the table where the patient is located. These options can be combined with a system that allows the beam to be rotated about the patient. Depth modulation can be achieved, for example, by varying the beam energy. These advanced techniques have been implemented at PROSCAN, a proton-therapy facility in Switzerland using a new purpose-built compact 250 MeV superconducting cyclotron with three beam lines providing treatment for different types of cancer.

The advantages of using protons over photons in practice is illustrated in Figure 9.13, which compares the treatment plans (i.e. simulations of the pattern of radiation that the patient would receive) for treating a case of advanced non-small-cell lung cancer. Figure 9.13a shows a plan using a photon beam shaped to closely match the area to be irradiated, a technique known as intensity-modulation. The amount of radiation received by nearby areas is seen to be a substantial fraction of the dose received by the region of the cancer. This is contrasted with the results of Figure 9.13b, which is for treatment using a single intensity modulated proton beam. Although there is some unwanted exposure at the input site (which could be lessened by a system of multiple beams or a rotating beam), the radiation energy is concentrated much more within the area of the tumour.

In his original paper, Wilson also suggested using heavy ions to deposit an even greater fraction of the available energy at the site of the tumour. For example, carbon ions at the beginning of their path in tissue have a low rate of energy loss more like an LET particle, but near the end of their range the local ionisation increases dramatically as it approaches the Bragg peak. The depth profile of the RBE can be tuned by varying the beam energy, as shown in Figure 9.14. Thus an even greater fraction

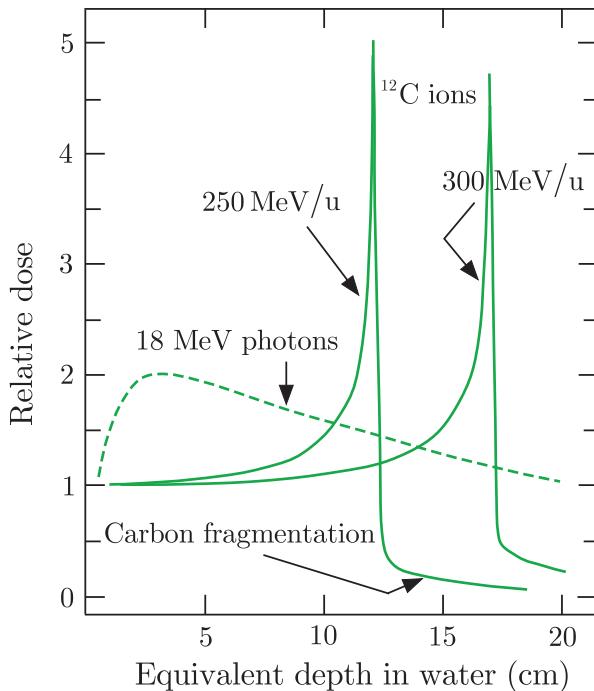


Figure 9.14 Depth profiles of ^{12}C ions at for two energies, 250 MeV/u and 300 MeV/u, compared with that for 18 MeV photons. Source: Adapted from Kraft (2000). Copyright (2000) Elsevier, reprinted with permission.

of the beam energy can be deposited at a precise depth without the danger of massive destruction of healthy tissue in the entrance channel. Experiments have shown that for carbon ions the section of particle track with increased RBE coincides with the few centimetres up to the Bragg peak. The damage caused at the tumour site is extensive and irreparable, whereas the damage caused in the entrance channel is relatively slight and is largely repairable. Thus carbon ions are an ideal projectile for cancer therapy. There is also interest in using helium ions, which, being lighter, would require a smaller accelerator.

Another potential advantage of using carbon ions is that nuclear interactions along the path length will convert a small fraction of the nuclei to radioactive positron-emitting isotopes that could then be used to image the irradiated region (using the PET technique to be discussed in Section 9.4.3(b)) to high precision, thus enabling the millimetre precision of the focussed carbon ion beam to be exploited and hence increase the effectiveness of the treatment programme. A slight disadvantage is that there would be some fragmentation of the beam that would produce a radiation dose of approximately 10–20% immediately downstream of the target. There has been considerable interest in turning these ideas into practice and a small number of facilities are currently operational, with others being considered. To date about 20 000 patients have been treated with carbon ion therapy.

The use of carbon ions may not be confined to treating tumours. Studies on animals have shown that carbon beams offer a possible treatment of cardiac arrhythmias, where abnormal electrical patterns can lead to

sudden heart failure or permanent damage as a result of a stroke. This is a major, and increasing, problem in developed countries. The procedure induces specific changes to cardiac tissue that prevents the transmission of electrical signals and permanently interrupts the propagation of disruptive impulses. If established in humans, this would mean the usage of drugs and ablation of heart tissue via catheters, with their attendant risks, could be avoided.

Looking further into the future, physicists at CERN have investigated the possibility of using antiprotons in cancer therapy. These can be delivered with the same precision as protons, but have the added advantage that at the end of their range they will annihilate with nucleons to deposit additional energy. An initial experiment at CERN (ACE – Antiproton Cell Experiment) has suggested that antiprotons are about four times more effective as protons at destroying cells. Progress in this area is likely to be slow, because at present only CERN has an antiproton beam of sufficiently low energy and high quality to study this type of application.

Although they have great potential, a major problem with using particle beams is the practical one of access to suitable accelerators, which are large and very expensive, although there are considerable efforts being made to reduce the size and cost of such machines by working in collaboration with accelerator physicists to implement new technologies. To date only 67 centres exist worldwide offering proton therapy, with a similar number under construction, although it is estimated this will rise to over 100 by 2020. All are in developed countries. Although these are important advances, they have to be taken in context. Current usage has concentrated on specific classes of patients, for example particularly difficult paediatric cases, and by 2016 approximately 150 000 patients in total worldwide had been treated by this method. Although that number is increasing very rapidly, and ways of using the techniques to treat a wider range of tumours may well be developed, it is to be compared to several million patients who are treated by conventional photon radiation *each year*.

9.4.3 Medical imaging using ionising radiation

Techniques for producing images for diagnostic purposes usually use photons. There are several methods available and in this section we will describe the principles of some of the main ones, but without technical details.²⁹

9.4.3(a) Imaging using projected images

The use of an *external* source of radiation for medical imaging is of long standing and well known. Basically, the system consists of a source of photons in the X-ray energy region placed some distance in front of the

²⁹A short useful review of the whole field is by Hendee (1997).

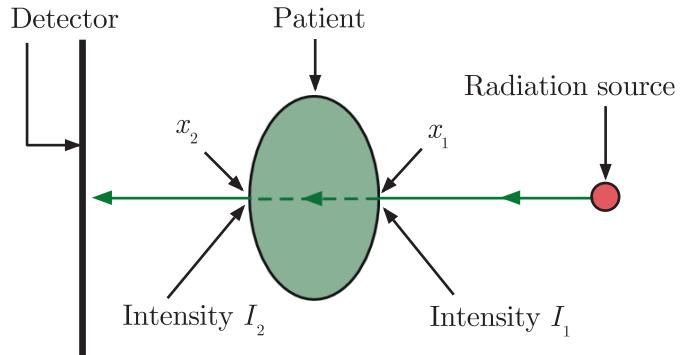


Figure 9.15 Basic layout for imaging using an external source.

patient and a detector (traditionally a special type of sensitive film, but now usually electronic so that the picture can be viewed directly on a computer screen) placed immediately behind the patient. Because the radiation is absorbed according to the exponential law (4.21), a measurement of the intensities just before and after the patient yields information on the integrated attenuation coefficient $\mu \equiv 1/\lambda$ of the photons in the body, where λ is the mean free path.

Thus, referring to Figure 9.15, we have for the ray shown, using (4.21),

$$\ln(I_1/I_2) = \int_{x_1}^{x_2} \mu(x) dx. \quad (9.64)$$

The full image reveals variations of this integral only in two dimensions and thus contains no depth information. A three-dimensional effect comes from overlapping shadows in the two-dimensional images and part of the skill of a radiologist is to interpret these effects. The attenuation coefficient is dependent on the material and is greater for elements with high Z than for elements with low Z . Thus X-rays are good at imaging bone (which contains calcium with $Z = 20$), but far less useful for imaging soft tissue (which contains a high proportion of water). The technique is commonly used to determine the extent and type of bone fractures, but it can also be used to visualise the structure of some soft issue organs by using a radio-opaque contrast medium, which is introduced into the tissue of the organ either by ingestion or injection. Thus, for example, a barium solution can be ingested and used to help in diagnosing ulcers and some types of colon cancer in the stomach and intestines. This technique is useful in situations where more advanced diagnostic techniques are unavailable.

In a type of external imaging called *fluoroscopy*, images of internal structures are produced in a similar fashion to that above, including the use of contrast media, but using a constant input of X-rays. In early applications, the unattenuated X-rays interacted with atoms in a screen placed behind the patient via the photoelectric effect, resulting in a small fraction of visible light producing an image. Modern devices dispense with the screen and instead use an X-ray image intensifier, which is a device that greatly enhances the intensity of the image, coupled directly

to a CCD camera system so that the images can be easily viewed in normal light conditions. Because the images are produced in real time, the technique can be used in a variety of applications, such as guiding catheters, placing metalwork in orthopaedic surgery, and angiography of the leg, heart, etc.

Images can also be obtained using an *internal* source of radiation. This is done by the patient ingesting, or being injected with, a substance containing a radioactive γ -emitting isotope. As modern photon detectors are very sensitive, the concentration of the radioisotope can be very low and any risk to the patient is further minimised by choosing an isotope with a short lifetime. If necessary, the radioisotope can be combined in a compound that is known to be concentrated preferentially in a specific organ if that is to be investigated, for example iodine in the thyroid. In practice more than 90% of routine investigations use the first excited state of $^{99}_{43}\text{Tc}$ (technetium) as the radioisotope. This has a lifetime of about 6 hours and is easily produced from the beta decay of $^{99}_{42}\text{Mo}$, which has a lifetime of 67 hours. The usefulness of this metastable state (written $^{99}\text{Tc}^m$) is that it emits a single 140 keV photon with negligible β -decay modes, decaying to the very long-lived (2×10^5 yr) ground state.

Because the radiation is emitted in all directions, a different technique is used to detect it. The patient is stationary and is scanned by a large-area detector consisting of a collimated single-crystal scintillator, usually NaI, the output from which is viewed by an array of photomultipliers (PMTs) via a light guide (see Section 4.4.2). A schematic diagram of such a gamma camera is shown in Figure 9.16. The output from the scintillator

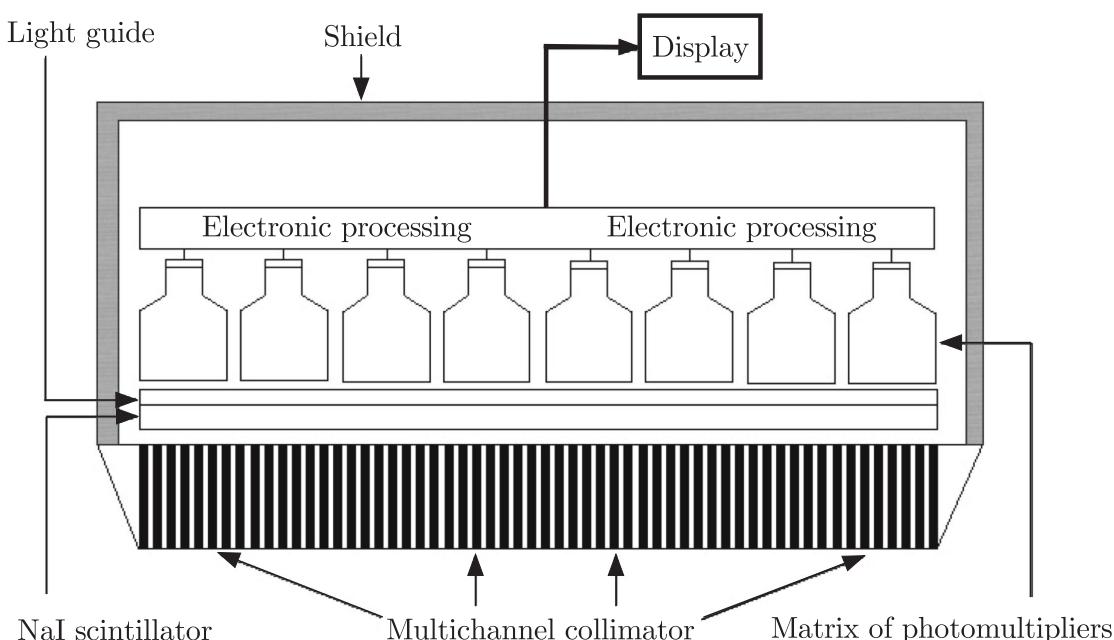


Figure 9.16 Schematic diagram of a gamma camera.

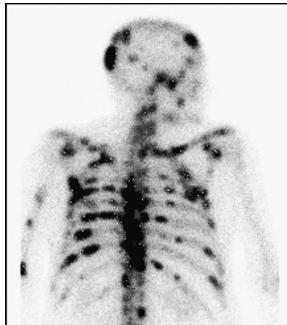


Figure 9.17 Part of a whole-body skeletal image obtained using $^{99}\text{Tc}^m$ MDP. Source: Image courtesy of Prof. R.J. Ott, Royal Marsden Hospital, London.

is received in several PMTs and the relative intensities of these signals depend on the point of origin. The signals can be analysed to locate the point to within a few millimetres. The collimator restricts the direction of photons that can be detected and combined with the information from the PMTs, the overall spatial resolution is typically of order 10 mm, provided the region being examined has an attenuation coefficient that differs by at least 10% from its surroundings.

Radioisotope investigations principally demonstrate function rather than anatomy, in contrast to X-ray investigations that show mainly anatomical features. Thus better images of soft tissue, such as tumours, can be obtained than those obtained using external X-rays, because the ability of the tumour to metabolise has been exploited, but the exact location of the tumour with respect to the anatomy is often lost or poorly defined. Figure 9.17 shows part of a whole body skeletal image of a patient who had been injected with a compound MDP that moves preferentially to sites of bone cancer, labelled with the isotope $^{99}\text{Tc}^m$. The image clearly shows selective take-up of the isotope in many tumours distributed throughout the body.

9.4.3(b) Computed tomography

A radiographic image is a two-dimensional display of a three-dimensional structure and although the overlapping images give a useful three-dimensional effect, details are always partially obscured by the superposition of information from underlying and overlying planes. The result is loss of contrast. Thus while images from the projection methods have good transverse spatial resolution, they have poor resolution in depth. A major advance that addresses this problem was made with the introduction of a new scanning technique called *computed tomography* (CT).³⁰ This enables a series of two-dimensional sections to be imaged as small as a millimetre across, even when the attenuation coefficient differs by less than 1% from its surroundings.

The principle behind the CT technique is the observation that all the information needed to construct an image of a two-dimensional section of tissue is contained in the one-dimensional projections that cover all possible directions within the plane of the slice. Thus, for example, if the slice is in the xy plane, a projected image of the slice contains information on $\mu(x, y)$ in the form of a set of line integrals of μ taken through the region in a particular direction. As the angle in the plane of the slice is varied, a different representation of $\mu(x, y)$ is obtained in the form of a different set of line integrals. Once a complete set of line integrals has been obtained there are mathematical methods (including some that have been used by particle physicists to reconstruct events from high-energy collisions)

³⁰The CT system was devised by Godfrey Hounsfield (later knighted) and first successfully introduced into medical use in 1971.

that allow the required two-dimensional function to be reconstructed.³¹ Modern high-speed computers are able to perform this construction very rapidly, so that images can now be obtained in real time and motion as fast as heartbeats can be captured.

Computed tomography may be used in conjunction with both external and internal radiation. As an example, the arrangement for a CT X-ray scan is shown schematically in Figure 9.18. In this example the patient remains stationary within a ring of several hundred detectors (solid-state scintillators are frequently used). Within this ring there is an X-ray source that moves on another ring and provides a fan of X-rays. Each alignment of the source and a detector in the ring defines a line through the patient and the recorded count rate enables a line integral to be computed from (9.64). By moving the source through its full angular range, a complete set of such line integrals is generated, enabling a two-dimensional section to be computed through the patient. This type of scanner is relatively expensive in both capital and maintenance costs and another type is more common. This differs from Figure 9.18 in having a single bank of detectors opposite the source and both source and detectors are rotated to cover the full angular range. Although the CT method can produce scans of soft tissue better than conventional X-ray projections (for example, it is widely used in making images of the pulmonary arteries to diagnose possible pulmonary embolisms), the images are achieved at the expense of the patient receiving a higher dose of potentially harmful radiation. An example of a CT X-ray scan is shown in Figure 9.19a.

CT can also be used to construct images obtained from projections from internal radiation using radioisotopes that emit a single gamma ray. This technique is called single-photon emission computed tomography (SPECT). The arrangement is in some sense the ‘inverse’ of that in Figure 9.15. Thus the source is now within the patient and the fixed ring of detectors is replaced by one or more gamma cameras designed so that they can rotate in a circle about the patient. An example of an image obtained using SPECT is shown in Figure 9.19b.

For a number of technical reasons, including the fact that the emitted radiation is isotropic, there are more stringent requirements on the gamma cameras and SPECT images have a resolution of only about 10 mm. However, although not suitable for accurate quantitative measurement of anatomy, they are of great use for clinical diagnostic work involving function. For example, the technique is used to make quantitative measurements of the functioning of an organ, i.e. clearance rates in kidneys, lung volumes, etc.

Since radionuclide imaging provides functional and physiological information, it would be highly desirable to be able to image the

³¹In the mid-1960s, Allan Cormack had independently worked on the mathematical theory and this was incorporated into Hounsfield’s scanner. They were jointly awarded the 1979 Nobel Prize in Physiology and Medicine for their work.

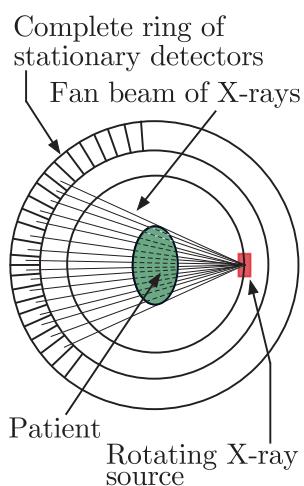
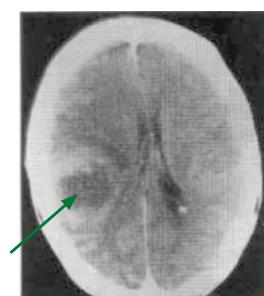
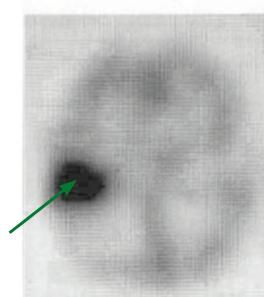


Figure 9.18 Schematic diagram of the arrangement for a CT X-ray scanner.



(a)



(b)

Figure 9.19 (a) X-ray CT scan of the brain and (b) SPECT brain scan using a $^{99}\text{Tc}^{\text{m}}$ labelled blood flow tracer, showing high perfusion in the tumour (indicated by arrows).
Source: Image courtesy of Prof. R.J. Ott, Royal Marsden Hospital, London.

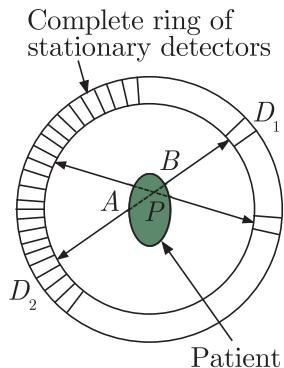


Figure 9.20 Schematic diagram of the arrangement of a PET scanner.

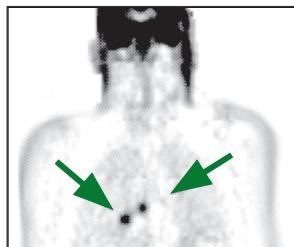


Figure 9.21 Part of a whole-body PET scan showing uptake of the chemical FDG (labelled by $^{99}\text{Tc}^m$) in lung cancer.
Source: Image courtesy of Prof. R.J. Ott, Royal Marsden Hospital, London.

concentrations of elements such as carbon, oxygen, and nitrogen that are present in high abundances in the body. The only radioisotopes of these elements that are suitable for imaging are short-lived positron emitters: ^{11}C (half-life ~ 20 min), ^{13}N (~ 10 min), and ^{15}O (~ 2 min). For these emitters, the radiation detected is the two gamma rays emitted when the positron annihilates an electron. This occurs within a few millimetres from the point of production of the positron, whose initial energy is typically less than 0.5 MeV. The photons each have energies equal to the rest mass of an electron, i.e. 0.511 MeV and emerge ‘back-to-back’ to conserve momentum. This technique is called positron emission tomography (PET) and was mentioned earlier in connection with radiation treatment using heavy ions.

The arrangement of a PET scanner is shown in Figure 9.20. If the detectors D_1 and D_2 detect photons of the correct energy in coincidence, then the count rate is a measure of the integral of the source activity within the patient along the line AB passing through P . The ring of detectors defines a plane through the patient and the complete set of data from all combinations of detector pairs contains all the information needed to generate the set of line integrals that can be converted into a two-dimensional image of the source using standard CT image reconstruction techniques. An example of an image using the PET technique is shown in Figure 9.21.

Progress in diagnostics continues. For example, work is underway on building a whole-body PET scanner. PET imaging is an important complementary method to MRI and CT scanning because it images function rather than anatomy. Conventional PET scanners only cover a small region of the body and to build up a picture of the whole body, for example to see the spread of cancer, requires multiple scans, which is inefficient and greatly increases the patient’s exposure to radiation.

This account of medical imaging has ignored many technical points. For example, there are a number of corrections that have to be made to the raw data, particularly in the SPECT technique, and the most useful radioisotopes used in PET are produced in a cyclotron, so the scanner has to be near such a facility, which considerably limits its use. The interested reader is referred to specialised texts for further details.³²

9.4.4 Magnetic resonance imaging

We conclude this discussion of imaging with an account of a remarkable technique that in a relatively short time has become one of the most sophisticated tools for clinical diagnostic work and medical research. It is not only capable of producing images of unprecedented clarity, but it does so without using potentially harmful ionising radiation.

³²See, for example, Dendy and Heaton (1999).

Magnetic resonance imaging (MRI) is based on the phenomenon of nuclear magnetic resonance (NMR).³³ It uses the fact that the quantum spin states of nuclei (strictly their associated magnetic moments) can be manipulated by magnetic fields. A brief overview of the method is as follows. Firstly, nuclear spins in tissue are aligned by a powerful static magnetic field, in the range 0.5–4 T (although fields up to 10 T are available for medical research), usually supplied by a superconducting magnet. As living tissue is dominantly water, the spins in question are mainly those of protons. Secondly, oscillating magnetic field pulses at radio frequency are applied in a plane perpendicular to the magnetic field lines of the static field, which causes some of the protons to change from their aligned positions. After each pulse, the nuclei relax back to their original configuration and in so doing they generate signals that can be detected by coils wrapped around the patient. Differences in the relaxation rates and associated signals are the basis of contrast in MRI images. For example, water molecules in blood have different relaxation rates than water molecules in other tissues.

There are several different types of MRI scan, each with its own specialised procedures and the full mathematical analysis of these is complex. We will therefore give only a rather general account concentrating on the basic physics. The interested reader is referred to more detailed texts at an appropriate level.³⁴

The proton has spin 1/2 and magnetic moment μ_p . In the absence of an external magnetic field, the two states corresponding to the two values of the magnetic quantum number $m_s = \pm 1/2$ are equally populated and the net magnetisation \mathbf{M} (i.e. the average magnetic moment per unit volume) is zero. In the presence of a static magnetic field \mathbf{B} , taken to be in the z direction, there is an interaction energy ($-\mu_p \cdot \mathbf{B}$) and the two states have different energies with different probabilities given by the Boltzmann distribution. The energy difference between the states is $\Delta E = 2\mu_p B = hf$, where f is the *Larmor* (or nuclear resonance) frequency, which is the frequency of a photon that would correspond to a transition between the two nuclear spin states. The energy difference is small. For example, for a field of 1 T, $\Delta E \approx 1.8 \times 10^{-7}$ eV and f is about 43 MHz, i.e. in the radio region of the electromagnetic spectrum. Although there is a net magnetisation in the z direction, the resultant magnetisation \mathbf{M}_0 is too small to be measured.

³³The original discovery is due to Isidor Rabi, who received the 1944 Nobel Prize in Physics for his ‘resonance method for recording the magnetic properties of atomic nuclei’. Felix Bloch and Edward Purcell shared the 1952 Nobel Prize in Physics for developing the method and their subsequent researches. Although the term NMR is still used in research environments, magnetic resonance imaging (MRI) is preferred in clinical environments to prevent patients associating the technique with ‘harmful nuclear radiation’.

³⁴See, for example, Dendy and Heaton (1999), McRobbie et al. (2003), and Hobbie (1997).

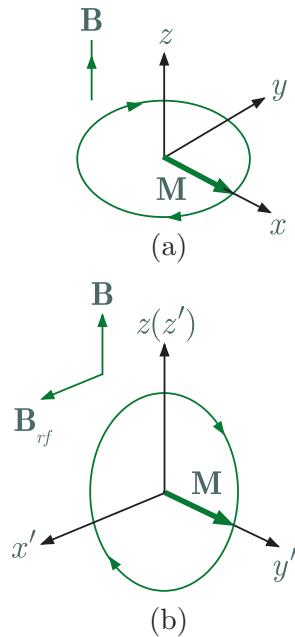


Figure 9.22 (a) Precession of the magnetisation \mathbf{M} in the xy plane under the action of a torque $\mathbf{M} \times \mathbf{B}$ resulting from an external field \mathbf{B} . (b) Motion viewed in a frame of reference (x', y', z') rotating at the Larmor frequency about the z axis. The rf pulse \mathbf{B}_{rf} applied in the x' direction has rotated \mathbf{M} so that it points in the y' direction.

The situation changes, however, if \mathbf{M} no longer points along the z axis, and a signal is generated if the magnetisation has a component in the plane orthogonal to \mathbf{B} . This is illustrated in Figure 9.22. In this figure, \mathbf{M} has been rotated to lie in the xy plane and since there is an angular momentum associated with the magnetisation, \mathbf{M} will precess about \mathbf{B} under the action of the torque $\mathbf{M} \times \mathbf{B}$.³⁵

The rotation can be achieved by applying an alternating rf magnetic field \mathbf{B}_{rf} to the sample at right angles to \mathbf{B} and at the Larmor frequency. As \mathbf{M} precesses about \mathbf{B} , one component of \mathbf{B}_{rf} rotates in phase with it. The resulting motion is complicated and is best viewed in a frame of reference rotating at the Larmor frequency about the z axis, which we label by (x', y', z') with z' parallel to z . This is shown in Figure 9.22b. The full mathematical analysis is given, for example, in Hobbie (1997) and we will just quote the result. This is that the magnetisation vector can be rotated through an arbitrary angle depending on the strength and duration of the rf pulse. In particular, it is possible to rotate it through 90° so that the magnetisation vector precesses about the x' axis, i.e. rotating with a frequency that depends on the magnitude of the rf field. As the rf pulse forces all the protons to precess exactly in phase, there will be a component of magnetisation along the y axis in the rotating frame. When the rf pulse is turned off, the system returns to equilibrium with \mathbf{M} aligned along the z axis by re-emitting the energy absorbed from the rf pulse. As it does so, the external field due to \mathbf{M} will vary with time with the same frequency and can be detected as an induced emf in a coil surrounding the patient. This is the basic MRI signal. Crucially, the frequency of the external rf field must exactly match the Larmor frequency of the protons to be excited.

The induced signal will decay as equilibrium is restored. If \mathbf{B} were uniform throughout the selected region, all the protons would precess at the same frequency and remain in phase. In that case the interaction of the proton spins with the surrounding lattice, the so-called spin-lattice interactions, would cause \mathbf{M} to relax to its equilibrium state \mathbf{M}_0 parallel to \mathbf{B} . Under reasonable assumptions, the radiated signal is proportional to the difference $(\mathbf{M}_0 - \mathbf{M})$ and decreases exponentially with a characteristic spin-lattice, or longitudinal, relaxation time T_1 . Typical spin-lattice relaxation times are of the order of a few hundred milliseconds and are significantly different for different materials, such as muscle, fat, and water. However, because there are always small irregularities in the field due to local atomic and nuclear effects, individual protons actually precess at slightly different rates and the signal decays because the component of \mathbf{M} orthogonal to \mathbf{B} (i.e. in the xy plane) decreases as the individual moments

³⁵A particle with magnetic moment μ placed in an external magnetic field \mathbf{B} will experience a torque \mathbf{T} given by $\mathbf{T} = \mu \times \mathbf{B} = \gamma \mathbf{J} \times \mathbf{B}$, where \mathbf{J} is the angular momentum vector and γ is the gyromagnetic ratio. As a result, the angular momentum vector precesses about the external field axis with a frequency $f = \mu B / J\hbar$, the *Larmor frequency*, where B , J , etc., are the moduli of the corresponding vector quantities.

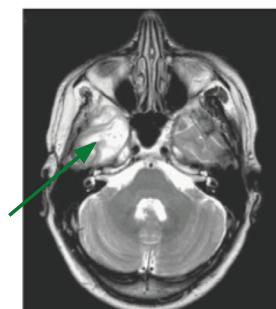
lose phase coherence. This decrease is characterised by a second time T_2 , called the spin–spin, or transverse, relaxation time. This is normally much shorter than T_1 , but again varies with material. Both relaxation times can be measured.

The above assumes that the external field \mathbf{B} is perfectly uniform, but of course the ideal is not realised in practice. The effects of macroscopic inhomogeneities in the magnetic field can be eliminated by generating so-called *spin echoes*, which may crudely be described as making two ‘orthogonal’ measurements such that the unwanted effects cancel out exactly in the sum. Many MRI imaging sequences use this technical device and again we refer the interested reader to McRobbie et al. (2003) and Hobbie (1997) for further details.

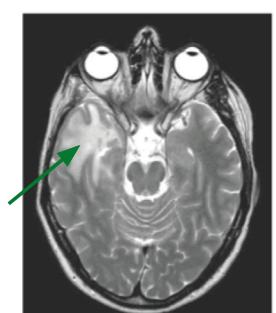
All the above assumes we are scanning the whole body. The original development of the method as a medical diagnostic technique is due to the realisation that gradients in the static magnetic field could be used to encode the signal with precise spatial information and be processed to generate two-dimensional images corresponding to slices through the tissue of the organ being examined.³⁶ The patient is placed in the fixed field \mathbf{B} pointing along the z direction. A second static field \mathbf{B}_g parallel to z , but with a gradient in the z direction, is then applied so that the total static field is a function of z . This means that the Larmor frequency (which is proportional to the magnetic field) will vary as a function of z . Thus when the rf field \mathbf{B}_{rf} is applied with a narrow band of frequencies about f_{rf} , the only protons to be resonantly excited will be those within a narrow slice of thickness dz at the particular value of z corresponding to the narrow band of frequencies. The field \mathbf{B}_{rf} is applied until the magnetisation in the slice has been rotated through either 90° or 180° depending on what measurements are to be taken. Both \mathbf{B}_{rf} and \mathbf{B}_g are then turned off.

The final step is to obtain a spatial image of the magnetisation as a function of x and y . This entails encoding the MRI signal with information linking it to a point of origin in real space. There are many ways this can be done (one utilises the CT method encountered earlier) and again we refer the interested reader to the specialised texts quoted earlier for the details. The outcome is that \mathbf{M} and the two relaxation times can both be measured. All three quantities vary spatially within the body and can give valuable biomedical information. For example, relaxation times are usually different for tumour tissue compared to normal tissue. In some areas MRI scans have considerable advantages over other forms of imaging. Thus, the contrast of soft tissue is much better than CT scans, leading to very high quality images, especially of the brain. Examples of such images are shown in Figure 9.23.

Progress in the MRI technique in medicine has been rapid. For example, a variation known as functional MRI (fMRI) has been developed



(a)



(b)

Figure 9.23 Two MRI scans of a brain: (a) T1-weighted, (b) T2-weighted, showing a frontal lobe tumour. Source: Images courtesy of the MRI Unit of the Royal Marsden NHS Foundation Trust, London.

³⁶This discovery was first made by Paul Lauterbur and an analysis of the effect was first made by Sir Peter Mansfield. They shared the 2003 Nobel Prize in Physiology and Medicine for their work in establishing MRI as a medical diagnostic technique.

that exploits the paramagnetic behaviour of deoxyhaemoglobin in red blood cells. When in a magnetic field, a blood vessel containing deoxyhaemoglobin distorts the field in its immediate environs, with the degree of distortion increasing with the concentration of deoxyhaemoglobin. This distortion affects the behaviour of water protons in the environs and, consequently, the magnetic-resonance signal arising from these protons. Neural activation of a region of the brain stimulates increased arterial flow of oxygenated blood, thereby decreasing the concentration of deoxyhaemoglobin in the region. Changes in the magnetic resonance signal can be detected and displayed as functional-MRI images. These so-called BOLD (blood-oxygen-level dependent) images enable studies to be made of the way the brain works by taking MRI images in real time while the patient is performing specific tasks. In this way areas of the brain can be studied that are associated with particular activities or sensations.

As another example, the gases ${}^3\text{He}$ and ${}^{129}\text{Xe}$ have the magnetic properties needed for MRI and the atomic structure needed to retain their polarisation for hours at a time. They can be introduced into lungs, allowing MRI studies of lung function. Because of the strong signal provided by the polarised nuclei in the gas atoms, the MRI scans are short and can be synchronised with breathing. Developments are also being made towards general high-speed imaging, which would be useful for claustrophobic patients and children who are unable to be in the confined environment of a conventional MRI magnet for sometimes up to an hour.

MRI is believed to be intrinsically safe at the field intensities currently used.³⁷ Although there is no direct proof of this, no contrary evidence has emerged during the years that MRI has been in use. However, because of the presence of high magnetic fields, care must be taken to keep all magnetic objects away from the scanner. This means that patients with heart pacemakers, or other implants incorporating magnetic materials, cannot in general be scanned and care has to be taken to screen out people who have had an occupational exposure to microscopic fragments of steel (such as welders) as these may well have lodged in critical organs such as the eyes, which could be seriously damaged if the fragments moved rapidly under the action of the very strong magnetic field.

The available magnetic field strength in MRI continues to increase, with the largest to date being 9.4 T in a machine installed at the Jülich Research Centre in Germany. This machine is also able to take PET images, so that information on structure (from the MRI image) and function (from the PET image) can be acquired simultaneously. Perhaps the greatest potential of all lies in the imaging of nuclei other than hydrogen, particularly the phosphorus nucleus. Phosphorus is a major constituent of the molecules adenosine triphosphate (ATP) and phosphocreatine, which

³⁷At higher magnetic fields there are limitations due to heat generation from the absorption of rf energy.

mediate the transfer of energy in living cells. From knowledge of such concentrations it is possible to infer the metabolic status of internal organs, and it may eventually be possible to add this capability to an image.

Although MRI systems have been used exclusively for imaging, there is a range of potential applications beyond this. For example, an experiment in 2007 has shown that the three orthogonal gradient coils inside the bore of a conventional MRI system, typically used for slice selection and signal encoding, can also induce a three-dimensional directional magnetic force sufficient to propel an object made of ferromagnetic material. This has been demonstrated by navigating a small metal sphere in an artery of a laboratory animal. Such a technique could be used, for example, to clear restricted arteries, as an alternative method to treat aneurysms, or to precisely deliver drugs.

The future will undoubtedly see both an improvement in the quality of conventional MRI images and a growing diversity of applications in clinical practice.

9.5 Further applications

In the introduction to this chapter, we explained that for reasons of space and relevance we would only discuss a few important applications in detail. However, the range of applications following from developments in nuclear and particle physics is very broad, and so we will conclude with a short overview of some other applications. This is not a comprehensive survey, but just a small sample to illustrate the wide range of applications to areas that include: health and life sciences, geosciences, archaeology, energy production, national security, finance, and many others. It is divided into sections for convenience, but they overlap because one section often uses more than one technique derived from nuclear and particle physics.

9.5.1 Computing and data analysis

Particle physics experiments generate unprecedented amounts of data that require advanced computing technology to analyse. To quickly process these data, more than two decades ago particle physicists pioneered the construction of low-cost computing ‘farms’, a group of connected servers housed in one location, which evolved into the modern Internet. To deal with the increased computing demands of the LHC experiments, particle physicists have created the world’s largest Grid computing system. This combines the strength of hundreds of thousands of individual computers, spanning more than 100 institutions in 36 countries and pushing the boundaries of global networking and distributed computing.

These techniques have been exploited by other disciplines that also have large databases. One area is medicinal chemistry, where very large numbers of chemical compounds are investigated for their potential use as medicinal drugs. For example, the molecular structure of 46 million

compounds were investigated in producing a new drug for malaria, a condition that still kills about 600 000 people worldwide annually and temporarily incapacitates another 200 million. Medical researchers also have very large databases of symptoms and diagnoses that need to be ‘mined’ rapidly so that diseases can be quickly identified and optimal treatment plans formulated. The Grid technology has many other uses, including enabling large databases of images, such as architectural plans and financial information, to be managed and rapidly searched.

9.5.2 Archaeology and geophysics

A well-known application in archaeology is the use of the radioactive decay law (2.62) to date ancient specimens using the properties of radioactive nuclei. For organic specimens, such as wood and bone, carbon is usually used. Carbon-14 is a radioactive isotope of carbon that is produced by the action of cosmic rays on nitrogen in the atmosphere. If the flux of cosmic rays remains roughly constant over time, then the ratio of ^{14}C to the stable most abundant isotope ^{12}C reaches an equilibrium value of about $1 : 10^{12}$. Both isotopes will be taken up by living organisms in this ratio, but when the organism dies there is no further interaction with the environment and the ratio slowly changes with time as the ^{14}C nuclei decay to ^{14}N with a lifetime of 8.27×10^3 yr. Thus, if the ratio of ^{14}C to ^{12}C is measured, the age of the specimen may be found.³⁸ The actual measurements can be made very accurately because modern mass spectrometers developed in nuclear physics can directly measure very small differences in the concentrations of ^{14}C and ^{12}C using only milligrams of material. In principle, this technique gives absolute dates. Nevertheless, in practice, corrections are made to agree with independent relative datings, by comparing with other estimates if available, because cosmic ray activity is not strictly constant with time.

The use of the radioactive decay law can also be used to give absolute dates for nonorganic specimens, principally rocks and fossils, by choosing different radioactive nuclei. Examples are the isotopes ^{40}K , with a half-life of 1.3×10^6 yr, and ^{235}U , with a half-life of 0.7×10^6 yr.

Another important problem in geophysics is to understand the nature of the heat source within the Earth. One contribution to this is radioactive decay, but the details are still sketchy. In principle, it should be possible to understand this better by detecting *geoneutrinos* – the neutrinos that are products of radioactive decays and which travel from the Earth’s core to the surface and beyond. Their detection could give important information about how much radioactive material exists in the core, and collaborations of physicists and geologists are starting to look at this question using neutrino detectors, including the KamLAND detector described in

³⁸This method of using radioactive carbon to date ancient objects was devised by Willard Libby, for which he received the 1960 Nobel Prize in Chemistry.

Chapter 4. At present very few geoneutrinos have been detected, but in a few years time detectors with much greater masses of detector material will become available, including JUNO in China, which will have 20 times the detector mass as KamLAND.

An elementary particle that has been used in archaeology is the muon, which has been used to probe for subterranean structures. The technique was first used almost 50 years ago by Luis Alvarez (a Nobel Laureate in Physics) to search for hidden cavities within one of the great pyramids of Egypt. Since then it has been used for many different applications, including searching for hidden cavities in other pyramids, and the technique has been refined by using arrays of plastic scintillators, of the type developed for use in particle physics, to enable cavities as small as 2–5 m to be detected. Similar techniques have been used to estimate the lava content of volcanoes as part of the process of determining when they might erupt.

9.5.3 Accelerators and detectors

Detectors have wide use in medicine. For example, in PET scanning discussed earlier the technology arose directly from detectors developed for particle physics experiments, and silicon tracking detectors are now used in neuroscience experiments to investigate the workings of the retina for development of retinal prosthetics for artificial vision. The reconstruction of MRI images borrows from computer image reconstruction techniques pioneered in particle physics.

In earlier sections we discussed the use of cyclotrons and linacs to produce beams of particles for use in cancer therapy, but another of their uses is to produce radioactive isotopes for both diagnostic and treatment purposes.³⁹ For example, by inserting such radioisotopes as ¹⁴C and tritium, it is possible to obtain information on how molecules move through the body, what types of cells contain receptors⁴⁰ and what kinds of compounds bind to these receptors. Radioisotopes are indispensable tools for the forensic technique of DNA fingerprinting and in the Human Genome Project. Intense beams of X-rays from accelerators are also used to determine the precise 3D structure of viruses and mutations that cause some diseases and so help develop effective therapies, for example by screening potential drug candidates for their suitability in particular circumstances, as mentioned above. Physicists are working in collaboration with medical scientists to improve the safety of artificial heart valves by designing a new material bombarded with silver ions from an accelerator. The treated surface of the material keeps the body from identifying the valve as an invader and surrounding it with potentially dangerous extra tissue.

³⁹A review of the history of the contributions of particle physics to medicine applications is given in Dosanjh (2017).

⁴⁰A receptor is a protein molecule inside the cell that causes a response in the cell when it receives a chemical signal from outside the cell.

Accelerator and detector technology has many applications in industry. Some examples are: weld inspection in aircraft manufacture; chemical investigation of the detailed structure of the polymer material used in sanitary devices, resulting in a better superabsorbent product; the use of electron beams to produce scratch- and stain-resistant furniture; and using such beams to produce ‘cross-linking’ chemical bonds in materials used for paving so that they do not degrade under the action of cold or heat. Accelerators are also widely used in the food industry to produce the strong, heat-shrinkable film that fruits and vegetables, and other edibles are frequently wrapped in. Radiation from accelerators is used to sterilise surfaces and equipment used in food production (this is also done in medical environments).

Of particular interest is the use of neutrino detectors to monitor the activity of nuclear reactors from outside the containment region surrounding the core. This can provide data on the power output and fissile content of the reactor, information that is useful in the context of the important problem of nuclear proliferation. Several test devices have been built and are currently in operation.

9.5.4 Industrial applications

Power transmission using cables made of superconducting material are far more efficient than conventional cables due to their minimal power losses. Superconducting technology advances in particle physics will help advance this industry, offering an opportunity to meet continued power needs in densely populated areas where underground copper transmission lines are near their capacity. Medical physicists also use experience in the design of large superconducting magnets gathered in nuclear and particle physics laboratories to improve PET and MRI scanners. Magnetically levitating trains use powerful superconducting magnets to suspend and propel the train and are seen as a fast and quiet method of transport. The superconducting magnet technology developed for particle physics will play a key role in the future of this mode of travel.

Turbulence is a dominant factor determining the performance of virtually all fluid systems, including long-distance oil pipelines, fuel injection systems, and models for global weather prediction, although it remains poorly understood and modelled. Silicon strip detectors and low-noise amplifiers developed for particle physics are used to detect light scattered from microscopic tracer particles in a turbulent fluid. This technique has permitted detailed studies of turbulence on microscopic scales and has considerably advanced understanding in this important problem.

Problems 9

- 9.1** The fission of ^{235}U is induced by a neutron and the fission fragments are $^{92}_{37}\text{Rb}$ and $^{140}_{55}\text{Cs}$. Use the SEMF to calculate the energy released (in MeV)

per fission. Ignore the (negligible) contributions from the pairing term. The reaction is used to power a 100 MW nuclear reactor whose core is a sphere of radius 100 cm. If an average of one neutron per fission escapes the core, what is the neutron flux at the outer surface of the reactor in $\text{m}^2 \text{s}^{-1}$? The core is surrounded by 1.3 m^3 of ideal gas maintained at a pressure of $1 \times 10^5 \text{ Pa}$ and a temperature of 298 K. All neutrons escaping the reactor core pass through the gas. If the interaction cross-section between the neutrons and the gas is 1 mb, calculate the rate of neutron interactions in the gas.

- 9.2** A neutron with nonrelativistic laboratory speed v collides elastically with a nucleus of mass M . If the scattering is isotropic, show that the average kinetic energy of the neutron after the collision is

$$E_{\text{final}} = \frac{M^2 + m^2}{(M + n)^2} E_{\text{initial}},$$

where $m \equiv m_n$. Use this result to estimate the number of collisions necessary to thermalise neutrons from the fission of ^{235}U using a graphite moderator. (Assume the latter is pure ^{12}C .)

- 9.3** A thermal fission reactor uses natural uranium. The energy released from fission is 200 MeV per atom of ^{235}U and the total power output is 500 MW. If all neutrons captured by ^{238}U lead to the production of ^{239}Pu , calculate the rate of production of plutonium in kilograms per year. The cross-sections at the relevant neutron energy are

$$\sigma_c \equiv \sigma_{\text{capture}} = 3 \text{ b} \quad \text{and} \quad \sigma_f \equiv \sigma_{\text{fission}} = 600 \text{ b}$$

and the relative abundance of ^{238}U to ^{235}U in natural uranium is 138:1.

- 9.4** In a particular ADS system, the spallation reaction generates an average of 30 neutrons per 1 GeV proton incident on a heavy target contained in a subcritical reactor assembly with a multiplication factor (9.3) of $k = 0.95$. Calculate the fraction F of these neutrons that go on to induce fission, and hence estimate the energy gain of the system, if the number of neutrons produced per fission is $N = 2.5$ and each fission releases 200 MeV of energy.
- 9.5** If the Sun was formed 4.6 billion years ago and initially consisted of 9×10^{56} hydrogen atoms and since then has been radiating energy via the PPI chain at a detectable rate of 3.86×10^{26} watts, how much longer will it be before the Sun's supply of hydrogen is exhausted (assuming that the nature of the Sun does not change)?
- 9.6** In the PPI cycle, helium nuclei are produced by the fusion of hydrogen nuclei and 6.55 MeV of electromagnetic energy is produced for every proton consumed. If the electromagnetic radiation energy at the surface of the Earth is $8.4 \text{ J cm}^{-2} \text{ s}^{-1}$ and is due predominantly to the PPI cycle, what is the expected flux of solar neutrinos at the Earth in $\text{cm}^{-2} \text{ s}^{-1}$?
- 9.7** In a plasma of equal numbers of deuterium and tritium atoms (in practice deuteron and triton nuclei) at an energy $kT = 10 \text{ keV}$, the Lawson criterion is just satisfied for a total of 5 seconds. Estimate the number density of deuterons.
- 9.8** A thermal power station operates using inertial confinement fusion. If the 'fuel' consists of 1 mg of pellets of a frozen deuterium–tritium mixture,

how many would have to be supplied per second to provide an output of 750 MW if the efficiency for converting the material is 25%?

- 9.9** In some extensions of the standard model (to be discussed in Chapter 10) the proton is unstable and can decay, e.g. via $p \rightarrow \pi^0 + e^+$. If all the energy in such decays is deposited in the body and assuming that an absorbed dose of 5 Gy per annum is lethal for humans, estimate the limit that the existence of life places on the proton lifetime.
- 9.10** The main decay mode of ^{60}Co is the emission of two photons, one with energy 1.173 MeV and the other with 1.333 MeV. In an experiment, an operator stands 1 m away from an open source of 40 kBq of ^{60}Co for a total period of 18 hr. Estimate the approximate whole-body radiation dose received.
- 9.11** A specimen consists of a bone of thickness b cm that is surrounded by tissue with a uniform thickness of t cm. It is irradiated with γ -rays of energy 140 keV. The intensities through the bone (I_b) (surrounded by tissue) and through a specimen of the same thickness but of tissue only (I_t) are measured and their ratio $R \equiv I_b/I_t$ is found to be 0.7. If the attenuation coefficients $\mu = 1/\lambda$ of bone and tissue at this energy are $\mu_b = 0.29 \text{ cm}^{-1}$ and $\mu_t = 0.15 \text{ cm}^{-1}$, calculate the thickness of the bone.
- 9.12** The flux of relativistic cosmic ray muons at the surface of the Earth is approximately $250 \text{ m}^{-2} \text{ s}^{-1}$. Use Figure 4.10 to make a rough estimate of their rate of ionisation energy loss as they traverse living matter. Hence estimate in Grays (Gy) the annual human body dose of radiation due to cosmic ray muons.
- 9.13** Use the formula for the nuclear magnetic resonance frequency given in Section 9.4.4 (Footnote 35) to calculate the nuclear magnetic resonance frequency for the nucleus ^{55}Mn in a field of 2 T if its magnetic dipole moment is $3.46 \mu_N$.
- 9.14** A 1 g sample taken from an organic artefact is found to have a β count rate of 2.1 counts per minute, which are assumed to originate from the decay of ^{14}C with a mean lifetime of 8270 years. If the abundance of ^{14}C in living matter is currently 1.2×10^{-12} , what can you deduce about the approximate age of the artefact?
- 9.15** On planet X it is found that the isotopes ^{205}Pb ($\tau = 1.53 \times 10^7 \text{ y}$) and ^{204}Pb (stable) are both present and have abundances n_{205} and n_{204} , with $n_{205}/n_{204} = 2 \times 10^{-7}$. If at the time of the formation of planet X both isotopes were present in equal amounts, how old is the planet?

10

Some outstanding questions and future prospects

In this chapter we shall describe a few of the outstanding questions in both nuclear and particle physics and future prospects for their solution. The list is by no means exhaustive, but the examples show that nuclear and particle physics, and their applications, continue to be exciting and vibrant subjects, with many interesting phenomena being discovered and questions awaiting answers.

10.1 Overview

Despite more than a century of research, nuclear physics is by no means a ‘closed’ subject. Even the basic strong nucleon–nucleon force is not fully understood at a phenomenological level, let alone in terms of the fundamental quark–gluon strong interaction. Indeed one of the outstanding problems of nuclear physics is to understand how models of interacting nucleons and mesons are related to the quark–gluon picture of quantum chromodynamics (QCD), and where these two descriptions merge. A related question is how the nuclear environment modifies the quark–gluon structure of hadrons. It follows from our lack of knowledge in these areas that the properties of nuclei cannot at present be calculated from first principles, although some progress has been made in this direction. Meanwhile, in the absence of a fundamental theory to describe the nuclear force, we have seen in earlier chapters that specific models are used to interpret phenomena in different areas of nuclear physics. Current nuclear physics models must break down at very high energy-densities, and at sufficiently high temperatures the distinction between individual nucleons in a nucleus should disappear and we enter the region of a quark–gluon plasma that was discussed in Section 5.3.2. This is the regime that is believed to have existed in the very early times of the Universe and is of great interest to

cosmologists. In addition, nuclear physics has wide applications in fields such as medicine and power production, which were discussed in Chapter 9, and we will discuss some recent developments in these areas.

Unlike nuclear physics, particle physics does have a comprehensive theory – the standard model. However, even within this framework, many important questions remain to be answered. These include the determination of neutrino masses discussed in Section 3.1.6, the existence and nature of exotic states of hadrons, discussed in Section 5.3, the exploration of the Higgs sector discussed in Section 6.5, the detailed characterisation of *CP* violation discussed in Chapter 7, as well as others discussed throughout the book. Here we will concentrate on topics commonly referred to as ‘beyond the standard model’. Although the standard model is very successful at explaining a wide range of phenomena, it contains a large number of free parameters and embodies many assumptions for which a deeper explanation is needed. For example, it gives no explanation of why electric charge is quantised in units of $e/3$, or why the charge on the proton and electron are equal in magnitude. There are also suggestions from astronomy of phenomena that lie outside the standard model. In particular, there is strong evidence that the particles of the standard model can account for only a small fraction of the matter in the Universe, and in the context of the standard model, it is difficult to understand why the Universe is composed almost entirely of matter, with relatively little antimatter. In addition, the success of the standard model has spurred physicists to construct theories that incorporate the strong interaction, in addition to electromagnetic and weak interactions, in wider unification schemes. A full discussion of these topics is beyond the scope of this book, but in this chapter we will review some of these questions and also look briefly at the rapidly growing field of particle astrophysics, which has relevance for many of them.¹ Firstly, however, we turn to some of the newer topics in nuclear physics research.²

10.2 Hadrons and nuclei

In this section we will look at a number of topics around the common theme of the properties and structure of hadrons and nuclei.

10.2.1 Hadron structure and the nuclear environment

In the standard model, the structure of nucleons is specified in terms of quarks and gluons, but questions remain. One concerns the spin of the proton. The simplest suggestion is that this is formed from the spins of

¹Reviews of several of the topics we discuss may be found in the reviews of the Particle Data Group (2018).

²Useful overviews are the reports of the Nuclear Physics European Collaboration Committee (NuPECC) of the European Science Foundation. These are available from the website: <http://www.nupecc.org/index.php?display=pub/publications>.

its constituent quarks. Deep inelastic electron–proton scattering experiments performed in the late 1980s, of the type described in Section 5.5, but using polarised proton targets, showed the surprising result that the spins of all the quarks and antiquarks together contribute less than 30% to the total spin of the proton, with later experiments refining this figure to 24%. This is referred to as the *proton spin crisis* (there is a similar problem with the spin of the neutron).³ The conclusion was based on the assumption that all quarks and antiquarks, both valence quarks and those in the ‘sea’, behave in the same way in nucleons, i.e. flavour symmetry holds. However, a number of experiments have shown that this assumption is incorrect. For example, scattering longitudinally polarised electrons from a polarised proton target has enabled the individual flavour contributions to be extracted, and whereas u quarks have their spins preferentially aligned in the proton spin direction, the spins of d quarks are preferentially aligned in the opposite direction. Moreover, the polarisation of the sea quarks is essentially zero.

These conclusions ignore the spins of the gluons and the orbital motions of both quarks and gluons. To explore the gluon contribution, in 2009 experiments were carried out at the relativistic heavy ion collider (RHIC), colliding two beams of polarised protons with their spins aligned and then antialigned. Analysis of the resulting data initially suggested that the gluon spin contribution was very small, but a reanalysis of the data in 2014 claimed that it could be as large as 50% of the proton spin, although this would still leave about 30% unaccounted for. There is some information that the orbital angular momentum contributions of the quarks play an important role, but very little is known about the contribution of the orbital angular momentum of the gluon. Both are very difficult to measure. The debate about the origin of the proton spin continues and has prompted suggestions for additional experimental programmes to explore these questions.⁴

Nucleons are the building blocks of nuclear matter, but there is no guarantee that the properties of nucleons in nuclei are identical to those exhibited as free particles. According to QCD, the properties of hadrons are strongly influenced by the sea of quark–antiquark pairs and gluons that we have seen in Chapter 5 are always present around confined quarks due to quantum fluctuations. These influences could well be different in the case of closely spaced nucleons in nuclear matter to those for a free nucleon. Indeed, there are theoretical predictions that the probability of finding a $q\bar{q}$ pair decreases as the density of the surrounding nuclear matter increases. If such effects could be established, they would have a profound influence on our understanding of quark confinement.

One consequence of these predictions is that the properties of hadrons will in general change in nuclear matter, including their masses, sizes,

³A review is given in Aidala et al. (2013).

⁴See, for example, Xu (2015).

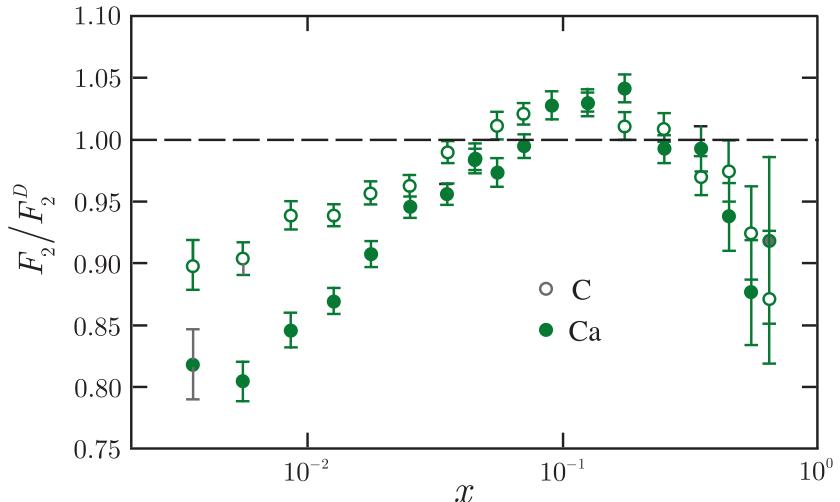


Figure 10.1 The ratios of the F_2 structure function found from nuclear targets to that found from deuterium, as a function of the scaling variable x . Source: Carbon data from Arneodo et al. (1995), calcium data from Amaudruz et al. (1995).

and interactions. There is already some evidence in favour of this suggestion from deep inelastic scattering from nucleons, where the structure functions obtained using targets of light and heavier nuclei differ slightly, even after allowing for calculable effects such as nuclear binding energies and the internal Fermi motion of the nucleons. This is the so-called ‘EMC effect’, named after the group that first discovered it. It is illustrated in Figure 10.1, which shows the ratios F_2^{Ca}/F_2^D and F_2^C/F_2^D , i.e. the F_2 nucleon structure function deduced from calcium and carbon targets divided by the structure function deduced using a deuterium target. There is a clear difference, particularly at very small values of x .

It is also important to study how the interactions of hadrons change when they are embedded in nuclear matter. For example, there is considerable interest in the interactions of hadrons containing a strange valence quark. (One reason is that they may play an important role in the high-density matter present in neutron stars.) The lightest mesons that contain a strange valence quark or antiquark are the kaons, and these can be implanted in nuclei by nuclear reactions that substitute a strange quark for an up or down quark. The result is an example of a so-called ‘hypernucleus’. Experiments at CEBAF and other laboratories will provide information on the interaction of implanted, negatively charged kaons with the surrounding nucleons in a nucleus.

The facilities at CEBAF and RHIC will enable a range of new experimental possibilities to be explored, in addition to those above. One is the existence and properties of glueballs and hybrid quark–gluon mesons, discussed in Section 5.3.1, which is important for the theory of confinement in QCD. The results may well help to find a solution to one of the central questions posed in Section 10.1: how are the properties of the strong nuclear force related to the standard model formulation in terms of quarks and gluons?

10.2.2 Nuclear structure

There are questions to be answered in the realm of nuclear structure, many with implications elsewhere. For example, can the properties of nuclei be related to those of an underlying nucleon–nucleon interaction and can they be derived from many-body theory? At present we have a good knowledge from scattering experiments of the long-range part of the nucleon–nucleon force in terms of meson exchanges (see Section 8.1), but models that fit data differ about the short-range part. This is not surprising, because at separations of less than 1 fm a description in terms of quarks and gluons is necessary and the interface with QCD is critical. Experiments on meson production in nucleon–nucleon collisions are sensitive to the short-range part of the forces and should provide information about this region. On the theoretical side, advances in computer power and computational techniques have enabled the binding energies of all light nuclei to be successfully calculated using the best available parameterisation of the nucleon–nucleon force. However, this is only possible by explicitly including a weaker three-nucleon force, which has to be adjusted to obtain the correct binding energies. A satisfactory theory of the three-body force between nucleons is lacking. This work also needs to be extended to heavier nuclei, but present computer power is inadequate to the task.

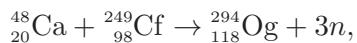
One approach to the latter problem is to work within the framework of the shell model, where each nucleon moves in the average potential (the mean field) generated by its interactions with all the other nucleons in the nucleus. We have seen the success of this approach in simple applications in Section 8.3. When combined with further computational improvements, it has enabled nuclear structure calculations to be extended to $A = 56$. This is an important point for astrophysics, because the details of the nuclear reactions of iron control the critical process occurring in the collapse of a supernova.

Fewer than 300 stable nuclei occur naturally, but the region outside the valley of stability is largely uncharted and there are many fundamental questions to be answered. For example, theoretical descriptions of nuclei far from the valley of stability suggest that their structures are different from those seen in stable nuclei. Other important questions include: what are the limiting conditions under which nuclei can remain bound, and do new structures emerge near these limits? Nuclei far from stability also play an important role in astrophysics, for example in understanding the processes in supernovas and how elements are synthesised in stars. We discussed some aspects of the latter in Section 9.2.3.

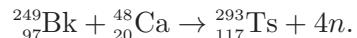
One way to explore these questions is to look for very rare decay modes that are predicted to exist for some nuclei in this region. Consider, for example, one-proton radioactivity. In nearly all cases this mechanism is absolutely forbidden by energy conservation. However, in the 1960s it was pointed out that bombarding a target having $N \approx Z$ using a projectile also with $N \approx Z$ could produce a heavy proton-rich nucleus also with $N \approx Z$, which is a very unstable configuration (see Figure 2.13), and in some cases

proton emission may be energetically possible. This was confirmed in 1982 in an experiment that bombarded a target of ^{96}Ru with a beam of ^{58}Ni nuclei. A study of the resulting particle spectrum showed evidence for the decay $^{151}\text{Lu} \rightarrow ^{150}\text{Yb} + p$, and one-proton decay has subsequently been observed in a small number of other nuclei. There are even a few nuclei where, although one-proton decay is energetically forbidden, in principle two-proton decay is allowed. The first observation of two-proton decay was made in 2002 for the nucleus ^{45}Fe and the mode has subsequently been seen in the decays of ^{48}Ni and ^{54}Zn . Other experiments plan to look at heavier nuclei.

Another interesting question concerns the possible existence of relatively stable super-heavy elements. The heaviest commonly occurring element in nature is uranium with $Z = 92$ and a half-life of 4.5×10^9 years, where by the lifetime of an element we mean the lifetime of its most stable isotope.⁵ Over the years a continuing search for new elements created at accelerators led to the discovery of nuclei with all values of Z up to and including $Z = 118$. The heaviest known element is called oganesson (Og) and was discovered at the Joint Institute for Nuclear Research (JINR) near Moscow in 2002 in the fusion reaction



despite the extremely low cross-section for such fusion processes, while the latest to be discovered, also at JINR in 2010, was tennessine (Ts), with $Z = 117$, in the reaction



In general, the half-lives of the known transuranic elements decrease rapidly with increasing Z , with values of 57 and 0.7 ms for tennessine and oganesson, respectively. It is therefore remarkable that shell model calculations predict a new group of super-heavy elements, stabilised by magic numbers, with slightly larger Z values and much longer lifetimes. Unfortunately the precise location of this *island of stability* varies somewhat depending on the assumptions made, but values in the region $Z = 120 - 122$ and neutron number about 200 are typical. Predicted lifetimes vary even more, with values ranging from minutes to many years. Strenuous efforts are being made to reach this predicted new island of relative stability.

Another limiting area that is expected to yield interesting information is that of angular momentum. Super-deformed nuclei have been discovered

⁵Traces of neptunium, with $Z = 93$ and a half-life of 2×10^6 yr, and plutonium, with $Z = 94$ and a half-life of 8×10^7 yr, have also been found in natural uranium, presumably arising from some uranium 238 atoms capturing neutrons emitted in the fission of other uranium atoms. The heaviest stable element is lead, with $Z = 82$, while bismuth, with $Z = 83$, has a half-life of 2×10^{19} yr, i.e. many orders of magnitude longer than the lifetime of the Universe.

with highly elongated shapes and very rapid rotational motion. The states associated with these shapes are extremely stable. Further investigation of these is also expected to yield important information about nuclear structure.

To explore some of the questions posed above will require facilities that can produce nuclear beams of exotic short-lived isotopes. In Section 4.2.3 we described how in principle such beams may be constructed, and there is much development work going on in this area.⁶ Unexpected nuclear structure phenomena can be explored at radioactive beam facilities. For example, in some neutron-rich light nuclei, the valence neutrons can have a spatial extent that greatly exceeds the expected nuclear dimensions. These are the *halo nuclei* briefly mentioned in Section 2.2.2 in the context of measuring nuclear radii, and their properties are difficult to explain in conventional theories. An extreme case is ^{11}Li (with three protons and eight neutrons), which is nearly as large as ^{208}Pb . To explain this it is necessary to assume a three-body model where ^{11}Li is viewed as an inert core of ^9Li together with two very loosely bound neutrons, the wavefunctions of which extend well beyond the nuclear potential. Radioactive beam facilities would also enable the study of very short-lived rare isotopes that play decisive roles in many astrophysical processes.

10.3 Unification schemes

As noted in Section 10.1, the successes of the standard model have led to various attempts to extend electroweak unification to include the strong interaction, and even gravity, in larger unification schemes. In this section we discuss some of these ideas and their experimental consequences.

10.3.1 Grand unification

Theories that attempt to include the strong interaction in a unification scheme with the electroweak interaction are called *grand unified theories* (GUTs). We have seen that unification of the weak and electromagnetic interactions does not manifest itself until energies of the order of the W and Z masses. At presently attainable energies, the strong coupling is much larger than the electroweak couplings, but coupling strengths depend on the magnitude of the associated squared energy-momentum scale μ^2 , as we saw in Section 5.2 for the strong interaction. To get some idea of the energy scale of a grand unified theory, we show in Figure 10.2 the couplings⁷

$$g \equiv 2\sqrt{2}g_W, \quad g' \equiv 2\sqrt{2}g_Z \quad (10.1)$$

⁶A review of the physics of a rare isotope accelerator is given in Geesaman et al. (2006). This article also discusses in detail the motivation for studies of rare radioisotopes.

⁷Recall that the electromagnetic coupling e is related to these couplings by the unification condition (6.52).

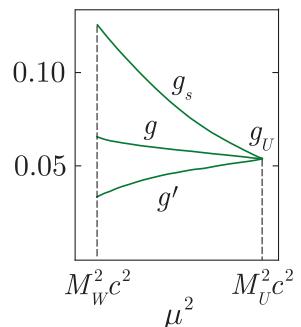


Figure 10.2 Idealised behaviour of the strong and electroweak coupling as functions of the squared energy-momentum transfer μ^2 in a simple grand unified theory.

and the strong coupling g_s (this is related to α_s by $\alpha_s = g_s^2/4\pi$) as functions of μ^2 , the squared energy-momentum transfer in a typical GUT. A naïve extrapolation in μ^2 (using, for example, (5.9) and its analogues for the more slowly varying electroweak interactions) from the region where these couplings are presently known, suggests that they become approximately equal at the enormous value $\mu^2 = M_u^{-2}c^4$, where M_u , the so-called *unification mass*, is of order 10^{15} GeV/c².⁸

In GUTs, all three interactions are united into a single interaction characterised by a single coupling constant g_U at the unification mass; differences between them emerge as one interpolates downwards to currently available energies. Of course, this interpolation assumes that nothing unexpected will emerge between energies of order 10^2 and 10^{15} GeV that could spoil the predictions. This is a very optimistic assumption given the history of particle physics. Nevertheless, this assumption is central to grand unified theories.

There are many ways in which GUTs may be constructed so that they contain the predictions of both QCD and the unified electroweak theory at currently attainable energies. The earliest, and simplest, is due to Georgi and Glashow in 1974, and incorporates the known quarks and leptons into common families. For example, in the standard model, the three colour states of the down quark, which are conveniently denoted

$$(d_r, d_g, d_b), \quad (10.2a)$$

can be converted into each other by gluon emissions, as illustrated in Figure 10.3a, while the positron and antineutrino

$$(e^+, \bar{\nu}_e) \quad (10.2b)$$

can be converted into each other by W^\pm emissions, as illustrated in Figure 10.3b. In the Georgi-Glashow model the three coloured d -quarks and the doublet $(e^+, \bar{\nu}_e)$ (strictly their right-handed components) are assigned to a single family

$$(d_r, d_b, d_g, e^+, \bar{\nu}_e). \quad (10.3)$$

In addition to the known QCD interaction in (a) and the electroweak interaction in (b), there are now two new interactions represented by (c) and (d) involving the emission or absorption of two new gauge bosons X and Y with electric charges $-4e/3$ and $-e/3$, respectively, and masses of order M_u . In this theory, all the processes of Figure 10.3 are characterised by a single GUT coupling given by

$$\alpha_U \equiv \frac{g_U^2}{4\pi\hbar c} \approx \frac{1}{42}, \quad (10.4)$$

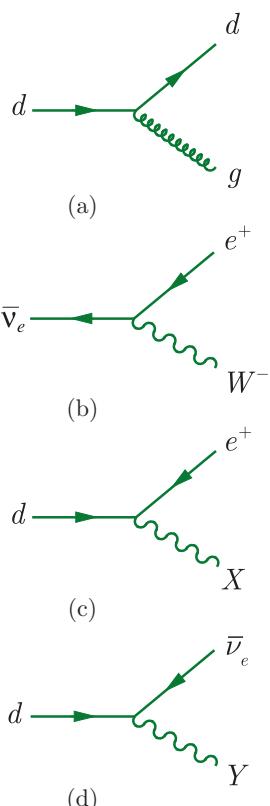


Figure 10.3 Fundamental vertices that can occur for the multiplet of particles in (10.2a).

⁸In practice, which couplings are assumed to become equal at the unification scale depends on the particular grand unified theory in question. In the simplest, to be discussed below, they are g_s , g and $(5/3)^{1/2}g'$.

which is found by extrapolating the known coupling of the standard model to the energy $E = M_U c^2$. However, at energies $E \ll M_X c^2, M_Y c^2$, processes involving the exchange of the X and Y bosons are heavily suppressed because of their large masses, in the same way that W^\pm exchange processes are suppressed relative to electromagnetic ones at energies $E \ll M_W c^2$ in the unified electroweak theory. Because of this, processes involving the exchange of X and Y particles are difficult, but perhaps not impossible, to observe at presently attainable energies, as we shall see later. Firstly, however, we comment on two other striking predictions of this simplest of all grand unification schemes.

10.3.1(a) Quark and lepton charges

The simple Georgi–Glashow model has a number of attractive features. For example, it can be shown that the sum of the electric charges of all the particles in a given multiplet is zero. Therefore, using the multiplet $(d_r, d_b, d_g, e^+, \bar{\nu}_e)$, it follows that

$$3q_d + e = 0, \quad (10.5)$$

where q_d is the charge of the down quark. Thus $q_d = -e/3$ and the fractional charges of the quarks are seen to originate in the fact that they exist in three colour states and that this is not just an empirical assignment of the standard model. By a similar argument, the up quark has charge $q_u = 2e/3$ and so with the usual quark assignment $p = uud$, the proton charge is given by

$$q_p = 2q_u + q_d = e. \quad (10.6)$$

Thus, we also have an explanation of the long-standing puzzle of why the proton and positron have precisely the same electric charge.

10.3.1(b) The weak mixing angle

When the strong and electroweak interactions are extrapolated to the unification mass M_U , they are characterised by a single coupling constant g_U , as shown in Figure 10.2. Conversely, the three effective low-energy couplings of the standard model shown in Figure 10.2 can be expressed in terms of the two parameters g_U and M_U by reversing this extrapolation. Consequently, in GUTs one of the three coupling constants can be predicted at lower energies, given the values of the other two. It is conventional to convert this result into a prediction of the weak mixing angle θ_W , which is related to the coupling constants (10.1) by (6.52), and in the Georgi–Glashow model this can be shown to give the value

$$\sin^2 \theta_W = 0.21. \quad (10.7)$$

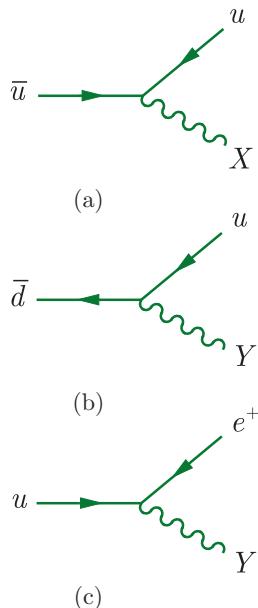


Figure 10.4 The three fundamental vertices involving X and Y bosons that are predicted by the Georgi–Glashow model, in addition to those shown in Figure 10.3c,d.

This value is encouragingly close to the measured value 0.23155 ± 0.00005 , but is not in precise agreement with it.⁹

10.3.1(c) Proton decay

The most striking prediction of GUTs is that the proton is unstable, and in the Georgi–Glashow model it can decay by a variety of processes involving the exchange of X and Y bosons and their antiparticles \bar{X} and \bar{Y} . These are built from a number of basic vertices, two of which are shown in Figure 10.3c and d. If we restrict ourselves to the first generation and their antiparticles, the model predicts three more such vertices, which are shown in Figure 10.4, and there are also another five that can be obtained from Figures 10.3c and d and 10.4 by replacing all particles by their antiparticles. These give rise to proton decays such as

$$p \rightarrow \pi^0 + e^+ \quad \text{and} \quad p \rightarrow \pi^+ + \bar{\nu}_e, \quad (10.8)$$

by mechanisms like those shown in Figure 10.5. In all such processes, both the baryon and lepton numbers are not conserved, but the combination

$$B - L \equiv B - \sum_{\alpha} L_{\alpha} \quad (\alpha = e, \mu, \tau) \quad (10.9)$$

is conserved, as may be verified directly from the vertices of Figures 10.3 and 10.4.

Although the proton is predicted to decay via processes like (10.8), its predicted lifetime is extremely long, and before quoting detailed values it is useful to make a qualitative estimate to understand why this is so. The first step is to estimate the magnitude of the effective coupling constant G for processes involving the exchange of X and Y bosons at energies $E \ll M_U c^2$. At these energies, X and Y exchanges may be approximated by point interactions as shown in Figure 10.6, where on the left-hand side we have approximated the couplings of the X and Y bosons to the lepton and quarks by their value g_U at the unification mass M_U , and ignored their dependence on the range of the interaction. The effective low-energy coupling is then given by

$$\frac{G}{(\hbar c)^3} \approx \frac{\sqrt{2}}{\hbar c} \left(\frac{g_U}{Mc^2} \right)^2 = \frac{4\pi\sqrt{2}\alpha_U}{(Mc^2)^2}, \quad (10.10)$$

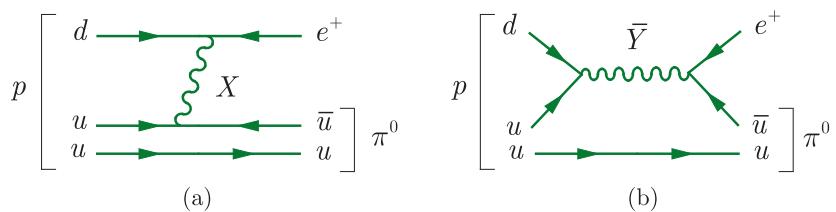


Figure 10.5 Examples of processes that contribute to the proton decay reaction $p \rightarrow \pi^0 + e^+$.

⁹Another way of putting this is to say that if the measurement is imposed, the three curves in Figure 10.12 fail to meet at the same point.

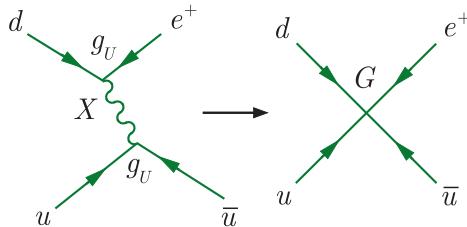


Figure 10.6 The zero-range approximation to an X boson exchange process.

by analogy with (1.54) for the Fermi coupling G_F , where $M = M_{X,Y}$, so that $G/(\hbar c)^3$ has the dimensions of inverse energy squared. Since the width Γ associated with any X or Y exchange process has the dimensions of energy and is proportional to G^2 , a simple dimensional estimate gives

$$\Gamma \approx \frac{G^2}{(\hbar c)^6} E^5 \approx \left(\frac{4\pi\alpha_U}{(Mc^2)^2} \right)^2 E^5,$$

where $E \ll Mc^2$ is some energy characteristic of the process. A reasonable estimate for proton decay is $E = m_p c^2$, and so we finally arrive at

$$\tau = \frac{\hbar}{\Gamma} \approx \frac{\hbar(Mc^2)^4}{(4\pi\alpha_U)^2(m_p c^2)^5} \quad (10.11)$$

for the lifetime of the proton. Taking $M \approx M_U = 10^{15} \text{ GeV}/c^2$ and α_U from (10.4) gives an estimate of approximately 10^{30} years, in agreement with the range of values

$$\tau = 10^{29} - 10^{30} \text{ years} \quad (10.12a)$$

obtained from more detailed and reliable calculations. This value is, however, very sensitive to the value chosen for the unification mass M_U , as can be seen from (10.11); in other GUTs in which M_U is somewhat larger, the lifetime can be as long as

$$\tau = 10^{32} - 10^{33} \text{ years.} \quad (10.12b)$$

In addition, neutrons bound in stable nuclei that cannot decay by beta decay because of energy conservation, would be expected to decay by baryon number violating decays, with similar lifetimes. For comparison, the age of the Universe is believed to be of order 10^{10} years.

If proton decays occur at all, they are clearly very rare events. For example, 300 tons of iron would only yield about one proton decay per year if the lifetime were of order 10^{32} years. The problems involved in searching for such rare events, and their resolution, have been discussed in general in Section 4.5 and have been exploited to search for many possible proton decay modes. The results of these searches are expressed as lower limits on the *partial lifetimes*, which are the inverse of the decay widths for any particular decay mode $p \rightarrow f$ and are given by

$$\tau(p \rightarrow f) = \tau_p / B(p \rightarrow f), \quad (10.13)$$

where τ_p is the proton mean lifetime and $B(p \rightarrow f)$ the corresponding branching ratio. The most stringent limit has been obtained for the decay $p \rightarrow \pi^0 + e^+$ (Nishino et al. 2009) using the SuperKamiokande detector discussed in Section 3.1.5(a) and shown in Figure 3.3.

The SuperKamiokande detector is essentially a huge Čerenkov detector, with a central volume containing 50 kilotons of extremely pure water. If a $p \rightarrow \pi^0 + e^+$ decay occurs, the positron would give rise to a cone of Čerenkov radiation, whose radius depends on the velocity of the particle, while the particle energy can be inferred from the number of photons detected. The neutral pion would decay almost immediately into two photons, which would in turn interact to produce electrons and two further cones of Čerenkov radiation. A number of corrections had to be taken into account, including the fact that most of the protons in water are bound in oxygen nuclei, giving rise to Fermi motion. The experiment ran for several years, giving a total exposure of 141 kiloton-years. No decay events were observed, putting a limit of

$$\tau(p \rightarrow \pi^0 + e^+) = \frac{\tau_p}{B(p \rightarrow \pi^0 + e^+)} > 8.2 \times 10^{33} \text{ yr} \quad (10.14a)$$

on the partial lifetime at a 90% confidence level.¹⁰ Limits on other partial lifetimes have also been obtained in this and other experiments, including

$$\tau(n \rightarrow \pi^0 + \nu) > 1.1 \times 10^{33} \text{ yr} \text{ and } \tau(p \rightarrow K^0 + e^+) > 1.0 \times 10^{33} \text{ yr}, \quad (10.14b)$$

which together suggest lower limits in the range $(10^{32} - 10^{33})$ yr on the proton lifetime.

In the Georgi–Glashow model, the branching ratio $B(p \rightarrow \pi^0 e^+)$ is predicted to be about 0.3, so that (10.14a) is clearly incompatible with the predicted lifetime (10.12a). However, other models, which retain the successes of the Georgi–Glashow model, predict longer lifetimes (10.12b) and are compatible with the proton decay data, as we shall see in the following section.

10.3.2 Supersymmetry

The problems with the predictions for the weak mixing angle (10.7) and the proton lifetime (10.12a) in GUTs can be resolved by incorporating a new proposed symmetry between bosons and fermions called *supersymmetry* (SUSY). According to this, every known elementary particle has a supersymmetric partner (called a *superpartner*), which is like it in all respects, including having the same quantum numbers, except for its spin.

¹⁰Plans are being formulated to build a two-cylinder Hyper-Kamiokande detector, which among other things would be used to study proton decay, the first cylinder of which would have a volume more than 20 times that of Super-Kamiokande and a timescale of 10 years to build.

Spin-1/2 fermions, leptons, and quarks, have spin-0 superpartners, while spin-1 bosons, like photons, have spin-1/2 superpartners. To distinguish between a spin-1/2 particle and its superpartner, an ‘s’ is attached to the front of its name in the latter case. Thus the superpartner of the electron is called the selectron, and so on. The situation for the Higgs boson is more complicated, since if one simply associates a superpartner with the standard model Higgs boson, the problems that were eliminated by the unification and anomaly conditions, discussed in Section 6.4.1, reappear and the theory becomes inconsistent. Supersymmetry requires additional spin-0 Higgs bosons, as well as their spin-1/2 superpartners, called *Higgsinos*. The simplest version of supersymmetry is called the *Minimal Supersymmetric Standard Model (MSSM)*, because it contains the minimum number of new particles that are required for a consistent theory. It has three neutral and two charged spin-0 Higgs bosons, together with two neutral and two charged Higgsinos. The *photino*, *zino*, and the neutral Higgsinos are all spin-1/2 particles that interact by electroweak forces only. They are expected to mix together in the same way as neutrinos to form four new particles, called *neutralinos*. In the same way, the charged Higgsinos can mix with the winos to form four particles, called *charginos*. Finally, a useful way of characterising particle states as particles or superparticles is provided by a new quantum number, called *R parity*, defined by

$$R = (-1)^{3(B-L)+2S}, \quad (10.15)$$

where B is the baryon number, L is the lepton number, and S is the spin. From this, one easily verifies that all the particles of the standard model have $R = +1$, while the superparticles have $R = -1$. The full set of elementary particles and their superpartners in the MSSM is shown in Table 10.1.

If supersymmetry were exact, a particle and its supersymmetric partner would have exactly the same mass. This is obviously not realised in nature or superparticles would have been detected long ago. Supersymmetry is thus at best only an approximate symmetry. Because of this, the

Table 10.1 The particles of the MSSM and their superpartners. In this model there are two charged and three neutral Higgs bosons

Particle	Symbol	Spin	Superparticle	Symbol	Spin
Quark	q	1/2	Squark	\tilde{q}	0
Electron	e	1/2	Selectron	\tilde{e}	0
Muon	μ	1/2	Smuon	$\tilde{\mu}$	0
Tauon	τ	1/2	Stauon	$\tilde{\tau}$	0
W boson	W	1	Wino	\tilde{W}	1/2
Z boson	Z	1	Zino	\tilde{Z}	1/2
Photon	γ	1	Photino	$\tilde{\gamma}$	1/2
Gluon	g	1	Gluino	\tilde{g}	1/2
Higgs bosons	H	0	Higgsinos	\tilde{H}	1/2

masses of the predicted superparticles are unknown and even the simplest models incorporating supersymmetry breaking contain large numbers of free parameters whose values are unknown. However, in GUTs, a rough limit on the degree of symmetry breaking is suggested by the so-called *hierarchy problem*. In quantum field theory, the square of the physical mass M_H of a Higgs boson – i.e. the mass that can be measured – can be expressed in the form

$$M_H^2 = M_0^2 - \Delta M^2, \quad (10.16)$$

where M_0 is the ‘bare mass’ of a hypothetical noninteracting Higgs boson and ΔM^2 is an additional contribution generated by radiative corrections. In the standard model, the lowest-order contributions to the latter are of order $g_W^2 \lambda^2 / 16\pi^2$, where λ is a cutoff parameter above which the theory is assumed to break down. In GUTs, λ must be at least as large as the unification scale. Hence ΔM^2 must be extremely large and a correspondingly huge Higgs mass can only be avoided by rather artificial ‘fine-tuning’ of the unknown bare mass to almost entirely cancel it. While this does not violate any known physical principles, it is not very natural. However, an attractive feature of supersymmetric theories is that fine-tuning is avoided by automatic cancellations between the contributions of particles and their superpartners to the radiative correction ΔM^2 , provided that the differences between their masses are not too large.¹¹ This is assumed to be the case in many ‘weak-scale’ supersymmetric theories in which the supersymmetric particles are assumed to have masses of order $1 \text{ TeV}/c^2$ or less, since above this mass fine-tuning begins to reappear. Furthermore, when supersymmetric particles assumed to be in this mass range are taken into account, it can be shown that the extrapolation of Figure 10.2 is modified in such a way that the grand unification mass M_u is increased to a value of order $10^{16} \text{ GeV}/c^2$, while the value of the grand unified coupling constant g_u remains relatively constant. As a consequence of this, the proton lifetime increases, as expected from (10.11), to a value of order $10^{32} - 10^{33}$ years, which is just consistent with present experimental limits. At the same time, the prediction (10.7) of the weak mixing angle is slightly modified to yield a value of 0.23155 ± 0.00005 , in agreement with the experimental value.

10.3.2(a) The search for supersymmetry

A direct test of supersymmetry would be to detect the predicted superparticles. Unfortunately, as noted earlier, it is difficult to say very much about their masses because the nature of supersymmetry breaking is unknown and even the simplest supersymmetric models contain many unknown parameters. However, according to most versions of supersymmetry,¹²

¹¹Fine-tuning can be avoided in other more complicated ways, but supersymmetry is usually considered the simplest.

¹²For a fuller introduction to supersymmetry, see, for example, Dawson (1997), and references therein.

including the MSSM, R parity is conserved, so that superparticles can only be created or destroyed in pairs. Hence the decay of a superparticle must yield at least one superparticle in the final state, and the lightest such particle must be stable.¹³ There are several candidates for the identity of the lightest superparticle, but most models assume it is the lightest *neutralino*, denoted $\bar{\chi}_0$, which is a mixture of a photino, neutral Higgsinos, and a zino, as we noted above. If we accept this, then a simple reaction that can be studied at an e^+e^- collider is

$$e^+ + e^- \rightarrow \tilde{e}^+ + \tilde{e}^-, \quad (10.17a)$$

followed by the decays

$$\tilde{e}^\pm \rightarrow e^\pm + \bar{\chi}_0, \quad (10.17b)$$

giving an overall reaction

$$e^+ + e^- \rightarrow e^+ + e^- + \bar{\chi}_0 + \bar{\chi}_0. \quad (10.17c)$$

The theory makes three clear predictions for this reaction sequence:

- (i) the cross-section for producing selectron pairs via (10.17a) should be comparable to that for producing ordinary charged particle pairs of the same mass;
- (ii) the selectrons decay very rapidly by (10.17b) before they can reach the detector;
- (iii) the neutralinos should escape detection because they interact extremely weakly with ordinary matter.

Hence only the final-state electron and positron should be detected in (10.17a) to (10.17c), and the events would be characterised by the fact that on average these will account for only half the energy of the initial state. Moreover, the electron and positron will not in general be emitted in opposite directions in the centre-of-mass frame, as they would be in the case of a two-body final state. Thus there are clearly defined criteria for the identification of selectron pair-production events, provided of course that there is enough energy to produce them. A similar discussion applies to the production of other pairs of charged superparticles in electron–positron collisions. The highest energy electron–positron collider to date was the Large Electron–Positron collider LEPII at CERN, which had a maximum energy of 208 GeV when it closed in 2002. No charged superparticles were detected, setting lower limits that are typically of order 100 GeV/c^2 on their masses.

Both charged and neutral superparticles can be produced at hadron colliders. At the large hadron collider (LHC), the ATLAS and CMS experiments have studied a wide range of reactions, but to date have found no evidence for the existence of superparticles in the mass ranges studied.

¹³Because of this, the lightest superparticle is a possible contributor to the so-called ‘dark matter’ in the universe, to be discussed in Section 10.5.2.

This failure to detect any such events sets lower limits on the masses of the different superparticles, which vary depending on the model. Here we shall just consider typical results arising from variations on the simple MSSM model with R parity conservation, in which the lightest supersymmetric particle is a neutralino. For the masses of the squarks, which play a crucial role in solving the hierarchy problem, the associated lower limits are typically of order $1 \text{ TeV}/c^2$ for the first two generations and somewhat lower for the third generation, while the lower limit on the mass of the gluino is also of order $1 \text{ TeV}/c^2$. In contrast, the lower limit obtained on the mass of the lightest neutralino, which has only weak interactions, is almost zero in the MSSM without further assumptions, and typically about $50 \text{ GeV}/c^2$ in a variety of favoured models.

As mentioned above, the MSSM model predicts the existence of several Higgs bosons, rather than the single one predicted by the standard model. In the MSSM¹⁴ these are: two $CP = +1$ neutral bosons, denoted h and H ; one $CP = -1$ neutral boson, denoted A ; and two charged bosons, denoted H^\pm . Ignoring radiative corrections, the mass of the lightest Higgs boson is given by

$$M_H = M_Z |\cos 2\beta| \leq M_Z, \quad (10.18a)$$

where, as the parameter $\beta \rightarrow 0$, $M_H \rightarrow M_Z$, and the lightest Higgs boson becomes identical to the Higgs boson of the standard model. The other Higgs bosons become increasingly heavy and essentially decouple from the theory at presently attainable energies. When radiative corrections ε_H are included, (10.18a) is modified to

$$M_H = M_Z |\cos 2\beta| + \varepsilon_H < (120 - 130) \text{ GeV}/c^2, \quad (10.18b)$$

where the uncertainty in the limit reflects the uncertainty in the calculation of the radiative corrections. The mass of the Higgs boson is close to the limit (10.18b) and the agreement, within errors, between its observed properties and those predicted in the standard model are a severe constraint upon supersymmetric theories. As more data are taken and the errors shrink, either the agreement will be maintained and the scale of supersymmetry breaking is pushed to still higher values, or deviations from the predictions of the standard model will be firmly established for the first time. Of course, a breakdown of the standard model does not in itself prove that it is due to supersymmetry, but it would be a prime candidate.

There are also indirect ways of finding evidence for superparticles by looking for discrepancies between the predictions of the standard model and experiment that might be explained by their existence. For example, the exchange of virtual superparticles would contribute to the deviation of the muon magnetic dipole moment from its Dirac value, although it would be difficult to separate these contributions from other corrections.

¹⁴For a fuller account of the Higgs sector in the MSSM, see Haber (2011) or Gunion et al. (1990).

Another example is the details of CP violation in the B -meson sector, which has not yet been fully explored. The latter will be the subject of intensive investigation at the LHC accelerator, particularly in the LHCb experiment, which has been designed to study the physics of B -particles. Finally, contributions from superparticles would change the prediction for the electric dipole moment (EDM) of the electron and this will be discussed in Section 10.5.3.

10.3.3 Strings and things

Undeterred by the lack of immediate success of supersymmetry, proposals have been made to include gravity in even larger unification schemes. The problems here are mathematically formidable, not least of which is that the divergences encountered in trying to quantise gravity are far more severe than those in either QCD or the electroweak theory and there is at present no successful ‘stand-alone’ quantum theory of gravity analogous to the former two. An unusual problem is the extreme difficulty of making unambiguous predictions that could be tested by experiment. The theories that have been proposed that include gravity invariably replace the idea of point-like elementary particles with tiny one-dimensional quantised *strings* and are formulated in many more dimensions (usually 10, including one time dimension) than we observe in nature. Such theories have a single free parameter – the string tension. However, we live in a four-dimensional world and so the extra dimensions have to be ‘compactified’, i.e. reduced to an unobservably small size. It was originally hoped that in doing this the Standard Model, with its 19 free parameters (masses of quarks, coupling constants, mixing angles, etc.) would emerge from string theory as a unique low-energy four-dimensional theory and thus the precise values of the 19 parameters of the model would be explainable in terms of just a single parameter, the string tension.

Early optimism has not been sustained. In the particle picture, the structure (i.e. the Hamiltonian or Lagrangian) of the corresponding quantum field theory (such as QED or QCD) is known and physical predictions may be obtained using the appropriate Feynman rules. However, in string theory the Lagrangian is not known, and there are five sets of possible Feynman rules, each operating in a ten-dimensional space–time continuum. In fact string theorists have discovered that far from being unique, there is a vast ‘landscape’ of at least 10^{500} possible low-energy theories that could result after compactification, each corresponding to a Universe with a different set of fundamental particles, interactions, and parameters. Unless there is a method of choosing between the vast possibilities offered by this ‘landscape’, string theories have little or no real predictive power. For this reason they have generated a lively philosophical debate as to whether they should even be considered as scientific theories, although their proponents claim that string theories are being judged by standards that historically have not been applied to other emergent theories.

One controversial approach to the question of choice has been to invoke the so-called ‘anthropic principle’. This states that what we can expect to

observe must be restricted by the conditions necessary for our presence as observers. In other words, the world is observed to be the way it is because that is the only way that humans could ever be here to consider such questions in the first place. This somewhat circular-sounding ‘principle’ has been invoked by cosmologists to explain the apparent improbable values of some cosmological constants, but is by no means generally accepted as a way forward for string theories and other theorists believe that some form of dynamical selection will eventually be possible. All one can say at present is that there is no consensus on how the problem of choice is to be solved.

The self-consistency of string theories in 10 dimensions has been shown to imply the existence of higher-dimensional objects, called *branes* (short for membranes), and it has been conjectured that using these it will be possible to construct an even more fundamental theory in 11 dimensions in which all five supersymmetric string theories are unified. This is called *M-theory*, although no-one knows if the conjecture is true, or how to construct such a theory. Nevertheless, string theory has provided some powerful theoretical tools that have contributed to a better understanding of gauge theories and their relation to gravity.

Another problem with string theories, leaving aside their formidable theoretical complexity and the ambiguity of their predictions, is that they apply at an energy scale where gravitational effects are comparable to those of the gauge interactions, i.e. at energies defined by

$$E_P = M_P c^2 = c^2(\hbar c/G)^{1/2} = 1.2 \times 10^{19} \text{ GeV}, \quad (10.19)$$

where M_P is the so-called *Planck mass* and G is the gravitational constant.¹⁵ This energy is so large that it is difficult to think of a way that the theories could be ruled out by experimental tests at currently accessible energies, or even indeed at energies accessible in the conceivable future, although some theorists believe that information produced even at the ‘low’ energies of the LHC (for example, the discovery of superparticles) may help to test string theories.

The appeal of string theories at present is mainly the mathematical beauty and ‘naturalness’ that their sponsors claim for them, although the history of physics is littered with failed theories whose authors doubtless considered them both beautiful and natural. Needless to say, experimentalists will remain sceptical until definite experimental tests can be suggested and carried out.

10.4 The nature of the neutrino

The neutrino is a particle whose properties are difficult to study experimentally because it only interacts via the weak interaction. It was not until

¹⁵This implies that strings have dimensions of order $x_P \sim \hbar/M_P c = 1.6 \times 10^{-35} \text{ m}$.

about 25 years after Pauli first postulated the neutrino that its existence was proved experimentally, and it took a further 50 years to establish that it had a nonzero mass. Another piece of history is the assumption that the neutrino is a so-called ‘Dirac particle’, i.e. its wavefunction obeys the Dirac equation (1.1) with neutrino and antineutrino being distinct entities. In this case there are two neutrino states

$$\nu_L, \quad \nu_R \quad (L = 1)$$

and two antineutrino states

$$\bar{\nu}_L, \quad \bar{\nu}_R \quad (L = -1), \quad (\text{Dirac neutrinos}),$$

with opposite values of the lepton number L , which is conserved in the standard model. However, in Section 1.3.2 we noted that for neutral states, a particle and its antiparticle do not necessarily have to be distinct, and in principle this could be true for neutrinos. In this case there are only two states, called *Majorana* neutrinos, denoted by

$$\nu_L, \quad \nu_R \quad (\text{Majorana neutrinos}).$$

In both cases, the subscripts R and L denote right- and left-handed helicity states as usual. In this section we consider this possibility and discuss how the nature of the neutrino, Dirac or Majorana, might be determined experimentally.

In the original formulation of the standard model, neutrino masses were assumed to be zero and it is not possible to experimentally distinguish between Dirac and Majorana neutrinos for two reasons. Firstly, for Dirac neutrinos with zero mass, the weak interaction only couples the left-handed neutrinos ν_L and right-handed antineutrinos $\bar{\nu}_R$, as we saw in Chapter 7. Secondly, in the zero-mass limit, the Dirac equation decouples into two two-component equations describing the states $(\nu_L, \bar{\nu}_R)$ and $(\nu_R, \bar{\nu}_L)$, respectively. The second pair can therefore be completely eliminated from the theory, which becomes indistinguishable from an analogous Majorana theory with the replacements $(\nu_L, \bar{\nu}_R) \rightarrow (\nu_L, \nu_R)$.

In contrast, as we saw in Section 7.2.2, the weak interaction couples to both helicity states of the electron, which is not massless. Similarly, for non-zero masses, the weak interaction would also couple to the other two Dirac neutrino states, albeit with relative couplings which are typically of order $(m_\nu c^2/E)^2$, where E is the neutrino energy. Because of this, differences of the same order will emerge between the two descriptions of the neutrino, which can, at least in principle, be detected experimentally. In particular, in the Majorana formalism, lepton number conservation for reactions involving neutrinos is an ‘accident’ arising from the spin structure of the weak interaction for zero mass neutrinos, and tiny deviations from it are expected for nonzero neutrino masses. In most cases these will be too small to observe, and the most promising prospect for detecting them, if they exist, is by detecting neutrinoless double-beta decay. This

will be discussed in Section 10.4.1. First, however, we digress briefly to mention a way in which very heavy Majorana neutrinos may play a role.

At present we have no theoretical understanding of why quarks and leptons have the particular masses they do. However, the three known neutrinos have much lower masses than the other fundamental fermions, and a possible explanation for this has been suggested in the context of GUTs. In these, it is possible for both types of neutrino to co-exist, and for a very small neutrino mass to emerge from the mixing of a zero-mass neutrino with a very heavy Majorana neutrino. The corresponding mass matrix for the right-handed neutrino is essentially of the form

$$\mathbf{M} = \begin{pmatrix} 0 & m_D \\ m_D & m_M \end{pmatrix}, \quad (10.20)$$

where the off-diagonal term m_D , called the Dirac mass, is of the order of the electroweak scale and m_M , the Majorana mass, is of the order of the GUT scale. Since $m_M \gg m_D$, the eigenvalues of \mathbf{M} are

$$|\lambda_+| \approx m_M \quad \text{and} \quad |\lambda_-| \approx m_D^2/m_M. \quad (10.21)$$

If λ_- is associated with the observed neutrinos, we have a natural explanation for a very small mass of order 1 eV or less, while the other eigenvalue implies a very heavy neutrino, yet to be discovered. This is called the ‘see-saw mechanism’, because from (10.21) we see that as one mass goes up, the other goes down.

10.4.1 Neutrinoless double beta decay

The double beta-decay process

$$(Z, A) \rightarrow (Z + 2, A) + 2e^- + 2\bar{\nu}_e \quad (10.22a)$$

(denoted by the label $\beta\beta2\nu$), in which two neutrons decay simultaneously to give two electrons and two antineutrinos, is a second-order weak interaction allowed in the standard model, as illustrated in Figure 10.7a. In contrast, the neutrinoless double beta-decay process (labelled $\beta\beta0\nu$)

$$(Z, A) \rightarrow (Z + 2, A) + 2e^- \quad (10.22b)$$

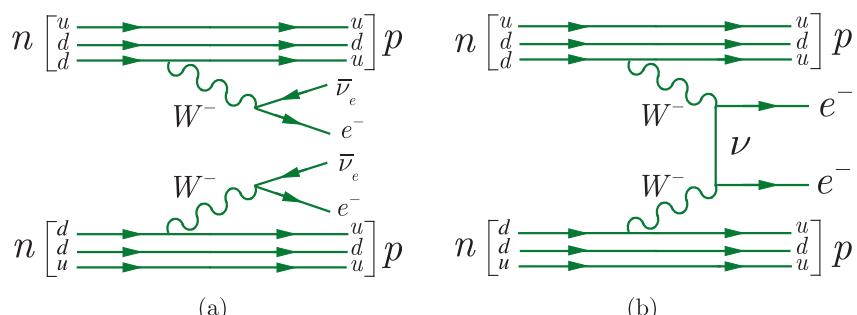


Figure 10.7 (a) Double β decay $\beta\beta2\nu$ (allowed in the standard model); (b) neutrinoless double β decay $\beta\beta0\nu$ (forbidden in the standard model).

is not allowed because it violates lepton number. However, it can occur for Majorana neutrinos of nonzero mass, which are their own antiparticles, by the mechanism illustrated in Figure 10.7b. Observation of the reaction (10.22b) would therefore be strong evidence for the existence of Majorana neutrinos, although in principle the reaction could also proceed by other mechanisms beyond the standard model, involving new hypothetical particles such as ones predicted by varieties of supersymmetry. On the other hand, the $\beta\beta2\nu$ process of (10.22a) is allowed for both Dirac and Majorana neutrinos.

Because the double β -decay processes (10.22b) is second-order in the weak interaction, with a rate of order G_F^4 , it can in practice only be observed if the single β -decay process

$$(Z, A) \rightarrow (Z + 1, A) + e^- + \bar{\nu}_e \quad (10.23)$$

is energetically forbidden, or highly suppressed. For an even- Z , even- $N = A - Z$ nucleus, double β -decay leads to another even-even nucleus ($Z + 2, A$), while single β -decay leads to an odd-odd nucleus ($Z + 1, A$). Since odd-odd nuclei are less tightly bound than even-even nuclei, due to the pairing term (2.55) in the semi-empirical nuclear mass formula, a sequence of states can result in which single β -decay is forbidden by energy conservation, while double β -decay is allowed. If the parent nucleus (Z, A) is also stable against α - and γ -decays, it will decay by double β -decay. Such decays were first directly observed in 1987, and have now been established for 12 isotopes, with mean lifetimes in the range $10^{18} - 10^{24}$ yr.

The observation of $\beta\beta0\nu$ decays would be a major breakthrough in particle physics because it would demonstrate the Majorana nature of neutrinos and the violation of total lepton number. It can in principle be distinguished from $\beta\beta2\nu$ decays by measuring the energies of the emitted electrons. In $\beta\beta2\nu$ decays, energy is carried away by the undetected neutrinos, resulting in a continuous spectrum for the combined energy of the electrons, whereas in $\beta\beta0\nu$ decays the electrons carry off all the available energy, resulting in a sharp line in their combined energy, as shown in Figure 10.8. However, a major problem is that the rate for $\beta\beta0\nu$ decay is expected to be much smaller than that for $\beta\beta2\nu$ decays, even for Majorana neutrinos, and for zero-mass neutrinos it would actually vanish.

For non-zero mass Majorana neutrinos, the decay rate Γ is given by

$$\Gamma = am_{\beta\beta}^2, \quad (10.24)$$

where the constant a is proportional to the coupling constant G_F^4 and depends upon the particular decaying nucleus. The quantity $m_{\beta\beta}$ is the ‘effective Majorana mass’, which reduces to

$$m_{\beta\beta} = |U_{e1}|^2 m_1 + |U_{e2}|^2 m_2 + |U_{e3}|^2 m_3. \quad (10.25)$$

Here $U_{e1} = c_{12}c_{13}$, $U_{e2} = s_{12}c_{13}$, and $U_{e3} = s_{13}e^{-i\delta}$ are the elements of the neutrino mixing matrix (7.83) that defines the electron-neutrino state; m_1 ,

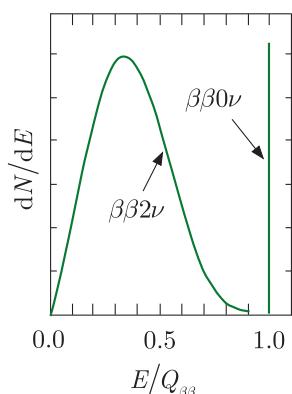
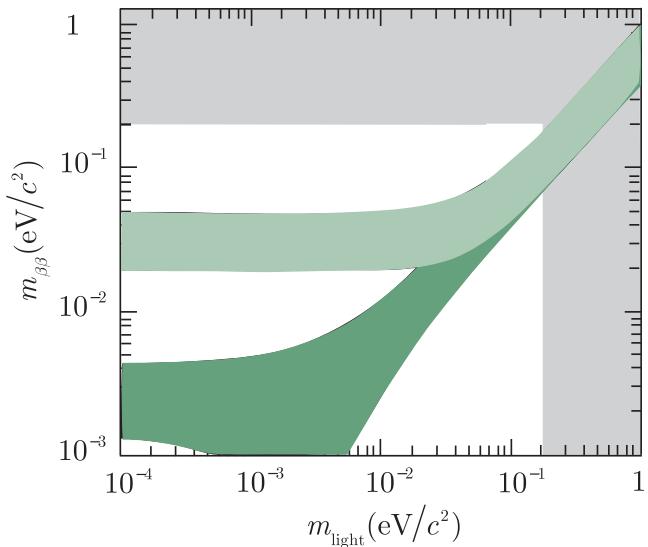


Figure 10.8 Energy spectra for the two electrons in $\beta\beta2\nu$ and $\beta\beta0\nu$ decays as a function of $E/Q_{\beta\beta}$, where E is their combined energy and $Q_{\beta\beta}$ is the energy released.

Figure 10.9 The predicted values of the effective neutrino Majorana mass $m_{\beta\beta}$, as a function of the lightest neutrino mass, m_{light} . The dark and light green bands correspond to the ‘normal’ and ‘inverted’ hierarchies of Figure 3.7, respectively. Two rectangular excluded regions are shown in grey. The vertical region comes from the cosmological bound $m_{\text{light}} < 0.22$ eV, to be discussed in Section 10.5.2(a), and the horizontal one from experimental constraints on neutrinoless double beta decay described in the text. Source: Adapted from the review of Gómez-Cadenas and Martín-Albo (2014).



m_2 , and m_3 refer to the mass eigenstates ν_1 , ν_2 , and ν_3 ; and for simplicity, we have neglected CP -violating phases in the neutrino sector. From this equation it is clear that $m_{\beta\beta}$ depends on the neutrino oscillation parameters and the unknown ordering of neutrino masses. Thus, if the three known neutrinos are indeed Majorana particles, and there are no other neutrinos, then values of $m_{\beta\beta}$ can be found for different assumed values for the lightest neutrino mass, given the measured values of the neutrino squared mass splittings Δm_{ij}^2 and mixing angles θ_{ij} discussed in Chapter 3. Figure 10.9 shows $m_{\beta\beta}$ as a function of the lightest neutrino mass m_{light} obtained in this way for both the ‘normal’ and ‘inverted’ hierarchies shown in Figure 3.7. The widths of the bands are due to uncertainties in the neutrino oscillation parameters, and also the unknown CP -violating phases, which have been neglected in (10.25). Also shown is the upper bound on m_{light} obtained from cosmological studies, as discussed later in Section 10.5.2(a), which is much more stringent than the bound of about 0.7 eV/ c^2 implied by direct measurements of tritium decay, together with the measured values of the squared mass splittings from oscillation experiments (cf. (3.44)). If this astrophysical bound is correct, then it is clear that the effective neutrino mass $m_{\beta\beta}$ is predicted to be at most of order 50 meV/ c^2 .

Assuming the exchange of light Majorana neutrinos is the dominant mechanism for neutrinoless double beta decay, the half-life may be written

$$(T_{1/2}^{0\nu})^{-1/2} = G_F^4 \left(\frac{m_{\beta\beta}}{m_e} \right)^2 G^{0\nu} |M^{0\nu}|^2, \quad (10.26)$$

where $G^{0\nu}$ is a density of states (or phase space) factor, $M^{0\nu}$ is the nuclear matrix element for the transition, and we have factored out the explicit dependence on $m_{\beta\beta}$ and the Fermi coupling constant arising from the latter term. Unfortunately, $M^{0\nu}$ can only be calculated using specific nuclear

models, giving rise to significant uncertainties, which must be taken into account when translating a given value or limit on the lifetime into a value or limit for the effective mass.¹⁶ In addition, because of the factor G_F^4 , and because the neutrino masses of interest are so small, the predicted lifetimes are extremely long, and like all low-energy weak interactions they increase rapidly with $Q_{\beta\beta}$, the energy released in the decay. The actual values vary from nucleus to nucleus, but for an effective neutrino mass of 50 meV/ c^2 , typical half-lives of order $10^{26} - 10^{27}$ years are obtained, i.e. 17 orders of magnitude greater than the age of the Universe. This presents huge difficulties for experiments, since it implies that to detect one $\beta\beta0\nu$ decay per year, assuming perfect conditions of 100% detection efficiency and no interfering background, would require a mass of the decaying isotope of approximately 100 kg. The situation confronting a real experiment is even more severe than this, because the detection efficiency is not perfect, and there are also many processes than can mimic a $\beta\beta0\nu$ signal, including ‘ordinary’ $\beta\beta2\nu$ decays. The latter can occur if the energy resolution of the experiment is insufficient to distinguish the tail of the $\beta\beta2\nu$ decay from the peak of the $\beta\beta0\nu$ decay. There could also be signals arising from naturally occurring radioactivity in the material of the detectors that could easily overwhelm the signal being sought. For this reason it is crucial to work with materials with minimal radioactive content, in addition to siting the experiment deep underground to shield it from cosmic rays.

Despite these considerable difficulties, a number of experiments have emerged using different techniques. These can be roughly divided into ‘internal-source systems’, in which the decaying material is contained in the detector, and ‘external-source systems’, in which it is not. The former has been implemented using many types of detector, such as scintillators, bolometers, solid-state devices, and gaseous chambers. If the decaying isotope is a semiconductor, the isotope can be both the sample and a solid-state detector that will measure the energy released in a given decay. In the bolometer type of experiment, the sample and detector are still combined, but the energy released in the decay is turned into heat and detected. Alternatively, in the external source category, the decaying sample may be surrounded by a separate detector that can observe and identify the electron tracks from the decays, as well as measure their energies. This has the advantage that a single detector can be used to study a range of different decaying isotopes. Three leading experiments of the internal source type current are EXO-200, KamLAND-Zen, and GERmanium Detector Array (GERDA). An experiment of the external source type is NEMO-3.

EXO-200 (Enriched Xenon Observatory) uses xenon, enriched to contain 81% of the double beta emitter isotope ^{136}Xe , which has a natural abundance of about 9%. This has a high $Q_{\beta\beta}$ value in a region that can

¹⁶Better nuclear models are also needed to improve determinations of the parameters of neutrino oscillations, the experiments for which nuclear targets are often used.

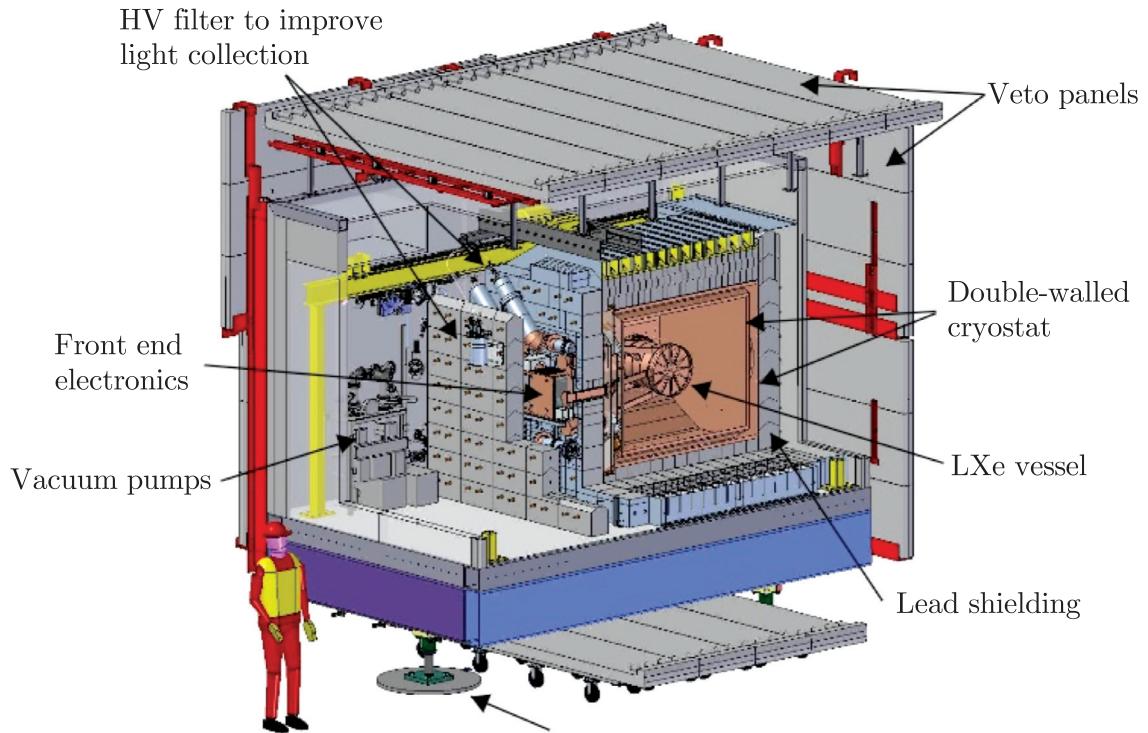


Figure 10.10 A sketch of the EXO-200 detector, cryostat and shielding. Source: Adapted from Auger et al. (2012).

have low contamination for radioactive backgrounds. It can also be dissolved in liquid scintillators, or used as a gas, to realise a homogeneous detector giving both scintillation and ionisation signals. The detector is shown in Figure 10.10. It consists of a cylindrical time projection chamber (TPC), about 40 cm in diameter and 44 cm in length, containing 200 kg of enriched liquid xenon (LXe). The TPC measures the coordinates and energy of ionisation products by simultaneously collecting scintillation light and charge. The xenon is held inside a thin copper vessel immersed in a cryofluid that also shields the detector from external radioactive backgrounds. The cryofluid is maintained at approximately 167 K inside a vacuum-insulated copper cryostat. Further shielding is provided by at least 25 cm of lead in all directions. Based on a fiducial volume containing 76.5 kg of LXe and an exposure of 100 kg yr, the EXO-200 collaboration has reported a lower limit on the half-life for the $\beta\beta0\nu$ decay of ^{136}Xe , and from (10.26), limits on the effective Majorana neutrino mass using a variety of calculated nuclear matrix elements. These are shown in Table 10.2.

The KamLAND-Zen experiment also uses xenon, dissolved in a liquid scintillator. The basic detector is shown in Figure 4.24, except for this application a balloon containing 383 kg of ^{136}Xe enriched xenon diluted in liquid scintillator is suspended inside the original plastic balloon, that

Table 10.2 Results from the EXO-200, KamLAND-Zen, GERDA and NEMO-3 experiments. The range of values shown for the upper bound on the Majorana neutrino mass is mainly due to uncertainties in the nuclear matrix elements used in (10.26)

Experiment	Isotope	Lower bound on $T_{1/2}$ (90% CL)	Upper bound on Majorana neutrino mass (90% CL)
EXO-200 ^a	^{136}Xe	1.10×10^{25} yr	(0.19 – 0.45) eV
KamLAND ^b	^{136}Xe	1.07×10^{26} yr	(0.061 – 0.165) eV
GERDA ^c	^{76}Ge	3.5×10^{25} yr	(0.15 – 0.33) eV
NEMO-3 ^d	^{100}Mo	1.1×10^{24} yr	(0.33 – 0.62) eV
NEMO-3 ^e	^{48}Ca	2.0×10^{22} yr	(6 – 26) eV
NEMO-3 ^f	^{150}Nd	2.0×10^{22} yr	(1.6 – 5.3) eV

^aAlbert et al. (2014), ^bGando et al. (2016), ^cAgostini et al. (2017), ^dArnold et al. (2015),
^eArnold et al. (2016a), ^fArnold et al. (2016b).

itself contains 1 kilotonne of pure liquid scintillator. Results based on an exposure of 110 kg yr are shown in Table 10.1.

The third experiment, GERDA, uses the nucleus ^{76}Ge . This isotope has long been used in $\beta\beta0\nu$ experiments because it can be embedded in solid-state detectors using a calorimetric approach with high-purity germanium diodes. However, there are disadvantages as well. As in the EXO-200 experiment, the isotope used here has a quite low natural abundance of about 7% and so isotope enrichment is necessary before the diodes can be constructed; also the $Q_{\beta\beta}$ value is not very high and lies in a region where background contamination is possible. Therefore, experiments using this isotope must pay special attention to screening all the materials close to the detectors, among other precautions. It does this by using ultrapure shielding against environmental radiation. The germanium detectors are suspended in strings into a cryostat where 64 m³ of liquid argon is used both as a coolant and shield. The vessel is surrounded by a large tank filled with high-purity water, which further shields the inner volumes from locally produced radiation. Results based on an exposure of 34.3 kg yr are shown in Table 10.2.

Finally, NEMO-3 is shown in Figure 10.11. This uses a tracker-calorimeter design, with the detector consisting of thin foils containing double beta isotopes sandwiched between the tracking detectors. Wire-chamber trackers measure particles' trajectories, while the segmented calorimeter provides energy and timing information. A unique feature of NEMO-3 is its ability to study double beta-decay processes for seven isotopes simultaneously. Results for the three nuclei, ^{100}Mo , ^{48}Ca , and ^{150}Nd , are shown in Table 10.2.

Comparing the experimental results in Table 10.2 with the predicted values shown in Figure 10.9 shows that there is a long way to go before neutrinoless double beta-decay experiments can confront theoretical predictions obtained, assuming neutrinos are indeed Majorana particles. All

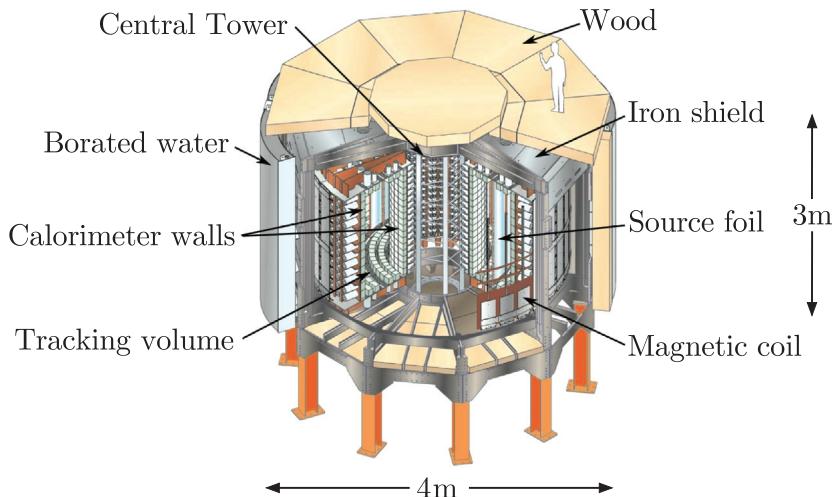


Figure 10.11 The NEMO-3 double beta decay detector.
Source: Courtesy of the NEMO-3 collaboration.

four experiments discussed above are currently being upgraded with this aim, and a number of new experiments are being designed or are under construction, using a range of detection methods and isotopes, with the same goal: to reach the region predicted by the inverted hierarchy. This will require increasing the effective mass of the decaying isotope by at least an order-of-magnitude to the 100 kg region, plus a similar-sized reduction in the background. If a signal is still not found, isotope mass in the 1000 kg range will be required, with another reduction in the background. The logistic and financial challenges to obtain such large qualities of isotopes, which are usually rare on Earth, will be truly formidable. On the other hand, neutrinoless beta decay may occur by other mechanisms ‘beyond the standard model’ or, in the case of Dirac neutrinos, may not occur at all.

10.5 Particle astrophysics

Particle physics and astrophysics interact in an increasing number of areas and the resulting field of particle astrophysics is a rapidly growing one. The interactions are particularly important in the field of cosmology, where, for example, the detection of neutrinos can provide unique cosmological information. Another reason is because the conditions in the early Universe implied by standard cosmological theory (the Big Bang model) can only be approached today, however remotely, in high-energy particle collisions. At the same time, these conditions occurred at energies that are relevant to the grand unified and SUSY theories of particle physics and so offer a possibility of testing the predictions of such theories. This is important because, as mentioned above, it is difficult to see other ways of testing such predictions. For reasons of space, we will discuss only a few examples of particle astrophysics.

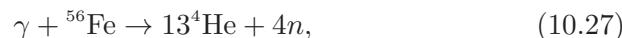
10.5.1 Neutrino astrophysics

We have seen in Chapter 3 that cosmic rays and emissions from the Sun are important sources of information about neutrinos and have led us to revise the view that neutrinos are strictly massless, as was originally assumed in the standard model. At the same time, there is considerable interest in studying ultrahigh-energy neutrinos as a potential source of information about galactic and extra-galactic objects and hence cosmology in general.

10.5.1(a) Supernovas and the neutrino mass

One of the first neutrino astrophysics experiments was the observation of neutrinos from a supernova. A simple description of the mechanism that produces a supernova is as follows. If a star has a mass greater than about 11 solar masses, it can evolve through all stages of fusion, ending in a core of iron surrounded by shells of lighter elements. Because energy cannot be released by the thermonuclear fusion of iron, the core will start to contract under gravity. Initially this is resisted by the electron degeneracy pressure¹⁷ of the dense gas of degenerate electrons in the core, but as more of the outer material is burned and more iron deposited in the core, the resulting rise in temperature makes the electrons become increasingly relativistic. When the core mass reaches about 1.4 solar masses (the so-called *Chandrasekhar limit*), the electrons become ultrarelativistic and they can no longer support the core. At this point the star is on the brink of a catastrophic collapse.

The physical reactions that lead to this are as follows. Firstly, photodisintegration of iron (and other nuclei) takes place,



which further heats the core and enables the photodisintegration of the helium produced, i.e.



As the core continues to collapse, the energy of the electrons present increases to a point where the weak interaction



becomes possible and eventually the hadronic matter of the star is predominantly neutrons. This stage is therefore called a *neutron star*. The

¹⁷The Pauli principle forbids the electrons occupying identical quantum states. Attempts to force them closer together results in some of them occupying higher energy levels, with larger particle momenta. The resulting force that resists the compression is called the ‘electron degeneracy pressure’.

collapse ceases when the gravitational pressure is balanced by the neutron degeneracy pressure, analogous to electron degeneracy pressure mentioned above. At this point the radius of the star is typically just a few kilometres. The termination of the collapse is very sudden and as a result the core material produces a shock wave that travels outwards through the collapsing outer material, leading to a supernova (a so-called Type II supernova). Initially there is an intense burst of ν_e with energies of a few MeV from reaction (10.29). This lasts for a few milliseconds because the core rapidly becomes opaque even to neutrinos and after this the core material enters a phase where all its constituents (nucleons, electrons, positrons, and neutrinos) are in thermal equilibrium. In particular, all flavours of neutrino are present via the reactions

$$\gamma \rightleftharpoons \ell^+ \ell^- \rightleftharpoons \nu_\ell \bar{\nu}_\ell \quad (\ell = e, \mu, \tau) \quad (10.30)$$

and these will eventually diffuse out of the collapsed core and escape. Neutrinos of all flavours, with average energies of about 15 MeV, will be emitted in all directions over a period of 0.1–10 seconds. Taken together, the neutrinos account for about 99% of the total energy released in a supernova. Despite this, the output in the optical region is sufficient to produce a spectacular visual effect.

The first experiments that detected neutrinos from a supernova were an earlier version of the Kamiokande experiment, described in Chapter 3 (see Figure 3.3), and the IMB collaboration, which also used a water Čerenkov detector. Both had been constructed to search for proton decay as predicted by GUTs, but by good fortune both detectors were ‘live’ in 1987 at the time of a spectacular supernova explosion (now named SN1987A) and both detected a small number of antineutrino events. The data are shown in Figure 10.12. The Kamiokande experiment detected 12 $\bar{\nu}_e$ events and the IMB experiment 8 events, both over a time interval of approximately 10 seconds and with energies in the range 0–40 MeV. These values are consistent with the estimates for the neutrinos that would have been produced by reaction (10.30) and then diffused from the supernova after the initial pulse.

The data can be used to make an estimate of the neutrino mass as follows. The time of arrival on Earth of a neutrino i is given by

$$t_i = t_0 + \left(\frac{L}{c} \right) \left(1 + \frac{m^2 c^4}{2E_i^2} \right), \quad (10.31)$$

where t_0 is the time of emission from the supernova at a distance L and (m, E_i) are the mass and total energy of the neutrino. Thus,

$$(\Delta t)_{ij} \equiv t_i - t_j = \frac{L m^2 c^4}{2c} \left(\frac{1}{E_i^2} - \frac{1}{E_j^2} \right). \quad (10.32)$$

Using data for pairs of neutrinos, (10.32) leads to the result

$$m_{\bar{\nu}_e} \leq 20 \text{ eV}/c^2, \quad (10.33)$$

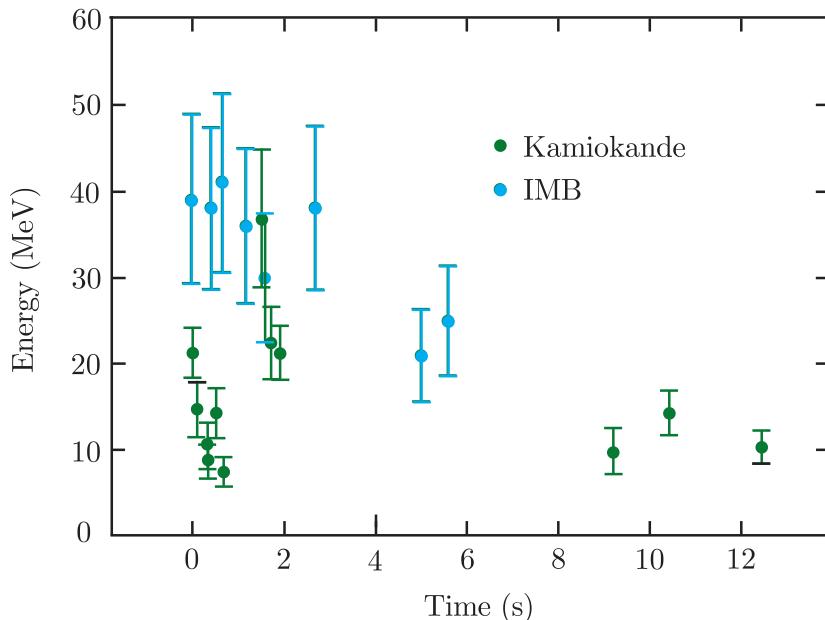


Figure 10.12 Data for neutrinos from SN1987A detected in the Kamiokande and IMB experiments. The thresholds for detecting neutrinos in the experiments are 6 MeV (Kamiokande) and 20 MeV (IMB). In each case the first neutrino detected is assigned the time zero.

which although larger than the value from tritium decay, is still a remarkable measurement.

10.5.1(b) Ultrahigh-energy neutrinos

The neutrinos from SN1987A were of low energy, but there is also a great interest in detecting ultrahigh-energy neutrinos. For example, it is known that there exist point sources giving bursts of γ -rays with energies in the TeV range, many of which have their origin within so-called ‘active galactic nuclei’. It is an open question whether this implies the existence of point sources of neutrinos with similar energies, which might help shed light on the problem of the origin of cosmic rays. The neutrinos to be detected would be those travelling upwards through the Earth, as the signal from downward travelling particles would be swamped by neutrinos produced via pion decay in the atmosphere above the detector. Like all weak interactions the intrinsic rate would be very low, especially so for such rare high-energy events, although this is partially compensated by the fact that the ν -nucleon cross-section increases with energy, as discussed in Section 5.5.4.

To detect neutrinos in the TeV energy range using the Čerenkov effect in water requires huge volumes, orders-of-magnitude larger than used in the SuperKamiokande detector. An ingenious solution to this problem is to use the vast quantities of water available in liquid form in the oceans, or frozen in the form of ice at the South Pole, and several experiments have been built, or are being built, using these sources. The largest so far is the IceCube detector, which is sited at the geographical South Pole. A schematic diagram of this detector is shown in Figure 10.13.

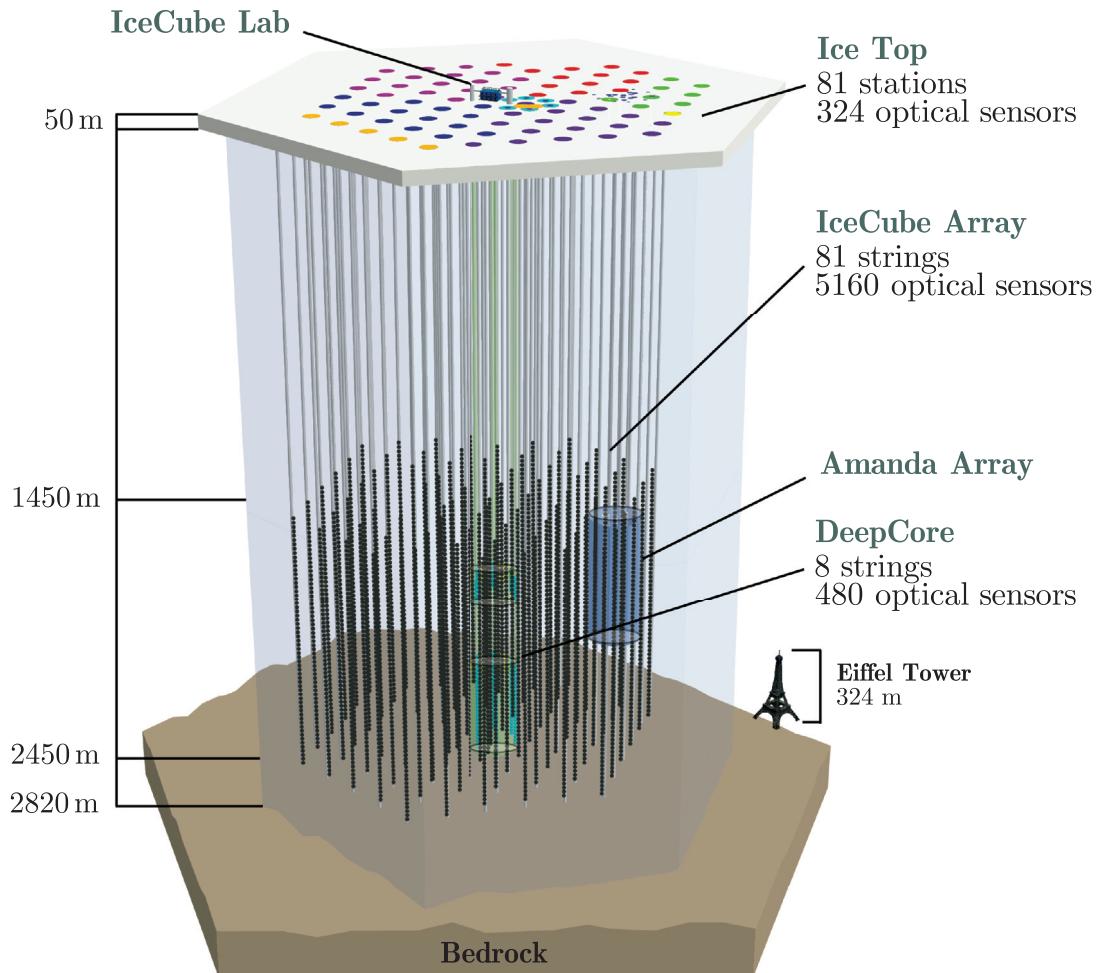


Figure 10.13 Schematic diagram of the IceCube neutrino detector. Source: Courtesy of the IceCube Collaboration.

Referring to Figure 10.13, the in-ice component of IceCube consists of 5160 digital optical modules (DOMs), each with a 10-inch photomultiplier tube, plus associated electronics. When neutrinos pass through the ice, they interact to produce charged leptons and these in turn produce Čerenkov radiation. The photomultiplier tubes convert the Čerenkov radiation to electrical signals. The modules are located in the ice at great depths by using a novel hot-water boring device. The ice then refreezes around them. They are attached to 81 vertical ‘strings’, and arrayed over a cubic kilometre from 1450 m to 2450 m in depth. The strings are deployed on a hexagonal grid with 125 m spacing and each holds 60 DOMs where the vertical separation of the DOMs is 17 m. They are most sensitive in detecting neutrinos with energies in the range $10^{11} - 10^{21}$ eV. Eight of these strings at the centre of the array are deployed more compactly, with a horizontal separation of 70 m and a vertical DOM spacing of 7 m. This

denser configuration forms the DeepCore subdetector, which lowers the neutrino energy threshold to below 100 GeV, creating the opportunity to study neutrino oscillations.¹⁸ There is also a smaller array of strings within the ice that constitute an earlier ‘proof of concept’ detector called AMANDA, but it is no longer active. On the surface of the glacier there is a km² sized Čerenkov array, called IceTop, consisting of 81 stations located on top of the same number of IceCube strings. These detect air showers from primary cosmic rays in the 300 TeV to 1 EeV energy range and also act as a veto, because if a muon is observed going through IceTop it cannot have originated from a neutrino interacting within the ice.

Several aspects of neutrino physics have been investigated by the IceCube Collaboration, including measuring the parameters of a neutrino oscillation, and searching for new particles that could be candidates for so-called ‘dark matter’ that is discussed in Section 10.5.2 below. Although no such particles have been detected, the results have significantly constrained the ranges of masses where such particles could exist. One of the most interesting results is the detection of a very small number of neutrinos that can be identified as having an extragalactic origin, although they are not coincident with observed γ -ray bursts. Like other detectors we have met, it is planned to upgrade IceCube to explore this in more detail.

Detection of even higher energy neutrinos is being pursued in several other innovative experiments. One of these is ANITA (Antarctic Impulsive Transient Antenna). Its primary interest is to address the nature of ultrahigh-energy cosmic rays ($E > 10^{19}$ eV) that have been observed over many years, but the origin of which is unknown. It seeks to do this by detecting the associated neutrinos produced by the interaction of the cosmic rays with the cosmic microwave photons that pervades the Universe. (The origin of this radiation is discussed in Section 10.5.2.) To do this it exploits an effect similar to Čerenkov radiation. In this case, a particle travelling faster than the speed of light in a dense radio-transparent medium produces a shower of charged particles that contain a charge anisotropy, and emits a cone of coherent radiation in the radio or microwave part of the electromagnetic spectrum. This is the *Askaryan effect*, discussed in Section 4.4.4(b). The neutrinos cascading through the Antarctic ice sheet lead to a strong electromagnetic pulse that propagates through the ice, because the latter is transparent to radio waves up to a frequency of 1.5 GHz. The ice sheet is thus a converter of neutrino energy to radio waves. The experiment consists of a detector system mounted on a balloon platform at a height of about 40 km above the Antarctic ice shelf. The balloon traverses a circumpolar flight path due to the continuous wind circulation around the South Pole and ‘sees’ the

¹⁸Some of the neutrinos that pass through the Earth will go through the core and others will graze through while avoiding it. With enough data, this will allow investigation of the material properties of the Earth’s core, which as noted in Section 9.5, is difficult to sample.

ice below out to the horizon at about 700 km. Thus the effective telescope ‘lens’ has an area of approximately 1.5×10^6 m². The flights made to date have resulted in constraints on the spectrum and flux of ultrahigh-energy neutrinos. Experiments such as IceCube and ANITA, together with others (including those discussed in Section 10.4.1 devoted to measuring the mass of the neutrino from neutrinoless double beta decay), will ensure that neutrino physics will be at the forefront of particle physics research for many years to come.

10.5.2 Cosmology and dark matter

Modern theories of particle physics and cosmology are closely related. This is because the conditions in the early Universe implied by the standard cosmological theory can only be approached, even remotely, in high-energy particle collisions. At the same time, laboratory experiments cannot reach anything like the energies at which we might expect to see the full effects of unification, so that the conditions that existed shortly after the formation of the Universe constitute a unique laboratory for the study of unification models. Here we will concentrate on a few aspects of this connection.

The modern description of the Universe is based on the observation that it is expanding and assumes that the origin of this is a sudden explosion at some time in the past. For this reason the description is called the *Big Bang model*. Because the Universe appears isotropic at large distance scales, there can be no preferred points in space and the Big Bang must have occurred everywhere at once, thus ensuring that the expansion appears the same to all observers irrespective of their locations. Two pieces of evidence for this model are the existence of a cosmic background radiation (CMB),¹⁹ now known to be very accurately represented by a black-body spectrum at an effective temperature of 2.7 K, and the cosmic abundance of light elements.²⁰ Whether the expansion will continue indefinitely depends on the average density of the Universe ρ . The critical density ρ_c at present times, below which the expansion will continue indefinitely, and above which it will eventually halt and the Universe start to contract, can be written

$$\rho_c = \frac{3H_0^2}{8\pi G} \approx 0.85 \times 10^{-26} \text{ kg/m}^3 \approx 4.8 (\text{GeV}/c^2)/\text{m}^3, \quad (10.34)$$

where G is the gravitational constant and we have used the current value for Hubble’s constant H_0 to evaluate (10.34). In the most popular version of the model, called the *inflationary Big Bang model*, the relative density

$$\Omega \equiv \rho/\rho_c = 1. \quad (10.35)$$

¹⁹Arno Penzias and Robert Wilson shared half of the 1978 Nobel Prize in Physics for their discovery of the cosmic microwave background radiation.

²⁰For an accessible discussion of the Big Bang model and other matters discussed in this section see, for example, Roos (2015).

This is conveniently written as the sum of three components,

$$\Omega \equiv \Omega_{\text{total}} = \Omega_r + \Omega_m + \Omega_\Lambda, \quad (10.36)$$

where Ω_r is the contribution due to *radiation*, Ω_m is that due to *matter*, and Ω_Λ is called the *cosmological term*, or *dark energy*. When Einstein first applied his theory of general relativity to cosmology, he introduced a new term into the equation of motion, with a magnitude characterised by a new *cosmological constant* Λ . This produces a repulsion, which opposed the gravitational attraction of matter, and it was later realised that the existence of such a term would follow from a uniform, nonzero energy density in the vacuum state. Such a term is necessary to account for the increasing expansion rate of the Universe revealed by studies of distant supernovae and is the largest of the three terms contributing to the relative density (10.36).

Of these terms, only Ω_r has been accurately directly measured, from an analysis of the precisely known form of the CMB. Numerically it is $\Omega_r \approx 5 \times 10^{-5}$, i.e. a negligible contribution to Ω_{tot} . The first estimates of the relative matter density were obtained from the gravitational energy needed for consistency with the observations of the rotations of galaxies and the kinematics of large-scale structures in the Universe mentioned earlier, and gave estimates in the range 0.24–0.30. However, the best values of both the matter and dark energy densities are obtained from precise data on the tiny temperature variations of the black body radiation obtained by the WMAP space probe and the Planck satellite²¹ together with data on the so-called ‘baryon acoustic oscillations’.²² The current best values of relative mass density Ω_m lie in the range 0.31 ± 0.01 , while the corresponding range for the relative dark energy density Ω_Λ is 0.69 ± 0.01 .²³ These analyses usually assume (10.35), but this should be a very good approximation, since when this constraint is relaxed, the value

$$\Omega_{\text{total}} = 1.0005 \pm 0.0033 \quad (10.37)$$

is obtained.

There are two unsatisfactory features of the decomposition (10.36). One is that the origin of the largest term, Ω_Λ , is totally unknown. The second concerns the composition of the matter density Ω_m . The contribution

²¹John Mather and George Smoot shared the 2006 Nobel Prize in Physics for their discovery of the black body form and anisotropy of the cosmic microwave background radiation.

²²In the early Universe, collisions between the hot particles would have produced acoustic waves that bounced between them. Residual patterns of these waves, left after the particles were too far apart for the waves to bounce between them, can be observed in small variations in the galaxy distribution of the present day Universe and are called baryon acoustic oscillations (see, for example, Davis 2014).

²³Numerical values are from Lahav and Liddle (2017) and Lesgourges and Verde (2017), where a more detailed account of the assumptions made and data used may be found.

of luminous baryonic matter²⁴ is obtained from the observed matter in the form of stars and intergalactic gas and dust and is about 0.01, so that the matter density is dominated by dark matter, as noted at the beginning of this section. There could also be sources of nonluminous baryonic matter, for example in the form of brown dwarfs and small black holes the size of planets, and there is experimental evidence that such ‘massive, compact halo objects’ (*MACHOs*) do indeed exist in the halo of our galaxy. However, it is thought that they alone cannot account for much of the ‘missing’ matter and the total baryonic contribution to Ω_m may be inferred from knowledge of how nuclei are formed in the Universe (nucleosynthesis – discussed in Section 9.2.3). Its value is $\Omega_b \approx 0.05$. Since Ω_m lies between 0.29 and 0.32, we are forced to conclude that the bulk of matter in the Universe is nonbaryonic.

The conclusion that there had to be a substantial quantity of nonluminous matter in the Universe had already been deduced by the astronomer Fritz Zwicky in the 1930s. He noted that there was a discrepancy between the mass estimates of clusters of galaxies based on luminosity calculations and those resulting from dynamical estimates based on the assumption that they were gravitationally bound systems. The latter resulted in much larger masses than those based on luminosity calculations, and Zwicky concluded that this implied there must be some form of nonluminous matter present in the clusters that dominated their overall mass content. This observation was reinforced by observations on the rotational motion of stars within galaxies, which were only consistent with the laws of gravity if the visible stars were surrounded by a so-called ‘halo’ of nonluminous matter that contained the bulk of the galaxies total mass.

The nature of dark matter is unknown, although several candidates have been suggested. In the following sections we will discuss two possible candidates, while a third is considered in Section 10.5.4 below.

10.5.2(a) Neutrinos

Neutrinos are one possible dark matter candidate. Such particles would have to be heavy enough to have been nonrelativistic in the early stages of the Universe (so-called *cold dark matter*), because if they were relativistic (*hot dark matter*) they would have rapidly dispersed, giving rise to a uniform energy distribution in space. Calculations suggest that in this case there would have been insufficient time for the observed galaxies to have formed. The contribution of massive neutrinos to the matter term can be calculated once the number of species and their masses are known. For masses in the range 5×10^{-4} eV to 1 MeV, the contribution of neutrinos to Ω_m is

$$\Omega_\nu = \sum m_\nu / 42 \text{ eV}/c^2, \quad (10.38)$$

²⁴Cosmologists use the term ‘baryonic matter’ to mean ordinary matter including electrons as well as nucleons.

which immediately gives a limit $\Omega_\nu < 0.14$ for the contribution of the three known neutrinos to the matter density, since from (3.44) their masses are each less than about $2 \text{ eV}/c^2$. However, this bound can be greatly improved. This is because neutrinos with masses even as small as $0.1 \text{ eV}/c^2$ can have an observable effect on the formation of large-scale structure in the Universe by dampening the growth of perturbations. This enables an upper limit to be placed on Ω_ν/Ω_m . Using the standard cosmological model, together with data on the cosmic microwave background, gives the value

$$\Omega_\nu < 0.016 \quad (10.39)$$

at the 95% confidence level, so that neutrinos can at most make a very small contribution to the relative density of nonbaryonic dark matter required by the cosmological data. In addition, one sees from (10.38) that this corresponds to a limit of

$$\sum m_\nu \leq 0.67 \text{ eV}/c^2 \quad (10.40)$$

for the sum of neutrino masses. This value is much lower than the upper bound of about $2 \text{ eV}/c^2$ obtained from tritium decay, and together with the very small measured values of the squared mass splittings from oscillation experiments implies a limit of about $0.22 \text{ eV}/c^2$ on all three neutrino masses.

10.5.2(b) WIMPs

The most popular hypothesis is that the bulk of dark matter consists of nonbaryonic cold dark matter in the form of hitherto unknown particles with masses in the range from about $10 \text{ GeV}/c^2$ to several TeV/c^2 whose interaction with ordinary matter is roughly the same strength as the familiar weak interaction. Such particles are called ‘weakly interacting massive particles’, or *WIMPs*. If they exist, they would have been produced in large numbers following the Big Bang, and calculations indicate that they might account for the bulk of the dark matter density observed today, provided they are stable or, at least, extremely long-lived. The problem is that, since *WIMPs* do not correspond to any known particles, new types of particle have to be postulated. One possibility is the lightest particle that appears in supersymmetric theories, which is stable if *R* parity is conserved. In most such theories this is the lightest neutralino, which interacts only via electroweak interactions. Its mass is not predicted, but we have seen in Section 10.3.2 that in the most popular supersymmetric theories it must be greater than about $50 \text{ GeV}/c^2$ to accord with current experimental limits. However, other theories beside supersymmetry suggest different possibilities, and we shall not assume any particular identity for WIMPs in what follows, except when explicitly stated.

If WIMPs are the dominant form of dark matter, their number can be inferred from the estimated density of dark matter. From estimates of this

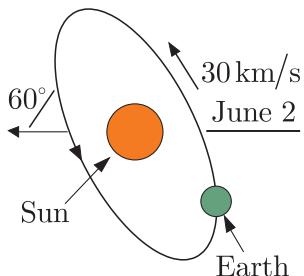


Figure 10.14 Motion of the Sun and Earth through the dark matter background.

in our own galaxy, it follows that of order 10^5 WIMPs per second will pass through every cm^2 of the Earth's surface, if their mass is assumed to be of order $60 \text{ GeV}/c^2$. The velocities of WIMPs relative to the galaxy will be similar to those of baryonic matter, which are of order $10^{-3} c$, and when they interact with nuclei, the nuclei will recoil with energies in the range 5 keV to about 50 keV. These recoils can be detected in several ways, as we shall see below. The WIMP–nucleon cross-sections are unknown, but could be of order 10^{-44} cm^2 , leading to just a few events per year per kilogram of target material. This is very small compared to the event rate from cosmic rays and naturally occurring radioactivity, including that in the materials of the detectors themselves. The former is minimised by working deep underground to shield the detector from cosmic rays and in areas with geological structures where radioactive rocks are absent; the latter is minimised by building detectors of extreme purity. These are similar requirements to those for experiments searching for neutrinoless double beta decay. Even with these precautions, the interpretation of possible events is difficult and can be controversial, as we will see. However, any WIMP signal obtained is expected to have a distinctive time dependence as a consequence of the Earth's motion around the Sun. The Sun moves in the galactic plane with a speed of order $2 \times 10^5 \text{ m s}^{-1}$. The Earth moves around the Sun with a speed of order $3 \times 10^4 \text{ m s}^{-1}$, in an orbit that is inclined at about 60° to the galactic plane, as seen in Figure 10.14. Hence the velocity of the Earth through dark matter varies by about $\pm 15\%$ over the course of the year. This leads to a corresponding modulation in the flux and energy of WIMP collisions, resulting in a maximum event rate in May/June and a minimum in November/December. The modulation is unfortunately only a few per cent in effect, and rather precise measurements will be required to detect it.

Several experiments have been built to detect WIMPs in the galactic halo by detecting the nuclear recoils when they collide with the material of a detector, as described above. This can be done in a variety of ways. The earliest experiments usually relied on the fact that in semiconductors such as GaAs, free charge will be produced that can be detected electronically. Alternatively, in a scintillator such as NaI or LXe, the emission of photons can be detected using photomultipliers, while cryostatic detectors exploit the fact that in crystals at low temperatures the energy can be converted to phonons and can be detected by a very small rise in temperature. Examples are the DAMA experiment, which is a scintillation detector containing 250 kg of sodium iodide crystals, and CRESST, which is a cryostatic detector that uses a similar amount of calcium tungstate crystals at a temperature of a few thousandths of a degree Kelvin. Both these experiments are located in the Gran Sasso Laboratory deep under the Italian Alps; both reported a WIMP signal that exhibited the characteristic time dependence described above, albeit in different kinematic regions. However, these results, and those of other experiments that suggest a possible signal, are controversial and seem to be in conflict with other experiments. In general, if WIMPs are assumed to saturate the

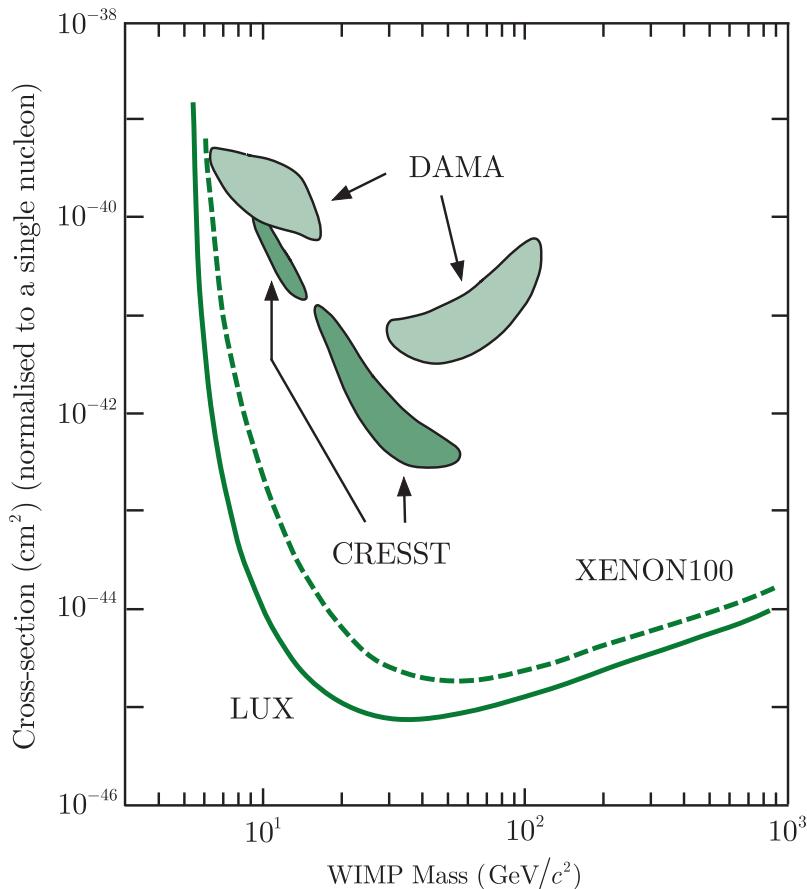


Figure 10.15 WIMP cross-section (normalised to a single nucleon) as a function of the WIMP mass. Source: Adapted from Drees and Gerbier (2015).

missing dark matter, the number density of WIMPs for any given mass is determined so that the number of events is determined by the WIMP mass and the WIMP–nucleon cross-section only. The failure to observe any events sets an upper limit on the WIMP–nucleon cross-section for any given WIMP mass. The most sensitive results obtained so far are from the XENON100 detector at Gran Sasso, which contains 161 kg of LXe, and the LUX detector at the SURF laboratory in the Homestake mine in South Dakota, which contains 170 kg of xenon in both gas and liquid form. The upper bounds on the cross-section obtained in these experiments are shown in Figure 10.15 and, as can be seen, the DAMA and CRESST results, mentioned above, lie well above them. Overall, there is no convincing evidence for WIMPS from direct observations²⁵ and the bounds already obtained impose serious constraints on any theory, such as supersymmetry, which purports to explain the origin of WIMPS. In addition, several further experiments are under development that will either detect WIMPS or considerably improve the existing bounds. These

²⁵For fuller reviews of dark matter searches, see Figueroa-Feliciano (2011), Drees and Gerbier (2015), and Roos (2015).

include the LUX-ZEPLIN experiment, which will use nine tonnes of LXe as its detector, and DEEP at the SNOLAB facility in Canada, which will use 3 tonnes of liquid argon to measure the recoil energy of struck nuclei and will be sensitive to higher mass values than the majority of earlier experiments (see Figure 10.15).

Not all experiments are direct searches. The Alpha Magnetic Spectrometer (AMS) is an experiment in orbit on the International Space Station. It uses a 7 tonne 0.15 T magnet to sort incoming particles by charge and momentum and searches for an excess of charged particles such as positrons in the energy range 500 MeV to 1 TeV, whose origin could be the interaction between WIMPs. It has reported an excess of positrons over electrons that decreases from 1 to 10 GeV, in accordance with theoretical expectations, but then increases from 10 to 250 GeV. This confirms results from an earlier experiment, but in both cases other conventional interpretations have been suggested. Other experiments include searches for high energy neutrinos emitted from the Sun. WIMPs would be expected to be slowed down and captured in the Sun over a long period of time, and the resulting higher WIMP density would then lead to a WIMP-enhanced pair annihilation rate to particles, which will decay, among other things, to high-energy neutrinos. Hence, if WIMPs exist, one would expect a larger rate of high-energy neutrinos incident on the Earth from the Sun, compared to neighbouring directions, which could be observed in detectors such as IceCube, although it was designed to investigate other phenomena. No WIMP signal has yet been detected by IceCube but the interpretation of the results is complicated, since it requires an understanding of WIMP capture in the Sun and of WIMP annihilations. However, within theories such as supersymmetry, the results are a powerful constraint on WIMP parameters and are an important complement to direct searches, especially at high WIMP masses.

10.5.3 Matter–antimatter asymmetry

One of the most striking facts about the Universe is the paucity of antimatter compared to matter. There is ample evidence for this. For example, cosmic rays are overwhelmingly composed of matter and what little antimatter is present is compatible with its production in intergalactic collisions of matter with photons. Neither do we see intense outbursts of electromagnetic radiation that would accompany the annihilation of clouds of matter with similar clouds of antimatter. The absence of antimatter is completely unexpected, because in the original Big Bang it would be natural to assume a total baryon number $B = 0$.²⁶ Then during the period when kT was large compared to hadron energies, baryons and

²⁶One could of course simply bypass the problem by arbitrarily assigning an initial nonzero baryon number to the Universe, but it would have to be exceedingly large to accommodate the observed asymmetry, as well as being an unaesthetic ‘solution’.

antibaryons would be in equilibrium with photons via reversible reactions such as



and this situation would continue until the temperature fell to a point where the photons no longer had sufficient energy to produce $p\bar{p}$ pairs and the expansion had proceeded to a point where the density of protons and antiprotons was such that their mutual annihilation became increasingly unlikely. The critical temperature is $kT \approx 20$ MeV and at this point the ratios of baryons and antibaryons to photons ‘freezes’ to values that can be calculated to be

$$N_B/N_\gamma = N_{\bar{B}}/N_\gamma \sim 10^{-18}, \quad (10.42)$$

with of course $N_{\bar{B}}/N_B = 1$. These ratios would then be maintained in time, whereas the actual observed ratios are

$$N_B/N_\gamma \sim 10^{-9}, \quad N_{\bar{B}}/N_\gamma \sim 10^{-13}, \quad (10.43)$$

with $N_{\bar{B}}/N_B \sim 10^{-4}$. The simple Big Bang model fails spectacularly.

The conditions whereby a baryon–antibaryon asymmetry could arise were first stated by Sakharov. It is necessary to have: (i) an interaction that violates charge conjugation C and the combined symmetry CP ; (ii) an interaction that violates baryon number; and (iii) a nonequilibrium situation must exist at some point to ‘seed’ the process. The reason for the first condition is that a baryon excess cannot be generated if the production of any particular particle (e.g. the proton) is balanced by an equal production of the corresponding antiparticle (the antiproton), as required by C or CP conservation. Baryon number violation is also obviously necessary and seems at first to be incompatible with the standard model in which, under present conditions, baryon number is conserved and the symmetry breaking between the electromagnetic and weak interactions is characterised by an energy scale of order 100 GeV. However, immediately after the Big Bang, thermal energies were very large compared to the energy scale of electroweak symmetry breaking and, if current theory is correct, the Universe underwent a phase transition about 10^{-7} s after the Big Bang from a state in which the electroweak symmetry was fully realised to the present state in which it is badly broken. Furthermore, it can be shown that during this transition, in which the Universe was in a nonequilibrium state, satisfying requirement (iii), sufficient baryon number violations occurred to enable the observed matter–antimatter asymmetry to be understood, provided there is sufficient C and CP violations. Unfortunately, while both C and CP violations occur in the standard model, the CP violation is much too small to explain the observed asymmetry. In contrast, supersymmetric theories, with their host of new particles, contain several unknown mixing angles and CP -violating phases, and by adjusting these parameters it is possible to induce much larger CP -violating effects than those predicted in the standard model. However, until these, or CP -violating effects from some other source ‘beyond the standard model’, are

experimentally detected, the origin of the matter–antimatter asymmetry in the Universe remains an unsolved problem.

10.5.3(a) *CP violation and electric dipole moments*

One way to detect hitherto unknown sources of *CP* violation is to search for non-zero EDMs of leptons and hadrons. Consider the nonrelativistic Hamiltonian H_{EDM} for the interaction of an EDM \mathbf{d} with an electric field \mathbf{E} . For an elementary particle, or a hadron, its spin \mathbf{J} is the only vector available to define a direction and so \mathbf{d} must be collinear with \mathbf{J} . Thus we can write

$$H_{\text{EDM}} = -\mathbf{d} \cdot \mathbf{E} = -\frac{d}{J} \mathbf{J} \cdot \mathbf{E},$$

where $d = |\mathbf{d}|$ and $J = |\mathbf{J}|$. However, \mathbf{E} is even under time-reversal, while \mathbf{J} is odd, and thus H_{EDM} is odd under time-reversal. It is also odd under parity. Thus a non-zero EDM can only exist if both parity (P) and time-reversal (T) invariance are violated, and by implication any non-zero EDM detected would imply *CP* violation, assuming that *CPT* invariance holds.

The magnitudes of predicted EDMs depend upon the assumed mechanism that causes *CP* violation. For example, in the standard model, a nonzero EDM could be generated by P , T violating weak radiative corrections to the P , C , and T conserving electromagnetic interaction. For the neutron and the electron, which are the two most extensively studied, the estimates from the standard model are

$$d_n \sim 10^{-32} \text{ e cm} \quad \text{and} \quad d_e \leq 10^{-38} \text{ e cm}, \quad (10.44)$$

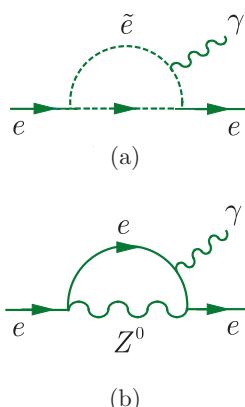


Figure 10.16 (a) An example of a diagram involving superparticles that can lead to a nonzero electron electric dipole moment in supersymmetric theories; (b) a corresponding diagram without superparticles, which is *CP*-conserving in the standard model.

where e is the magnitude of the charge on the electron. For the electron EDM, for example, this is 16 orders of magnitude smaller than the electron magnetic moment and is far too small to be measured in the foreseeable future. However, in supersymmetric theories much larger *CP*-violating effects can be generated by diagrams involving virtual superparticles, like Figure 10.16a for the electron. In comparison, the corresponding particle diagram in Figure 10.16b does not violate *CP* invariance in the standard model, so a nonzero electron EDM can only arise in higher orders in perturbation theory and is consequently much smaller. Unfortunately, the magnitude of the EDM in supersymmetric theories is difficult to pin down, because it depends on the nature of supersymmetry breaking, which inevitably involves many unknown parameters. However, ‘weak scale’ supersymmetric theories, which also attempt to resolve cosmological problems like dark matter and matter–antimatter asymmetry, usually predict values in the ranges $(10^{-25} - 10^{-28}) \text{ e cm}$ and $(10^{-27} - 10^{-30}) \text{ e cm}$ for the neutron and electron EDMs, respectively. Other theoretical speculations also lead to much larger values than those predicted in the standard model, and the detection of EDMs significantly larger than those given in (10.44) would be clear evidence for new *CP*-violating effects, whatever their origin.

For neutral spin-1/2 particles, nonzero EDMs can in principle be detected by making use of the fact that in collinear \mathbf{E} and \mathbf{B} fields, the particle's spin will undergo a Larmor precession²⁷ about the field direction. The frequency f of this precession is given by

$$\hbar f = (2\mu B \pm 2dE), \quad (10.45)$$

where μ is the magnetic moment, d is the EDM, and the \pm sign refers to E being parallel (antiparallel) to B . Thus, a nonzero d_n will be revealed by an electric field-dependent shift in f . The experiments are technically very demanding, because of the relative sizes of μ and d , and innovative techniques are required to detect the frequency shift. For example, ultracold neutrons with kinetic energies less than about 10^{-7} eV are stored in magnetic traps for many hundreds of minutes while the experiments are performed.²⁸ Several experiments are planned that develop this basic idea and an improvement of a factor of about 100 on the present limit seems possible. The same technique cannot be used directly for charged particles, such as the electron, because the interaction with the electric field will accelerate the particle out of the observed region. However, the EDM of a neutral atom or molecule with an unpaired electron can be measured, and it can be shown that in this case the effective field experienced by d is enhanced by a factor $R = d_a/d$ in (10.45), where d_a is the EDM of the atom. The results can then be used to deduce the EDM for electrons and nucleons. This leads to the best current experimental limits (at 90% confidence level):

$$|d_n| < 2.9 \times 10^{-26} \text{ e cm}, \quad |d_p| < 5.0 \times 10^{-25} \text{ e cm}, \quad |d_e| < 8.7 \times 10^{-29} \text{ e cm}. \quad (10.46)$$

Finally, we should mention that in the standard model, EDMs of leptons are approximately proportional to the particle mass and so from the present limit on the EDM of the electron, we can predict $|d_\mu| < 2.6 \times 10^{-25}$ e cm for the muon, which is already smaller than the predictions from some extensions of the standard model discussed in Section 10.3.1. The present experimental limit is $|d_\mu| < 1 \times 10^{-20}$ e cm and was obtained in a storage ring experiment that was designed primarily to measure the difference between the magnetic dipole moment of the muon and its Dirac value. The technique is being developed to improve this measurement and also to measure the EDM for the proton and deuteron.

10.5.4 Axions and the strong CP problem

There is another reason for measuring EDMs. In the standard model, the theory of strong interactions (QCD) is assumed to satisfy both gauge

²⁷See Footnote 35 in Chapter 9.

²⁸The principle is the same as in the traps used for making accurate mass measurements of very short-lived isotopes, discussed in Section 2.1.3.

invariance and CP invariance. However, it is possible to add a term to the interaction that satisfies gauge invariance, but violates CP . Noninvariance would permit a range of particle decays, for example $\eta^0 \rightarrow \pi^+ \pi^-$, none of which have ever been seen. Whether such a term is present or not is often called ‘the strong CP problem’. If it is present, its magnitude is determined by a coupling constant $\bar{\theta}\alpha_s$, the size of which is not specified by the theory, and its contribution to the EDM of the neutron is $d_n \approx 3 \times 10^{-16} \bar{\theta} \text{ e cm}$. Hence the detection of a neutron EDM, which is too large to be explained by the standard model, could indicate either a nonzero value for $\bar{\theta}$ or some other completely new physics, while a large EDM for the electron would definitely be proof of the latter. On the other hand, the experimental limit on the neutron EDM given in (10.46) implies that $\bar{\theta} \leq 1 \times 10^{-10}$, an extraordinarily small value for which there is no obvious explanation. However, it can be shown that by introducing an appropriate new symmetry into the theory, a new particle with $J^P = 0^-$, named the *axion*, is predicted and in turn $\bar{\theta}$ becomes zero.²⁹ There is an analogy here with the introduction of the Higgs boson, although the motivation is nowhere near as strong. The mass of the axion is unknown, but is given approximately by³⁰

$$m_a \approx \left(\frac{6 \times 10^6 \text{ GeV}}{f_a} \right) \text{ eV}/c^2, \quad (10.47)$$

where f_a is an unknown energy scale at which the new symmetry is broken. Axions would mainly couple via the electromagnetic interaction to photons, and would convert to two photons in the presence of a strong magnetic field. The strength of this coupling is given by

$$g_{a\gamma\gamma} = \alpha g_\gamma / \pi f_a, \quad (10.48)$$

where α is the fine structure constant and g_γ is a model-dependent dimensionless constant of order unity. Equations (10.47) and (10.48) imply a relation between the axion mass m_a and $g_{a\gamma\gamma}$, which is represented by the diagonal band in Figure 10.17. A light axion would be an excellent dark matter candidate as its density relative to the critical density ρ_c of the Universe is given by

$$\Omega_a \approx \left(\frac{6 (\mu\text{eV}/c^2)}{m_a} \right)^{7/6}. \quad (10.49)$$

An axion of mass $m_a \approx 20 (\mu\text{eV}/c^2)$ would therefore account for the entire dark matter contribution $\Omega_m \approx 0.3$. Much smaller values of the mass would produce too large a value for Ω_m , so a strong lower bound on the mass is $m_a > 1 (\mu\text{eV}/c^2)$. On the other hand, much larger masses would make a negligible contribution to the mass density, but would

²⁹The axion is an example of a class of proposed particles called WISPs (Weakly Interacting Slim Particles).

³⁰See, for example, Cerrillo (2015).

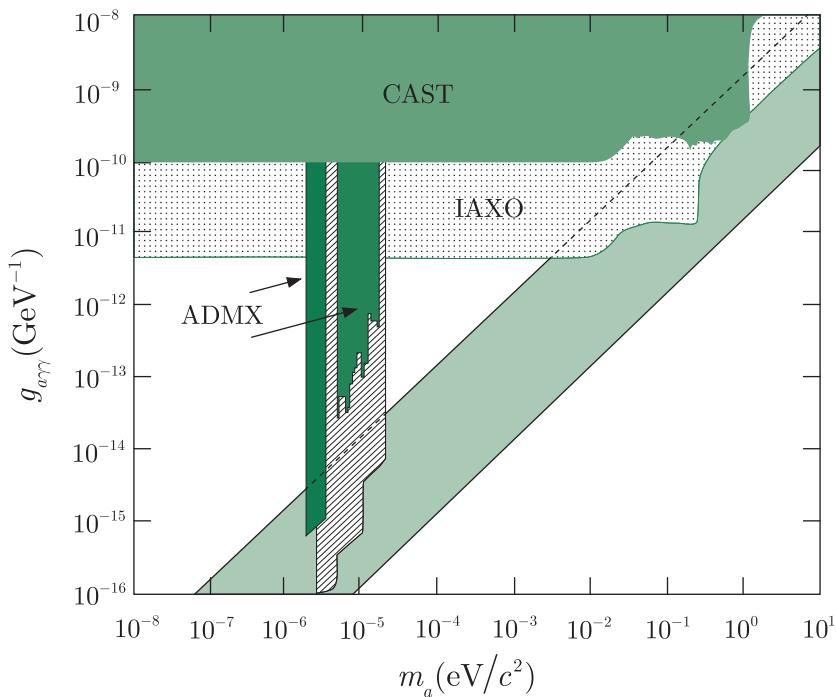


Figure 10.17 Plot of the various possible regions for dark-matter axions in the $g_{a\gamma\gamma} - m_a$ plane. The pale green diagonal band shows the expected region of viable dark matter from conventional axion models (see text). The green shaded areas labelled CAST and ADMX show the regions in the $g_{a\gamma\gamma} - m_a$ plane excluded by these two experiments, while the dotted and cross-hatched areas show the regions that will be probed by the planned IAXO and ADMX experiments. Source: Adapted from Dafni and Iguax (2015).

still be relevant to the strong CP problem. In addition, various astrophysical data seem to exclude masses between about $2 \times 10^{-2} \text{ eV}/c^2$ and $10^{-5} \text{ eV}/c^2$, so that the favoured window for axion masses is now $10^{-6} (\text{eV}/c^2) < m_a < 10^{-5} (\text{eV}/c^2)$. However, these various bounds apply to ‘conventional’ axions, in which the mass and photon couplings are related by (10.48) and (10.49), and do not exhaust all possibilities. For example, all string-like theories have a myriad of axion and axion-like particles, although they are predicted to have masses of order $10^{-9} \text{ eV}/c^2$ and we will not consider them further here.

Axions light enough to contribute significantly to dark matter were initially thought to be so weakly coupled as to be undetectable in conventional experiments. However, it was later shown that axions present in the Milky Way could be resonantly converted into an accurately monochromatic microwave signal in a microwave cavity permeated by a strong magnetic field. The conversion power depends on a number of quantities outside the experimenter’s control, such as the axion mass, its coupling to photons, and the local axion density in the galactic halo. It also depends on quantities that can be controlled, such as the strength of the magnetic field and the characteristics of the cavity. The experiments were conducted at very low temperatures of a few millikelvin to minimise spurious noise, and the searches were conducted by tuning the cavity in small overlapping steps. Several experiments have used variants of the cavity resonance technique to explore regions of $g_{a\gamma\gamma} - m_a$ space, but to date no definite signals have been reported. The region of most interest in the $g_{a\gamma\gamma} - m_a$ plane is that corresponding to conventional axions light enough

to contribute to dark matter. One experiment – the Axion Dark Matter Experiment (ADMX) located at the University of Washington – has already begun to penetrate this region and will do so more extensively in the future, as can be seen in Figure 10.17.

Axions could also be produced in the solar interior by conversion of plasma photons in the Coulomb field of charged particles, and these can be efficiently converted back to photons in the presence of a strong electromagnetic field. Dedicated ‘helioscope’ experiments have searched for axions using powerful magnets to effect the conversion. The best results to date come from the CERN Solar Axion Telescope (CAST) that uses an LHC prototype magnet to produce a magnetic field of 9 T over a length of 9.3 m. CAST has been operating since 2003 and its latest bounds are shown in Figure 10.17. A significant improvement over these results is expected when a new facility, called the International Axion Observatory (IAXO), becomes operational. This will have a signal-to-noise ratio five orders-of-magnitude better than that in CAST, and will considerably extend the region of the $g_{a\gamma\gamma} - m_a$ plane that can be explored, as shown in Figure 10.17.

Finally, we stress that the above are just a sample of the experiments performed or planned, not all of which rely on cosmological or astrophysical sources.³¹ The search for axions is very much a work in progress.

³¹For a recent review of axion searches, see, for example, Graham et al. (2015).

A

Some results in quantum mechanics

In this appendix we briefly discuss several aspects of quantum mechanics that are used in various sections of the main text.

A.1 Barrier penetration

Consider the one-dimensional potential shown in Figure A.1a. Free particles of mass m and energy E represented by plane waves are incident from the left and encounter the rectangular barrier of constant height V , where $V > E$.

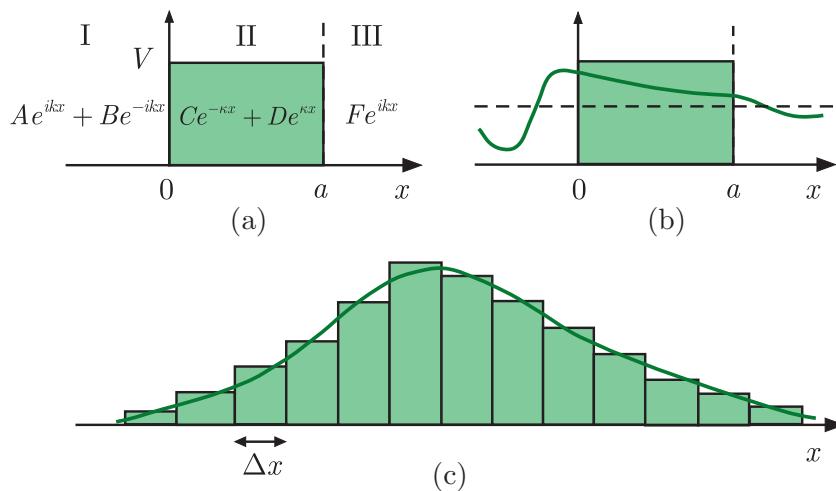


Figure A.1 (a) Rectangular barrier with wavefunction solutions. (b) Form of the real parts of the incoming and outgoing waves. (c) Modelling an arbitrary smooth barrier as a series of rectangular barriers.

In region I ($x < 0$), there is an incoming wave e^{ikx} , with the wavenumber k given by

$$\hbar^2 k^2 = 2mE, \quad (\text{A.1})$$

and also a wave reflected at the barrier travelling from right to left of the form e^{-ikx} . Thus the total wavefunction in region I is

$$\psi_1(x) = A e^{ikx} + B e^{-ikx}, \quad (\text{A.2})$$

where A and B are complex constants. Within the barrier, region II ($0 < x < a$), the solution of the Schrödinger equation is a decaying exponential, plus an exponential wave reflected from the boundary at $x = a$, so the total wavefunction is

$$\psi_2(x) = C e^{-\kappa x} + D e^{\kappa x}, \quad (\text{A.3})$$

where C and D are complex constants and κ is given by

$$\hbar^2 \kappa^2 = 2m(V - E). \quad (\text{A.4})$$

Finally, in region III ($x > a$) to the right of the barrier, there is only an outgoing wave of the form

$$\psi_3(x) = F e^{ikx}, \quad (\text{A.5})$$

where again F is a complex constant.

We are interested in the *transmission coefficient* T , defined by

$$T \equiv |F/A|^2. \quad (\text{A.6})$$

The values of F and A are found by imposing continuity of the wavefunction and its first derivative, i.e. matching the values of these quantities, at the two discontinuous boundaries $x = 0$ and $x = a$. The algebra may be found in any introductory book on quantum mechanics.¹ The result is

$$T = \left| \frac{2k\kappa e^{-ika}}{2k\kappa \cosh(\kappa a) - i(k^2 - \kappa^2) \sinh(\kappa a)} \right|^2. \quad (\text{A.7})$$

The corresponding incident and transmitted waves are shown in Figure A.1b (the reflected waves are not shown).

For large κa , which corresponds to small penetrations, we can make the replacement

$$\sinh(\kappa a) \approx \cosh(\kappa a) \approx e^{\kappa a}/2, \quad (\text{A.8})$$

and hence

$$T \approx \left(\frac{4k\kappa}{k^2 + \kappa^2} \right)^2 e^{-2\kappa a}. \quad (\text{A.9})$$

The first factor is due to the reflection losses at the two boundaries $x = 0$ and $x = a$; the decreasing exponential describes the amplitude decay within the barrier. The first factor is slowly varying with energy and is usually neglected.

¹See, for example, Chapter 2 of Mandl (1992).

The result (A.9), ignoring the first factor, may be used to find the transmission coefficient for an arbitrary smoothly varying barrier by modelling it as a series of thin rectangular barriers. This is illustrated in Figure A.1c. Thus by making the replacement

$$2\kappa a \rightarrow 2 \sum \kappa(x)\Delta x,$$

and taking the limit of small Δx , the summation goes over to an integral, i.e.

$$2\kappa a \rightarrow 2 \int dx \left\{ \frac{2m}{\hbar^2} [V(x) - E] \right\}^{1/2} \quad (\text{A.10})$$

and

$$T \approx \exp \left[-2 \int dx \left\{ \frac{2m}{\hbar^2} [V(x) - E] \right\}^{1/2} \right]. \quad (\text{A.11})$$

This is the essence of what is known as the WKB approximation in quantum mechanics. Equation (A.11) was used in Section 8.6 to discuss α decay.

A.2 Density of states

Consider a spinless particle of mass m confined within a cube of sides L and volume $V = L^3$, oriented so that one corner is at the origin $(0, 0, 0)$ and the edges are parallel to the x , y , and z axes. If the potential is zero within the box, then the walls represent infinite potential barriers and the solutions of the Schrödinger equation must therefore vanish on all faces of the cube. It is straightforward to show that the solutions of the Schrödinger equation satisfying these boundary conditions are standing waves of the form

$$\psi(x, y, z) = C \sin(k_x x) \sin(k_y y) \sin(k_z z), \quad (\text{A.12})$$

where C is a constant and the components of the wavenumber $\mathbf{k} = (k_x, k_y, k_z)$ take the values

$$k_x = n_x \pi / L, \quad k_y = n_y \pi / L, \quad k_z = n_z \pi / L, \quad (n_x, n_y, n_z) = 1, 2, 3, \dots \quad (\text{A.13})$$

The energy of the particle is given by

$$E = \frac{\hbar^2}{2m} (k_x^2 + k_y^2 + k_z^2) = \frac{\hbar^2 k^2}{2m} = \frac{1}{2m} \left(\frac{\hbar \pi}{L} \right)^2 (n_x^2 + n_y^2 + n_z^2), \quad (\text{A.14})$$

where $k \equiv |\mathbf{k}| = p/\hbar$ and p is the particle's momentum. Negative values of the integers do not lead to new states since they merely change the sign of the wavefunction (A.12) and the phase factors have no physical significance.

The allowed values of \mathbf{k} form a cubic lattice in the quadrant of ‘ \mathbf{k} -space’, where all the values of (n_x, n_y, n_z) are positive. Since each state

corresponds to one combination of (n_x, n_y, n_z) , the number of allowed states is equal to the number of lattice points. The spacing between the lattice points is (π/L) , so the density of points per unit volume in \mathbf{k} -space is $(L/\pi)^3$. The number of lattice points $n(k_0)$, with k less than some fixed value k_0 , is the number contained within a volume that for large values of k_0 may be well approximated by the quadrant of a sphere of radius k_0 , i.e.

$$n(k_0) = \frac{1}{8} \cdot \frac{4}{3} \pi k_0^3 \left(\frac{L}{\pi} \right)^3 = \frac{V}{(2\pi)^3} \frac{4\pi k_0^3}{3}. \quad (\text{A.15})$$

Hence the number of points with k in the range $k_0 < k < (k_0 + dk_0)$ is

$$dn(k_0) = \frac{V}{(2\pi)^3} 4\pi k_0^2 dk_0. \quad (\text{A.16})$$

The *density of states* is defined as $\rho(k_0) \equiv dn(k_0)/dk_0$ and so is given by

$$\rho(k_0) = \frac{V}{(2\pi)^3} 4\pi k_0^2. \quad (\text{A.17})$$

Thus $\rho(k_0) dk_0$ is the number of states with k between k_0 and $k_0 + dk_0$, or equivalently

$$\rho(p) dp = \frac{4\pi V}{(2\pi\hbar)^3} p^2 dp \quad (\text{A.18})$$

is the number of states with momentum between p and $p + dp$. This is the form used in (8.1) when discussing the Fermi energy in the Fermi gas model. Equation (A.18) can also be written in terms of energy using $E = p^2/2m$, when it becomes

$$\rho(E) dE = \frac{4\pi V}{(2\pi\hbar)^3} mp dE. \quad (\text{A.19})$$

Although the above derivation is for a particle confined in a box, the same technique can be used for scattering problems. In this case we can consider a large volume $V = L^3$ and impose ‘periodic’ boundary conditions

$$\psi(x + L, y, z) = \psi(x, y + L, z) = \psi(x, y, z + L) = \psi(x, y, z). \quad (\text{A.20})$$

Instead of standing waves, the solutions of the Schrödinger equation consistent with (A.20) are the travelling waves

$$e^{i\mathbf{k}\cdot\mathbf{r}} = \exp(ik_x x)\exp(ik_y y)\exp(ik_z z), \quad (\text{A.21})$$

where

$$k_x = \frac{2n_x\pi}{L}, \quad k_y = \frac{2n_y\pi}{L}, \quad k_z = \frac{2n_z\pi}{L}, \quad n_x, n_y, n_z = 0, \pm 1, \pm 2, \dots \quad (\text{A.22})$$

The density of lattice points in \mathbf{k} -space now becomes $(L/2\pi)^3$, but unlike the standing wave case, permutations of signs in (A.22) do produce new states and the whole quadrant of lattice points has to be considered. Thus these two effects ‘cancel out’ and we arrive at the same result for the density of states (A.18) and (A.19).

In scattering problems it is also useful to consider the number of states within the momentum space volume

$$d^3\mathbf{p} = p^2 dp d\Omega, \quad (\text{A.23})$$

corresponding to momenta with magnitude in the range p to $p + dp$ and contained within the cone of solid angle $d\Omega$ (see Figure 1.8). Using (A.22), we have

$$\rho(p)dp = \left(\frac{L}{2\pi\hbar}\right)^3 d^3\mathbf{p} = \frac{V}{(2\pi\hbar)^3} p^2 dp d\Omega. \quad (\text{A.24})$$

This may be expressed in terms of energy E by using

$$\rho(E)dE = \rho(p) \frac{dp}{dE} dE, \quad (\text{A.25})$$

where, nonrelativistically,

$$dp/dE = 1/v \quad (\text{A.26})$$

and v is the velocity of the scattered particle. Thus,

$$\rho(E) = \frac{V}{(2\pi\hbar)^3} \frac{p^2}{v} d\Omega. \quad (\text{A.27})$$

This result is used in Chapter 1 to derive an expression for the differential cross-section, (1.69).

All of the above is for spinless particles. If the particles have spin, then the density of states must be multiplied by the appropriate spin multiplicity factor, taking account of the Pauli principle as necessary. Thus, for example, for spin-1/2 particles, with two spin states, (A.19) becomes

$$\rho(E)dE = \frac{8\pi V}{(2\pi\hbar)^3} mp dE. \quad (\text{A.28})$$

A.3 Perturbation theory and the Second Golden Rule

Without detailed proof, we will outline the derivation from perturbation theory of the important relationship between the transition probability per unit time for a process and its matrix element.²

In perturbation theory, the Hamiltonian at time t may be written in general as

$$H(t) = H_0 + V(t), \quad (\text{A.29})$$

²The derivation follows that given in Chapter 9 of Mandl (1992).

where H_0 is the unperturbed Hamiltonian and $V(t)$ is the perturbation, which we will assume is small. The solution for the eigenfunctions of H starts by expanding in terms of the complete set of energy eigenfunctions $|u_n\rangle$ of H_0 , i.e.

$$|\psi(t)\rangle = \sum_n c_n(t) |u_n\rangle \exp(-iE_n t/\hbar), \quad (\text{A.30})$$

where E_n are the corresponding energies. If $|\psi(t)\rangle$ is normalised to unity, then the squared coefficient $|c_n(t)|^2$ is the probability that at time t the system is in a state $|u_n\rangle$. Substituting (A.30) into the Schrödinger equation leads to a differential equation for the transition coefficients:

$$i\hbar \frac{dc_f(t)}{dt} = \sum_n V_{fn}(t) \exp(i\omega_{fn}t) c_n(t), \quad (\text{A.31})$$

where the matrix element $V_{fn} \equiv \langle u_f | V(t) | u_n \rangle$ and the angular frequency $\omega_{fn} \equiv (E_f - E_n)/\hbar$. If we assume initially ($t = 0$) that the system is in a state $|u_i\rangle$, then $c_i(0) = \delta_{ni}$ and the solutions for $c_f(t)$ are found by substituting this result into the right-hand side of (A.31), giving, *to first-order in V*

$$c_i(t) = 1 + \frac{1}{i\hbar} \int_0^t V_{ii}(t') dt' \quad (\text{A.32a})$$

and

$$c_f(t) = \frac{1}{i\hbar} \int_0^t V_{fi}(t') \exp[i\omega_{fi}t'] dt' \quad (f \neq i). \quad (\text{A.32b})$$

For $f \neq i$, the quantity $|c_f(t)|^2$ is the probability, in first-order perturbation theory, that the system has made a transition from state i to state f .

The above is for a general time-dependent perturbation $V(t)$, but the results can also be used to describe other situations, for example where the perturbation is zero up to some time t_0 and a constant thereafter. In this case, the integrals in (A.32a) and (A.32b) can be evaluated and, in particular, (A.32b) gives, again to first-order in V ,

$$c_f(t) = \frac{V_{fi}}{\hbar\omega_{fi}} (1 - e^{i\omega_{fi}t}) \quad (\text{A.33})$$

and hence the probability of the transition $i \rightarrow f$ is

$$P_{fi}(t) = |c_f(t)|^2 = \frac{4|V_{fi}|^2}{\hbar^2} \left[\frac{\sin^2(\frac{1}{2}\omega_{fi}t)}{\omega_{fi}^2} \right]. \quad (\text{A.34})$$

The function in the square brackets in (A.34) is shown in Figure A.2.

For sufficiently large values of t , (A.34) has the form of a large central peak with much smaller side oscillations. In this case P_{fi} is only appreciable if

$$\hbar|\omega_{fi}| = |E_f - E_i| \leq 2\pi\hbar/t \quad (\text{A.35})$$

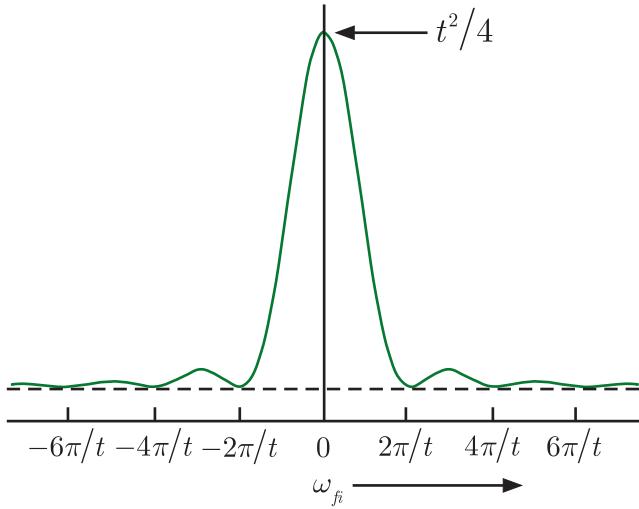


Figure A.2 The function $(\sin^2(\frac{1}{2}\omega_{fi}t)/\omega_{fi}^2)$.

and then the square bracket can be replaced by a Dirac delta function,³ when

$$\lim_{t \rightarrow \infty} \frac{\sin^2(\frac{1}{2}\omega_{fi}t)}{\omega_{fi}^2} = \frac{1}{2}\pi\hbar t\delta(E_f - E_i), \quad (\text{A.36})$$

where the external factors are to preserve the normalisation. Then

$$P_{fi}(t) = t\frac{2\pi}{\hbar}|V_{fi}|^2\delta(E_f - E_i) \quad (\text{A.37})$$

and the transition probability per unit time is

$$\frac{dP_{fi}(t)}{dt} = \frac{2\pi}{\hbar}|V_{fi}|^2\delta(E_f - E_i). \quad (\text{A.38})$$

The above assumes that the final state is discrete, but it is more common for the final states to form a continuum defined by the density of states $\rho(E)$ derived in Section A.2 above. In this case, since $\rho(E)dE$ is the number of states with energy between E and $E+dE$, we can write the

³The delta function is defined by the two conditions:

(i) $\delta(x' - x) = 0$, if $x \neq x'$ and (ii) $\int_{-\infty}^{+\infty} \delta(x' - x) dx' = 1$.

It follows that if $f(x)$ is a function continuous in the interval $x_1 < x < x_2$, then

$$\int_{x_1}^{x_2} f(x') \delta(x' - x) dx' = f(x) \text{ if } x_1 < x < x_2, \text{ or } 0 \text{ if } x < x_1 \text{ or } x > x_2.$$

The delta function was introduced by Dirac and was the first so-called ‘generalised function’.

transition rate per unit time dT_f/dt to a group of states f with energies in this range as

$$\frac{dT_f}{dt} = \int \frac{dP_f(t)}{dt} \rho(E_f) dE_f = \frac{2\pi}{\hbar} [|V_f|^2 \rho(E_f)]_{E_f=E_i}, \quad (\text{A.39})$$

where the integral has been evaluated using the properties of the delta function. The result (A.39) is called the *Second Golden Rule* (sometimes Fermi's Second Golden Rule, although strictly the result is not due to Fermi, but Dirac) and has been used in several places in this book, for example in Chapter 1 when discussing potential scattering and Chapter 8 when discussing nuclear β decay.

A.4 Isospin formalism

The mathematical formalism of isospin is identical to that of angular momentum in quantum mechanics and so we will draw on that analogy in this section.⁴

A.4.1 Isospin operators and quark states

If $\hat{\mathbf{I}}$ is an isospin operator with components $\hat{I}_{x,y,z}$, then many of the properties we have used elsewhere in this book follow from the commutation relations

$$[\hat{I}_i, \hat{I}_j] = i \hat{I}_k, \quad (\text{A.40})$$

where i, j, k is a cyclic permutation of x, y, z . An equivalent form of (A.40) is

$$[\hat{I}_+, \hat{I}_-] = 2\hat{I}_3, \quad [\hat{I}_3, \hat{I}_\pm] = \pm \hat{I}_\pm, \quad (\text{A.41})$$

where

$$\hat{I}_\pm \equiv \hat{I}_x \pm i\hat{I}_y, \quad \hat{I}_3 \equiv \hat{I}_z. \quad (\text{A.42})$$

The operators \hat{I}_\pm are called *ladder* (or *shift*) *operators*, with \hat{I}_+ being referred to as a *raising operator* and \hat{I}_- as a *lowering operator*. These equations are extremely useful in applications, and in particular they can be used to show (cf. Problem A.2) that the $(2I+1)$ states $|I, I_3\rangle$ have the same energy if isospin is conserved, i.e. if

$$[H, \hat{I}_\pm] = [H, \hat{I}_3] = 0. \quad (\text{A.43})$$

The origin of these names will become clear presently.

We start by considering the action of \hat{I}_\pm and \hat{I}_3 on the basic set of quark states $q \equiv |q, \Psi\rangle$, where Ψ specifies the spin and space properties. For brevity these will be omitted in future, but it is assumed they remain unchanged under the action of an isospin operator. If we assume that the

⁴The discussion is based on that given in Appendix C of Martin and Shaw (2017).

states u and d are analogous to the ‘up’ and ‘down’ states of a spin-1/2 particle, the action of \hat{I}_3 is straightforward. Thus,

$$\hat{I}_3 u = \frac{1}{2} u, \quad \hat{I}_3 d = -\frac{1}{2} d \quad (\text{A.44a})$$

and for the antiquarks

$$\hat{I}_3 \bar{u} = -\frac{1}{2} \bar{u}, \quad \hat{I}_3 \bar{d} = \frac{1}{2} \bar{d}. \quad (\text{A.44b})$$

The action of \hat{I}_{\pm} is less obvious, but can again be deduced by analogy with spin-1/2 operators, by explicitly constructing the analogous ladder operators from the Pauli spin matrices.⁵ The results are

$$\hat{I}_+ u = 0, \quad \hat{I}_+ d = u, \quad \hat{I}_- u = d, \quad \hat{I}_- d = 0 \quad (\text{A.45a})$$

and for the antiquarks

$$\hat{I}_+ \bar{u} = -\bar{d}, \quad \hat{I}_+ \bar{d} = 0, \quad \hat{I}_- \bar{u} = 0, \quad \hat{I}_- \bar{d} = -\bar{u}. \quad (\text{A.45b})$$

In addition,

$$\hat{I}_{\pm} r = \hat{I}_3 r = 0 \quad (\text{A.45c})$$

for all other quarks and antiquarks r .

To discuss hadrons, we have to consider the action of the operators (A.42) on an arbitrary state $|A\rangle$ composed of quarks and antiquarks. Such a state may be written

$$|A\rangle = \sum_i \beta_i |A_i\rangle, \quad (\text{A.46})$$

where $|A_i\rangle$ is any combination of quarks or antiquarks and β_i are arbitrary constants. Since the isospin operators are linear,

$$\hat{I}_{\alpha} |A\rangle = \sum_i \beta_i \hat{I}_{\alpha} |A_i\rangle, \quad (\text{A.47})$$

where $\alpha = +, -, 3$. The right-hand side of (A.47) may then be evaluated using (A.45a) to (A.45c) by exploiting the distributive laws

$$\hat{I}_{\alpha}(ab) = (\hat{I}_{\alpha}a)b + a(\hat{I}_{\alpha}b) \quad (\text{A.48a})$$

and

$$\hat{I}_{\alpha}(abc) = (\hat{I}_{\alpha}a)bc + a(\hat{I}_{\alpha}b)c + ab(\hat{I}_{\alpha}c), \quad (\text{A.48b})$$

where a , b , and c are any quark or antiquark. It is straightforward to show from the above equations that the operators \hat{I}_{\pm} and \hat{I}_3 do indeed satisfy the commutation relations (A.41).

⁵See, for example, Sections 5.2 and 5.8 of Mandl (1992).

By analogy with angular momentum, the commutation relations (A.41) lead to the existence of a family of $(2I+1)$ states $|I, I_3\rangle$, which are simultaneous eigenstates of the operators \hat{I}_3 and

$$\hat{I}^2 \equiv \hat{I}_x^2 + \hat{I}_y^2 + \hat{I}_z^2 = \frac{1}{2}(\hat{I}_+ \hat{I}_- + \hat{I}_- \hat{I}_+) + \hat{I}_3^2, \quad (\text{A.49a})$$

with eigenvalues given by

$$\hat{I}^2 |I, I_3\rangle = I(I+1) |I, I_3\rangle \quad (\text{A.50a})$$

and

$$\hat{I}_3 |I, I_3\rangle = I_3 |I, I_3\rangle, \quad (\text{A.50b})$$

where

$$I_3 = I, I-1, \dots, -I. \quad (\text{A.51})$$

Equivalent, and in many applications more useful, relations are

$$\hat{I}_{\pm} |I, I_3\rangle = C_{\pm}(I, I_3) |I, I_3 \pm 1\rangle, \quad (\text{A.52a})$$

with

$$C_{\pm}(I, I_3) = +[(I \mp I_3)(I \pm I_3 + 1)]^{1/2}. \quad (\text{A.52b})$$

Equation (A.52a) shows why \hat{I}_{\pm} are called raising and lowering operators because they increase or decrease the values of I_3 by one unit. These equations are very useful in applications. For example, in Problem A.2 they can be used to show that the $(2I+1)$ states $|I, I_3\rangle$ all have the same energy if isospin is conserved, that is if

$$[H, \hat{I}_{\pm}] = [H, \hat{I}_3] = 0. \quad (\text{A.52c})$$

A.4.2 Hadron states

We have seen in Chapter 3 that hadron states exist in multiplets, which we will write as

$$\begin{pmatrix} |a; I, I_3 = I\rangle \\ |a; I, I_3 = I-1\rangle \\ \vdots \\ \vdots \\ |a; I, I_3 = -I\rangle \end{pmatrix}, \quad (\text{A.53})$$

where a labels the particle type, for example the pion π . In this notation, the basic quark and antiquark isodoublets are

$$\begin{pmatrix} u \\ d \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \bar{d} \\ -\bar{u} \end{pmatrix}, \quad (\text{A.54})$$

so that

$$u = \left| q; \frac{1}{2}, \frac{1}{2} \right\rangle, \quad d = \left| q; \frac{1}{2}, -\frac{1}{2} \right\rangle \quad (\text{A.55a})$$

and

$$\bar{u} = - \left| \bar{q}; \frac{1}{2}, -\frac{1}{2} \right\rangle, \quad \bar{d} = \left| \bar{q}; \frac{1}{2}, \frac{1}{2} \right\rangle. \quad (\text{A.55b})$$

All other quarks and antiquarks are assigned to isosinglets with $I = I_3 = 0$. Since the strange quark is such a state, combining it with the states (A.54) does not change their isospin properties. Thus we can deduce immediately that the states

$$\begin{pmatrix} u\bar{s} \\ d\bar{s} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} s\bar{d} \\ -s\bar{u} \end{pmatrix} \quad (\text{A.56})$$

are also isodoublets. With the usual quark assignments, this means that the hadron states

$$\begin{pmatrix} K^+ \\ K^0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \bar{K}^0 \\ -K^- \end{pmatrix} \quad (\text{A.57})$$

form isodoublets.

For the case of mesons that are bound states of u and d quarks and their antiparticles, we start from (A.55a) and (A.55b). This gives the unique identification

$$|I = 1, I_3 = 1\rangle = -u\bar{d}, \quad (\text{A.58a})$$

where the negative sign is chosen to agree with the usual phase convention. The other $I = 1$ states then follow by using the lowering operator \hat{I}_- and (A.52a) to (A.52c). Thus,

$$\hat{I}_- |I = 1, I_3 = 1\rangle = \sqrt{2} |I = 1, I_3 = 0\rangle \quad (\text{A.59})$$

and

$$\hat{I}_-(-u\bar{d}) = -(\hat{I}_- u) \bar{d} - u(\hat{I}_- \bar{d}) = -d\bar{d} + u\bar{u}, \quad (\text{A.60})$$

where we have used (A.45a) to (A.45c) and (A.48a) and (A.48b). Comparing (A.59) and (A.60) gives

$$|I = 1, I_3 = 0\rangle = (u\bar{u} - d\bar{d})/\sqrt{2}. \quad (\text{A.58b})$$

Acting again with \hat{I}_- leads to the remaining $I = 1$ state

$$|I = 1, I_3 = -1\rangle = d\bar{u}. \quad (\text{A.58c})$$

Finally, there remains the state

$$|I = 0, I_3 = 0\rangle = \alpha u\bar{u} + \beta d\bar{d}, \quad (\text{A.61a})$$

where the form of the right-hand side follows from the fact that $I_3 = 0$. In addition, we can use the ladder operator \hat{I}_+ to give

$$\hat{I}_+ |I = 0, I_3 = 0\rangle = 0, \quad (\text{A.61b})$$

and combining this with the normalisation condition $\alpha^2 + \beta^2 = 1$, we have $\alpha = \beta = 1/\sqrt{2}$ so that

$$|I = 0, I_3 = 0\rangle = \frac{1}{\sqrt{2}}(u\bar{u} + d\bar{d}), \quad (\text{A.62})$$

where we have chosen the overall phase to agree with the standard convention. Finally, identifying pions with the states (A.59), we obtain the isotriplet

$$\begin{pmatrix} -\pi^+ \\ \pi^0 \\ \pi^- \end{pmatrix}, \quad (\text{A.63})$$

with the quark assignments

$$\pi^+ = u\bar{d}, \quad \pi^0 = \frac{1}{\sqrt{2}}(u\bar{u} - d\bar{d}), \quad \pi^- = d\bar{u}. \quad (\text{A.64})$$

The isosinglet (A.62) is identified with the η -meson. Similar arguments can be used to deduce baryon multiplets.

The above formulas have been used in a number of places in the book, including Section 3.3 when we discussed various aspects of hadron spectroscopy in the quark model.

Problems A

- A.1** Use the relations (A.40) and (A.42) to verify the commutation relations (A.41).
- A.2** Use the relations (A.52a) to (A.52c) and (A.53) to show that the $2I+1$ members of an isomultiplet $|I, I_3\rangle$ all have the same energy if isospin is conserved.
- A.3** Use the definitions (A.44a), (A.44b), and (A.45a) to (A.45c) to verify that (A.50a) and (A.50b) and (A.52a) to (A.52c) are satisfied for the quark and antiquark assignments (A.54).
- A.4** A resonance $X^0(1520)$ decays via the strong interaction to the final states $n\pi^0$ and $p\pi^-$ with branching ratios of approximately 18% and 36%, respectively. What is its isospin?
- A.5** The $\Lambda(1405)$ resonance decays by the strong interaction to $\Sigma\pi$ final states with a branching ratio of 42%. What are the individual branching ratios for the charged modes $\Sigma^+\pi^-$, $\Sigma^0\pi^0$, and $\Sigma^-\pi^+$ if the $\Lambda(1405)$ is an isosinglet state?

B

Relativistic kinematics

In particle physics, most scattering interactions take place between particles whose speeds are comparable to the speed of light c . This is often true even in decays, particularly if light particles are emitted. The requirements of special relativity therefore cannot be ignored. In nuclear physics, accurate predictions can also often only be obtained if relativistic effects are taken into account. In this appendix we review (usually without proof) some relativistic kinematical results and the use of invariants to simplify calculations.¹

B.1 Lorentz transformations and four-vectors

Consider a particle of *rest mass* m in an inertial frame of reference S , i.e. one in which Newton's laws hold. Its coordinates are $(t, \mathbf{r}) \equiv (t, x, y, z)$ and its speed is $u = |\mathbf{u}|$, where \mathbf{u} is its velocity. In a second inertial frame S' its coordinates are $(t', \mathbf{r}') \equiv (t', x', y', z')$ and its speed is $u' = |\mathbf{u}'|$ where \mathbf{u}' is its velocity. If S and S' coincide at $t = 0$ and S' is moving with uniform speed v in the positive z direction with respect to S , so that the axes in the two frames remain parallel, then the two sets of coordinates are related by the *Lorentz transformation*

$$x' = x; \quad y' = y; \quad z' = \gamma(v)(z - vt); \quad t' = \gamma(v)(t - vz/c^2), \quad (\text{B.1})$$

where $\gamma(v) = (1 - \beta^2)^{-1/2}$ is the *Lorentz factor* and $\beta \equiv v/c$. As $v \rightarrow 0$, the transformations (B.1) approach the Galilean transformations. The most general Lorentz transformation has its simplest form in terms of

¹For a more complete coverage of relativistic kinematics see, for example, Hagedorn (1964).

four-vectors, whose general form is $a = (a_0, a_1, a_2, a_3) = (a_0, \mathbf{a})$. Then (B.1) becomes

$$a'_0 = \gamma (a_0 - va_3/c); a'_1 = a_1; a'_2 = a_2; a'_3 = \gamma (a_3 - va_0/c). \quad (\text{B.2})$$

For example, the space-time four-vector is $x = (ct, \mathbf{x})$ and when used in (B.2) reproduces (B.1). The scalar product of two four-vectors a and b is defined as

$$ab \equiv a_0 b_0 - \mathbf{a} \cdot \mathbf{b} \quad (\text{B.3})$$

and is an *invariant*, i.e. is the same in all inertial frames of references.

The basic four-vector in particle kinematics is the *four-momentum*, defined by

$$P \equiv mu, \quad (\text{B.4a})$$

with

$$u = \gamma(v)(c, \mathbf{v}), \quad (\text{B.4b})$$

where u is the *four-velocity*, \mathbf{v} is the usual three-velocity, and $v \equiv |\mathbf{v}|$. In terms of the *total energy* E (i.e. including the rest mass) and the three-momentum \mathbf{p} ,

$$P = (E/c, \mathbf{p}). \quad (\text{B.5})$$

Thus for two four-momentum P_1 and P_2 , the invariant scalar product is

$$P_1 P_2 = E_1 E_2 / c^2 - \mathbf{p}_1 \cdot \mathbf{p}_2 \quad (\text{B.6})$$

and for $P_1 = P_2 = P$,

$$P^2 = E^2 / c^2 - \mathbf{p}^2. \quad (\text{B.7})$$

However, from (B.3), (B.4a), and (B.4b), we have $u^2 = c^2$ and hence $P^2 = m^2 c^2$, so combining this with (B.7) gives

$$E^2 = \mathbf{p}^2 c^2 + m^2 c^4. \quad (\text{B.8})$$

It follows that

$$E = \gamma(v) mc^2, \quad \mathbf{p} = \gamma(v) m \mathbf{v}, \quad \mathbf{v} = c^2 \mathbf{p}/E. \quad (\text{B.9})$$

The Lorentz transformations for energy and momentum follow from these definitions and (B.2). Thus, in S' we have

$$E' = mc^2 \gamma(u') = \gamma(v) (E - vp) \quad (\text{B.10a})$$

and

$$p' = mu' \gamma(u') = \gamma(v) (p - vE/c^2), \quad (\text{B.10b})$$

where $p = |\mathbf{p}|$ and $p' = |\mathbf{p}'|$. For a set of N noninteracting particles,

$$p'_z = \gamma(v) (p_z - vE/c^2); \quad p'_x = p_x; \quad p'_y = p_y; \quad (\text{B.11a})$$

and

$$E' = \gamma(v) (E - vp_z), \quad (\text{B.11b})$$

where

$$E = \sum_{i=1}^N E_i \quad \text{and} \quad \mathbf{p} = \sum_{i=1}^N \mathbf{p}_i. \quad (\text{B.11c})$$

In the general case where the relative velocity \mathbf{v} of the two frames is in an arbitrary direction, the transformations (B.10a) and (B.10b) become

$$\mathbf{p}' = \mathbf{p} + \gamma \mathbf{v} \left(\frac{\gamma}{\gamma+1} \mathbf{v} \cdot \mathbf{p} - E \right) \frac{1}{c^2}, \quad E' = \gamma(E - \mathbf{v} \cdot \mathbf{p}). \quad (\text{B.12})$$

B.2 Frames of reference

The two most commonly used frames of reference for particle kinematics are the *laboratory system* (LS) and the *centre-of-mass system* (CMS). We will start by discussing these in the context of two-particle scattering. In the laboratory system, a moving projectile a in a beam strikes a target particle b at rest, i.e.

$$P_a = (E_a/c, \mathbf{p}_a), \quad P_b = (m_b c, \mathbf{0}). \quad (\text{B.13})$$

In the CMS, the three-momenta of the two particles a and b are equal and opposite, so that the total three-momentum is zero,² i.e.

$$P_a = (E_a/c, \mathbf{p}_a), \quad P_b = (E_b/c, \mathbf{p}_b), \quad (\text{B.14a})$$

with

$$\mathbf{p}_a + \mathbf{p}_b = \mathbf{0}. \quad (\text{B.14b})$$

In a colliding beam accelerator, these two views become mixed. The colliding particles are both moving, but only if they have equal momenta and collide at a zero crossing angle is the system identical to the CMS.

The four-vectors of the initial-state particles in the two systems may be written (L = laboratory, T = target) as

$$P_a = (E_L/c, 0, 0, p_L), \quad P_T = (m_T c, 0, 0, 0) \quad \text{LS,} \quad (\text{B.15a})$$

with $E_L^2 = m_B^2 c^4 + p_L^2 c^2$ (B = beam), and as

$$P_a = (E_a/c, 0, 0, p), \quad P_b = (E_b/c, 0, 0, -p) \quad \text{CMS,} \quad (\text{B.15b})$$

with $E_a^2 = m_B^2 c^4 + p^2 c^2$ and $E_b^2 = m_T^2 c^4 + p^2 c^2$. The Lorentz transformations between them are

$$p = \gamma(p_L - v E_L/c^2), \quad E_a = \gamma(E_L - vp_L), \quad (\text{B.16})$$

²Although the ‘centre-of-mass’ system is the most frequently used name, some authors refer to this as the ‘centre-of-momentum’ system. Logically, a better name would be ‘zero-momentum’ frame.

where

$$v = \frac{c^2 p_L}{E_L + m_T c^2}, \quad \gamma = \frac{E_L + m_T c^2}{c^2 \sqrt{s}}, \quad v\gamma = \frac{p_L}{\sqrt{s}}, \quad (\text{B.17})$$

and s is the *invariant mass squared* of the system defined by

$$s \equiv (p_a + p_b)^2/c^2 = [(E_a + E_b)^2 - (\mathbf{p}_a c + \mathbf{p}_b c)^2]/c^4. \quad (\text{B.18})$$

In particular, in the LS,

$$s = m_T^2 + m_B^2 + 2m_T E_L/c^2, \quad (\text{B.19a})$$

whereas in the CMS,

$$s = E_{\text{CM}}^2/c^4, \quad (\text{B.19b})$$

where E_{CM} is the total CMS energy. Comparing these equations leads directly to the relation (4.5) between the CMS and LS energies, which was used in Chapter 4 when comparing the relative merits of fixed-target and colliding beam accelerators. Substituting (B.17) into (B.16) gives

$$p = \frac{p_L m_T}{\sqrt{s}}, \quad E_a = \frac{m_B^2 c^2 + m_T E_L}{\sqrt{s}} \quad (\text{B.20a})$$

and similarly for particle b :

$$p = \frac{p_L m_T}{\sqrt{s}}, \quad E_b = \frac{m_T^2 c^2 + m_T E_L}{\sqrt{s}}. \quad (\text{B.20b})$$

Finally, we state without proof the transformation of scattering angles for the specific case of laboratory and centre of mass systems. Consider the general scattering reaction

$$B(E_L, \mathbf{p}_L) + T(m_T c^2, \mathbf{0}) \rightarrow P(E, \mathbf{q}) + \dots, \quad (\text{B.21})$$

where B is a beam particle incident on a target particle T at rest in the laboratory system and P is one of a number of possible particles in the final state. If \mathbf{p}_L is taken along the z direction, then

$$\mathbf{p}_L = (0, 0, p_L) \quad \text{and} \quad \mathbf{q} = (0, q \sin \theta_L, q \cos \theta_L), \quad (\text{B.22})$$

where θ_L is the scattering angle in the laboratory system, i.e. the angle between the beam direction and \mathbf{q} . In the CMS,

$$\mathbf{p}'_B + \mathbf{p}'_T = \mathbf{0}, \quad (\text{B.23})$$

where \mathbf{p}'_B and \mathbf{p}'_T are the CMS momenta of the beam and target, respectively. The relation between the scattering angle θ_C in this system and θ_L is

$$\tan \theta_L = \frac{1}{\gamma(v)} \frac{q' \sin \theta_C}{q' \cos \theta_C + v E'/c^2}, \quad (\text{B.24})$$

where

$$E' = m_P c^2 \gamma(u) \quad \text{and} \quad q' = m_P u \gamma(u), \quad (\text{B.25})$$

and u is the magnitude of the velocity of P in the centre-of-mass frame.

It is instructive to consider the form (B.24) at high energies. From (B.17) the velocity of the transformation is

$$v = p_L c^2 (E_L + m_T c^2)^{-1}, \quad (\text{B.26})$$

so at high energies, where $E_L^2 \approx p_L c \gg m_B c^2, m_T c^2$, the velocity $v \approx c(1 - m_T c/p_L) \approx c$ and

$$\gamma(v) \approx \left(\frac{p_L}{2m_T c} \right)^{1/2}. \quad (\text{B.27})$$

Substituting (B.25), (B.26), and (B.27) into (B.24) gives

$$\tan \theta_L \approx \left(\frac{2m_T c}{p_L} \right)^{1/2} \frac{u \sin \theta_C}{u \cos \theta_C + c}. \quad (\text{B.28})$$

Thus, unless $u \approx c$ and $\cos \theta_C \approx -1$, the final-state particles will lie in a narrow cone about the beam direction in the laboratory system. Similarly, when a high-energy particle decays, its decay products will emerge predominantly at small angles to the initial beam direction.

B.3 Invariants

The transformations between the LS and CMS for energy and momentum have been derived explicitly above, but a more efficient way is to work with quantities that are invariants, i.e. have the same values in all inertial frames. We have already met one of these: s the invariant mass squared, defined in (B.18). We will find expressions for the energy and momentum in terms of invariants for both the LS and the CMS.

In the LS, from (B.13), we have

$$\mathbf{p}_B = \mathbf{0}, \quad E_B = m_B c^2, \quad (\text{B.29})$$

but from (B.19a) and (B.19b),

$$s = m_B^2 + m_T^2 + 2m_T E_L / c^2, \quad (\text{B.30})$$

that is

$$E_L = \frac{(s - m_B^2 - m_T^2) c^2}{2m_T}, \quad (\text{B.31})$$

and so

$$p_L^2 = \frac{E_L^2}{c^2} - m_B^2 c^2 = \frac{(s - m_B^2 - m_T^2)^2 c^2 - 4m_B^2 m_T^2 c^2}{4m_T^2}. \quad (\text{B.32})$$

This can be written in the useful compact form

$$p_L = \frac{c}{2m_T} \lambda^{1/2} (s, m_B^2, m_T^2), \quad (\text{B.33a})$$

where the *triangle function* λ is defined by

$$\lambda(x, y, z) \equiv (x - y - z)^2 - 4yz. \quad (\text{B.33b})$$

This function is invariant under all permutations of its arguments and in particular (B.33a) can be written in the form

$$p_L = \frac{c}{2m_T} \left\{ [s - (m_T + m_B)^2] [s - (m_T - m_B)^2] \right\}^{1/2}. \quad (\text{B.34})$$

In a similar way, it is straightforward to show that in the CMS,

$$p = \frac{c}{2\sqrt{s}} \left\{ [s - (m_T + m_B)^2] [s - (m_T - m_B)^2] \right\}^{1/2}, \quad (\text{B.35})$$

from which it follows that

$$E_a = \frac{(s + m_B^2 - m_T^2)c^2}{2\sqrt{s}}, \quad E_b = \frac{(s - m_B^2 + m_T^2)c^2}{2\sqrt{s}}. \quad (\text{B.36})$$

The above formulas have many applications. For example, if we wish to produce particles with a certain mass M , the minimum laboratory energy of the beam particles is, from (B.31),

$$E_L(\min) = \frac{M^2 c^2 - m_B^2 c^2 - m_T^2 c^2}{2m_T}. \quad (\text{B.37})$$

In the case of the decay of a particle A to a set of final-state particles $i = 1, 2, 3, \dots, N$, i.e.

$$A \rightarrow 1 + 2 + 3 + \dots + N, \quad (\text{B.38})$$

the invariant mass W of the final-state particles is given by

$$W^2 c^4 = \left(\sum_i E_i \right)^2 - \left(\sum_i \mathbf{p}_i c \right)^2 = E_A^2 - (\mathbf{p}_A c)^2 = M_A^2 c^4. \quad (\text{B.39})$$

Hence the mass of the decaying particle is equal to the invariant mass of its decay products. The latter can be measured if the lifetime of the particle is too short for its mass to be measured directly.

A related example is the identification of resonances from three-particle final states in reactions of the type $1 + 2 \rightarrow 3 + 4 + 5$, using so-called Dalitz plots.³ From measurements of the kinematic variables for particles 3, 4, and 5, we can form the squared invariant masses m_{34}^2 , m_{45}^2 , and m_{35}^2 . If we plot the first two along the x and y axes, then the third is a constant along lines at 45° because the sum of the squared invariant masses is a constant for a given initial energy. The physically allowed kinematic variables lie inside a well-defined region on the plot and in the absence of resonances it can be shown that this region is uniformly populated with

³There are several varieties of these. The original form was published by Dalitz (1953) and used to study the so-called ' $\tau - \theta$ puzzle' – see Footnote 1 of Chapter 7.

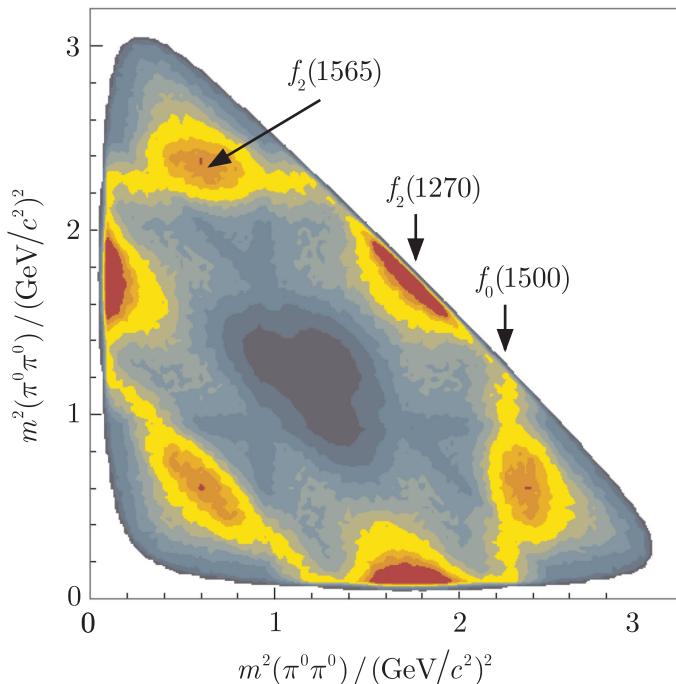


Figure B.1 Dalitz plot for the reaction $p + \bar{p} \rightarrow \pi^0 + \pi^0 + \pi^0$ using data obtained from an experiment at the LEAR accelerator at CERN. Source: Adapted from Amsler (1998). Copyright (1998) the American Physical Society, reprinted with permission.

events.⁴ Resonance behaviour in two of the final-state particles gives rise to a band of higher density parallel to one of the coordinate axes, or along a 45° line. This is illustrated in Figure B.1, which shows a high-statistics Dalitz plot for 712,000 final-state events in the reaction

$$p + \bar{p} \rightarrow \pi^0 + \pi^0 + \pi^0,$$

obtained by the Crystal Barrel collaboration using the LEAR accelerator at CERN in an experiment studying meson spectroscopy. The plot has a high degree of symmetry because the three final-state particles are identical and therefore each event is entered six times. Clear enhancements (the orange and red areas) due to the presence of meson resonances can be seen, three of which are labelled with the appropriate state.

Problems B

- B.1** The *Mandelstam variables* s , t , and u are defined for the reaction $A + B \rightarrow C + D$ by

$$s = (p_A + p_B)^2/c^2, \quad t = (p_A - p_C)^2/c^2, \quad u = (p_A - p_D)^2/c^2,$$

where p_A , etc. are the relevant energy-momentum four-vectors.

⁴For a proof, see, for example, Burcham and Jobes (1995, p. 310).

(a) Show that

$$s + t + u = \sum_{j=A,B,C,D} m_j^2.$$

(b) In the case of elastic scattering show that $t = -2p^2(1 - \cos \theta)/c^2$, where $p \equiv |\mathbf{p}|$, \mathbf{p} is the centre-of-mass momentum of particle A and θ is its scattering angle in the CMS.

- B.2** A pion travelling with speed $v \equiv |\mathbf{v}|$ in the laboratory decays via $\pi \rightarrow \mu + \nu$. If the neutrino emerges at right angles to \mathbf{v} , find an expression for the laboratory angle θ at which the muon emerges.
- B.3** A pion at rest decays via $\pi \rightarrow \mu + \nu$. Find the speed of the muon in terms of the masses involved.
- B.4** A neutral particle X^0 decays via $X^0 \rightarrow A^+ + B^-$. The momentum components of the final-state particles are measured to be (in GeV/c):

	p_x	p_y	p_z
A^+	-0.488	-0.018	2.109
B^-	-0.255	-0.050	0.486

Test the hypotheses that the decay is (a) $D^0 \rightarrow \pi^+ + K^-$ or (b) $\Lambda \rightarrow p + \pi^-$.

- B.5** Calculate the minimum laboratory energy E_{\min} of the initial proton for the production of antiprotons in a fixed-target experiment using the reaction $pp \rightarrow ppp\bar{p}$. If the protons are bound in nuclei, show that taking the internal motion of the nucleons into account leads to a smaller minimum energy given by

$$E'_{\min} \approx (1 - p/m_P c) E_{\min},$$

where p is the modulus of the average internal longitudinal momentum of a nucleon. Use a typical value of p to calculate E'_{\min} .

- B.6** A particle A decays at rest via $A \rightarrow B + C$. Find the total energy of B in terms of the three masses.
- B.7** A meson M decays via $M \rightarrow \gamma\gamma$. Find an expression for the angle in the laboratory between the momentum vectors of the two photons in terms of the photon energies and the mass of M .
- B.8** A photon is Compton scattered from a stationary electron through a scattering angle of 60° and its final energy is half its initial energy. Calculate the value of the initial energy in MeV.

C

Rutherford scattering

C.1 Classical physics

In Chapter 1 we commented on the experiments of Rutherford and his co-workers, Geiger and Marsden, that provided evidence for the existence of the nucleus. They scattered low-energy α particles from thin gold foils and observed that sometimes the projectiles were scattered through large angles, in extreme cases close to 180° . If we start by ignoring the fact that there is a Coulomb interaction present, then it is easy to show that this behaviour is incompatible with scattering from light particles such as electrons.

Consider the nonrelativistic elastic scattering of an α particle of mass m_α and initial velocity \mathbf{v}_i from a target of mass m_t stationary in the laboratory. If the final velocities of the projectile and target are \mathbf{v}_f and \mathbf{v}_t , respectively, then we have the situation as shown in Figure C.1.

Conservation of linear momentum and kinetic energy are

$$m_\alpha \mathbf{v}_i = m_\alpha \mathbf{v}_f + m_t \mathbf{v}_t \quad (\text{C.1})$$

and

$$m_\alpha v_i^2 = m_\alpha v_f^2 + m_t v_t^2, \quad (\text{C.2})$$

where $v_i = |\mathbf{v}_i|$, etc. Squaring (C.1) we get

$$m_\alpha v_i^2 = m_\alpha v_f^2 + \frac{m_t^2}{m_\alpha} v_t^2 + 2m_t (\mathbf{v}_f \cdot \mathbf{v}_t) \quad (\text{C.3})$$

and hence, from (C.2),

$$v_t^2 \left(1 - \frac{m_t}{m_\alpha} \right) = 2\mathbf{v}_f \cdot \mathbf{v}_t. \quad (\text{C.4})$$

Thus, if the target is an electron, with $m_t = m_e \ll m_\alpha$, the directions of motion of the outgoing α particle and the recoiling target are essentially

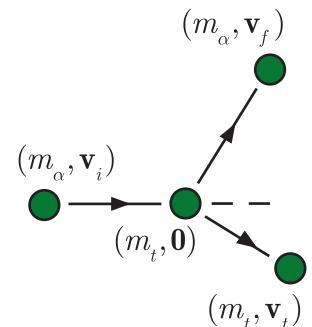


Figure C.1 Kinematics of the Geiger and Marsden experiment.

along the direction of the initial α particle and no large-angle scatterings are possible. Such events could, in principle, be due to multiple small-angle scattering, but the thinness of the gold foil target rules this out.¹ If, however, $m_t = m_{Au} \gg m_\alpha$, then the left-hand side of (C.4) will be negative and large scattering angles are possible.

The above only makes plausible the existence of a heavy nucleus, because it has ignored the existence of the Coulomb force, so we now have to take this into account. We will do this firstly using nonrelativistic classical mechanics.

Consider the nonrelativistic Coulomb scattering of a projectile of mass m and electric charge ze from a target particle of mass M and electric charge Ze . The kinematics of this are shown in Figure C.2. The target mass is assumed to be sufficiently large that its recoil may be neglected. The initial velocity of the projectile is \mathbf{v} and it is assumed that in the absence of any interaction it would travel in a straight line and pass the target at a distance b , called the *impact parameter*. The derivation follows from considering the implications of linear and angular momentum conservation.

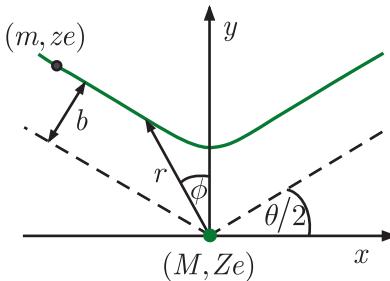


Figure C.2 Kinematics of Rutherford scattering.

Angular momentum conservation implies that

$$mvb = mr^2 \frac{d\phi}{dt}, \quad (\text{C.5})$$

where as usual $v = |\mathbf{v}|$. Since the scattering is symmetric about the y axis, the component of linear momentum in the y direction is initially $p = -mv \sin(\theta/2)$ and changes to $+mv \sin(\theta/2)$ after the interaction, i.e. the total change in momentum in the y direction is

$$\Delta p = 2mv \sin(\theta/2). \quad (\text{C.6})$$

The change in momentum may also be calculated by integrating the impulse in the y direction due to the Coulomb force on the projectile. This gives

$$\Delta p = \int_{-\infty}^{+\infty} \frac{zZe^2}{4\pi \epsilon_0 r^2} \cos \phi dt, \quad (\text{C.7})$$

¹For completeness one should also show that the observations cannot be due to scattering from the diffuse positive charge present. This was done by the authors of the original experiment.

where we have taken $t = 0$ to coincide with the origin of the x axis. Using (C.5) to change variables, (C.7) may be written as

$$2mv \sin(\theta/2) = \frac{zZe^2}{4\pi\varepsilon_0} \left(\frac{1}{bv} \right) \int_{-(\pi-\theta)/2}^{(\pi-\theta)/2} \cos \phi \, d\phi. \quad (\text{C.8})$$

Thus,

$$b = \frac{zZe^2}{8\pi\varepsilon_0 E_{\text{kin}}} \cdot \frac{1}{E_{\text{kin}}} \cot(\theta/2), \quad (\text{C.9})$$

where $E_{\text{kin}} = \frac{1}{2}mv^2$ is the kinetic energy of the projectile.

Finally, we need to calculate the differential cross-section. If the initial flux of projectile particles crossing a plane perpendicular to the beam direction is J , then the intensity of particles having impact parameters between b and $b+db$ is $2\pi b J db$, which is equal to the rate dW at which particles are scattered into a solid angle $d\Omega = 2\pi \sin \theta d\theta$ between θ and $\theta + d\theta$. Thus

$$dW = 2\pi b J db. \quad (\text{C.10})$$

However, from (1.60) and considering a single target particle,

$$dW = J \frac{d\sigma}{d\Omega} d\Omega = 2\pi J \sin \theta d\theta \frac{d\sigma}{d\Omega}, \quad (\text{C.11})$$

i.e.

$$\frac{d\sigma}{d\Omega} = \frac{b}{\sin \theta} \cdot \frac{db}{d\theta}. \quad (\text{C.12})$$

The right-hand side of (C.12) may be evaluated from (C.9) and gives

$$\frac{d\sigma}{d\Omega} = \left(\frac{zZe^2}{16\pi\varepsilon_0 E_{\text{kin}}} \right)^2 \operatorname{cosec}^4(\theta/2) = \frac{(\hbar c)^2 (\alpha z Z)^2}{4m^2 v^4 \sin^4(\theta/2)}. \quad (\text{C.13})$$

where α is the fine structure constant. This is the final form of the Rutherford differential cross-section for nonrelativistic scattering.

C.2 Quantum mechanics

While (C.13) is adequate to describe the α -particle scattering experiments, in the case of electron scattering we need to take account of both relativity and quantum mechanics. This may be done using the general formalism for the differential cross-section in terms of the scattering potential that was derived in Chapter 1. We will neglect spin factors.

The starting equation is (1.69), which in the present notation is

$$\frac{d\sigma}{d\Omega} = \frac{1}{4\pi^2 \hbar^4} \frac{p'^2}{uv'} |\mathcal{M}(\mathbf{q}^2)|^2, \quad (\text{C.14})$$

where \mathbf{v} and \mathbf{p} are the velocity and momentum, respectively, of the projectile (which for convenience we take to have a unit negative charge),

because the target is assumed to be heavy, with $v = |\mathbf{v}|$, $p = |\mathbf{p}|$ and the primes refer to the final-state values. The matrix element is given by

$$\mathcal{M}(\mathbf{q}) = \int V(\mathbf{r}) e^{i\mathbf{q}\cdot\mathbf{r}/\hbar} d^3\mathbf{r}, \quad (\text{C.15})$$

where $\mathbf{q} = \mathbf{p} - \mathbf{p}'$ is the momentum transfer. $V(\mathbf{r})$ is the Coulomb potential

$$V(\mathbf{r}) = V_C(\mathbf{r}) = -\frac{\alpha Z(\hbar c)}{r}, \quad (\text{C.16})$$

where $r = |\mathbf{r}|$, and Ze is the charge of the target nucleus. Inspection of the integral in (C.15) shows that it diverges at large r . However, in practice, charges are always screened at large distances by intervening matter and so we will interpret the integral as

$$\mathcal{M}_C(q) = \lim_{\lambda \rightarrow 0} \int \left(-\frac{Z\alpha(\hbar c) e^{-\lambda r}}{r} \right) e^{i\mathbf{q}\cdot\mathbf{r}/\hbar} d^3\mathbf{r}. \quad (\text{C.17})$$

To evaluate this, take \mathbf{q} along the z axis, so that in spherical polar coordinates $\mathbf{q} \cdot \mathbf{r} = qr \cos \theta$. The angular integration may then be done and yields

$$\mathcal{M}_C(q) = -\frac{4\pi(\hbar c)Z\alpha\hbar}{q} \lim_{\lambda \rightarrow 0} \int_0^\infty e^{-\lambda r} \sin(qr/\hbar) dr. \quad (\text{C.18})$$

The remaining integral may be done by parts (twice) and taking the limit $\lambda \rightarrow 0$ gives

$$\mathcal{M}_C(\mathbf{q}) = -\frac{4\pi(\hbar c)Z\alpha\hbar^2}{q^2}. \quad (\text{C.19})$$

Finally, substituting (C.19) into (C.14) gives

$$\frac{d\sigma}{d\Omega} = 4Z^2\alpha^2(\hbar c)^2 \frac{p'^2}{vv'q^4}, \quad (\text{C.20})$$

which is the general form of the differential cross-section. If we neglect nuclear recoil, we may substitute the approximations

$$v = v', \quad p = p' = mv\gamma, \quad (\text{C.21})$$

together with the kinematic relation for the scattering angle

$$q = 2p \sin(\theta/2), \quad (\text{C.22})$$

into (C.20) to obtain

$$\frac{d\sigma}{d\Omega} = \frac{z^2 Z^2 \alpha^2 (\hbar c)^2}{4\beta^4 E^2 \sin^4(\theta/2)}. \quad (\text{C.23})$$

For $z = 1$ this reduces to the differential cross-section (2.24) used in Chapter 2, while in the limit of nonrelativistic electrons, $E = mc^2$, it reduces to the Rutherford cross-section (C.13).

Problems C

- C.1** Calculate the differential cross-section in mb/sr for the scattering of a 20 MeV alpha particle through an angle of 20^0 by a nucleus $^{209}_{83}\text{Bi}$, stating any assumptions made. Ignore spin and form factor effects.
- C.2** Show that in Rutherford scattering at a fixed impact parameter b , the distance of closest approach d to the nucleus, is given by $d = b[1 + \text{cosec}(\theta/2)]/\cot(\theta/2)$, where θ is the scattering angle.
- C.3** Find an expression for the impact parameter b in the case of small-angle Rutherford scattering. A beam of protons with speed $v = 4 \times 10^7 \text{ m s}^{-1}$ is incident normally on a thin foil of $^{194}_{78}\text{Pt}$ of thickness 10^{-5}m (density = $2.145 \times 10^4 \text{ kg m}^{-3}$). Estimate the proportion of protons that experience double scattering, where each scattering angle is at least 5^0 .

D

Gauge theories

The term ‘gauge theory’ has been used in many places in this book to describe the theories of the standard model. These theories have a particular type of symmetry, called gauge invariance. In this section we will give a short introduction to gauge invariance, then briefly describe the modifications necessary to bring the predictions of the gauge theory of electroweak interactions into agreement with experiment.¹ *To keep the equations simple looking, and because the discussion will be only qualitative, uniquely in this appendix we will use natural units, $\hbar = c = 1$, and also set $\varepsilon_0 = 1$.*

D.1 Gauge invariance and the standard model

The concept of gauge invariance has its origins in electromagnetism, so we start there. Then we introduce the gauge principle and discuss qualitatively its application to the theories of the standard model – QCD and the electroweak interaction.

D.1.1 Electromagnetism and the gauge principle

Consider the electromagnetic interaction of a particle of mass m and electric charge q moving nonrelativistically in an electromagnetic field

¹These topics are usually discussed in more advanced treatments, where advantage can be taken of students’ knowledge of relativistic quantum field theory to present the formalism in an elegant way using explicitly covariant Lagrangians. This approach may be found in many more advanced books, for example Chapters 17 and 18 of Mandl and Shaw (2010) or Chapter 14 of Halzen and Martin (1984). A discussion at the level of the present book that does not use covariant notation, but is based on the relevant equations of motion, is given in Appendix D of Martin and Shaw (2017). Our discussion is based on that appendix, but is more qualitative.

described by the electric and magnetic field vectors $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{B}(\mathbf{r}, t)$, respectively, or equivalently by the scalar and vector potentials $\phi(\mathbf{r}, t)$ and $\mathbf{A}(\mathbf{r}, t)$, defined by

$$\mathbf{E} \equiv -\nabla\phi - \frac{\partial\mathbf{A}}{\partial t} \quad \text{and} \quad \mathbf{B} \equiv \nabla \times \mathbf{A}. \quad (\text{D.1})$$

These definitions do not uniquely determine the potentials ϕ and \mathbf{A} , because if $f(\mathbf{r}, t)$ is an *arbitrary* scalar function, the transformations

$$\phi \rightarrow \phi' = \phi + \frac{\partial f}{\partial t}, \quad \mathbf{A} \rightarrow \mathbf{A}' = \mathbf{A} - \nabla f \quad (\text{D.2})$$

leave the fields \mathbf{E} and \mathbf{B} , and hence the underlying physics, unchanged, as can easily be demonstrated by direct substitution into (D.1). The relations (D.2) are examples of a *gauge transformation* and a theory whose physical predictions remain unchanged by such a transformation is said to be *gauge invariant*.

We also need to consider the equation of motion for the wavefunction ψ of the particle. Nonrelativistically, this is

$$i \left(\frac{\partial}{\partial t} + iq\phi \right) \psi = -\frac{1}{2m} (\nabla - iq\mathbf{A})^2 \psi, \quad (\text{D.3})$$

which is obtained by using the usual quantum mechanical substitutions in the classical Hamiltonian²

$$H = \frac{1}{2m} (\mathbf{p} - q\mathbf{A})^2 + q\phi. \quad (\text{D.4})$$

Applying the transformations (D.2) and (D.3), and leaving ψ unchanged, leads to an equation for the primed quantities that is not in the same form as (D.3); thus (D.3) is not gauge invariant. This may be overcome by extending the definition of a gauge transformation by assuming that (D.2) implies that the wavefunction simultaneously undergoes a transformation of the form

$$\psi(\mathbf{r}, t) \rightarrow \psi'(\mathbf{r}, t) = \exp[-iqf(\mathbf{r}, t)]\psi(\mathbf{r}, t). \quad (\text{D.5})$$

Then, using (D.2) and (D.5), it is straightforward to show that ψ' obeys (D.3), but where the potentials are primed; the equation is now gauge invariant.

The equations of motion for the potentials themselves follow from Maxwell's equations for \mathbf{E} and \mathbf{B} in free space and are

$$\left(\frac{\partial^2}{\partial t^2} - \nabla^2 \right) \phi - \frac{\partial}{\partial t} \left(\frac{\partial \phi}{\partial t} + \nabla \cdot \mathbf{A} \right) = 0 \quad (\text{D.6a})$$

²This Hamiltonian leads to the classical equation of motion for a charged particle moving with velocity \mathbf{v} in electric and magnetic fields given by (D.1) (cf. Problem D.1).

and

$$\left(\frac{\partial^2}{\partial t^2} - \nabla^2 \right) \mathbf{A} + \nabla \left(\frac{\partial \phi}{\partial t} + \nabla \cdot \mathbf{A} \right) = \mathbf{0}. \quad (\text{D.6b})$$

To interpret these equations, we impose a convenient subsidiary condition, without changing any of the physical predictions of the theory. In practice, for any set of potentials (ϕ, \mathbf{A}) satisfying Maxwell's equations we can always find a transformed set that satisfies the so-called *Lorentz condition*

$$\frac{\partial \phi}{\partial t} + \nabla \cdot \mathbf{A} = 0, \quad (\text{D.7})$$

and then (D.6a) and (D.6b) become

$$\left(\frac{\partial^2}{\partial t^2} - \nabla^2 \right) \phi = 0 \quad \text{and} \quad \left(\frac{\partial^2}{\partial t^2} - \nabla^2 \right) \mathbf{A} = \mathbf{0}. \quad (\text{D.8})$$

Comparing these with the Klein–Gordon equation (1.3) for an exchanged particle of mass m , it can be seen that they are of exactly the same form, and imply that the associated particles, photons, have zero mass.

Photons with nonzero masses m_γ (or any other massive spin-1 particles) would be described by the *Proca equations*

$$\left(\frac{\partial^2}{\partial t^2} - \nabla^2 \right) \phi - \frac{\partial}{\partial t} \left(\frac{\partial \phi}{\partial t} + \nabla \cdot \mathbf{A} \right) + m_\gamma^2 \phi = 0 \quad (\text{D.9a})$$

and

$$\left(\frac{\partial^2}{\partial t^2} - \nabla^2 \right) \mathbf{A} + \nabla \left(\frac{\partial \phi}{\partial t} + \nabla \cdot \mathbf{A} \right) + m_\gamma^2 \mathbf{A} = \mathbf{0}. \quad (\text{D.9b})$$

These equations differ from the gauge-invariant Maxwell equations (D.6a) and (D.6b) by the addition of ‘mass terms’ $m_\gamma^2 \phi$ and $m_\gamma^2 \mathbf{A}$, which are obviously not invariant under (D.5). Hence the Proca equations are only gauge invariant in the case $m_\gamma = 0$, when they reduce to Maxwell’s equations. In other words, gauge invariance requires the photon to be massless.

We can conclude two things from the above discussion: (a) if the theory is to be gauge invariant, the gauge transformation must include the transformation of the wavefunction and (b) the equations are only gauge invariant if the associated spin-1 mesons (photons in the case of electromagnetism) have zero mass.

In discussing the theories of the standard model – QCD and the electroweak interaction – we reverse the sequence of arguments above to examine the consequences of *assuming* that the gauge invariance of the wavefunction is fundamental. This is called the *principle of minimal gauge invariance*, or simply the *gauge principle*. To illustrate this we will continue to consider the electromagnetic interaction, but now for a relativistic spin-1/2 particle.

In the absence of interactions, such a particle obeys the Dirac equation (1.1), where the wavefunction of (D.3) is replaced by the spinor ψ . Using the Hamiltonian (1.2),³ we have

$$i \frac{\partial \psi}{\partial t} = -i \boldsymbol{\alpha} \cdot \boldsymbol{\nabla} \psi + \beta m \psi. \quad (\text{D.10})$$

However, this equation is not invariant under the gauge transformation (D.5), now taken as fundamental, as is easily verified by direct substitution. This can be remedied by adding to (D.10) terms involving the electromagnetic potentials. Specifically, if we make the so-called *minimal substitutions*

$$\frac{\partial}{\partial t} \rightarrow \frac{\partial}{\partial t} + iq\phi \quad \text{and} \quad \boldsymbol{\nabla} \rightarrow \boldsymbol{\nabla} - iq\mathbf{A} \quad (\text{D.11})$$

in (D.10), we arrive at the equation

$$i \left(\frac{\partial}{\partial t} + iq\phi \right) \psi = -i \boldsymbol{\alpha} \cdot (\boldsymbol{\nabla} - iq\mathbf{A}) \psi + \beta m \psi, \quad (\text{D.12})$$

which for the case of an electron, where $q = -e$, is the equation used in quantum electrodynamics (QED). It is straightforward to show by direct calculation, using (D.3) and (D.5), that (D.12) is gauge invariant and thus we conclude that imposing gauge invariance leads to the form of the interaction, which can then be compared with experiment, in the case of QED with spectacular success.

D.1.2 The standard model

An analogous procedure to that above can be applied to the strong interactions. Here the overall state of any relativistic quark is the product of a Dirac spinor $\psi(\mathbf{r}, t)$ and a colour wavefunction χ^c , i.e.

$$\Psi \equiv \psi(\mathbf{r}, t)\chi^c. \quad (\text{D.13})$$

Associated with the colour wavefunctions are eight colour charges $\hat{F}_i (i = 1, 8)$. These act upon χ^c and serve as the source of the strong interaction, in the same way that electric charge is the source of the electromagnetic interaction. Apart from trivial normalisation constants, the eigenvalues of two of these operators \hat{F}_3 and \hat{F}_8 are just the *colour isospin* and *colour hypercharge* I_3^C and Y^C introduced in Section 5.1, while the other operators change coloured quarks into quarks of another colour (cf. Problem 5.2). The gauge transformation in this case involves not only

³Recall that we are using natural units.

the phase of the wavefunction but also the colour state and by analogy with (D.5) is written

$$\Psi(\mathbf{r}, t) \rightarrow \Psi'(\mathbf{r}, t) = \exp \left[-ig_s \sum_{i=1}^8 \hat{F}_i \omega_i(\mathbf{r}, t) \right] \Psi(\mathbf{r}, t), \quad (\text{D.14})$$

where g_s is the strong coupling, related to α_s by $\alpha_s = g_s^2/4\pi$, and ω_i are a set of eight arbitrary gauge functions⁴. Proceeding as in the electromagnetic case and demanding invariance under this gauge transformation, leads to an equation of motion for Ψ with interaction terms that correspond to the observed interactions between quarks and the spin-1 bosons of the theory, the gluons, and in addition implies, by colour conservation, the existence of gluon–gluon interactions.⁵

The application of these ideas to the weak interaction is complicated by the fact that there exist both charged and neutral currents. Consider, for simplicity, the case where we have just electrons and electron neutrinos. Proceeding by analogy with the discussion of QCD above, we replace the colour wavefunction by a flavour wavefunction χ^f and the colour charge operators by three *weak isospin* operators \hat{I}_i^W ($i = 1, 2, 3$). The neutrino and electron are eigenstates of \hat{I}_3^W with eigenvalues $1/2$ and $-1/2$ respectively, and are said to form a *weak isodoublet* (ν_e, e), with eigenvalues, respectively, in analogy with strong isodoublets, like (u, d), etc., while the other weak isospin operators transform electrons and neutrinos into each other.⁶ Since there are three weak isospin operators, the gauge transformation analogous to (D.14) is

$$\Psi(\mathbf{r}, t) \rightarrow \Psi'(\mathbf{r}, t) = \exp \left[-ig \sum_{i=1}^3 \hat{I}_i^W f_i(\mathbf{r}, t) \right] \Psi(\mathbf{r}, t), \quad (\text{D.15})$$

where g is an arbitrary constant and f_i is a set of arbitrary gauge functions. The application of the gauge principle then leads to interactions of the form

$$e^- \rightarrow \nu_e W^-, \quad \nu_e \rightarrow e^- W^+, \quad e^- \rightarrow e^- W^0, \quad \nu_e \rightarrow \nu_e W^0,$$

in which gauge bosons $W^{+,-,0}$ of the appropriate charge are emitted or absorbed. The charged bosons can be identified with the observed W^\pm particles, leading to charged current weak interactions of the type observed. However, if we identify the neutral boson with the observed Z^0 , we are

⁴The parallel between (D.5) and (D.14) would be even more transparent if we wrote $q = eQ$ in (D.5), where e characterises the strength of the electromagnetic interaction and $Q = 0, -1/3, 2/3$ or -1 , the charge of the quark or lepton concerned.

⁵The proof of these statements, and analogous statements to follow about the application of the gauge principle to the weak interaction, may be found in Appendix D, Sections D.6 and D.7, respectively, of Martin and Shaw (2017).

⁶The weak isospin operators are so-called because they have exactly the same mathematical structure as the ordinary isospin operators discussed in Appendix A.4.1. There is, however, no physical connection between the two.

forced to conclude that the theory predicts weak neutral currents with essentially the same strength as charged currents, which we have seen in Chapter 6 is not what is observed experimentally.

The solution to this problem lies in the unification of the weak interactions with the electromagnetic interaction. This is done firstly by incorporating electric charge into the theory by introducing the *weak hypercharge* Y^W , defined by

$$Q \equiv I_3^W + Y^W, \quad (\text{D.16})$$

where Q is the electric charge in units of e , and then requiring gauge invariance under both (D.15) and the *additional* transformation

$$\Psi(\mathbf{r}, t) \rightarrow \Psi'(\mathbf{r}, t) = \exp[-ig'Y^W\omega(\mathbf{r}, t)]\Psi(\mathbf{r}, t), \quad (\text{D.17})$$

where g' is a constant to be determined and $\omega(\mathbf{r}, t)$ is an arbitrary function. Then, just as for electromagnetism, where the gauge principle led to interactions of the form $e^- \rightarrow e^-\gamma$, in this case it leads to interactions of the form

$$e^- \rightarrow e^-B^0, \quad \nu_e \rightarrow \nu_eB^0,$$

where B^0 is a new gauge boson. Next we write the photon γ and the Z^0 boson as mixtures of the W^0 and B^0 bosons:

$$\gamma = B^0 \cos \theta_W + W^0 \sin \theta_W \quad (\text{D.18a})$$

and

$$Z^0 = -B^0 \sin \theta_W + W^0 \cos \theta_W, \quad (\text{D.18b})$$

where θ_W is a weak mixing angle, and choose the couplings g and g' so that the combination (D.18a) has all the properties of the observed photon. It turns out that this can be achieved if

$$g \sin \theta_W = g' \cos \theta_W,$$

which guarantees that there is no direct coupling to neutrinos, and

$$\frac{1}{2}(g \sin \theta_W + g' \cos \theta_W) = e,$$

which ensures that the coupling of the electromagnetic field has the strength required in QED. These two conditions are equivalent to the unification condition⁷

$$e = g \sin \theta_W = g' \cos \theta_W \quad (\text{D.19a})$$

or alternatively (cf. (6.52), remembering that we are setting $\varepsilon_0 = 1$)

$$\frac{e}{2\sqrt{2}} = g_W \sin \theta_W = g_Z \cos \theta_W, \quad (\text{D.19b})$$

⁷This calculation is done explicitly in Martin and Shaw (2017), Appendix D, Section D.7.3.

where

$$g_W \equiv \frac{g}{2\sqrt{2}} \quad \text{and} \quad g_Z \equiv \frac{g'}{2\sqrt{2}}. \quad (\text{D.19c})$$

Thus the neutral current interactions are specified in terms of the electric charge and the weak mixing angle θ_W , and the successes of QED are preserved. However, there is one other complicating factor that we still need to consider.

In the above discussion, we have ignored parity violation in weak interactions. As we saw in Chapter 7, when fermion masses are neglected, the charged weak current interactions involve left-handed spin states only, in violation of parity conservation. For zero-mass fermions, this can be easily taken into account because the Dirac equation (D.10) can be shown to separate into two independent equations⁸

$$i \frac{\partial \psi^L}{\partial t} = -i \boldsymbol{\alpha} \cdot \nabla \psi^L, \quad i \frac{\partial \psi^R}{\partial t} = -i \boldsymbol{\alpha} \cdot \nabla \psi^R, \quad (\text{D.20})$$

where $\psi^{L,R}$ are Dirac spinors corresponding to the left- and right-handed spin states of Figure 7.3, respectively. Hence, in discussing the spin structure of weak interactions, we can treat the left- and right-handed spin states completely separately. We therefore assume that the preceding results, arising from the gauge transformations (D.15) and (D.17) with $I_3^W = 1/2, -1/2$ for neutrinos and electrons, respectively, and $Y^W = -1/2$ in both cases by (D.16), apply to left-handed spin states only.

It now remains to incorporate the right-handed spin states, which are instead assigned to weak isosinglets with $I_i^W = 0$ ($i = 1, 2, 3$). In this case, both electrons and neutrinos are described by separate Dirac spinors $\psi_e^R(\mathbf{r}, t), \psi_\nu^R(\mathbf{r}, t)$ and the gauge transformations (D.15) reduce to

$$\psi_e^R \rightarrow \psi'_e^R = \psi_e^R, \quad \psi_\nu^R \rightarrow \psi'_\nu^R = \psi_\nu^R.$$

In other words, the wavefunctions remain unchanged and no interactions with W -bosons are generated by the gauge principle for right-handed spin states. However, by (D.16), $Y^W = (0, -1)$ for the neutrino and the electron, respectively, so that the gauge transformations (D.17) become

$$\psi_\nu^R \rightarrow \psi'_\nu^R = \psi_\nu^R, \quad \psi_e^R \rightarrow \psi'_e^R = e^{ig'\omega} \psi_e^R.$$

From this we see that the right-handed neutrino wavefunction is invariant under the gauge transformation (D.17), so that again no interaction is generated by the gauge principle. On the other hand, the right-handed electron wavefunction is not invariant, and an interaction with the B^0 field is required to restore gauge invariance. Using the mixing equations (D.18a) and (D.18b) and the unification conditions (D.19a) to (D.19c), this can be shown to correspond to the usual electromagnetic interaction for photons, together with a weak neutral current interaction.

⁸See, for example, Section 5.6 of Kane (1987).

In other words, for electrons (but not neutrinos) both right-handed and left-handed spin states are predicted to have neutral current interactions.

In this way, one arrives at the final predictions for the electroweak interactions of electrons and neutrinos, and the results may be extended to all three lepton generations and also to quarks by the use of lepton–quark symmetry and quark mixing. Essentially, the gauge transformations are chosen throughout to reproduce the known properties of the electromagnetic and weak charged current interactions. The form of the neutral current interactions, including their spin structure, is predicted uniquely in terms of the mixing angle θ_w , and the agreement between these predictions and the measured properties of neutral current interactions is one of the great triumphs of the unified theory.

This analysis is based on the fact that the Dirac equation can be separated into two equations in (D.20) in the limit of the zero fermion mass. This enabled us to ascribe different gauge transformation properties to the left- and right-handed states, which are treated completely separately. If finite masses are assumed, then the left- and right-handed states do not decouple and the analysis breaks down, so that not only gauge boson masses but also lepton and quark masses are required to be zero if parity violation is to be incorporated into the theory. These results can, however, be avoided if the fermion masses, together with the W^\pm and Z^0 masses, arise from interactions with the Higgs field.

D.2 Particle masses and the Higgs field

One remaining question that has to be addressed is the prediction that gauge invariance is only exact if the spin-1 bosons in the theory are massless, which we demonstrated in Section D.1.1 for the case of the photon. While this is not a problem for QED or QCD, it is a serious problem for weak interactions where the relevant bosons – the W^\pm and Z^0 – are very massive. This problem is overcome by assuming that the various particles interact with a new type of scalar field, called the *Higgs field*. The interactions of the Higgs field with the gauge bosons are gauge invariant, but the Higgs field differs from other fields in its behaviour in the so-called *vacuum state*, which contains no particles of any kind. Other fields, such as the electromagnetic field, are assumed to be zero in the vacuum state, as one would naïvely expect. However, the Higgs field has a non-zero value in the vacuum and this value is not invariant under a gauge transformation. Because of this, the theory is no longer gauge invariant, and the gauge bosons are no longer required to have zero mass. This form of symmetry breaking, in which the gauge invariance of the interaction (as opposed to the gauge invariance of the vacuum) remains exact, is called *spontaneous symmetry breaking*.

Spontaneous symmetry breaking occurs whenever the vacuum state, defined as the state of lowest energy, is not unique and is observed in

other branches of physics, one of the commonest being a simple bar magnet. When heated above its Curie temperature, its elementary magnetic domains are randomly oriented, leading to a zero net magnetic moment. On cooling, the domains will set in a particular direction, that of the resultant moment \mathbf{M} , and the rotational symmetry is spontaneously broken. However, the rotational invariance of the interaction still manifests itself in that all the other properties of the magnet are independent of the direction of \mathbf{M} .

To see how such effects can occur in particle physics, we consider the simple case of a complex scalar field $\eta(\mathbf{r}, t)$, which we write in terms of two real fields $\eta_1(\mathbf{r}, t)$ and $\eta_2(\mathbf{r}, t)$, i.e.

$$\eta(\mathbf{r}, t) = \eta_1(\mathbf{r}, t) + i\eta_2(\mathbf{r}, t). \quad (\text{D.21})$$

Let us assume that the potential energy density of $\eta(\mathbf{r}, t)$ is given by

$$V(\eta) = \mu^2|\eta(\mathbf{r}, t)|^2 + \lambda|\eta(\mathbf{r}, t)|^4. \quad (\text{D.22})$$

Here λ and μ^2 are real parameters and the interactions are invariant under the transformation

$$\eta(\mathbf{r}, t) \rightarrow \eta(\mathbf{r}, t)e^{i\beta}, \quad (\text{D.23})$$

where β is an arbitrary phase parameter. The vacuum state can be identified by minimising the potential energy density, since the kinetic energy density is either positive or zero. There are two cases to consider, and in both we require $\lambda > 0$ so that the potential energy density is bounded from below. If $\mu^2 > 0$, both terms in $V(\eta)$ are positive definite. The corresponding potential energy density curve $V(\eta)$ is sketched as a function of $\eta(\mathbf{r}, t)$ in Figure D.1a; $V(\eta)$ has a unique minimum at $\eta(\mathbf{r}, t) = 0$. If $\mu^2 < 0$, the potential energy curve is shown in Figure D.1b and there is a circle of absolute minima at

$$\eta(\mathbf{r}, t) = \eta_0 = \left(\frac{-\mu^2}{2\lambda} \right)^{1/2} e^{i\theta}, \quad 0 \leq \theta \leq 2\pi, \quad (\text{D.24})$$

where the phase angle θ specifies a direction in the complex η -plane. This arbitrariness in the direction of θ is analogous to that of the direction of the magnetisation \mathbf{M} of a ferromagnet. By analogy with the latter case, spontaneous symmetry breaking corresponds to taking a particular value of θ to represent the ground state. The actual value chosen is not significant, because of the invariance of the interactions under a phase transformation (D.23), and it is conventional to choose $\theta = 0$, so that the vacuum state corresponds to

$$\eta_0 = \frac{v}{\sqrt{2}} \equiv \left(-\frac{\mu^2}{2\lambda} \right)^{1/2} \geq 0, \quad (\text{D.25})$$

where v is a constant and the symmetry is spontaneously broken.

To see how spontaneous symmetry breaking affects gauge theories, consider the introduction of electromagnetic interactions into the above

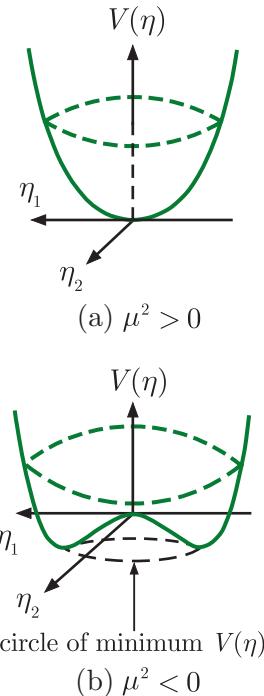


Figure D.1 The potential energy density $V(\eta)$, as given by (D.22), for $\lambda > 0$. (a) $\mu^2 > 0$; (b) $\mu^2 < 0$.

model in such a way that the equations of motion remain invariant under a gauge transformation of the standard form (D.5), i.e.

$$\eta(\mathbf{r}, t) \rightarrow \eta'(\mathbf{r}, t) = \exp[-iqf(\mathbf{r}, t)]\eta(\mathbf{r}, t). \quad (\text{D.26})$$

Since the electromagnetic field (ϕ, \mathbf{A}) vanishes in the vacuum state, this has no effect on the above discussion of the scalar field value in the vacuum and we can rewrite the theory in terms of a new field $\sigma(\mathbf{r}, t)$, defined by

$$\eta(\mathbf{r}, t) \equiv \frac{1}{\sqrt{2}}[v + \sigma(\mathbf{r}, t)], \quad (\text{D.27})$$

so that unlike $\eta(\mathbf{r}, t)$, $\sigma(\mathbf{r}, t)$ does vanish in the vacuum state. For any given set of fields (ϕ, \mathbf{A}, η) it is always possible to find a gauge function $f(\mathbf{r}, t)$ such that the transformed scalar field, and hence the new field $\sigma(\mathbf{r}, t)$, is real. In this case Maxwell's equations in free space (D.6a) and (D.6b) are replaced by⁹

$$\left(\frac{\partial^2}{\partial t^2} - \nabla^2 \right) \phi - \frac{\partial}{\partial t} \left(\frac{\partial \phi}{\partial t} + \nabla \cdot \mathbf{A} \right) = -2q^2 \eta^2 \phi \quad (\text{D.28a})$$

and

$$\left(\frac{\partial^2}{\partial t^2} - \nabla^2 \right) \mathbf{A} + \nabla \left(\frac{\partial \phi}{\partial t} + \nabla \cdot \mathbf{A} \right) = -2q^2 \eta^2 \mathbf{A}. \quad (\text{D.28b})$$

Substituting (D.27) into these equations gives

$$\left(\frac{\partial^2}{\partial t^2} - \nabla^2 \right) \phi - \frac{\partial}{\partial t} \left(\frac{\partial \phi}{\partial t} + \nabla \cdot \mathbf{A} \right) + q^2 v^2 \phi = -2q^2 v \sigma \phi - q^2 \sigma^2 \phi \quad (\text{D.29a})$$

and

$$\left(\frac{\partial^2}{\partial t^2} - \nabla^2 \right) \mathbf{A} + \nabla \left(\frac{\partial \phi}{\partial t} + \nabla \cdot \mathbf{A} \right) + q^2 v^2 \mathbf{A} = -2q^2 v \sigma \mathbf{A} - q^2 \sigma^2 \mathbf{A} \quad (\text{D.29b})$$

as the new equations of motion for the electromagnetic field. On comparing with (D.9a) and (D.9b), we see that (D.29a) and (D.29b) are identical with the Proca equations for ‘photons’ with nonzero mass

$$m_\gamma = qv, \quad (\text{D.30})$$

whose interactions with the scalar field $\sigma(\mathbf{r}, t)$ are specified by the terms on the right-hand side of (D.29a) and (D.29b). In addition, because $\sigma(\mathbf{r}, t)$ is real, like the electromagnetic field, the corresponding quanta – the Higgs bosons – are electrically neutral. This mechanism, whereby a gauge boson acquires mass as a result of its interaction with a nonvanishing vacuum field, is called the *Higgs mechanism*.

⁹These equations are derived in Section D.5.2 of Martin and Shaw (2017).

We will only outline briefly the application of the Higgs mechanism to the standard electroweak theory.¹⁰ It is much more complicated than in the simple model above because the forms of gauge invariance used are much more complicated, and it is necessary to introduce a two-component Higgs field in general. However, gauge invariance can be exploited to set one component to zero, while the other is real. The latter can again be expanded in the form (D.27), and since it is real, its quanta, the Higgs bosons, are again electrically neutral. Furthermore, as in the simple model above, the non-zero vacuum expectation value of the Higgs field breaks the electroweak gauge symmetry and can be shown to give rise to nonzero W^\pm and Z^0 masses

$$M_W = g v / 2 \text{ and } M_W / M_Z = \cos \theta_W. \quad (\text{D.31})$$

On the other hand, provided the unification conditions (D.19a) to (D.19c) is imposed, the gauge invariance is not completely broken and the usual gauge invariant electromagnetic interaction is retained with zero mass photons. Finally, it is possible to introduce a gauge invariant interaction of the Higgs field with each lepton flavour $f = e, \mu, \tau, u, d, s, \dots$, with an associated coupling constant g_f . On substituting (D.27) into this interaction, a nonzero fermion mass m_f and a coupling between the Higgs lepton fields arise from the first and second terms in (D.27), respectively, and so are proportional to each other. Specifically, the lepton mass $m_f = g_f v / \sqrt{2}$ and the Higgs boson-lepton coupling constant g_{Hf} is given by

$$g_{Hf} = g_f / \sqrt{2} = m_f / v = \sqrt{2} g_W (m_f / M_W), \quad (\text{D.32})$$

where we have used (D.31) and (D.19c) to eliminate the vacuum expectation value constant v in favour of the usual weak coupling g_W and the W -boson mass.

Problems D

D.1 Show that Hamilton's equations of motion

$$\dot{x}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial x_i} \quad (i = 1, 2, 3)$$

lead to the equation of motion

$$m \frac{d\mathbf{v}}{dt} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

for a charged particle in an electromagnetic field if the Hamiltonian is of the form (D.4), where $\mathbf{r} = (x_1, x_2, x_3) \equiv (x, y, z)$.

¹⁰A more detailed discussion is given in Section D.5 of Martin and Shaw (2017).

D.2 Show that Maxwell's equations in free space

$$\text{div } \mathbf{B} = 0, \quad \text{curl } \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

and

$$\text{div } \mathbf{E} = 0, \quad \text{curl } \mathbf{B} = \frac{\partial \mathbf{E}}{\partial t}$$

reduce to (D.6a) and (D.6b) when expressed in terms of the potentials (D.1).

D.3 Show that for any electromagnetic potentials (ϕ, \mathbf{A}) there are many possible choices of gauge-transformed potentials

$$\phi = \tilde{\phi} + \frac{\partial f}{\partial t}, \quad \mathbf{A} = \tilde{\mathbf{A}} - \nabla f,$$

such that the Lorentz condition (D.7) is satisfied.

D.4 Verify that the Dirac equation in the presence of an electromagnetic field (D.12) is gauge invariant, as stated in the text.

D.5 Show that the gauge transformation (D.5) used in our discussion of QED can be written as two successive electroweak gauge transformations of the form (D.15) and (D.17).

E

Short answers to selected problems

Full solutions to all problems are available on the book's website. Here we give the final results for some selected problems where they are very brief.

Problems 1

- 1.1** Boson if $(A + Z)$ is even and fermion if $(A + Z)$ is odd.
- 1.5** (a) $C_\gamma = -1$; (b) from invariance of Maxwell's equations, $\mathbf{B} \rightarrow \mathbf{B}$ under P and $\mathbf{B} \rightarrow -\mathbf{B}$ under T .
- 1.8** $P_{\pi^-} = -1$.
- 1.9** (a) $\bar{\nu}_e + e^+ \rightarrow \bar{\nu}_e + e^+$; (b) $p + p \rightarrow p + p + \pi^0 + \pi^0$;
(c) $\bar{p} + n \rightarrow \pi^- + \pi^0 + \pi^0$, $\pi^- + \pi^+ + \pi^-$.
- 1.14** $\theta' = \text{atan}(\frac{k'}{k}) = \text{atan}(\frac{mc}{mc+k})$.
- 1.15** $\sigma = 161 \text{ mb}$.
- 1.16** $\sigma = 665 \text{ mb}$.

Problems 2

- 2.2** The reduction $F^2 = 0.69$.
- 2.4** $\sqrt{\langle r^2 \rangle} = 4.9 \text{ fm}$.
- 2.6** $\tau \approx 1.06 \times 10^{16} \text{ yr}$.
- 2.7** $t_{\max} = 265 \text{ min}$.
- 2.8** $\tau = 1.06 \times 10^{11} \text{ yr}$.
- 2.9** $a = 0.720 \text{ MeV}$.
- 2.10** $E = 154 \text{ MeV}$
- 2.11** About 10%.
- 2.12** $m_{\text{Hs}} c^2 = 269.13 \text{ u}$.
- 2.13** 7.43 kg.

2.16 1.55 days.

2.17 $\Gamma_n = 0.35 \times 10^{-3}$ eV and $\Gamma_\gamma = 9.65 \times 10^{-3}$ eV.

Problems 3

3.1 Forbidden: (a), (b), (c), (f). Allowed: (d), (e).

3.6 $\lambda \approx 7 \times 10^{12}$ km.

3.7 $\sigma(\pi^+ p \rightarrow \Sigma^+ K^+) : \sigma(\pi^- p \rightarrow \Sigma^- K^+) : \sigma(\pi^- p \rightarrow \Sigma^0 K^0)$
 $= |\mathcal{M}_{3/2}|^2 : \frac{1}{9} |\mathcal{M}_{3/2} + 2\mathcal{M}_{1/2}|^2 : \frac{2}{9} |\mathcal{M}_{3/2} - \mathcal{M}_{1/2}|^2.$

3.9 An octet of $J^P = 1/2^+$ states and a singlet $J^P = 3/2^+$ state.

(2, 1, 0, 1, 0) = cuu, (0, 1, -2, 1, 0) = css, (0, 0, 1, 0, -1) = b \bar{s} ,
3.10 (0, -1, 1, 0, 0) = $\bar{s}d\bar{u}$, (0, 1, -1, 1, 0) = csd, (-1, 1, -3, 0, 0) = sss.

3.12 Strong (a), (d), (f); weak (e); electromagnetic (c); forbidden (b).

3.13 Lifetime in range $(10^{-7} - 10^{-13})$ s.

3.17 $\mu_{\Sigma^0} = 0.83\mu_N$; $\mu_{\Omega^-} = -1.83\mu_N$.

3.19 Only (1,0,0,1,1) and (-1, 1, -2, 0, -1) are compatible with a simple quark model.

Problems 4

4.1 $E_L = 1.26 \times 10^4$ GeV.

4.2 $L = 1$ m.

4.3 $f = 6.15$ MHz and $R = 62.3$ cm.

4.5 4.63 m.

4.6 $L = 6.43 \times 10^{31}$ cm $^{-2}$ s $^{-1}$.

4.7 71 cm.

4.8 $E = 1.51$ GeV.

4.9 Intensity of elastically scattered neutrons is 0.765 s $^{-1}$; flux at 5 m is 2.44×10^{-3} m $^{-2}$ s $^{-1}$.

4.10 $l = 2.17$ m.

4.11 848 photons per second.

4.12 $E_p = 324$ keV.

4.14 $2^{21.7} = 3.41 \times 10^6 \approx 10^{6.53}$.

4.15 $L \approx 4.8$ m.

4.16 7.5 m.

4.19 $L\sigma t = 8.68 \times 10^4$.

Problems 5

5.4 $V(0.001 \text{ fm}) / V(0.1 \text{ fm})$ is of order 1/3.

5.5 $\alpha_s = 0.31$ from hadronic decay and $\alpha_s = 0.32$ from radiative decay.

5.11 $R(E_{\text{CM}} = 2.8 \text{ GeV}) \approx 2.17$, $R(E_{\text{CM}} = 15 \text{ GeV}) \approx 3.89$.

5.15 $a = 2b = 2.19$; $[xV(x)]_{\max} = 0.78$ at $x = 0.143$; fraction of the proton's momentum carried by valence quarks is 0.33.

5.18 (a) $\sigma_{\max} = 760 \text{ nb}$,

$$(b) \sigma_{p\bar{p}} = \frac{\pi Mc^2\Gamma}{s}\sigma_{\max} \int_0^1 \frac{u(x_u)d(M^2c^4/x_us)}{x_u} dx_u.$$

Problems 6

6.2 q^2 of order (a) $-4 \times 10^{-2} (\text{GeV}/c)^2$ and (b) $-4 \times 10^4 (\text{GeV}/c)^2$.

6.4 Rate increases by factor 4096.

6.8 (a) Cabibbo-allowed, (b) forbidden, (c) forbidden, and (d) Cabibbo-suppressed.

6.9 (a) and (c) forbidden by the $\Delta S = \Delta Q$ rule (6.31) for semi-leptonic decays; (e) forbidden by the $\Delta S = 0, \pm 1$ rule (6.33) for purely hadronic decays; (b), (d), and (f) are allowed.

6.10 $BR(b \rightarrow c + e^- + \bar{\nu}_e) = 1/9$; $BR(\tau^- \rightarrow e^- + \bar{\nu}_e + \nu_\tau) = 1/5$.

6.12 Predicted $R = 2$ (measured $R = 1.8$).

6.13 $r = 1.7 \times 10^{-2} \text{ fm}$; $t = 2.7 \times 10^{-24} \text{ s}$.

6.16 Instantaneous luminosity $= 3.3 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$.

6.18 $B(H^0 \rightarrow 4\ell) \approx 1.2 \times 10^{-4}$; $B(H^0 \rightarrow 2\ell 2\nu) \approx 9.8 \times 10^{-3}$; $N = 60$.

Problems 7

7.1 $\alpha = -1$.

7.2 Most probable energy $E_e = m_\mu c^2/2$; $\Gamma = \frac{G_F^2(m_\mu c^2)^5}{192\pi^3(\hbar c)^6}$; lifetime $\tau = \hbar/\Gamma = 2.19 \times 10^{-6} \text{ s}$.

7.3 $|q^2|_{\max} = 0.051(\text{GeV}/c)^2$, $\Delta(\sin^2\theta_W) \approx 0.002$, which is about 1%.

7.4 $A_{FB} = 3G(s)/8F(s)$; no.

7.5 $\Gamma_\nu = 501 \text{ MeV}$; $\Gamma_q = 547 \text{ MeV}$.

7.9 Distance travelled $1.4 \times 10^{-2} \text{ cm}$.

7.10 (a) Yes, (b) Yes, (c) No, (d) Yes.

7.13 Relative importance of Figures 7.14a and b is $\approx 1/10$.

Problems 8

8.1 For ${}^3\text{Li}$, the configuration is: protons: $(1s_{1/2})^2(1p_{3/2})^1$, neutrons: $(1s_{1/2})^2(1p_{3/2})^2$; $J^P = 3/2^-$; and magnetic moment $m = 3.8\mu_N$.

8.2 $j = 7/2$, $l = 3$.

8.4 ${}^{93}_{41}\text{Nb}$: $j^P = 9/2^+$, $\mu = 6.8\mu_N$; ${}^{33}_{16}\text{S}$: $j^P = 3/2^+$, $\mu = 1.14\mu_N$.

8.6 $a \approx 6.85 \text{ fm}$, $b \approx 5.82 \text{ fm}$.

8.7 $t_{1/2}({}^{228}_{90}\text{Th}) = 1.72 \text{ yr}$.

8.9 $t_{1/2} = 1.39 \text{ min}$.

- 8.10** $T_{\max} = 4.47 \text{ MeV}.$
8.11 Fraction $= 4.25 \times 10^{-11}.$
8.14 $\tau(\Sigma^0) = 5.8 \times 10^{-20} \text{ s}.$
8.15 $(2.35 \times 10^{-14})E_{\gamma}^7 A^2 \text{ eV}$ (E_{γ} in MeV).

Problems 9

- 9.1** Flux $= 3.13 \times 10^{17} \text{ s}^{-1} \text{ m}^{-2}$; interaction rate $= 0.989 \times 10^{12} \text{ s}^{-1}.$
9.2 $N \approx 121.$
9.3 Rate $\approx 135 \text{ kg yr}^{-1}.$
9.4 Energy gain $= 44.$
9.5 73 billion years.
9.6 Flux $\approx 4 \times 10^{12} \text{ cm}^2 \text{s}^{-1}.$
9.7 $n_d = 6.8 \times 10^{18} \text{ m}^{-3}.$
9.8 Rate ≈ 9 per second.
9.9 Lifetime $\geq 10^{16} \text{ yr}.$
9.10 $0.30 \mu\text{Sv}.$
9.11 2.5 cm.
9.12 $3.8 \times 10^{-4} \text{ Gy}.$
9.13 $f = 21.1 \text{ MHz}.$
9.14 $\approx 15600 \text{ yr}.$
9.15 $2.4 \times 10^8 \text{ yr}.$

Problems A

- A.4** $I = 1/2.$
A.5 All equal at 14%.

Problems B

- B.2** $\tan \theta = (m_{\pi}^2 - m_{\mu}^2)/(2\beta\gamma^2 m_{\pi}^2).$
B.3 $v = (m_{\pi}^2 - m_{\mu}^2)/(m_{\pi}^2 + m_{\mu}^2).$
B.5 $E_{\min} = 7m_p c^2 = 6.6 \text{ GeV}.$
B.7 $\cos \theta = 1 - M^2 c^4 / 2E_1 E_2.$
B.8 $E_{\gamma} = 2mc^2 = 1.02 \text{ MeV}.$

Problems C

- C.1** $d\sigma/d\Omega = 98.3 \text{ b/sr}.$
C.3 $b = \frac{zZe^2}{2\pi\varepsilon_0 mv^2\theta} = 2zZ \left(\frac{e^2}{4\pi_0\hbar c} \right) \frac{\hbar c}{mc^2} \frac{1}{(v/c)^2\theta},$
 $P_2 = m^2 \exp(-m)/2! \approx 1.3 \times 10^{-3}.$

References

- Aad, G. et al. (ATLAS Collaboration) (2013). *Physics Letters* **B726**: 88.
- Aaij, R. et al. (2015). *Physical Review Letters* **115**: 072001.
- Abasov, V.M. et al. (D0 Collaboration) (2016). *Physical Review Letters* **117**: 022003.
- Abe, K. et al. (2013). *Physical Review Letters* **111**: 211803.
- Abe, Y. et al. (Double Chooz Collaboration) (2012). *Physical Review* **D86**: 052008.
- Ablikim, M. et al. (BESIII Collaboration) (2013). *Physical Review Letters* **110**: 252001.
- Adachi, I. et al. (Belle Collaboration) (2012). *Physical Review Letters* **108**: 171802; arXiv:1201.4643.
- Agafonova, N. et al. (Opera Collaboration) (2015). *Physical Review Letters* **115**: 121802.
- Agostini, M. et al. (GERDA Collaboration) (2017). *Nature* **544**: 47.
- Aidala, C.A. et al. (2013). *Reviews of Modern Physics* **85**: 655.
- Ajaib A. et al. (2014) (Higgs Working Group Report 2014); arXiv.1310.8361
- Ajzenberg-Selove, F. (1990). *Nuclear Physics* **A506**: 1.
- Albert, J.B. et al. (EXO-200 Collaboration) (2014). *Nature* **510**: 229; arXiv:1402.6956.
- Amaudruz, P. et al. (1995). *Nuclear Physics* **B441** (1995): 3.
- Amsler, C. (1998). *Reviews of Modern Physics* **70**: 1293.
- Amsler, C. et al. (2017). *Quark Model*, Section 15.6 of Tanabashi et al. (Particle Data Group 2018).
- An, F.P. et al. (Daya Bay Collaboration) (2013). *Chinese Physics* **C37**: 011001.
- Andrade, E.N.da,C. (1964). *Rutherford and the Nature of the Atom*. New York: Doubleday.
- Angelopoulos, A. et al. (CPLEAR Collaboration) (2001). *European Physical Journal* **C22**: 55.
- Anthony, P.L. et al. (SLAC Collaboration E158) (2005). *Physical Review Letters* **95**: 081601; arXiv:hep-ex/0504049.
- Arneodo, M. et al. (1995). *Nuclear Physics* **B441**: 12.
- Arnold, R. et al. (2015). *Physical Review* **D92**: 072011.
- Arnold, R. et al. (2016a). *Physical Review* **D93**: 112008.
- Arnold, R. et al. (2016b). *Physical Review* **D94**: 072003.
- Ashie, Y. et al. (SuperKamiokande Collaboration) (2004). *Physical Review Letters* **93**: 101801.
- Aubert, B. et al. (2006). *Physical Review* **D73**: 012004.
- Auger, M. et al. (EXO Collaboration) (2012). *Journal of Instrumentation* **7**: P05010; arXiv:1202.2192.
- Barnes, V. et al. (1964). *Physical Review Letters* **12**: 204.
- Barrett, R.C. and Jackson, D.F. (1977). *Nuclear Sizes and Structures*. Clarendon Press.
- Bazin, D. (2012). *Nature* **486**: 330.
- Behnke, O.A., Geiser, A., and Lisovyi, M. (2015). *Progress in Particle and Nuclear Physics* **84**: 1; arXiv:1506.07519.
- Bellard, J.B. et al. (1967). *Physical Review Letters* **19**: 527.
- Bernstein, J. (2007). *Plutonium. A History of the World's most Dangerous Element*. Joseph Henry Press.
- Bertulani, C.A. (2007). *Nuclear Physics in a Nutshell*. Princeton University Press.
- Bethke S., Dissertori G., and Salam G.P. (2017). Section 9 of Tanabashi et al. (Particle Data Group 2018)
- Bjorken, J.D. and Drell, S.D. (1964). *Relativistic Quantum Mechanics*. New York: McGraw-Hill.
- Blatt, J.M. and Weisskopf, V.F. (1952). *Theoretical Nuclear Physics*. Wiley.
- Blaum, K. (2006). *Physics Reports* **425**: 1.
- Blaum, K., Nagy, S. and Werth, G. (2009). *Journal of Physics B: Atomic, Molecular and Optical Physics* **42**: 154015; (arXiv:0904.4712v1).
- Bosch, H.S. and Hale, G.M. (1992). *Nuclear Fusion* **32**: 611.
- Brandelik, R. et al. (1978). *Physics Letters* **B76**: 361.
- Brandt, R. et al. (1963). *Physical Review* **131**: 2617.

- Brown, L.S. and Gabrielse, G. (1986). *Reviews of Modern Physics* **58**: 233.
- Burcham, W.E. and Jobes, M. (1995). *Nuclear and Particle Physics*. Longmans Scientific and Technical.
- Carena M. et al. (2013) Section 11 of K.A. Olive et al. (Particle Data Group 2014)
- Carter, A.A. et al. (1968). *Physical Review* **168**: 1457.
- Ceccucci A., Ligeti Z., and Sakai Y. (2018) Section 12 Tanabashi et al. (Particle Data Group 2018)
- Cerrillo A.R. (2015) arXiv:1512.03311v2[hep-ph]
- Chatrchyan, S. et al. (CMS Collaboration) (2011). *Journal of High Energy Physics* **10**: 007; arXiv:1108.0566.
- Chatrchyan, S. et al. (CMS Collaboration) (2012). *Physical Review Letters* **109**: 222301.
- Chatrchyan, S. et al. (CMS Collaboration) (2014). *Physical Review* **D89**: 092007.
- Chen, Y. et al. (2006). *Physical Review* **D73**: 014516.
- Close, F.E., Marten, M., and Sutton, C. (1987). *The Particle Explosion*. Oxford University Press.
- Cottingham, W.N. and Greenwood, D.A. (2001). *An Introduction to Nuclear Physics*, 2e. Cambridge University Press.
- Dafni T. and Iguaz F.J. (2015). Proceedings of Science (Technology and Instrumentation in Particle Physics), Amsterdam; arXiv:1501.01456
- Dalitz, R.H. (1953). *Philosophical Magazine* **44**: 1068.
- Davis T. (2014). Physics World July 2014, 24
- Dawson, S. (1997). SUSY and such. *NATO Advanced Study Institute Series B Physics* **365**: 33; arXiv:hep-ph/9612229.
- Dendy, P.P. and Heaton, B. (1999). *Physics for Diagnostic Radiology*, 2e. Institute of Physics Publishing.
- Dosanjh, M. (2017). *From Particle Physics to Medical Applications*. Institute of Physics Publishing.
- Drees, M. and Gerbier, G., Dark Matter (2015). Section 26 of Patrignani C. et al. (Particle Data Group) (2016). *Chinese Physics* **C40**: 100001.
- Enge, H.A. (1966). *Introduction to Nuclear Physics*. Addison-Wesley Publishing Company.
- Ferbel, T. (ed.) (1992). *Experimental Techniques in High-Energy Nuclear and Particle Physics*, 2e. World Scientific.
- Figueroa-Feliciano, E. (2011). *Progress in Particle and Nuclear Physics* **66**: 661.
- Fossan, D.B. et al. (1961). *Physical Review* **123**: 209.
- Foster, B. et al. (2015). Section 9 of Patrignani C. et al. (Particle Data Group) (2016). *Chinese Physics* **C40**: 100001.
- Foster, B. et al. (2018). Section 18 of Tanabashi et al. (Particle Data Group 2018)
- Frois, B. (1983). *Proceedings of the International Conference on Nuclear Physics*, vol. **2** (ed. P. Blasi and R.A. Ricci), 221. Florence. Bologna: Tipografia Compositori.
- Gando, A. et al. (KamLAND Collaboration) (2011). *Physical Review* **D83**: 052002.
- Gando, A. et al. (KamLAND-Zen Collaboration) (2016). *Physical Review Letters* **117**: 082503.
- Garber D.I. and Kinsey R.R. (1976) *Neutron Cross-sections*, Vol **2**, Brookhaven National Laboratory Report BNL-325, Upton, NY
- Gasiorowicz, S. (1974). *Quantum Mechanics*. Wiley.
- Geesaman, D.F. et al. (2006). *Annual Review of Nuclear Particle Science* **56**: 53.
- Ghosh, P.K. (1995). Ion physics. In: *International Series of Monographs on Physics*, vol. **90**. Clarendon Press.
- Giunti, C. and Kim, C.W. (2007). *Fundamentals of Neutrino Physics and Astrophysics*. Oxford University Press.
- Gjesdal, S. et al. (1974). *Physics Letters* **B52**: 113.
- Gómez-Cadenas J.J. and Martin-Albo J. (2014). *Proceedings of Science* (Gran Sasso Summer Institute 2014 Hands-On Experimental Underground Physics at LNGS), INFN-Gran Sasso (2014), arXiv:1502.00581 (2015)
- Gottfried, K. and Weisskopf, V.F. (1986). *Concepts of Particle Physics*, vol. **2**. Oxford University Press.
- Graham, P.W. et al. (2015). *Annual Review Nuclear Particle Science* **65**: 485.
- Green, D.R. (2000). *The Physics of Particle Detectors*, Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology No.12. Cambridge University Press.
- Griffiths, D. (1987). *Introduction to Elementary Particles*. Wiley.
- Grupen, C. and Shwartz, B. (2011). *Particle Detectors*, Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology No.26. Cambridge University Press.
- Gunion, J.F. et al. (1990). *The Higgs Hunter's Guide*. Cambridge, MA: Perseus Publishing.
- Haber H.E. (2011). Theory of the Higgs Boson: The Standard Model and Beyond, Lectures at the Idpasc Higgs School, Foz do Arelho, Portugal; <http://www.idpasc.lip.pt/file.php/1/higgs2011/lectures/Haber.pdf>
- Hagedorn, R. (1964). *Relativistic Kinematics: A Guide to the Kinematic Problems of High Energy Physics*. Benjamin.
- Halzen, F. and Martin, A.D. (1984). *Quarks and Leptons*. Wiley.
- Hardy, J.C. and Towner, I.S. (2009). *Physical Review* **C79**: 055502.
- Heinemeyer S. et al. (2013) (LHC Higgs Cross Section Working Group), CERN-2013-004, arXiv:1307.1347 (hep-ph)
- Hendee, W.R. (1997). Physics and applications of medical imaging. *Reviews of Modern Physics* **71**: S444.

- Hintenberger, F. et al. (1968). *Nuclear Physics* **A115**: 570.
- Hively, L.M. (1983). *Nuclear Technology – Fusion* **3**: 1999.
- Hobbie, R.K. (1997). *Intermediate Physics for Medicine and Biology*, 3e. Springer and AIP Press.
- Hodgson, P.E., Gadioli, E., and Gadioli Erba, E. (1997). *Introductory Nuclear Physics*. Oxford University Press.
- Jackson, J.D. (1975). *Classical Electrodynamics*, 2e. Wiley.
- Kane, G.L. (1987). *Modern Elementary Particle Physics*. Addison-Wesley.
- Karliner M. and Skwarnicki T. (2016), Section 105 of Tanabashi et al. (Particle Data Group 2018)
- Klein O. (1938). *Proceedings of the Symposium on Les Nouvelles Théories de la Physique*, Warsaw
- Klein, O. (1948). *Nature* **161**: 897.
- Kleinknecht, K. (1986). *Detectors for Particle Radiation*, 1e. Cambridge University Press.
- Kleinknecht, K. (1998). *Detectors for Particle Radiation*, 2e. Cambridge University Press.
- Knoll, G.F. (1999). *Radiation Detection and Measurement*, 3e. New York: Wiley.
- König, M. et al. (1995). *International Journal of Mass Spectrometry and Ion Processes* **142**: 95.
- Kraft, G. (2000). *Progress in Particle and Nuclear Physics* **45**: S473.
- Krane, K.S. (1988). *Introductory Nuclear Physics*. Wiley.
- Lahav, O. and Liddle A.R. (2017). Section 24 of Particle Data Group (2018)
- Leo, W.R. (1994). *Techniques of Nuclear and Particle Physics*. Springer-Verlag.
- Lesgourges, J. and Verde L. (2017). Section 25 of Tanabashi M. et al. (Particle Data Group) (2018).
- LHCb Collaboration (2016). CERN Report CERN-LHCb-CONF-2016-004 (2016)
- Lilley, J. (2001). *Nuclear Physics. Principles and Applications*. Wiley.
- Liu, Z.Q. et al. (Belle Collaboration) (2013). *Physical Review Letters* **110**: 252002.
- Major, F.G., Gheorghe, V.N., and Werth, G. (2005). *Charged Particle Traps, Physics and Techniques of Charged Particle Field Confinement*, Springer Series on Atomic, Optical and Plasma Physics, vol. **37**. Springer.
- Mandl, F. (1992). *Quantum Mechanics*, 2e. Wiley.
- Mandl, F. and Shaw, G. (2010). *Quantum Field Theory*, 2e. Wiley.
- Martin, B.R. and Shaw, G. (2015). *Mathematics for Physicists*. Wiley.
- Martin, B.R. and Shaw, G. (2017). *Particle Physics*, 4e. Wiley.
- McRobbie, D.W. et al. (2003). *MRI from Picture to Proton*. Cambridge University Press.
- Nishino, H. et al. (Super-Kamiokande Collaboration) (2009). *Physical Review Letters* **102**: 141801;. (arXiv:0903.0676).
- Olive, K.A. et al. (Particle Data Group) (2014). *Chinese Physics* **C38**: 090001.
- Olsen, S.L. (2015). *Frontiers of Physics* **10**: 101401.
- Orito S. (1979) *Proceedings of the 1979 International Symposium on Lepton and Photon Interactions at High Energies*, Fermilab
- Paganetti, H. (2017). *Proton Beam Therapy*. Institute of Physics Publishing.
- Pais, A. (1986). *Inward Bound*. Oxford University Press.
- Particle Data Group (2018). *Physical Review* **D98**: 030001.
- Patrignani, C. et al. (Particle Data Group) (2016). *Chinese Physics* **C40**: 100001.
- Perkins, D.H. (1987). *Introduction to High Energy Physics*. Cambridge University Press.
- Phillips, A.C. (1994). *The Physics of Stars*. Wiley.
- Poenaru, D.N. and Greiner, W. (eds.) (1997). *Experimental Techniques in Nuclear Physics*. Mouton de Gruyter.
- Pohm, A.V. et al. (1955). *Physical Review* **97**: 432.
- Povh, B. et al. (1999). *Particles and Nuclei. An Introduction to the Physical Concepts*, 2e. Springer.
- Prescott, C.Y. et al. (SLAC-E122 Collaboration) (1978). *Physics Letters* **B77**: 347.
- Prescott, C.Y. et al. (SLAC-E122 Collaboration) (1979). *Physics Letters* **B84**: 524.
- Roos, M. (2015). *Introduction to Cosmology*, 4e. Wiley-Blackwell.
- Rubbia, C. et al. (1995). CERN report CERN/AT/95–44(ET)
- Satchler, G.R. (1967). *Nuclear Physics* **A92**: 273.
- Schiff, L.I. (1968). *Quantum Mechanics*, 3e. McGraw-Hill.
- Schindler, R.H. et al. (1980). *Physical Review* **D21**: 2716.
- Segrè, E. (1980). *From X-Rays to Quarks*. W.H. Freeman and Company.
- Seo S.-H. (RENO Collaboration) (2013). TAUP2013 International Workshop, September 2013, Asilomar, California, USA
- Sick, I. et al. (1975). *Physical Review Letters* **35**: 910.
- Smith, M. et al. (2008). *Physical Review Letters* **101**: 202501.
- Sublette C. (2007). The Nuclear Weapons Archive: A Guide to Nuclear Weapons; <http://nuclearweaponarchive.org>

- Suzuki, Y. (2005). *Physica Scripta* **T121**: 33.
- Tanabashi, M. et al. (Particle Data Group) (2018). *Physical Review* **D98**: 030001.
- Trigg, G.L. (1975). *Landmark Experiments in 20th Century Physics*. New York: Crane, Russak and Co. Ltd.
- Weingarten D.H. (1996). *Scientific American* (February 1996) 104
- Williams, W.S.C. (1991). *Nuclear and Particle Physics*. Oxford: Clarendon Press.
- Wu, S.-L. (1984). *Physics Reports* **107**: 59.
- Xu, Q. (2015). *Frontiers of Physics* **10** (6): 101402.
- Zhang, X. et al. (2010). *International Journal of Radiation Oncology, Biology, Physics* **77**: 357.

Index

- absorption length 154
accelerator driven systems (ADS) 353–356
accelerators (see *particle accelerators*)
activation energy 75
activity 33, 67–68
ADS (see *accelerator driven systems*)
allowed transitions in beta decay 331–333, 335–336
 α decay 5, 67, 73, 324–327
 Gamow factor 326, 358
 Geiger-Nuttall relation 325–327
 α rays, discovery 16
amplification factor in gas detectors 162–164, 167
amplitude 27–29
 scattering 27, 31–33
angular momentum conservation 14
ANITA detector 431–432
anomaly condition 248
anthropic principle 417–418
antiparticles 11–13
 and CPT theorem 289
 relative parity of 16
antiscreening 194
Askaryan effect 172, 431
associated production 152
asymmetry term in SEMF 62–63, 310, 323
asymptotic freedom 189–194
ATLAS detector 149, 177–178
atomic mass number (see *mass number*)
atomic masses 3, 29, 42–43
atomic mass unit 37
atomic models
 Bohr-Rutherford model 3–4
 Thomson plum pudding model 2
atomic numbers 4, 41–42
atomic shell model 310–312
atomic weight 3
axions 441–444
- barn (unit of area) 36
barrier penetration/tunneling 445–447
 in α decay 325–327
baryon-antibaryon asymmetry in universe 439–440
- baryon number 108–109
 violation in grand unified theories 410–412
baryons
 colour wave function of 187
 exotic states 200–201
 magnetic moments 119–121
 mass splittings in quark model 121–122, 124–126
 nucleon resonances 35–36, 122–124
 pentaquarks 134–135, 200–201, 216
 supermultiplets 115–116
 tables of states 112
beams (see *particle beams*)
beauty quantum number 107
Becquerel 67
bending magnet 146
 β decay 21, 66, 69–72, 327–337
 continuous beta decay spectrum 5–6, 329–330
 double β decay 72, 420–426
 inverse beta decay 89–90
 neutrinoless double β decay 420–426
 parity violation in 272–273
beam fragmentation 219, 220
Bethe–Bloch formula 155–156
 B factories 294–295
Big Bang model 361–363, 432–440
binding energy 60
 experimental data 61, 63, 313
 semi-empirical mass formula (SEMF) 60–64
 shell model 312–315
biological effects of radiation 377–380
Bjorken scaling (see *scaling*)
 B - L quantum number 410
 B mesons 290
 $B_H^0 - B_L^0$ mass difference 295
 $B^0 - \bar{B}^0$ mixing and oscillations 291–295
 CP violation in B decays 290–291, 295–298, 300–301
Bohr-Rutherford model of atom 3–4
Born approximation 27, 32
bottomonium 126–127, 129–130
 OZI rule 128, 192
 spectrum of states 130
bottom quantum number 114

- Bragg curve 156
 Bragg peak 156, 382
 branching ratio 34
 isospin predictions of 111–113
 brane 418
 Breit–Wigner formula 34–36
 bremsstrahlung 157–158
 bubble chamber 8, 162
 event pictures 230, 242
- Cabibbo allowed/suppressed decays 237–238
 Cabibbo angle 236
 Cabibbo hypothesis 236
 Callan–Gross relation 211
 calorimeters 173–176
 carbon dating 396
 cascade baryon 241–242
 CEBAF accelerator 144–145, 308
 centre-of-mass energy 140, 449
 Čerenkov counters 170–172
 Čerenkov radiation 94, 170–171
 coherent 172
 CERN accelerator complex 148–149
 chain reaction 346–349
 Chandrasekhar limit 427
 charge conjugation 16–17
 nonconservation in muon decay 273–274
 charged and neutral weak currents 229–230
 charge distribution of nuclei 52–56
 charged weak current reactions
 quark mixing 235–238
 selection rules 239–242
 W boson decay 233–234, 238–239
 W -lepton vertices 232–234
 W -quark vertices 235–238
 charge independence of nuclear force 110–111
 charge/isospin multiplets 109–110
 charge symmetry of nuclear force 110
 charginos 413
 charmed quark 106–108, 225
 charmonium 126–130
 exotic states 197–200
 OZI rule 128–129, 192
 charm quantum number 114
 charm threshold 128
 chirality 279–281
 chiral symmetry 225
 CKM matrix 244–246, 299–302
 cloud chamber 7, 8
 CNO chain 364–365
 Cockcroft–Walton accelerator 142
 cold dark matter 434–438
 collective model 322–324
 colliding beam experiments 30, 140–141
 collision length 154
- colour 185–187
 charges 186–188
 confinement 186–187, 189–190
 evidence for in e^+e^- annihilation 205–207
 hypercharge 186–188, 474
 isospin 186, 474
 singlet 189
 complex nuclear potential 58
 compound nucleus 75, 77–80, 154, 344
 Compton scattering 158–159
 computed tomography 388–389
 continuous beta-decay spectrum 5–6, 329–300
 confinement
 colour confinement 106, 186–187, 189
 fusion reactors 368–370
 Penning trap 47–48
 confining potential 190
 conservation laws
 angular momentum 14
 baryon number 108–109
 B - L 410
 charge conjugation 16–17
 colour 186, 192
 CP 274–275
 lepton numbers 85–86
 parity 15–16
 constituent quarks 107, 224
 control rods 348–350
 cosmic microwave background radiation 432–433
 cosmic rays 6–8, 93–95, 172, 396, 431, 438
 cosmological constant 433
 COSY accelerator 147–148
 Coulomb barrier 75, 79, 325–326, 357–358
 Coulomb scattering 52, 466–468
 coupling constant 25
 C parity 16–17
 violation in weak interactions 273–275
 CP invariance 274–275, 278, 281–283
 CPT theorem 19–20, 285–286, 289
 CP violation
 and electric dipole moment 440–441
 and strong CP problem 441–443
 by mixing 285, 292
 direct CP violation 285, 290–291, 297–298, 300–301
 in B decays 290–291, 295–298
 in interference 295–298
 in K decays 283–285
 in neutrino sector 301–302
 in standard model 299–301
 critical mass 347
 cross-sections 29–33
 classification for nuclear reactions 76–79
 deep inelastic scattering 79, 208
 elastic scattering from nuclei 52–54
 lepton pair production (Drell–Yan process) 221–223

- Mott scattering formula 5
 neutron scattering from nuclei 78, 80–81, 344–345
 photon scattering 158–159
 pi-nucleon scattering 112–113, 153–154
 Rutherford scattering formula 53
 (see also *electron-positron annihilation*)
 CT (see *computed tomography*)
 Curie (unit of radioactivity) 67
 current quarks 224–226
 cyclotron 145–146
 cyclotron frequency 48
 Dalitz plot 462–463
 dark energy 433
 dark matter 434
 search for cold dark matter 434–438
 DC accelerators 142–143
 decay chains 67–69
 decay constant 33
 decay width 34
 deep inelastic neutrino scattering 215–217
 deep inelastic scattering 207–217
 EMC effect 404
 Gottfried sum rule 227
 nuclear 79, 214, 404
 parton distributions 211–214, 218–219
 deflection spectrometers 43–45
 degeneracy pressure 427
 delayed neutron emission 345–346, 349–350, 355
 $\Delta S = \Delta Q$ rule 240–241
 density of states 32–33, 308, 447–448
 in β decay 329–330
 Fermi gas model 308
 density of the universe 432–433
 detectors (see *particle detectors, particle detector systems*
 and *detectors for neutrino experiments*)
 detectors for neutrino experiments
 ANITA 431–432
 EXO-200 423–424
 GERDA 425
 IceCube 429–431
 KamLAND 99, 181–182
 KamLAND-Zen 424–425
 NEMO-3 425–426
 Superkamiokande 93–94
 detailed balance 19
 deuteron 98–99, 121, 306
 differential cross-section 31, 467
 Dirac delta function 35, 451
 Dirac equation 11–12
 Dirac magnetic moment 13
 direct reaction 77–80
 double beta decay 72, 420–426
 Drell–Yan mechanism 221–223
 drift chamber 164–165
 effective dose of radiation 378
 elastic scattering 20
 electric dipole moment 319, 440–441
 electric quadrupole moment 319–322
 electromagnetic interactions 10, 22–23, 25, 26, 29
 gauge invariance of 471–474
 electromagnetic showers 174–176
 electron capture 66, 71–72, 278
 electron–electron scattering 22, 193
 electron neutrino 84, 88–90, 96
 mass 89, 334–335
 oscillations 93–96, 96–99
 electron number 85
 electron-positron annihilation
 evidence for colour in 205–207
 Higgs boson production in 259–262
 jets in 219
 to hadrons 127–128, 205–206
 to muon pairs 127, 221, 251–252
 to photons 21
 electron–positron pair production 158–159
 electroweak reactions 251–252, 275–277
 electroweak unification 229–231, 247–250, 476–478
 EMC effect 404
 energy losses by particles in matter 152
 ionisation losses 154–157
 radiation losses 157–158
 short-range interactions with nuclei 153–154
 equivalent dose of radiation 378
 EXO-200 detector 423–424
 exotic hadrons 133, 194–201
 exotic quantum numbers 133–135, 196
 exponential decay law 33, 67
 FAIR accelerator complex 150
 fast breeder reactor 351–352
 Fermi coupling constant 28
 Fermi gas model 308–310
 Fermi screening factor 332–334
 Fermi’s golden rule (see *golden rule*)
 Fermi transitions 331
 Feynman diagrams 21–23
 electromagnetic vertex 232
 Higgs boson vertices 256–257
 order of 29
 QCD vertices 188
 weak boson vertices 232–233, 236–237, 246, 250–251
 Feynman rules 21–22, 26
 fine structure constant 26
 fissile nuclei (see *nuclear fission*)
 fixed-target experiments 29, 140

- flavour independence of strong interaction 109–113, 131–132, 187–188
 flavour oscillations 282
 in B decays 293–295
 in K decay 285–289
 forbidden transitions in beta decay 332
 form factors 53–54, 56
 fragmentation 105
 fusion (see *nuclear fusion*)
- γ decays 76, 337–340
 Gammasphere 178–181
 Gamow-Teller transitions 331–332
 gas amplification factor 162–163
 gaseous ionisation detectors 162–167
 gauge bosons 10, 87
 gauge theories
 electromagnetic interaction 187–188, 471–474
 electroweak interaction 247–250, 253–255, 476–478
 gauge principle 253–255, 473–478
 Higgs boson 252–253, 255–257
 photon mass 473
 QCD 187–194, 474–476
 unification condition 248–250, 476
 gauge transformations
 electromagnetic 253, 471–474
 electroweak 476–478
 quantum chromodynamics 474
 Geiger and Marsden experiments 3, 465–466
 Geiger–Müller counter 167
 Geiger–Müller region 167
 Geiger–Nuttall relation 325–327
 Georgi–Glashow model 408–412
 GERDA detector 425
 g-factor 316–317
 glueballs 189, 195–197
 gluino 413, 416
 gluons 10, 187–192
 momentum distribution function 218–219
 spin determination in e^+e^- annihilation 205
 golden rule 329, 449–452
 Gottfried sum rule 227
 grand unified theories (GUTs) 407–412
 see-saw mechanism for neutrino mass 420
 gray (unit of radiation dosage) 378
 GUTS (see grand unified theories)
- hadronic showers 174, 176
 hadrons
 allowed and exotic quantum numbers 133–135, 196
 baryon magnetic moments 119–121
 charge/isospin multiples 109–110
 colour wavefunction 47, 187
 decays and lifetimes 117–119, 128–129
 exotic states 133–135, 194–201
 heavy quarkonia 126–123
 mass plittings 121–126
 resonances 34–36, 112–113, 117
 supermultiplets 115–117
 half-life 2, 33–34
 halo nuclei 57, 407
 helicity 277–280, 418–419
 hierarchy problem 414
 Higgs boson 9, 231, 252
 discovery 259–265
 in supersymmetry 413–414, 416
 production and properties 257–262
 Higgs field and particle masses 252–256, 478–481
 Higgsinos 413
 Higgs mechanism 478–481
 hot dark matter 434
 hybrid states 189, 196–197, 198
 hypercharge 115
 hyperfine interaction 122
 hypernucleus 404
- IceCube detector 429–431
 impact parameter 466–467
 inelastic scattering 21, 79
 electron and muon scattering 207–215
 neutrino scattering 215–217
 inflationary Big Bang model 432–433
 integrated luminosity 30–31, 141
 interaction lengths 154
 interactions of photons in matter 158–159
 internal conversion 67, 339
 internal pair production 339
 internal quantum numbers 133–134
 international linear collider (ILC) 150
 intrinsic parity 14
 invariant mass 460–462
 inverse β decay 89–90
 inverse muon decay 232
 ionisation chamber 162–164
 ionisation energy losses 154–158
 ion trap 46–50
 isobar 41
 isospin 109–113
 commutation relations 452
 formal theory of 452–456
 isotone 41
 isotope 3–5, 41
 ITER 370
- JET 369–370
 jet chamber 105, 165
 jets 105
 evidence for quarks 105–106
 in e^+e^- annihilation 105, 204–207
 in deep inelastic scattering 219

- in pp collisions 220
 rapidity 219–220
 J/ψ meson 126–127, 129
- KamLAND detector 99, 181–182
 KamLAND-Zen detector 424–425
K-capture (see *electron capture*)
 Klein–Gordon equation 12, 25
K mesons (kaons) 8, 114 (see also *neutral K mesons*)
 Kurie plot 332–335
- large hadron collider (LHC) 148–149
 Larmor frequency 391–393, 441
 lattice gauge theory 189–190, 194–196, 202, 224
 Lawson criterion 367–368
 lepton 9, 83–90 (see also *neutrino*)
 decays 84–85, 87
 parity 16
 tables of properties 84
 universality 86–87
 lepton numbers 85
 conservation 85–86, 103–104
 lepton pair production
 by photons 158–159
 in e^+e^- collisions 127, 221, 251–252
 in pp collisions 221–223
 lepton–quark symmetry 234–238
 LET (see *linear energy transfer*)
 LHC accelerator 148–149
 linear accelerators (linacs) 143–145
 linear energy transfer (LET) 381–382
 liquid drop model (see *semi-empirical mass formula*)
 Lorentz condition 473
 Lorentz transformations 457–461
 luminosity 30–31, 141
- MACHOS 434
 magic numbers
 atomic 311–312
 nuclear 65–66, 312–313, 324, 406
 magnet designs 146
 magnetic deflection spectrometer 43–45
 magnetic moments
 baryons 119–121
 deuteron 121
 Dirac prediction 13
 nuclear 316–318, 324
 quark 119–120
 magnetic resonance imaging (MRI) 390–395
 Majorana neutrinos 419–420
 masses 420, 420–422
 mass deficit 60
 mass number 41
 mass spectroscopy (see *nuclear mass measurements*)
 matter–antimatter asymmetry in universe 438–440
- Maxwell-Boltzman distribution 359
 Maxwellian energy distribution 358–360
 Maxwell equations 472–473
 mean free path 154
 mean lifetime 33–34
 mean square charge radius of nuclei 56
 medical imaging 385–395
 carbon ion beams 383–385
 computed tomography (CT) 388–390
 γ camera 387–389
 magnetic resonance imaging (MRI) 390–395
 positron emission tomography (PET)
 389–390
 single-photon emission CT (SPECT) 389
 X-rays 385–387
 Mendeleev periodic table of the elements 4, 41–42
 Mesons
 charmonium and bottomonium 126–132
 exotic states 197–200
 glueballs and hybrids 195–197
 in quark model 113–115, 117–119
 masses in quark model 123–124
 microstrip detectors 166, 169
 Minimal Supersymmetric Standard Model (MSSM) 413, 415–416
 minimum ionisation region 155
 mirror nuclei 110–111
 moderator 151, 348–349
 molecular states 198
 Mott cross-section 52–53
 MRI (see *magnetic resonance imaging*)
 MSW effect 92–93, 99
M theory 418
 multipole radiation 338
 multiwire proportional chamber 164–165
 muon 7, 83–87
 decay asymmetries 273–275
 pair production 127, 206, 251–252
 muon neutrino oscillations 93–96
- natural units 37
 NEMO-3 detector 425–426
 neutral and unstable particle beams 150–152
 neutral *K* mesons 281–289
 neutralino 413, 415–416, 435
 neutral weak currents 229–230
 and unified theory 229–230
 neutrino 6, 83–86
 and dark matter 434–435
 Dirac or Majorana 419
 helicity of 277–279
 limit on number of 86–88
 mixing and oscillation 90–93
 oscillation experiments 93–101
 ultra high-energy 429–432

- neutrino astrophysics 427–432
 neutrino masses 83, 88–89, 101–103, 334–335
 and double β decay 421–422
 from cosmological models 435
 from supernova 427–429
 see-saw mechanism 420
 neutrino–nucleon scattering
 deep inelastic scattering 215–217
 inverse β decay 89–90
 neutron
 capture 75–76, 181, 312, 344–345, 352–353
 cross-sections 78, 80
 decay 21, 107, 119, 232, 327–328
 discovery 5
 magnetic moment 13, 121
 number 41
 neutron star 204–207
 nonspherical nuclei 319–323
 nuclear binding energy (see *binding energy*)
 nuclear chain reaction 346–348
 nuclear charge distribution 52–56
 mean square charge radius 56
 nuclear density 57–59
 nuclear fission 66, 72–76, 343–348
 nuclear fission reactors (see *nuclear power*)
 nuclear force (see *strong nuclear force*)
 nuclear fusion 357–361
 (see also *nuclear power*, *nuclear weapons* and
 nucleosynthesis)
 nuclear mass measurements 43–51
 nuclear matter distribution 57–59
 nuclear models
 collective 322–324
 Fermi gas 308–310
 liquid drop (see *semi-empirical mass formula*)
 shell model (see *nuclear shell model*)
 summary of properties 323–324
 nuclear power
 accelerator driven system (ADS) 353–356
 fast breeder reactor 351–352
 fission reactors 348–352
 fusion reactors 366–371
 radioactive waste 352–354
 nuclear radius 56
 nuclear reactions 21, 76–81
 nuclear shell model 63, 312–319
 fission 73–74
 γ decays 339–340
 magnetic moments 314, 316–318
 quadrupole moments 320–322
 summary 323–324
 superheavy nuclei 406
 nuclear weapons
 fission devices 371–374
 fission/fusion devices 374–377
 nucleon 6
 nucleon-nucleon potential 305–308
 boson exchange model 7, 23–26, 305
 charge symmetry 110–111, 306
 quark model interpretation 306–307
 nucleosynthesis
 in early universe 361–363
 in stellar evolution 363–366
 nucleus, discovery of 3
 nuclide 41
 oblate nuclei 320
 omega-minus particle 241–242
 one-proton decay 405–406
 optical model 57–59
 oxygen effect 380
 OZI rule 128, 192–193
 pairing hypothesis 316
 pairing term in SEMF 62
 parity 14–16
 $\tau - \theta$ puzzle 272
 violation in electroweak interaction 275–277
 violation in weak interaction 271–275
 partial width 34
 particle accelerators 142–149
 future accelerators 150
 neutral and unstable beams 150–152
 particle-antiparticle mixing
 for B mesons 291–295
 for K mesons 281–282, 286–289
 particle astrophysics
 cosmology and dark matter 432–438
 matter-antimatter asymmetry 438–440
 neutrino astrophysics 427–432
 particle beams 139–140
 carbon ion beams 383–385
 neutral and unstable particle beams 150–152
 secondary beams 147, 150–152
 particle detectors 160–176
 particle detector systems 176–182 (see also *detectors for neutrino experiments*)
 parton distributions 211–215, 218–219
 scaling violations 214
 particle interactions with matter 152–159
 parton model 210–211
 gluon spin determination 205
 neutrino scattering 215–217
 quark charge determination 217
 Paul trap 47
 penguin diagrams 300–301
 Penning trap 46–52
 pentaquarks 134, 200–201
 periodic table of the elements 4, 19, 41–43
 PET (see *positron emission tomography*)

- photino 413
 photoelectric effect 158–159
 photographic emulsion 7
 photomultiplier tube 94, 168
 photon
 C parity 16
 interactions with matter 158–159
 mass and gauge invariance 473
 pickup reaction 77
 pion–proton scattering 112–113, 153
 pions 7, 28, 114, 118, 280
 role in nuclear forces 24–25
 Planck mass 418
 plum pudding model of atoms 2
 positron 12
 positron emission tomography (PET) 389–390
 principle of detailed balance 19
 Proca equations 473
 prolate nuclei 320
 prompt neutron emission 345–346
 proportional chamber 164
 proton
 decay in grand unified theories 410–412
 magnetic moment 13, 120
 number 41
 spin crisis 402–403
 structure functions 208–210
 proton-proton cycle 363–364
 Prout’s law 3, 4
 pseudorapidity 220

 quadrupole moment 312, 319–322
 quadrupole transition 338
 quantum chromodynamics (QCD) 187–194
 cross-section for e^+e^- to hadrons 205–207
 current and constituent quarks 224–225
 gauge theory of 474–475
 glueball predictions 196
 jets in QCD 204–205, 220
 scaling violations in 211–214
 strong coupling constant 190–192
 quantum electrodynamics (QED) 21–23,
 26
 and gauge invariance 253–254, 471–474
 quantum fluctuations 193–194
 quark–antiquark potential 131–132
 quark confinement 106, 190
 quark–gluon plasma, 201–204
 quark mixing 236–238, 243–246, 250
 quark model 113–117
 hadron decays and lifetimes 117–119
 hadron masses 121–126
 heavy quarkonia 126–131
 magnetic moments 119–121
 quark-parton model (see *parton model*)
 quarks 8–9, 104–109, 243
 colour charges 186, 474
 constituent and current 224–226
 distribution functions 211–215, 218–219
 electric charge determination 216–217
 properties of top quark 243, 246–247
 running mass 225
 Q value 67, 119

 rad (unit of radiation dosage) 378
 radioactive dating 396
 radioactive decay (see α, β, γ decays and spontaneous fission)
 radioactive waste 352–354
 radiation
 and living matter (see biological effects of radiation)
 energy losses 157–158
 length 157, 159
 therapy 380–385
 range
 of particle exchange force 24–25
 of particles in matter 159–160
 zero-range approximation 28
 rapidity 219–220
 Rayleigh scattering 158–159
 relativistic kinematics 457–463
 relativistic wave equations 11–13, 473
 resistive plate chamber 166–167
 resonances (see also Breit–Wigner formula)
 hadron 34–36, 112–113, 117
 nuclear 78, 80
 RHIC accelerator 149
 R parity 413, 415
 running coupling constant of QCD 191–192
 running quark mass 225
 Rutherford cross-section 53
 Rutherford scattering 465–468

 Sargent’s rule 87, 336–337
 scaling 207–210
 scaling violations 209–210
 in QCD 214
 scattering amplitude 27–28
 scattering formulas 31–33
 Schmidt lines 317–318
 scintillation counter 161, 167–169
 sea quarks 212
 secondary particle beams 147, 150–152
 second golden rule 32, 329–330, 449–452
 ‘see-saw mechanism’ for neutrino masses 420
 Segrè plot 64–65
 selection rules
 in β decay 330–332
 in γ decays 337–339
 in weak interactions 239–242

- selectron 413
 semiconductor detector 169–170
 semi-empirical mass formula (SEMF) 60–64, 323
 and fission 72, 74–75
 fit to binding energy data 313
 use in analysing β decay 69–72
 separation energy 64, 80
 shell model (see *atomic shell model* and *nuclear shell model*)
 shower counters 174
 Sievert (unit of radiation dosage) 378
 silicon microstrip detector 169
 single-photon emission computed tomography (SPECT) 389–390
 smuon 413
 SNO detector 98–99
 solar neutrinos 96–99
 solid-state detectors 169
 spallation neutron sources 151
 spallation process 151, 353–356
 spark chamber 167
 spectator model 107–108
 spectrometer 161
 deflection 43–45
 γ ray 178–181
 spectroscopic notation 314
 spin dependence of nuclear forces 306–307
 spin echoes 393
 spin-lattice relaxation time 392
 spin-orbit potential 311, 313–315
 spinors 12, 477
 spin-spin interaction 122–123
 spin-spin relaxation time 393
 spontaneous symmetry breaking 255–256, 478–481
 squark 413, 416
 standard cosmological model (see *inflationary Big Bang model*)
 standard solar model 96
 STAR detector 166, 178, 179
 stauon 413
 stellar evolution
 CNO chain 364–365
 production of heavy elements 365–366
 proton-proton cycle 363–364
 storage rings 46, 147
 strangeness oscillations 286–289
 strange particles 8, 114
 streamer chamber 167
 string theories 417–418
 stripping reaction 77
 strong nuclear force
 boson exchange model 24–25, 305
 charge symmetry 110–111, 306
 quark model interpretation 306–307
 saturation 61, 306
 short-range repulsion 306–307
 three-body contribution 405
 strong coupling constant 190–192
 strong CP problem 441–442
 structure functions 208–210
 superallowed transitions in beta decay 336
 superheavy elements 66, 406
 Superkamiokande detector 93–94
 supernova 366, 427–429
 superparticles 413–414
 experimental searches for 414–417
 supersymmetry (SUSY) 412–414
 CP violation in 439–440
 electric dipole moments in 440
 experimental limits 414–417
 Minimal Supersymmetric Standard Model (MSSM) 413, 415–416
 WIMPs 435
 symmetries
 charge conjugation 16–17
 CP invariance 274–275, 278, 281–283
 gauge invariance 253–254, 471–474
 parity 15–16
 time reversal 17–20
 synchrotron accelerator 146–147
 synchrotron radiation 146
 tandem Van de Graaff 142
 target fragmentation 219
 tauon 83–85, 87
 Teller-Ulam configuration 375–376
 tetraquarks 134, 197–198
 thermal neutron 75, 80, 154
 thermal nuclear fission reactor 348–352
 radioactive waste disposal problem 352–354
 time ordering 22–23
 time projection chamber 165–166
 time reversal 17–20
 tokamak 369–371
 top quark 106–108, 246–247
 total cross-section 31
 tracking detectors 161
 transformation theory 2
 transition radiation 170, 172–173
 transition radiation detector 161, 173
 tree diagrams 300–301
 triangle function 462
 tunnelling (see *barrier penetration*)
 unification
 and gauge principle 253–255, 473–477
 condition 248–250, 476
 mass in GUTs 408
 unified electroweak interaction 229–231, 247–255, 475–478

- units 36–37
 universality of lepton interactions 86–87
- vacuum polarization effects 193–194
 $V\!-\!A$ interaction 279, 328–329
 valence quarks 212–213
 Van de Graaff accelerator 142–143
 Van der Waals force 26, 306
 vertex detector 169, 178
 virtual processes 24
- wavelength shifter 168
 W boson 10, 25, 229–231
 decays 233–234, 238–239
 exchange and Fermi coupling 23, 28
 lepton vertices 232–234
 quark vertices 235–238
 weak charged current reactions 231–242
 selection rules 239–242
 weak hypercharge 476
 weak interaction spin structure 277–281
 weak isospin 475
 weak mixing angle 248–250, 276–277, 409–410, 414, 476
- weak neutral current reactions 229–230, 247, 250–252
 weight diagrams 115–116
 Weisskopf approximation 339
 Weinberg angle (see *weak mixing angle*)
 WIMPs 435–438
 wino 413
 wire chambers 164–166
 Wolfenstein parameterisation 299–300
 Woods-Saxon potential 313–315
- X boson 408–411
- Y boson 408–410
 Yukawa potential 25–26
 Yukawa theory of nuclear forces 7, 23–24
- Z bosons 10, 25, 229–231
 decays 87–88
 lepton and quark vertices 250–251
 production in e^+e^- annihilations 252
 production in pp collisions 223
 zero-range approximation 26, 28
 zino 413

PHYSICAL CONSTANTS

Note: Entries are given to six significant figures where these are known. Otherwise the error is given in brackets.

Quantity	Symbol	Value
Speed of light in vacuum	c	$2.99792 \times 10^8 \text{ m s}^{-1}$
Planck's constant	h	$4.13567 \times 10^{-24} \text{ GeV s}$
Planck's constant reduced	$\hbar \equiv h/2\pi$	$6.58212 \times 10^{-25} \text{ GeV s}$
	$\hbar c$	$1.97327 \times 10^{-16} \text{ GeV m}$
	$(\hbar c)^2$	$3.89379 \times 10^{-32} \text{ GeV}^2 \text{ m}^2$
Electron charge (magnitude)	e	$1.60218 \times 10^{-19} \text{ C}$
Avogadro's number	N_A	6.02214×10^{23}
Boltzmann's constant	k_B	$8.61733 \times 10^{-11} \text{ MeV K}^{-1}$
Electron mass	m_e	$0.510999 \text{ MeV}/c^2$
Neutron mass	m_n	$0.939565 \text{ GeV}/c^2$
Proton mass	m_p	$0.938272 \text{ GeV}/c^2$
W boson mass	M_W	$80.379(0.012) \text{ GeV}/c^2$
Z boson mass	M_Z	$91.1876(0.0021) \text{ GeV}/c^2$
Higgs boson mass	M_H	$125.18(0.16) \text{ GeV}/c^2$
Atomic mass unit	$u \equiv (\frac{1}{12} \text{ mass } {}^{12}\text{C atom})$	$931.494 \text{ MeV}/c^2$
Bohr magneton	$\mu_B \equiv e\hbar/2m_e$	$5.78838 \times 10^{-11} \text{ MeV T}^{-1}$
Nuclear magneton	$\mu_N \equiv e\hbar/2m_p$	$3.15245 \times 10^{-14} \text{ MeV T}^{-1}$
Fermi coupling constant	$G_F/(\hbar c)^3$	$1.16638 \times 10^{-5} \text{ GeV}^{-2}$
Fine structure constant	$\alpha \equiv e^2/4\pi\varepsilon_0\hbar c$	$7.29735 \times 10^{-3} = 1/137.036$
Strong coupling constant	$\alpha_s(M_Z c^2)$	$0.1181(0.0011)$

CONVERSION FACTORS

$1 \text{ eV} = 1.6022 \times 10^{-19} \text{ J}$	$1 \text{ eV}/c^2 = 1.7827 \times 10^{-36} \text{ kg}$
$1 \text{ fermi} = 1 \text{ fm} \equiv 10^{-15} \text{ m}$	$1 \text{ barn} = 1 \text{ b} \equiv 10^{-28} \text{ m}^2$
$1 \text{ yr} = 3.156 \times 10^7 \text{ s}$	$k_B T \text{ (at 300 K)} = 0.025852 \text{ eV}$

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.